DESCRIPTION

TECHNICAL FIELD

[0001]   The invention relates to a method for isolating a complex nucleotide fragment comprising a known nucleotide sequence element, i.e. a sequence encoding a conserved active site or domain, the method being applicable i.e. to high throughput screening for DNA fragments containing a known sequence element.

BACKGROUND OF THE INVENTION

General introduction

[0002] Molecular diagnostics and other DNA based approaches have gained increasing focus in sectors such as Discovery, R&D, and in various branches of diagnostics. However, whatever target is being analysed by assays or screening, all DNA based approaches face the same challenge, namely to generate sufficient amounts of available target relative to background DNA.

[0003]   By increasing the presence of low abundance target DNA, at the expense of undesirable background DNA in a complex mixed DNA sample, traditional molecular methods, such as clone library construction, natural product discovery, PCR diagnostics, hybridizations, sequencing, metagenomics, and a variety of other molecular approaches are enabled.

[0004] The following describes methods currently used and the challenges of low abundance, mixed DNA samples.

PCR Diagnostics

[0005] In recent years, PCR tests have been extensively developed for routine diagnostics of infectious diseases in clinical microbiology. PCR is suited for rapid detection of bacteria directly in clinical specimens, allowing early, sensitive and specific laboratory confirmation of related diseases [1]. Moreover, it allows a rapid assessment of the presence of antibiotic resistance genes or gene mutations. An approach combining pathogen detection, their mechanisms of antibiotic resistance, their virulence factors and bacterial load in clinical samples could lead to profound changes in the care of these infected patients. Therefore, complex and multiplex PCR assays are currently being developed to enhance the field of molecular diagnostics.

[0006] However, PCR based diagnoses from mixed samples are often associated with challenges, as the presence of PCR products can arise from more than one DNA source in mixed complex samples. The presence of a specific antibiotic gene combined with the presence of specific bacterial strain does not necessarily mean that a resistant bacterial strain was present in the original sample. It only indicates that both a resistance gene and the bacterial strain were present in the sample, not necessarily that they arise from the same cell. Approaches to circumvent this problem have been suggested where the integration site of the antibiotic resistance gene was targeted by specific primers also known to target a specific bacterial strain. However, the exact integration site will then need to be known and this limits the use of the method.

DNA sequencing

[0007] Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as diagnostic, biotechnology, forensic biology, and biological systematics. The rapid speed of sequencing and decreasing cost attained with modern DNA sequencing technology has been instrumental in the sequencing of DNA sequences and the total amount of DNA sequenced worldwide is increasing rapidly.

[0008] While sequencing of pure samples is now standard procedure, sequencing of mixed samples of DNA is still challenging, costly and time consuming. When the target fragment is present at low frequency in a mixed nucleotide sample, for instance in a swab, faeces or blood sample, one must first make a clone or fragment library and then either sequence the complete library, or make smaller PCR fragments and sequence these. Sequencing of the complete library is expensive and time consuming and sequencing of PCR fragments is only possible if the sequence is already known to a very large extent and will return relatively short fragments which cannot be assigned to the same molecule or organism. If only part of the target sequence is known, PCR will not be possible and metagenome sequencing is needed.

[0009] There is therefore a need for a method for increasing the frequency of rare nucleotide molecules in a mixed sample of nucleotide molecules in order to decrease the cost and time of sequencing mixed samples.

Metagenomics

[0010] Metagenomics refers to a general sequencing approach, where a complex mixture of DNA is differentiated into small fractions and sequenced individually. The system does not rely on culturability and is, thus, applicable for both samples that can and cannot be cultivated in a laboratory.

[0011] Although metagenomics can be applied in some cases to describe complex samples, it

is frequently the case that researchers prefer a specific subset of the genome or mixture of genomes, rather than the entire sequence of a sample or mixed sample of genomes [2]. Thus, there exists a strong need for flexible targeting methods that are matched to the individual requirements of the various techniques and studies. Although such techniques do exist they are often based on hybridization assays, and have the prerequisite of extensive sequence knowledge or have low sensitivity.

[0012] Therefore, a strong demand is present, for specific enrichment of a pre-defined sub-fraction of a complex mixture, if metagenomics is to be applied for analyzing rare/low abundance DNA fragments or targets.

Discovery

[0013] Different industries have different motivations to explore the vast resource that lies within uncultivated microbial diversity. Currently, white (industrial) biotechnology seems to play a central role in the establishment of the sustainable modern society. It is a commonly accepted assumption that only a small sub-fraction of the natural microbial biodiversity is available for screening. In 1990, Torsvijk et al. [3] estimated that at best only 1% of the naturally occurring microbial diversity of a soil sample could be cultivated under laboratory conditions. Hence, a huge potential lies within the discovery of yet unknown natural products, and also in biotechnological techniques that enable access to the unknown majority of DNA present in natural and environmental samples. Unfortunately, enormous amounts of sequencing is required if the otherwise unavailable information is to be retrieved.

[0014] Targeted enrichment of DNA fragments encoding industrially relevant proteins or enzymes would substantially decrease the required amount of sequencing.

[0015] Lukyanov et al. [13] describes method for enriching a target nucleic acid in a complex DNA mixture to enable a subsequent sequencing step. WO 2004058987 [14] describes the sequential use of whole genome amplification steps to enrich target nucleic acids. Erlandsson et al. [15] describes the use of whole genome amplification steps followed by PCR.

SUMMARY OF THE INVENTION

[0016] The present invention provides *in vitro* method for enriching and/or isolating a target DNA molecule from a mixed polynucleotide sample comprising the steps:

1. a) providing a mixed polynucleotide sample containing said target DNA molecule, wherein said target DNA molecule comprises one or more unique consecutive sequence of at least 10 nucleotides,
2. b) serially diluting said mixed polynucleotide sample until the probability of detecting said target DNA molecule in a diluted sample is less than 0.75 preferably less than 0.50 or

even more preferably less than 0.25, and

3. c) replicating a sufficient number of the diluted sample until the probability of detecting said target DNA molecule in at least one of the replicate dilution samples is at least 0.75 preferably 0.80 to 0.95;

4. d) amplifying the DNA in said replicate diluted samples to increase the abundance of the DNA in each sample;

5. e) detecting the presence or absence of said target DNA molecule in said replicate dilution samples amplified in step (d), wherein the frequency of said target DNA molecule in said replicate dilution sample is increased compared to the mixed polynucleotide sample in step (a);

6. f) serially diluting at least one replicate dilution sample containing said DNA molecule until the probability of detecting said target DNA molecule in a diluted sample is less than 0.75, and repeating steps (c) to (e) or (f) at least once.

LEGENDS TO THE FIGURES

[0017]

Figure 1:
The figure shows a 2% agarose gel of nine negative PCR reactions from the initial mixture. The boxed area indicates where a positive PCR product would be present in the gel. The PCR products outside the box are unspecific products and are disregarded in this context.

Figure 2:
The figure shows a 2% agarose gel of ten PCR samples after the first round of PINS. Five Samples (A, B, C, D, & I) contain PCR products of the expected size whereas the remaining five samples do not. The correct product size is arrowed in the gel (left). The PCR products in other locations (not arrowed) are unspecific products and are disregarded in this context.

Figure 3:
The figure shows a 2% agarose gel of ten PCR samples following the re-amplification of sample #10. Six samples (A, C, E, F, I, & J) contain PCR products of the expected size whereas the remaining four samples do not. The correct product size is arrowed in the gel (left). The PCR products in other locations (not arrowed) are unspecific products and are disregarded in this context.

Figure 4:
The figure shows a 2% agarose gel of ten PCR samples after the second round of PINS using 2E-1 dilution prior to PCR. Five samples (C, D, E, G, & I) contain PCR products of the expected size whereas the remaining four samples do not. The correct product size is arrowed in the gel (left). The PCR products in other locations (not arrowed) are unspecific products and are disregarded in this context.

Figure 5:

The figure shows a schematic illustration of two rounds of PINS with the following parameters:

A: The initial sample - with 0,028 ng/µl.

B: 10 Phi samples, created from A, using 3.5 µl "A" as template.

C: One sample with all 3 (of three) PCR products were positive.

D: Sample "C" re-amplified by using 3,5 µl sample #10 in a total reaction volume of 50.

E: 10 Phi samples, created from "D" using 1.0µl as template in each reaction.

F & G: Two samples where all 3 (of three) PCR products were positive.

H: Pooling of samples "F" & "G"

I: Gain, calculated from initial concentration to the result after the first round of PINS.

J: Gain, calculated from first round of PINS to second round of PINS, using direct comparison at identical dilutions.

K: Gain, calculated from first round of PINS to second round of PINS, using comparison at different dilutions.

L: Gain, calculated from initial sample to finish (targets/ng).

Figure 6 The figure shows a 2% agarose gel of ten negative PCR reactions generated from an initial mixed polynucleotide sample comprising the target HPV18-DNA molecule. The boxed area indicates where a positive PCR product would be present in the gel. The PCR products outside the box are unspecific products and are disregarded in this context.

Figure 7 The figure shows a 2% agarose gel of the products of three PCR designed to detect the presence of the target HPV18-DNA molecule. Lane #1 is a positive control PCR reaction; lane #2 is the PCR product of the PINS (MDA-mediate) enriched mixed polynucleotide sample containing 2.5 target copies/µl; and lane #3 is a negative PCR control. The arrow indicates the relative mobility of a DNA molecule of -667 bp in the agarose gel.

Figure 8. The figure shows a 2% agarose gel of the PCR product obtained by re-amplification of the target HPV18-DNA molecule shown in Figure 7. The arrow indicates the relative mobility of a DNA molecule of -667 bp in the agarose gel

Figure 9 Alignment of the retrieved sequences for the PINS enriched target HPV18-DNA molecule with HPV18 sequence (HPU89349), showing that the sequence is perfectly aligned with zero mismatches.

Figure 10. Primers and partial alignments of the sequences of retrieved microbial genes encoding endoglucanase.

Figure 11. The figure illustrates the nucleotide sequence of a gene encoding endoglucanase

enriched by PINS. Restriction sites (*MboI, EcoRI,* and *Hind*III) used for RGW are inserted in boxes. -35 & -10 illustrates the position of the putative promoter location and SD is the position of the ShineDalgarno.

[0018] EndoGlu-Fw & EndoGlu-Re are the primers used during PINS. ORF indicates the position of the open reading frame.

DETAILED DESCRIPTION OF THE INVENTION

[0019] The present invention pertains to an *in vitro* method in which the frequency of the targeted nucleotide sequence containing the DNA fragment of interest is increased stepwise, by several rounds of 1) dilution of a sample containing the DNA fragment of interest in several replicates (separation), 2) randomly amplifying DNA in the replicates (concentration), 3) detecting the DNA fragment of interest in at least one of the diluted and amplified replicates (selection) and repeating steps 1) through 3) until the DNA fragment of interest can be sequenced by standard sequencing techniques.

[0020] The invention is based on the principle that if a selected DNA fragment of interest, present in a mixed sample, is located at a dilution where the probability of finding it is small, the frequency of the DNA fragment of interest will be higher at that dilution than in the mixed sample, and its frequency can be further increased by rounds of selection (as above), until it can be sequenced by standard methods such as Sanger sequencing or Pyro sequencing or similar detection of DNA sequence or by PCR, hybridization or other detection assays.

[0021] The method of the invention is surprisingly efficient, whereby the extent of screening can be reduced from screening of hundreds of thousands of clones in a clone library according to traditional screening methods [17], to screening less than a few hundred DNA samples (Table 1).

Table 1 Efficiency of PINS

Prevalence	PINS (no. of tests)	Traditional (no. of tests)	Screening Improvement
$1/10^4$	88	10,000	114x
$1/10^5$	108	100,000	909x
$1/10^6$	128	1,000,000	7,692x
$1/10^7$	148	10,000,000	66,667x
$1/10^8$	168	100,000,000	588,235x
$1/10^9$	188	1000,000,000	5,263,158x

[0022] Each cycle of PINS provides an estimated 10x increase in the prevalence of the genetic element of interest. The number of tests required to increase the prevalence from the starting level to the desired ending level can be calculated, i.e. 78 tests are required to increase the prevalence from $1/10^6$ to $1/10^2$ (based on estimated 10x/cycling).

I: PINS

[0023] The essential steps of the *in vitro* PINS method are further described below:

1. a) <u>A sample of mixed polynucleotides comprising a target DNA molecule</u>

   A sample of mixed polynucleotides known to comprise a target DNA molecule, is selected for performing PINS. One or more nucleotide sequences of at least 10 (or 15) unique nucleotides located within the target DNA molecule is selected for screening and detecting the DNA molecule by a desired method, such as PCR detection, DNA detection with hybridization probes or similar. Typically, the frequency of the target DNA molecule sample of mixed polynucleotides is less than 10^{-2}, it may for example lie between 10^{-3} and 10^{-7}.

2. b) <u>Primary dilution of the mixed polynucleotide sample</u>

   The mixed polynucleotide sample is serially diluted by a desired number of dilutions until the probability of detecting the target DNA molecule in a diluted sample is less than 0.75, (this sample is designated dilution (N)). Serial dilutions preferably have a dilution factor of greater than 1:1, preferably between 1:2 to 1:20, e.g., 1:10. Each diluted sample is placed in a separate container, such as a well in a microtiter plate, a plastic tube or similar.

   Various methods may be used to identify sample (N) in the dilution series in which the probability of detecting the target DNA molecule is less than 0.75. For example, the serial dilutions of the mixed polynucleotide sample may be amplified to increase the abundance of the DNA in each sample, followed by the step of detecting the presence or absence of said target DNA molecule in each amplified sample. The most dilute sample within the dilution series in which the target DNA molecule can be detected is designated "P". The next dilution in the dilution series, designated "N", is the least diluted sample in the dilution series where the target DNA molecule cannot be detected. The dilution factor of "N" relative to the DNA sample in a) is designated Dt, and the dilution factor between P and N is designated, D.

   Alternatively, the frequency and abundance of the target DNA molecule in the mixed polynucleotide sample may be determined by real time PCR, from which the dilution required to prepare sample (N) can be calculated (see example 1).

   Alternatively, the target DNA molecule may be detected by hybridization based assays or by assays detecting an RNA or protein product of the target sequence.

3. c) <u>Creating replicate dilution samples having dilution N</u>

   A sufficient number of replicates of the diluted sample (N) are generated until the probability of detecting the target DNA molecule in at least one of the replicate dilution

samples is one. A suitable number of replicates lies between 2 and 500, preferably at least 10 - 20 replicates. Typically, a suitable number of replicates will correspond to the dilution factor used to create the dilution series, i.e. if the dilution factor is 1:10, then about 10 replicates should be sufficient. Each diluted sample is placed in a separate container, such as a well in a microtiter plate, a plastic tube or similar.

4. d) <u>Genomic DNA amplification</u>

DNA in each replicated diluted sample is amplified using any method of total DNA amplification to increase the abundance of the DNA in each sample. Suitable amplification methods including Degenerate Oligonucleotide Primed PCR (DOP-PCR), Multiple Displacement Amplification (MDA)[4], randomly primed PCR or similar.

5. <u>e) Screening the replicate dilution sample (+/-) for target DNA</u>

The replicate samples (N) following genome amplification in step d) are screened for the presence of the target DNA molecule using the desired detection technique. In at least one or more screened samples that are shown to contain the target DNA molecule (sample$^+$), the frequency of the target DNA molecule will be increased compared to its frequency in the mixed polynucleotide sample in step (a).

One or more of these samples$^+$ are then serially diluted until the probability of detecting the target DNA molecule in a diluted sample is less than 0.75. Typically, the total dilution of this sample is Dt increased by factor D. Using this sample the steps (c) to (e) are repeated at least once, preferably until the target DNA molecule has been amplified to an extent where the nucleotide sequence of the target DNA molecule or parts thereof can be readily sequenced, as described below in f). The number of repetitions of steps (c) to (e) is generally at least 1, but is more likely to require 2, 3, 4, 5 or 6 or more repetitions, whereby the eventual frequency reached of the target DNA molecule is greater than 10^{-3}, preferably greater than 10^{-1}.

6. f) <u>Characterization of the enriched target DNA molecule</u>.

Once the frequency of the target DNA molecule is sufficiently increased, the DNA can be subjected to direct DNA sequencing. Sequencing of the target DNA molecule; both the PCR amplified target DNA fragment used for detection and the nucleotide sequences flanking the 5' and 3' direction of the target DNA fragment, is implemented using the desired DNA sequencing method, such as Sanger sequencing, pyro sequencing or similar methods.

[0024] An example of how the PINS method may be implemented is set out in Table 2.

Scheme for performing PINS

Original DNA sample designated Sample N0

[0025] Make 10-fold dilutions of Sample N0 in wells A1 - A8 of a microtiter plate. Perform MDA (optionally): N0 -> N0^{MDA}. Transfer aliquots of N0 -> N0^{MDA} to PCR plate and perform PCR to detect target DNA molecule.

Table 2

Dilution plate	PCR	PCR Plate
N0 x 10-0^{A1}		P
N0 x 10-1^{A2}	->	P
N0 x 10-2^{A3}		P

Dilution plate	PCR	PCR plate
N0^{MDA} x 10-0^{A1}		P
N0^{MDA} x 10-1^{A2}	->	P
N0^{MDA} x 10-2^{A3}		P

N0 x 10-3^{A4}		P
N0 x 10-4^{A5}		N
N0 x 10-5^{A6}		N
N0 x 10-6^{A7}		N
N0 x 10$^{-7\,A8}$		N

N0^{MDA} x 10-3^{A4}		P
N0^{MDA} x 10-4^{A5}		N
N0^{MDA} x 10-5^{A6}		N
N0^{MDA} x 10-6^{A7}		N
N0^{MDA} x 10$^{-7\,A8}$		N

Select N0 x 10-4 ^{A5} or N0^{MDA} x 10-4 ^{A5} = Sample N1 Make 20 replicate aliquots of sample N1 and add into A1 - C4 on dilution Plate (Table 3). Perform MDA on all replicates to amplify total DNA in each replicate, followed by PCR to detect the target DNA molecule in the amplified DNA.

Table 3

Cycle 1											
Dilution plate			**MDA**	**MDA plate**			**PCR**	**PCR plate**			
N1^{A1}	N1^{B1}	N1^{C1}		N1^{A1+MDA}	N1^{B1+MDA}	N1^{C1+MDA}		N	N	N	
N1^{A2}	N1^{B2}	N1^{C2}		N1^{A2+MDA}	N1^{B2+MDA}	N1^{C2+MDA}		N	N	P	
N1^{A3}	N1^{B3}	N1^{C3}		N1^{A3+MDA}	N1^{B3+MDA}	N1^{C3+MDA}		P	N	N	
N1^{A4}	N1^{B4}	N1^{C4}	->	N1^{A4+MDA}	N1^{B4+MDA}	N1^{C4+MDA}	->	N	N	N	
N1^{A5}	N1^{B5}			N1^{A5+MDA}	N1^{B5+MDA}			N	N		
N1^{A6}	N1^{B6}			N1^{A6+MDA}	N1^{B6+MDA}			N	P		
N1^{A7}	N1^{B7}			N1^{A7+MDA}	N1^{B7+MDA}			N	N		
N1^{A8}	N1^{B8}			N1^{A8+MDA}	N1^{B8+MDA}			N	N		
Select (N$^{1A3 + MDA}$ + N1$^{B6 + MDA}$ + N1$^{C2 + MDA}$) and dilute 10^{-5} = Sample N2											

[0026] Make 20 aliquots of sample N2 and add into A1 - C4 on dilution plate (Table 4). Perform MDA on all replicates to amplify total DNA in each replicate, followed by PCR to detect the target DNA molecule in the amplified DNA.

Table 4

Cycle 2											
Dilution plate			**MDA**	**MDR plate**			**PCR**	**PCR plate**			
N2^{A1}	N2^{B1}	N2^{C1}		N2^{A1+MDA}	N2^{B1+MDA}	N2^{C1+MDA}		N	N	N	
N2^{A2}	N2^{B2}	N2^{C2}		N2^{A2+MDA}	N2^{B2+MDA}	N2^{C2+MDA}		N	P	N	

Cycle 2										
Dilution plate			MDA	MDR plate			PCR	PCR plate		
$N2^{A3}$	$N2^{B3}$	$N2^{C3}$		$N2^{A3}$+MDA	$N2^{B3}$+MDA	$N2^{C3}$+MDA		N	N	N
$N2^{A4}$	$N2^{B4}$	$N2^{C4}$	->	$N2^{A4}$+MDA	$N2^{B4}$+MDA	$N2^{C4}$+MDA	->	N	N	N
$N2^{A5}$	$N2^{B5}$			$N2^{A5}$+MDA	$N2^{B5}$+MDA			P	N	
$N2^{A6}$	$N2^{B6}$			$N2^{A6}$+MDA	$N2^{B6}$+MDA			N	N	
$N2^{A7}$	$N2^{B7}$			$N2^{A7}$+MDA	$N2^{B7}$+MDA			N	N	
$N2^{A8}$	$N2^{B8}$			$N2^{A8}$+MDA	$N2^{B8}$+MDA			N	N	
Select ($N2^{A5}$ + MDA + $N2^{B2}$ + MDA) and dilute 10^{-6} = Sample N3										

[0027] Make 20 aliquots of Sample N2 and add into A1 - C4 on dilution plate (Table 5). Perform MDA on all replicates to amplify total DNA in each replicate, followed by PCR to detect the target DNA molecule in the amplified DNA.

Table 5

Cycle 3										
Dilution plate			MDA	MDA plate			PCR	PCR plate		
$N3^{A1}$	$N3^{B1}$	$N3^{C1}$		$N3^{A1}$+MDA	$N3^{B1}$+MDA	$N3^{C1}$+MDA		N	N	P
$N3^{A2}$	$N3^{B2}$	$N3^{C2}$		$N3^{A2}$+MDA	$N3^{B2}$+MDA	$N3^{C2}$+MDA		N	N	N
$N3^{A3}$	$N3^{B3}$	$N3^{C3}$		$N3^{A3}$+MDA	$N3^{B3}$+MDA	$N3^{C3}$+MDA		N	P	N
$N3^{A4}$	$N3^{B4}$	$N3^{C4}$	->	$N3^{A4}$+MDA	$N3^{B4}$+MDA	$N3^{C4}$+MDA	->	N	N	N
$N3^{A5}$	$N3^{B5}$			$N3^{A5}$+MDA	$N3^{B5}$+MDA			N	P	
$N3^{A6}$	$N3^{B6}$			$N3^{A6}$+MDA	$N3^{B6}$+MDA			N	N	
$N3^{A7}$	$N3^{B7}$			$N3^{A7}$+MDA	$N3^{B7}$+MDA			N	N	
$N31^{A8}$	$N3^{B8}$			$N3^{A8}$+MDA	$N3^{B8}$+MDA			N	N	

[0028] Select ($N3^{B3}$ + MDA + $N3^{B5}$ + MDA + $N3^{C1}$ + MDA) and determine the abundance of the target DNA molecule. If this is sufficient for analysis such as sequencing, the selected sample may be used directly; otherwise additional cycles of PINS can be applied.

II: Multiplex PINS

[0029] PINS can be adapted to perform multiplex PINS. Multiplex PINS employs additional features that are designed to detect a 2nd consecutive sequence of at least 10 (or 15)

nucleotides in the sample of mixed polynucleotides analysed, by amplification of this 2nd consecutive sequence with sequence specific primers to generate a 2nd target DNA fragment. If the 1st and 2nd consecutive sequence co-purify, at each cycle of PINS, then they must be located on the same MDA-amplified target DNA molecule.

III Samples analyzed by PINS and Multiplex PINS

III.i Sample of mixed polynucleotides

[0030] PINS may be applied to a sample of mixed polynucleotides known to comprise a target DNA molecule. A sample of mixed polynucleotides comprises a population of DNA molecules (e.g. chromosomal DNA molecules or plasmid DNA molecules) where the individual DNA molecules within the population differ with respect to a known consecutive sequence of at least 10 (or 15) nucleic acid base pairs in their DNA, such that a target molecule comprising the known consecutive sequence differs from and can be distinguished from non-target molecules in the sample. The sample of mixed polynucleotides may additionally comprise single stranded RNA or DNA polynucleotides. The population of DNA molecules in the sample of mixed polynucleotides comprises the target DNA molecule.

[0031] A target DNA molecule comprises one or more known consecutive sequence of at least 10 (or 15) unique nucleic acid base pairs (or nucleotides). A target DNA molecule can be selected from a sample of mixed polynucleotides, by selecting for a target DNA molecule comprising this consecutive sequence of at least 10 nucleic acid base pairs (or nucleotides). The target DNA molecule can also be selected from the sample of mixed polynucleotides, by selecting for a target DNA molecule comprising at least two consecutive sequences of at least 10 (or 15) nucleic acid base pairs (or nucleotides), wherein the two consecutive sequences are comprised within a DNA molecule of 50 to 100,000 nucleic acid base pairs, preferably 150 to 3,000 nucleic acid base pairs, more preferably 150 to 1500 nucleic acid base pairs.

[0032] The method of the invention is particularly suitable where the frequency of the target DNA molecule in the sample of mixed polynucleotides is less than 10^{-3}. The method of the invention is also suitable where the frequency of the target DNA molecule in the sample of mixed polynucleotides is 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, or lower. In many instances, the sample of mixed polynucleotides will be derived from a cell population comprising genomic DNA, while in other instances, the sample may be derived from samples where the polynucleotides are of diverse origin, such as samples collected from nature. Irrespective of its source, the frequency of the target DNA molecule is defined as the number of genomes or genome equivalents containing the target DNA divided by the number of total genome equivalents. The frequency of the target DNA molecule in the sample of mixed polynucleotides is determined by making a dilution series in triplicate, detecting the presence or absence of target and determining the number of targets using, for instance, most probable number methods. The concentration of

DNA is measured and the number of total genome equivalents is determined by dividing this concentration by the average molecular weight of the genome.

III.ii Source of the sample of mixed polynucleotides

[0033] According to one embodiment of the present invention, the target DNA molecule is derived from the genome of a cell, where the genome may be either chromosomal or extrachromosomal DNA. Further, the target DNA molecule may be derived from a cell, where the cell is selected from amongst a microbial cell, a plant cell, an animal cell, or a mammalian cell. The mammalian cell may be a human cell. The microbial cell may be a bacterial cell, a yeast cell or a fungal cell.
Furthermore, the target DNA molecule may be derived from a fungal mycelium or fungal spores.

[0034] When the target DNA molecule is derived from one or more cell, the cell(s) may be part of a multicellular tissue or multicellular organism.

[0035] Furthermore, the target DNA molecule may be derived from one or more viral particles, where the virus has an RNA or DNA genome. Alternatively the target DNA molecule may be derived from a host genome comprising integrated DNA derived from a virus. The target DNA molecule may also be derived from a bacteriophage.

[0036] Irrespective of the derivation of the target DNA or RNA molecule, the target DNA or RNA molecule is present in a sample of mixed polynucleotides, where the mixed polynucleotides may be derived from a sample collected from nature, for example a sample of soil, water or air. Alternatively, the sample may be derived from a multicellular organism, such as a mammal, for example an animal or a human subject. When the sample is derived from a mammal, the sample (for example a biopsy) may be derived from a body fluid (e.g. blood, plasma, serum, lymph and urine), from faeces or from a body tissue or organ. The multicellular organism from which the sample is derived may be a living or may be a dead organism.

III.iii Preparation of the sample of mixed polynucleotides

[0037] The sample of mixed polynucleotides comprising the target DNA molecule may be prepared from a sample collected from nature or from an organism (e.g. a biopsy). Methods for selective extraction of polynucleotides comprising DNA or RNA are known in the art [5]. When the target DNA molecule is derived from a cell, the step of cell disruption or cell permeabilisation is normally required in order to release total nucleic acid molecules (including DNA or RNA) from a cell, this step preceeding the subsequent step of selective extraction of polynucleotides comprising DNA or RNA.

[0038] Where the target DNA molecule is derived from an RNA genome, the RNA genome or parts thereof are first reverse transcribed to provide a cDNA molecule, where the nucleotide sequence of the cDNA corresponds to (is a reverse transcript of) the RNA genome.

IV Methods of Random Amplification of DNA suitable for PINS

[0039] A range of different approaches have been suggested for general amplification of DNA, such as randomly degenerate primed PCR, linker ligation PCR, or, Degenerate Oligonucleotide Primed (DOP) PCR and Multiple Displacement Amplification (MDA). MDA has proven efficient in performing whole-genome amplification (WGA) of even very small amounts of DNA [6]. Compared with more traditional PCR-based WGA methods, MDA generates DNA molecules with a higher molecular weight, having better genome coverage. MDA employs a strand displacement polymerase that possesses two enzymatic activities: DNA synthesis (polymerase) and an exonucleolytic activity that degrades single stranded DNA in the 3'- to 5'-direction, as exemplified by bacteriophage phi29 DNA polymerase, that belongs to eukaryotic B-type DNA polymerases (UniProtKB/TrEMBL: Q38545). Other useful polymerases include BstI polymerase.

V Sequence determination of the target DNA molecule

[0040] Isolation of the target DNA molecule by PINS is based on detection of one or more unique consecutive sequence of at least 15 nucleotides in said DNA molecule. When detection is based on PCR, where the one or more unique consecutive sequences are amplified to generate a target DNA fragment, the nucleotide sequence of this fragment can be determined. In addition, the nucleotide sequences flanking the target DNA fragment in the 5' and 3' direction can be determined by rapid genome walking (RGW)[7]. RGW is a simple, PCR-based method for determining sequences upstream or downstream in a larger DNA molecule starting from a known sequence, such as a target DNA fragment. RGW enables individual amplification of up to 6 kb in a large DNA molecule using PCR. The sequences can be extended simply by taking multiple cycles of RGW, using new primers based on the sequence obtained in previous cycles. Typically libraries are constructed from a purified sample of the large target DNA molecule, by digesting the DNA separately with four different restriction enzymes and ligating the products to a specially designed adaptor. The ligated DNA is then sequenced with primers annealing to the adaptors or to known sequences within the DNA, using the desired DNA sequencing method, such as Sanger sequencing, pyro sequencing, sequencing by synthesis, ligation or two base-coding sequencing or similar methods [8].

VI Application of PINS and multiplex pins

VI.i Research and development applications of PINS and Multiplex PINS

[0041] Use of PINS to isolate or enrich target DNA in mixed polynucleotide samples extracted from samples collected from nature, provides direct access to the genome, or parts thereof, of microorganisms that exist in nature and that frequently cannot be obtained as axenic cultures. PINS is particularly useful for isolating or enriching DNA encoding enzymes involved in producing chemical building blocks or active agents such as:

- acids (such as maleic-, aspartic-, malonic-, propionic-, succinic-, fumaric-, citric-, acetic-, glutamic-, itaconic-, levulinic-, acotinic-, glucaric-, gluconic-, and lactic-acid),
- amino acids (such as serine, lysine, threonine),
- alcohols (ethanol, butanol, propanediol, butanediol, arabitol) and
- other high value products (such as acetoin, furfural, and levoglucosan)
- antibiotics, anticancer compounds (e.g. peptide-polyketides; lactam analogues)

[0042] Thus PINS is particularly useful for isolation or enrichment of target DNA encoding enzymes selected among an oxidoreductase, transferase, hydrolase, lyase, isomerase or ligase. Multiplex PINS is particularly useful for isolation or enrichment of target DNA comprising an operon of two or more open reading frames encoding two or more enzymes, where the enzymes may be part of a multifunctional enzyme complex, as the technique can take advantage of high fidelity of polymerases (like Phi29) in the MDA amplification to generate large amplified DNA fragments of up to 100000 bp.

VI.ii Diagnostic applications of PINS and Multiplex PINS

[0043] PINS or Multiplex PINS may be used to analyse target DNA in samples derived from multicellular organisms, such as biopsy or sample of body fluid or faeces obtained from a subject (e.g. human or animal subject), for the diagnosis or monitoring the progress of a medical indication or disease.

[0044] Diagnosis of a wide range of medical indications in a subject such as a disease caused by an infectious agent (e.g. micro-organism or virus) can be assisted by the isolation or enrichment and detection of a target DNA or RNA molecule that is derived from the genome of the infectious agent by PINS or Multiplex PINS, where the target DNA molecule is detected in a sample of mixed polynucleotides derived from a biopsy or sample of body fluid obtained from a patient.

[0045] Use of multiplex PINS in target DNA molecule isolation provides the additional feature, that additional diagnostic features of the disease can be determined. For example, where the genome of the infectious agent comprises resistance genes that confer resistance to certain therapeutic agents, co-enrichment of a resistance gene and a sequence describing the

infectious agent in the PINS procedure will indicate that the infectious agent is carrying the resistance gene.

[0046] PINS or Multiplex PINS may also be used to assist diagnosis of a disease caused by, or originating from the presence of a viral agent in a subject. The presence and/or chromosomal integration of DNA that is derived from a viral genome, in a biopsy or sample of body fluid obtained from a patient can be detected by PINS by sequencing the entire DNA fragment comprising the viral DNA sequence. Using multiplex PINS the presence or absence of viral DNA at a known integration site can be determined by tracking the co-enrichment of the viral DNA and integration site DNA sequence.

VII PINS and bias in amplification

[0047] PINS is based on specifically selecting samples where amplification of a desired DNA region from a complex, mixed DNA sample has occurred. Although Phi29 based amplification (MDA) has been described repeatedly as the most reliable genome amplification currently available, it is known to introduce significant bias. Pan et al. [18] states in general terms that a highly specific whole genome amplification (WGA) of complex DNA pools which avoids amplification bias remains a challenge. Moreover, similar observations are seen with alternative amplification methods such as DOP-PCR and random priming PCR. These two amplification methods are described as being much less efficient at reproducing the locus representation [19], resulting in even more biased amplification products. While bias is seemingly unavoidable regardless of the amount of reaction template [20] present, the amount of template independent product (TIP) or bias introduced during amplification is seemingly correlated negatively to the amount of DNA template in the reaction and has in some studies been documented to represent 70-75% of the total yield [18]. Whole genome amplification is applied to amplify DNA in the PINS process and it would therefore be expected that these general challenges relating to bias in genome amplifications would also apply to PINS. As a procedure including multiple steps of WGA it would thus be expected that significant bias against the target DNA molecule should be observed. Hence a procedure such as PINS employing several steps of WGA would not have been considered as a method capable of enriching for a specific region of DNA in a mixed sample.

[0048] Surprisingly, the challenge of TIP/bias is not seen when applying the PINS technique even though the initial concentration of the target DNA molecule is very low. The minimal negative TIP/bias observed in the PINS system is significantly lower than the overall gain obtained from the amplification process and the net result is therefore a substantial enrichment of the target DNA molecule with each cycle.

[0049] The method does not require a dilute initial DNA template, as high concentrations of DNA also can efficiently be subjected to PINS. In this case only a few fold amplification by the WGA method is needed.

EXAMPLES

[0050] The following examples illustrate why PINS is superior or complementary to other methods, such as shot-gun cloning and metagenomics, for the enrichment and/or isolation of a target DNA molecule. Employing traditional metagenomics, instead of PINS used in example 1, would require the sequencing of $2*10^6$ genomes with an average size of $3*10^9$ bases. Thus, a total of at least $6*10^{15}$ base pairs would be required in order to obtain the sequence of the HPV18. After two rounds of PINS the frequency of the HPV18 target was increased by a factor of 79 and therefore only $4.47*10^{13}$ base pairs would now need to be sequenced. Additional rounds of PINS could further decrease the required amount of sequencing.

Materials and general methods

[0051] The following materials and methods were applied in the examples below:

Enzymes and reagents: If not stated otherwise enzymes were supplied by MBI Fermentas (Germany) and used according to the suppliers recommendations.

Amplification and detection methods: If not stated otherwise, the following procedures and conditions were employed for DNA fragment amplification by PCR; genome amplification by MDA; and extension of DNA fragment into flanking regions using RGW. The target DNA fragment was detected by 2% agarose gel electrophoresis.

PCR amplification:

[0052] PCR reactions were: carried out in 20μl reactions by mixing: nuclease free water (7 μl), 10 μl 2x SSO Advance Sybr TR-Mixture (BioRad), 5 μM fw-primer (1 μl), 6 μM Re-primer (1 μl), with 1μl DNA template. The composition of the 2x SSO Advance Sybr TR-Mixture is: dNTPs, Sso7d fusion polymerase, MgCl2, SYBR® Green I, and stabilizers.

[0053] PCR amplification conditions: were $(15sec/15sec/15sec)_{x25}$ at temperatures (94°C/65.8°C /72°C). The PCR reactions were performed in a BioRad Connect RT-PCR machine.

MDA amplification:

[0054] MDA was carried out according to the protocol described below:

Mix I: Denaturing & Re-annealing:

2,5µl 2x Annealing buffer (33 mM Tris-acetate (pH 7.9 at 37°C), 10 mM Mg-acetate, 66 mM K-acetate, 0.1 mg/mL BSA)

0,5µl Water

1µl Exo-resistant random hexamer Primers (Thermo Scientific)

(1µl DNA Template)*

Mix II - Amplification

7,5µl Water

4µl 2mM dNTP

2µl Phi29 buffer (Thermo Scientific)

1µl Pyrophosphotase 0,01U/µl (Thermo Scientific)

0,5µl Phi29 polymerase (Thermo Scientific)

Protocol

[0055]

1. 1. Mix 1 was initially prepared without the addition of DNA template* and 4 µl mix 1 was added to each micro centrifuge tube. DNA template was then added to the mix by placing 1 µl inside the lid of the microcentrifuge tube. The tubes were centrifuged, and then transferred to ice immediately after centrifugation.
2. 2. Mix 1 (from step 1) was then transferred from ice to 94°C (pre-warmed PCR machine) and kept at 94°C for three minutes and then transferred back to ice.
3. 3. 15 µl from mix 2 was then added to mix 1 (from step 2).
4. 4. The final mixture (mix1 + mix2) was incubated at 30°C for 16 hours (PCR machine). At the end of the incubation, the temperature of the final mixture was elevated to 65°C for 10 minutes to inactivate the polymerase.

DNA quantification:

[0056] DNA was quantified using Quantus™ fluorometer (Promega) as described by the

manufacturer. *Gel electrophoresis:* PCR results were evaluated on the basis of 2% agarose gel electrophoresis using BioRad SubCell Equipment. Gels were run at 80V in 1% agarose for 20-30 minutes. Visualization was done by casting Ethidium Bromide (0.5µg/ml) into the gels.

[0057] MDA amplification was verified by loading 2µl onto a 0.7% agarose gel. All MDA gels were run at 80V in 0.7% agarose for approximately 30 minutes.

[0058] The molecular marker on all agarose gels was Fermentas 1bk+ ladder, with the band sizes (75, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 7000, 10.000, 20.000).

Calculation of probability

[0059] The probability of detecting a target DNA molecule in a sample was calculated using hypergeometric distribution, where the population is the originalsample volume of liquid in which the target is either present (positive) or not present and the test sample is the volume of liquid withdrawn from the original sample. Because the nucleic acid molecules only contribute to a minor part of the volume, both the population volume and test sample volumes are multiplied by 1000 in the calculations.

[0060] For a 100 µl sample, the population size is 100000. If 100 µl sample contains four target DNA molecules, the probability of detecting a target DNA molecule in a diluted sample containing 3.5 µl (test sample = 3500) of the original 100 µl sample is:

P = 1-Hypgeom.dist(0;3500;4;100000) = 0.133

[0061] If 10 replicate dilution test samples of each 3.5 µl are analyzed from the same original sample, the probability of detecting the target DNA molecule in at least one of the replicate dilution samples is:

P = 1-Hypgeom.dist(0;35000;4;100000) = 0.821

[0062] The equation for the hypergeometric distribution is:

$$P(X=x) = h(x,n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

where:

x = number of positives in the (diluted) test sample

n = the volume of (diluted) test sample * 1000

M = number of target DNA molecules in the original sample (population)

N = the volume of the original sample (population) * 1000

Example 1: Method of enrichment of a polynucleotide encoding HPV-18 virus from a mixed sample of human DNA.

1.0 Creating the mixture for enrichment.

[0063] A mixed polynucleotide sample was prepared from reference DNA that was spiked with DNA derived from HeLa cells. The HeLa DNA was obtained from New England Biolabs (100µg/ml) as a purified DNA sample. The reference DNA was extracted from cells derived from a human volunteer. The target DNA molecule in the HeLa DNA was efficiently amplified by the selected target PCR primers [SEQ ID NO: 1 and SEQ ID NO: 2], whereas the human reference DNA did not yield an amplification product with this primer pair. Both types of DNA were amplified separately using the MDA protocol prior to setting up the mixture. Following the MDA amplification, the quantity of target DNA copies in MDA amplified HeLa DNA was determined to be 71750 copies/µl using most probable number (MPN) calculation on triplicate 10-fold dilutions of HeLa DNA. The target DNA was diluted by a factor of 1:2.500.000 in undiluted MDA amplified reference DNA (employing serial dilution) to create the final spiked mixture containing 0.029 targets copies/µl and having a total DNA concentration of 0.187 µg/µl. Thus, the initial quantity of target DNA copies in the initial mixed polynucleotide sample was calculated to be: (0.029 target copies/µl):(0.187µg/µl) = 0.153 target copies/µg DNA. The total volume of mixed DNA sample was 45 µl. The probability of detecting the target DNA molecule in 3.5 µl mixed polynucleotide sample is 0.078 (1-hypgeo(0;3500;1;45000)); i.e. less than 0.75. 3.5 µl is the volume of sample used in the first round of PINS.

1.1 Defining the primers for selection of the target DNA molecule encoding HPV18:

[0064] DNA sequence from HPV18 was obtained from GenBank (GQ180790), and the following 2 primers were designed manually to target a DNA fragment of 96bp in the HPV18 genome (GenBank Acc. No: GQ180790):

Fw primer: GTTTAGTGTGGGCCTGTGC [SEQ ID NO: 1]

Rev primer: GGCATGGGAACTTTCAGTGT [SEQ ID NO: 2]

1.2 Initial mixture - below limit of detection:

[0065] PCR analysis was carried out to verify that the number of target DNA molecules was below the limit of PCR detection in the initial polynucleotide mixture. Nine polynucleotide mixture samples (each with 1µl of initial mixture) were used as DNA template for PCR amplification with the primer pair [SEQ ID NO: 1 and 2], but no PCR products of the expected 96 bp were found in any of the samples. An agarose gel showing the individual negative PCR samples from the initial mixture can be found in Figure 1.

1.3 First round of PINS

[0066] Ten individual 3.5 µl samples of the mixed polynucleotide sample, prepared in step 1, were chosen for MDA amplification in a volume of 20 µl, as described above, where the probability of detecting the target DNA molecule in 3.5 µl mixed polynucleotide sample prior to amplification is 0.078 (1-hypgeo(0;3500;1;45000)); i.e. less than 0.75. 1 µl triplicate aliquots of each MDA amplified sample (#1 - #10) were screened for the presence of the target DNA molecule by PCR amplification, as described above using the target-specific PCR primer pair in 1.1., followed by agarose gel electrophoresis to visualize PCR products (Figure 2). Because ten replicate samples were analysed, the probability of detecting the target DNA molecule in at least one of the ten replicate dilution samples was 0.778 (1-hypgeo(0;35000;1;45000)) i.e. higher than 0.75. #10 was the only sample found to result in positive PCR products of the correct size in all three replicates. Ten additional samples (each with 1µl template) from #10 were analysed by PCR using identical conditions, and 5 (out of 10) resulted in a correct PCR product. By combining the results from the first screening 3 (of 3) and the following screening, where 5 (of 10) were positive, a total of 8 (of 13) were observed as positive, all originating from #10. The concentration of total DNA in sample #10 was quantified to be 0.222 µg/µl. An agarose gel showing 10 PCR products from the second screening can be found in Figure 2.

[0067] The abundance of the target DNA molecule after MDA amplification was calculated to: (8/13 target DNA copies/µl):(0.208 µg/µl) = 2.96 target DNA copies/µg. Consequently, the increase in frequency obtained by PINS in the first round of amplification is: (2.96 targets/µg): (0.153 targets/µg) = 19.36 fold.

Re-amplification of #10

[0068] To generate sufficient amount of template prior to progressing to the 2nd round of PINS, 3.5 µl template from #10 was amplified by using a 50 µl MDA reaction mixture. The result of re-amplification of #10 was verified by analysing ten 1 µl aliquots by PCR where, in 6 of 10 PCR reactions, the target DNA molecule was detected (0.60 target DNA copies/ µl). This

result corresponds well to the initial frequency of the target DNA molecule in #10 which was 8/13 (0.62). By dividing with frequency of the target copies/µl sample by the total DNA concentration, the relative abundance of the target DNA molecule is calculated to be: (0.6 targets/µl):(0.222 µg/µl) = 2.70 targets/µg.

[0069] Surprisingly, an amplification bias had apparently only occurred to a limited degree during re-amplification. An agarose gel showing the ten PCR reactions products from the re-amplification can be found in Figure 3.

1.4 Second round of PINS

[0070] Ten 1µl aliquots of sample the re-amplification of #10, derived from 1st round of PINS, were diluted 20 fold in the MDA reaction mixture thereby diluting the target DNA molecule to a frequency of 0.03 target DNA molecules/µl. The probability of detecting the target DNA molecule (HPV18) in a 1 µl aliquot of the diluted sample was 1-hypgeo(0;1000;30;50000) = 0.455 i.e. below 0.75. Each aliquot was amplified using the MDA protocol described above. Out of the ten MDA reactions, the target DNA molecule was detected in two reactions samples, #8 and #10, where 3 (of 3) replicated PCR reactions were positive for the target. The probability of detecting the target in at least one of the ten replicate samples was 1-hypgeo(0;10000;30;50000) = 0.9988 i.e. above 0.75.

[0071] Samples #8 and #10 were pooled in one sample and diluted 2*10^{-1} (5-fold). Analysis of 1 µl aliquots of this dilution by PCR, detected the target DNA molecule in 5 (of 10) PCR reactions, as seen in the gel in (Figure 4). Since the pooled sample was diluted 2*10^{-1} (5-fold), the quantity of targets is calculated to be: 5 targets /10 µl test sample volume * 5 = 2.5 targets/ µl. Since the total DNA content of the pooled sample was measured to 0.207 µg/µl, the frequency of the target DNA molecule is calculated to 2.5 targets/µl divided by 0.207 µg total DNA/µl; which corresponds to 12.08 targets/µg DNA.

[0072] An overview of the protocol followed through the two rounds of PINS is set out in Figure 5.

1.5 Calculated enrichment for target molecule DNA by two rounds of PINS

[0073] By calculating from the initial mixture (0.153 targets/µg) to the final sample (12.08 targets/µg) a total gain of 79.05 has been reached after two consecutive rounds of PINS amplification.

[0074] The abundance of the HPV18 target before first round of PINS was 1 / 1.97*10^6 HPV18 positive genome equivalent / total genome equivalents. The corresponding abundance after

first and second round of PINS was $1.02*10^5$ and $2.49*10^4$, respectively. This enrichment was achieved using 21 tests and the requirement for required number of traditional tests such as sequencing was decreased by a factor of 79 as can be seen in Table 1.

Example 2

[0075] The use of PINS to facilitate the sequencing of a target DNA molecule present in a mixed polynucleotide sample, where the amount of target DNA is below the limit of detection by PCR is demonstrated in this example. The selected target DNA molecule was a polynucleotide located within HPV18 genome. The amount of this target DNA molecule in the mixed sample corresponded to 0.0286 copies/µl. The mixed sample was prepared as described in Example 1. Using standard PCR protocols, the amount of template solution added to a PCR reaction is normally 1µl, and the minimum target DNA concentration required for its positive detection by PCR is 1 copy/µl. Accordingly, the amount of target DNA molecule in the mixed polynucleotide sample was approximately 35 times below that required for PCR amplification and subsequent sequencing. The example demonstrates how PINS can be used to determine the sequence of target DNA in a mixed sample, which has otherwise been unattainable by known standard PCR based procedures.

2.1 The target HPV18 DNA molecule in the mixed polynucleotide sample is not detectable by PCR

[0076] In order to verify that the number of copies of target DNA molecule per µl of the mixed polynucleotide sample (corresponding the mixed polynucleotide sample of Example 1, prior to MDA amplification) was below the detection limit for PCR, a series of ten duplicate samples were PCR amplified (according to protocol given in Material and General Methods), with the modification of using 50 cycles and using the optimized amplification parameters: 94°C/65.8°/72°C, where each temperature was maintained for 15 seconds. HPV-specific primers used during amplification were:

HPV-264f: 5'-GTGGTGTATAGAGACAGTATACC-3' [SEQ ID NO: 3]

HPV-911f: 5'- CCTTCTGGATCAGCCATTGT [SEQ ID NO: 4]

[0077] Positive amplification of the target DNA molecule in the HPV18 genome would generate a PCR product of 667 bp (Boxed in Figure 6). As expected, due to the low abundance none (of ten) reactions resulted in amplification of the correct size, verifying the low quantity of target DNA molecule in the mixed polynucleotide sample (Figure 6).

2.2 PINS facilitates enrichment of the target HPV18 DNA molecule in the mixed polynucleotide sample to allow its PCR detection

[0078] The initial copy number of the target DNA molecule in the mixed polynucleotide sample (0.0286 target copies/μl) was increased to a final quantity of 2.5 target copies/μl by the implementation of PINS as described in Example 1. The MDA amplification products, generated by performing 1st round PINS on the initial mixed polynucleotide sample, were analysed by PCR as described above using 1μl of the MDA amplified sample. A PCR product of the expected size (667 bp) was obtained as shown in Figure 7. Low levels of non-specific amplification of the initial mixed polynucleotide sample generated a low background detectable as a smear in lane #2 of Figure 7. The detectable product of ~667 bp in lane #2 was isolated by gel-out (GeneJet Gel-out kit, Thermo Fischer Scientific) and PCR re-amplified to increase the amount of the target DNA molecule to allow subsequence sequencing (lane #A, Figure 8).

2.3 Verification of the target HPV18 DNA molecule obtained by PINS

[0079] The purified product isolated from lane #A (Figure 8) was sequenced forward and reverse using both primers [SEQ ID NO: 3 and 4]. The retrieved sequences were aligned to the sequence of HPV18 (HPU89349) resulting in a perfect alignment with zero mismatches (Figure9). The example confirms the efficacy of PINS for enriching for rare sequences to a level where their sequence can be determined.

Example 3: Method of isolating a nucleotide fragment encoding an endoglucanase enzyme from a soil sample

3.1 Defining the primers for selection of the target DNA molecule encoding an endoglucanase:

[0080] Multiple microbial genes encoding endoglucanases and putative endoglucanases are retrieved from GenBank and their nucleotide sequences are aligned using ClustalW alignment for multiple sequences[9]. From the alignment, regions of conserved nucleic acid sequence are identified and a set of primers is designed to target a DNA molecule having one of these conserved DNA sequences.

[0081] Primers and partial alignments of the sequences of retrieved genes encoding endoglucanase are shown in Figure 10.

[0082] The selected primers are estimated to result in a PCR amplification product of 481bp. However, as the origin and sequence of the templates in a natural sample are unknown, the sizes of PCR products may deviate from the expected size.

3.2 Extraction of the DNA from the soil sample:

[0083] Samples (8) of forest soils are collected as a source of diverse soil microorganisms. DNA from the 8 different samples are extracted using Bead Beating, as described by Kvist et. al [10]. Each extraction is amplified using MDA Repli-g Mini kit, using MDA amplification conditions as defined above (Materials and general methods). The quality of the amplified DNA is verified by visualization on 0,7% agarose gels.

3.3 Determining of the presence and abundance of the target DNA molecule:

[0084] The initial PCR analysis is carried out to evaluate if the target DNA sequence could be observed using the selected primers [SEQ ID NO: 5 and 6] using the selected PCR conditions: 94°C(30 sec) + 60°C(30 sec) + 72°C(30 sec) and 30 cycles is carried out.

[0085] One (out of eight) MDA amplified sample resultes in a positive PCR amplification, whereas the remaining 7 show no sign of amplification. The PCR product is evaluated to have approximately the correct predicted size on an agarose gel. By performing Real Time PCR (RT-PCR), the abundance of the target DNA molecule (encoding a putative endoglucanase) is determined using primers [SEQ ID NO: 5 and 6] and compared to the abundance of 16S rRNA genes using primers [SEQ ID NO: 7 and 8] (where the 16S rRNA gene is used as a measure of total bacterial DNA abundance). It is found that 16S rRNA genes are approximately 550.000 higher in quantity compared to the putative endoglucanase. Quantification reveals that 16S rRNA genes are present in 9.4E6 copies, whereas endoglucanase was detected only in 17 copies. Thus, a dilution of 10^{-2} was required to generate samples (N), as each aliquot sample would then averagely contain 0.17 templates.

3.4 Application of the PINS procedure to enrich and isolate the targeted DNA molecule and its flanking regions:

[0086] DNA samples found to contain the targeted DNA molecule, are subjected to PINS as follows:

3.4.1: 1st round PINS

[0087] Based on the above detected quantification, a dilution series is created using a dilution factor (D) of 1:10, aimed at then setting up multiple replicate dilutions of (N), 10^{-2}. Twenty identical samples of 10^{-2} dilution are prepared and subjected to genomic amplification, using

Repli-g Mini kit (Qiagen). Each amplified sample is individually analysed using the PCR primers [SEQ ID NO: 5 and 6] and the RT-PCR conditions described above in 1.3. Three out of 20 samples results in positive PCR amplification, with a DNA product of the predicted size. The associated PINS analysis output data are listed below.

PINS data #1

PINS quant no: 24.3

Dilution factor: 10

Next dilution: 10^{-3}

Samples in next repetition: 20

3.4.2: 2nd round PINS

[0088] Based on the PINS-software analysis from the previous round (PINS#1), a total of at least 20 replicate samples are estimated for next round of replication. Thus, 20 samples of dilution 10^{-3} are prepared (i.e. a 10 fold dilution (D) increase over the previous total dilution of the sample (Dt)), and each of the individual samples are amplified using Repli-g MDA kit. Each sample is subsequently analyzed by RT-PCR using primers [SEQ ID NO: 5 and 6] which shows that one (out of 20) sample yielded a DNA product of the predicted size. This sample is selected for further progress. PINS analysis output data are listed below.

PINS data PINS#2

PINS quant no: 230

Dilution factor: 9.6

Next dilution: 10^{-4}

Samples in next repetition: 20

3.4.3: 3rd round PINS

[0089] Based on the PINS#2 calculated values, it is estimated that at least 20 samples are again required for the next round of enrichment. The dilution proposed by the software is 10^{-4} and 20 samples are prepared. The 20 samples are amplified using the Repli-g Mini Kit, and are subsequently subjected to RT-PCR amplification using primers [SEQ ID NO: 5 and 6], where four samples yields a DNA product of the predicted size. These four samples are pooled and used for next round of repetition. PINS analysis output data are listed below.

PINS data PINS#3

PINS quant no: 2840

Dilution factor: 12

Next dilution: $8.4*10^{-6}$

Samples in next repetition: 25

3.4.4: 4th round PINS

[0090] Based on the PINS#3 calculated values, a total of 25 samples are now created having the dilution $8,4*10^{-6}$. Each of the individual replicated dilution samples are amplified using Repli-g Mini Kit. RT-PCR analysis using primers [SEQ ID NO: 5 and 6] shows that two samples yield a DNA product of the predicted size. PINS analysis output data are listed below.

PINS data PINS#4

PINS quant no: 27200

Dilution factor: 12

Next dilution: $8.8*10^{-7}$

Samples in next repetition: 25

3.4.5: 5<u>th</u> round PINS

[0091] Based on the PINS#4 calculated values, a total of 25 samples are now created in dilution $8.8-10^{-7}$. And, each of the individual samples are amplified using Repli-g Mini Kit. RT-PCR analysis using primers [SEQ ID NO: 6 and 6] shows that two samples yield a DNA product of the predicted size.

PINS software results are listed below.

PINS data PINS#5

PINS quant no: 365000

[0092] The two positive samples from this round of analysis are pooled into one sample, and RT-PCR is carried out on the targeted gene using primers [SEQ ID NO: 5 and 6] and 16S rRNA genes using primers [SEQ ID NO: 7 and 8] in the same sample. Levels of total DNA (as measured by rRNA genes) following the 5 rounds of MDA amplification, have not changed markedly compared to the initial observations but the ratio of target DNA to total DNA is now calculated to be 365000/ 9200000 = approximately 1/25. Previous studies have shown that a ratio 1:100 is sufficient to obtain genome data by RGW. Each of the amplification primers are combined with the linker ligation primers of the RGW.

3.4.6: Sequencing of the target DNA molecule enriched by PINS

[0093] The 5 cycles of PINS serves to enhance the presence of the targeted DNA molecule to such extent that the entire gene and a potential operon comprising the targeted DNA molecule, placed can be sequenced by traditional sequencing methods. Consequently, the RGW technique [7] is implemented in which 3 random libraries of the DNA sample, enriched with target DNA molecule are generated by preparing 3 restriction digests (EcoRI, HindIII and BglII/MboI) and then annealing the digests each with RGW primers comprising compatible 5' overhangs, for example primer pair [SEQ ID NO: 5] and [SEQ ID NO: 6] to form an oligo-

cassette. The target DNA molecule specific primers used for screening [SEQ ID NO: 5 and 6] are used in combination with the RGW primer [SEQ ID NO: 5] in a PCR amplification to generate larger PCR products that span the target DNA molecule and its flanking regions.

Primer nucleotide sequence	Description
>WalkID: GCG CTG CAG GCA TGC GAG CTC [SEQ ID NO: 9]	Used as PCR primer together with specific primer
>WALKOL_EcoRI_I: GCGCTGCAGGCATGCGAGCTCCCAAGCTTGATCG [SEQ ID NO: 10]	Linkers used for EcoRI digested DNA
>WALKOL_EcoRI_II: AATTCGATCAAGCTTGGGAGCTCGCATGCCTGCAGCGC [SEQ ID NO: 11]	
>WALKOL_HindIII: GCGCTGCAGGCATGCGAGCTCCCA [SEQ ID NO: 12]	Linkers used for HindIII digested DNA
>WALKOL_HindIII_II: AGCTTGGGAGCTCGCATGCCTGCAGCGC [SEQ ID NO: 13]	
>WALKOL_BglII: GATCTGGGAGCTCGCATGCCTGCAGCGC [SEQ ID NO: 14]	Linkers used for BglII/MboI digested DNA
>WALKOL_BglII_II: GCGCTGCAGGCATGCGAGCTCCCAAGCTTGATCG [SEQ ID NO: 15]	

[0094] Initially, two DNA sequences are generated using the procedure, and additional design of primers based on the acquired sequence results in a complete assembly of one ORF constituted of a 3996 bp gene (Figure 11). The gene has an ATG start code placed 196 bp up-stream of the EndoGlu-fw (5') end. A ShineDalgarno and associated -10 & -35 boxes are be located in the 5'direction of the ATG start code. Moreover, no apparent linkage to genes in 3' direction was found, illustrating that the gene is probably not a part of a larger operon.

Example 4: Use of PINS to link multiple genetic elements of common genomic origin

[0095] This example illustrates the use of PINS to detect the presence of the *mecA* gene in a mixed sample (e.g. taken from a patient) and to determine whether the gene originates from the genome of *Staphylococci* present in the sample. This analysis relies on the use of PINS to monitor the simultaneous co-amplification of a *Staphylococcus* gene located approximately 13000 base pairs up-stream of the *mecA* gene. The experiment is designed to use the high fidelity of phi29 polymerase for multiple displacement amplification (MDA), allowing genome fragments up to +70000 bp to be generated during amplification. If, after implementing PINS, the levels of PCR products generated using primers specific for *mecA* gene and for

Staphylococcus aureus (SA) reach comparable quantities, this indicates that the PCR products originate from the same genome, namely MRSA, as they have been co-amplified within the range of MDA. Assembly of sequenced PCR products after PINS is used to confirm that the mecA originated from a *Staphylococcus* genome.

4.1 Primer design

[0096] Multiple sequences of microbial genes coding for *mecA* are retrieved from the GenBank database and these sequences are aligned using ClustalW for multiple sequence alignment [9]. Regions of the alignment with 100% sequence identity are located and used for primer design. The resulting primers targeting *mecA* are:

MR-fw: 5'-CAAACTACGGTAACATTGATCGCAAC-3' [SEQ ID NO: 16]

MR-Re: 5' - CAATATGTATGCTTTGGTCTTTCTGC-3' [SEQ ID NO: 17],

and which are predicted to generate a 126 bp PCR product.

[0097] Strain specific primers used to target *Staphylococcus aureus* are designed approximately 13,000 bp 5' up-stream from the SCCmec integration site of resistance (OrfX) [14]. Primer sequences are:

SA-fw: 5' - CGTGAAGAAACAGAACGAATGATTC - 3' [SEQ ID NO: 18],

SA-Re: 5'- GCCTCTAGAATATTTCATCGCATTG - 3' [SEQ ID NO: 19],

and which are predicted to generate a 267 bp PCR product.

4.2 Primary biological sample collection and DNA extraction and quantification

[0098] A sample suspected to contain MRSA is subjected to PINS analysis. The sample is collected from a patient by the use of a "cotton-stick-swab". Total DNA is extracted from the swab using A&A Biotechnology (Poland) DNA Swab extraction kit.

[0099] The extracted DNA is used as template for a 10-fold serial dilution. Each of the diluted tubes are directly analyzed by Real-Time (RT) PCR using 1µl template in a 25 µl RT-reaction mix with each of the two sets of primers MR (MR-fw and MR-Re) and SA (SA-fw and SA-re). From the RT-PCR analysis targeted at MR (*mecA*), the Ct value is measured and converted to absolute numbers/copies by comparison to internal standards running in parallel samples in the same cycle program. The quantity of MR (*mecA*) is calculated to be $1.70*10^2$ whereas SA (*Staphylococcus aureus*) is calculated to be present in a quantity of $1.34*10^3$. From the RT analysis, it is apparent, that the sample contains more *Staphylococcus* targets than *mecA* targets. The ratio of the two products is close to 1/8 (MR/SA).

4.3 Application of the PINS procedure to enrich and for target DNA molecules diagnostic of *mecA* and *Staphylococcus*

4.3.1: 1st round PINS #1

[0100]   Based on the detected levels of target DNA molecules in the initial sample, 10 replicates of sample N (dilution $1.7*10^{-3}$) are prepared. By applying Repli-g Ultra-fast to each of the ten replicated samples and by subsequently performing RT-PCR analysis to detect the MR target DNA molecule, it is found that one sample from the 10 samples ("Sample E") yields a large amount of MR-PCR product, whereas the 9 other samples does not comprise detectable levels of MR PCR product. In this MR positive sample, RT-PCR quantification shows that the abundance of MR-PCR product is $1.27*10^3$ and the abundance of the SA-PCR product in sample $1.31*10^3$. The ratio has now shifted towards comparable levels of the two target DNA molecules, from having an initial ratio of approximately 1:8, up to a ratio that is closer to 1:1.

4.3.2: 2nd round PINS #2

[0101]   Based on the positive "E" sample, having dilution $1.27*10^{-3}$ (2.3.1) another dilution series is set up and, again, 10 replicated samples of N (dilution $1,27*10^{-4}$) are prepared. Each of the samples are amplified using Repli-g Ultra-fast MDA kit. From the RT-PCR of the analysis, it is apparent, that MR can be detected in 9 out of 10 samples, and that abundance of MR in the positive samples is $1.32*10^4$. Comparable analysis reveals, that the level of the SA-PCR product has not increased, as the abundance of SA remains identical: $1.31*10^4$. The ratio of the two products are, again, very close to 1:1. Having employed PINS to obtain an additional amplification factor of x10 without obtaining any noticeable change in target abundance, we conclude that the quantity cannot be further increased, indicating the presence of a (close to) pure sample. Additional rounds of PINS will, thus, not result in a higher level of purification. Thus, the next step is to sequence the amplified template to evaluate the identity of the resistance gene *mecA* and the adjacent genes, and to verify the presence of MRSA rather than MR and SA as independent DNA molecules located in the same sample.

4.3.3: Sequencing *mecA* and adjacent genes in the MR target DNA molecule

[0102]   Primers designed to target known MRSA (GenBank, NCBI),) are used to generate overlapping PCR products used for sequencing and assembly. 24 PCR products of various lengths are generated and two continuous sequences of 2475 bp. and 19186 bp. are produced

by PCR, sequencing, and assembly.An additional 789bp are obtained by RGW [7] and sufficient overlap is established to combine the two assemblies into one large assembly of 22450 bp. Nucleotide BLAST [11] shows close to identical match (>99.9%) to multiple MRSA strains in the database.

Summary:

[0103] It was evaluated, that the abundance of MR had increased by a factor of 8 by the first round of PINS. By performing analysis on the non-targeted background, it is apparent that the ratio between MR and SA was clearly biased during the first run (PINS #1), whereas this ratio was maintained in the following round of amplification (PINS #2). Thus, it is concluded, that both target DNA molecules originate from the same DNA fragment, since the non-target SA is present in equal quantity even though only the MR target DNA molecule has been used during the rounds of PINS selection. It is apparent from the analysis, that although non-resistant *Staphylococci* presumably also are present in the initial mixture, it can be concluded that the MR resistance originated from a *Staphylococcus aureus* and not from other *mecA* containing bacteria such as coagulase-negative *Staphylococci* [12]. Although MR can be present in non-*Staphylococcal* bacterial strains, it was evidently amplified together with SA, due to the co-amplification of both targeted DNA molecules during MDA amplification. Thus, although the primed sequences are separated by 13000 kb in the genome, it is possible to assign both properties to the same fraction of the genome, sufficient to diagnose the presence of MRSA. In addition, sequencing verifies this observation, and it was evident that the DNA was close to identical (+99.9%) to known MRSA genome sequences.

REFERENCES

[0104]

1.
   Maurin, M., Real-time PCR as a diagnostic tool for bacterial diseases. Expert Rev Mol Diagn, 2012.12(7): p. 731-54.
2.
   Shendure, J. and H. Ji, Next-generation DNA sequencing. Nat Biotechnol, 2008. 26(10): p. 1135-45.
3.
   Torsvik, V., J. Goksoyr, and F.L. Daae, High diversity in DNA of soil bacteria. Appl Environ.Microbiol, 1990. 56(3): p. 782-787.
4.
   Lasken, R.S. and S. Huges, Multiple displacement amplification of genomic DNA, in Whole Genome Amplification, S. Huges, Editor 2005, Scion Publishing Ltd. 2005: Oxfordshire. p. 99-118.

5.

Sambrook, J. and D.W. Russell, Molecular Cloning a laboratory manual2001: Cold Spring Harbor Laboratory Press.

6.

Raghunathan, A., et al., Genomic DNA amplification from a single bacterium. Applied and Environmental Microbiology, 2005. 71(6): p. 3342-3347.

7.

Kilstrup, M. and K.N. Kristiansen, Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res, 2000. 28(11): p. e55.

8.

Liu, L., et al., Comparison of next-generation sequencing systems. J Biomed Biotechnol, 2012. 2012: p. 251364.

9.

Chenna, R., et al., Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res., 2003. 31(13): p. 3497-3500.

10.

Kvist, T., et al., Diversity of thermophilic and non-thermophilic crenarchaeota at 80 degrees C. FEMS Microbiol.Lett., 2005. 244(1): p. 61-68.

11.

Altschul, S.F., et al., Basic local alignment search tool. J.Mol.Biol., 1990. 215(3): p. 403-410.

12.

Huletsky, A., et al., Identification of methicillin-resistant Staphylococcus aureus carriage in less than 1 hour during a hospital surveillance program. Clin Infect Dis, 2005. 40(7): p. 976-81.

13. Lukyanov,K., et al., Molecule by molecule PCR amplification of complex DNA mixtures for direct sequencing: an approach to in vitro cloning, Nucleic Acids Research, 1996. 24(11): p. 2194-2195

14. WO2004058987 (MOLECULAR STAGING INC [US] QIAGEN GMBH [DE]) 15 July 2004)

15. Erlandsson, L., et al., Amp-PCR: Combining a Random Unbiased Phi29-Amplification with a Specific Real-Time PCR, Performed in One Tube to Increase PCR Sensitivity. PLOS ONE, 2010. 5(12): p. e15719

SEQUENCE LISTING

[0105]

<110> Samplix S.a.r.l
Kvist, Thomas

Mikkelsen, Marie J

<120> Probability-directed isolation of nucleotide sequences (PINS)

<130> P1384PC00

<150> PA 2012 70822
<151> 2012-12-21

<160> 19

<170> PatentIn version 3.5

<210> 1
<211> 19
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(19)
<223> HPV18 fw primer

<400> 1
gtttagtgtg ggcctgtgc   19

<210> 2
<211> 20
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(20)
<223> HPV18 Rev primer

<400> 2
ggcatgggaa ctttcagtgt   20

<210> 3
<211> 23
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(23)
<223> HPV-264f primer

<400> 3

gtggtgtata gagacagtat acc   23

<210> 4
<211> 20
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(20)
<223> HPV-911f primer

<400> 4

ccttctggat cagccattgt   20

<210> 5
<211> 22
<212> DNA
<213> Bacillus acidopullulyticus

<220>
<221> primer_bind
<222> (1)..(22)
<223> EndoGlu - Fw primer

<400> 5

gcacattgct ggtacagaga ta   22

<210> 6
<211> 22
<212> DNA
<213> Bacillus acidopullulyticus

<220>
<221> primer_bind
<222> (1)..(22)
<223> EndoGlu Re primer

<400> 6

gatctgcttc cattatgctc tg   22

<210> 7
<211> 19
<212> DNA
<213> Bacterium

<220>

<221> primer_bind
<222> (1)..(19)
<223> B-all-27f - bacterial 16S rRNA forward primer

<400> 7
gagtttgatc ctggctcag   19

<210> 8
<211> 18
<212> DNA
<213> Bacteria

<220>
<221> primer_bind
<222> (1)..(18)
<223> Eub338r - bacterial 16S rRNA reverse primer

<400> 8
gctgcctccc gtaggagt   18

<210> 9
<211> 21
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(21)
<223> WalkID primer

<400> 9
gcgctgcagg catgcgagct c   21

<210> 10
<211> 34
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(34)
<223> WALKOL_EcoRI_I primer

<400> 10
gcgctgcagg catgcgagct cccaagcttg atcg   34

<210> 11
<211> 38

<212> DNA

<213> Human papillomavirus type 18

<220>

<221> primer_bind

<222> (1)..(38)

<223> WALKOL_EcoRI_II primer

<400> 11

aattcgatca agcttgggag ctcgcatgcc tgcagcgc   38

<210> 12

<211> 24

<212> DNA

<213> Human papillomavirus type 18

<220>

<221> primer_bind

<222> (1)..(24)

<223> WALKOL_HindIII primer

<400> 12

gcgctgcagg catgcgagct ccca   24

<210> 13

<211> 28

<212> DNA

<213> Human papillomavirus type 18

<220>

<221> primer_bind

<222> (1)..(28)

<223> WALKOL_HindIII_II primer

<400> 13

agcttgggag ctcgcatgcc tgcagcgc   28

<210> 14

<211> 28

<212> DNA

<213> Human papillomavirus type 18

<220>

<221> primer_bind

<222> (1)..(28)

<223> WALKOL_BglII primer

<400> 14

agcttgggag ctcgcatgcc tgcagcgc   28

<210> 15
<211> 34
<212> DNA
<213> Human papillomavirus type 18

<220>
<221> primer_bind
<222> (1)..(34)
<223> WALKOL_BglII_II primer

<400> 15
gcgctgcagg catgcgagct cccaagcttg atcg   34

<210> 16
<211> 26
<212> DNA
<213> Bacterium

<220>
<221> primer_bind
<222> (1) .. (26)
<223> MR-fw primer (multi resistance gene)

<400> 16
caaactacgg taacattgat cgcaac   26

<210> 17
<211> 26
<212> DNA
<213> Bacterium

<220>
<221> primer_bind
<222> (1) .. (26)
<223> MR-Re primer (multi resistance gene)

<400> 17
caatatgtat gctttggtct ttctgc   26

<210> 18
<211> 25
<212> DNA
<213> Staphylococcus aureus

<220>
<221> primer_bind

<222> (1)..(25)
<223> SA-fw primer

<400> 18
cgtgaagaaa cagaacgaat gattc   25

<210> 19
<211> 26
<212> DNA
<213> Staphylococcus aureus

<220>
<221> primer_bind
<222> (1) .. (26)
<223> SA-Re primer

<400> 19
gcctctagaa tatttcatcg catttg   26

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO2004058987A [0015] [0104]
- PA201270822 [0105]

Non-patent literature cited in the description

- **MAURIN, M.** Real-time PCR as a diagnostic tool for bacterial diseasesExpert Rev Mol Diagn, 2012, vol. 12, 7731-54 [0104]
- **SHENDURE, J.H. JI** Next-generation DNA sequencingNat Biotechnol, 2008, vol. 26,

101135-45 [0104]

- **TORSVIK, V.J. GOKSOYR F.L. DAAE** High diversity in DNA of soil bacteria Appl Environ.Microbiol, 1990, vol. 56, 3782-787 [0104]
- Multiple displacement amplification of genomic DNA **LASKEN, R.S.S. HUGES** Whole Genome Amplification Scion Publishing Ltd.2005000099-118 [0104]
- **SAMBROOK, J.D.W. RUSSELL** Molecular Cloning a laboratory manual Cold Spring Harbor Laboratory Press 20010000 [0104]
- **RAGHUNATHAN, A. et al.** Genomic DNA amplification from a single bacterium Applied and Environmental Microbiology, 2005, vol. 71, 63342-3347 [0104]
- **KILSTRUP, M.K.N. KRISTIANSEN** Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer Nucleic Acids Res, 2000, vol. 28, 11e55- [0104]
- **LIU, L. et al.** Comparison of next-generation sequencing systems J Biomed Biotechnol, 2012, vol. 2012, 251364- [0104]
- **CHENNA, R. et al.** Multiple sequence alignment with the Clustal series of programs Nucleic Acids Res., 2003, vol. 31, 133497-3500 [0104]
- **KVIST, T. et al.** Diversity of thermophilic and non-thermophilic crenarchaeota at 80 degrees C FEMS Microbiol.Lett., 2005, vol. 244, 161-68 [0104]
- **ALTSCHUL, S.F. et al.** Basic local alignment search tool J.Mol.Biol., 1990, vol. 215, 3403-410 [0104]
- **HULETSKY, A. et al.** Identification of methicillin-resistant Staphylococcus aureus carriage in less than 1 hour during a hospital surveillance program Clin Infect Dis, 2005, vol. 40, 7976-81 [0104]
- **LUKYANOV,K. et al.** Molecule by molecule PCR amplification of complex DNA mixtures for direct sequencing: an approach to in vitro cloning Nucleic Acids Research, 1996, vol. 24, 112194-2195 [0104]
- **ERLANDSSON, L. et al.** Amp-PCR: Combining a Random Unbiased Phi29-Amplification with a Specific Real-Time PCR, Performed in One Tube to Increase PCR Sensitivity PLOS ONE, 2010, vol. 5, 12e15719- [0104]

Patentkrav

5   1. *In vitro*-fremgangsmåde til at berige et target DNA-molekyle fra en blandet polynukleotidprøve, hvilken fremgangsmåde omfatter trinnene:
a) at tilvejebringe en blandet polynukleotidprøve indeholdende target-DNA-molekylet, hvor target-DNA-molekylet omfatter en eller flere unikke konsekutive sekvenser på mindst 10 nukleotider,

10   b) at seriefortynde den blandede polynukleotidprøve, indtil sandsynligheden for detektering af target-DNA-molekylet i en fortyndet prøve af blandede polynukleotider er mindre end 0,75, foretrukket mindre end 0,50 eller endnu mere foretrukket mindre end 0,25, og
c) at fremstille et tilstrækkeligt antal af replikater af den fortyndede prøve af

15   blandede polynukleotider, indtil sandsynligheden for detektering af target-DNA-molekylet i mindst en af replikatfortyndingsprøverne er mindst 0,75, foretrukket 0,80 til 0,95;
d) at amplificere DNA'et i de replikatfortyndede prøver for at øge abundansen af DNA'et i hver prøve;

20   e) at detektere tilstedeværelsen eller fraværet af target-DNA-molekylet i replikatfortyndingsprøverne, der er amplificeret i trin (d), hvor hyppigheden af target-DNA-molekylet i replikatfortyndingsprøven øges sammenlignet med den blandede polynukleotidprøve i trin (a);
f) at seriefortynde mindst en replikatfortyndingsprøve detekteret i trin e) inde-

25   holdende DNA-molekylet, indtil sandsynligheden for detektering af target-DNA-molekylet i en fortyndet prøve er mindre end 0,75, og at gentage trin (c) til (e) eller (f) mindst en gang.

30   2. Fremgangsmåde ifølge krav 1, hvor hyppigheden af target-DNA-molekylet i forhold til non-target-DNA i en blandet polynukleotidprøve i trin a) er mellem 10^{-2} og 10^{-7}, foretrukket mellem 10^{-4} og 10^{-7}.

3. Fremgangsmåde ifølge krav 1 eller 2, hvor seriefortyndingerne af den

35   blandede polynukleotidprøve i trin (b) amplificeres for at øge abundansen af DNA'et i hver prøve, efterfulgt af trinnet med detektering af tilstedeværelsen

eller fraværet af target-DNA-molekylet i hver amplificeret prøve, og derved bestemmelse af antallet af seriefortyndinger, som er nødvendige for at opnå en fortyndet prøve, hvor sandsynligheden for detektering af target-DNA-molekylet er mindre end 0,75.

5

4. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 3, hvor DNA'et amplificeres i trin (d) med en teknik, som er valgt blandt vilkårligt degenereret primet PCR, linker ligation PCR, Degenerate Oligonucleotide Primed (DOP) PCR og Multiple Displacement Amplification.

10

5. Fremgangsmåde ifølge krav 1, yderligere omfattende trin g), hvor der udføres en nukleinsyresekvensanalyse af DNA-molekylet for at bestemme mindst en del af nukleotidsekvensen af DNA-molekylet.

15

6. Fremgangsmåde ifølge et hvilket som helst af kravene 1-3, hvor tilstedeværelsen af den ene eller flere unikke konsekutive sekvenser på mindst 10 nukleotider i DNA-molekylet detekteres ved PCR.

7. Fremgangsmåde ifølge et hvilket som helst af kravene 1-3, hvor tilstede-
20 værelsen af den ene eller flere unikke konsekutive sekvenser på mindst 10 nukleotider i DNA-molekylet detekteres ved hybridisering til DNA-molekylet.

8. Fremgangsmåde ifølge krav 1, hvor DNA-molekylet omfatter mindst to unikke konsekutive sekvenser på mindst 10 nukleotider, og hvor DNA-
25 molekylet omfatter 50 til 100.000 nukleinsyrebasepar, foretrukket 150 til 3.000 nukleinsyrebasepar, mere foretrukket 150 til 1500 nukleinsyrebasepar.

9. Fremgangsmåde ifølge et hvilket som helst af kravene 1-8, hvor den samlede fortynding af den fortyndede prøve i trin (f) øges med en faktor 2 til 20
30 for hver gentagelse af trin (c) til (e).

10. Fremgangsmåde ifølge et hvilket som helst af kravene 1-9, hvor antallet af replikatfortyndingsprøver, som er fremstillet i trin (c), er 2 til 500.

11. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 8, hvor target-DNA-molekylet er afledt af genomet af en celle.

12. Fremgangsmåde ifølge krav 11, hvor cellen udvælges fra en bakteriecelle, en svampecelle og en pattedyrscelle.

5

13. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 8, hvor target-DNA-molekylet er afledt af et virusgenom eller et pattedyrsgenom eller en kombination deraf.
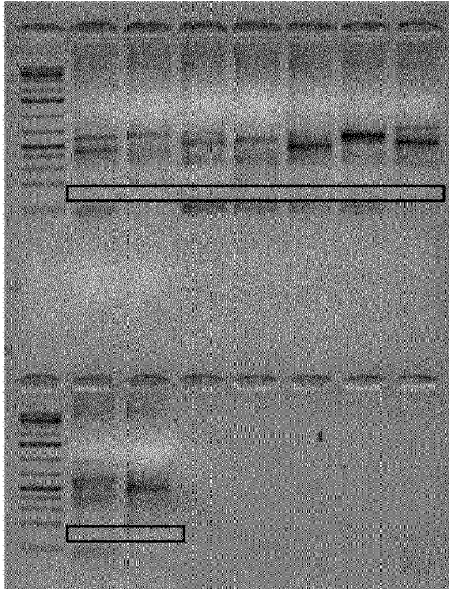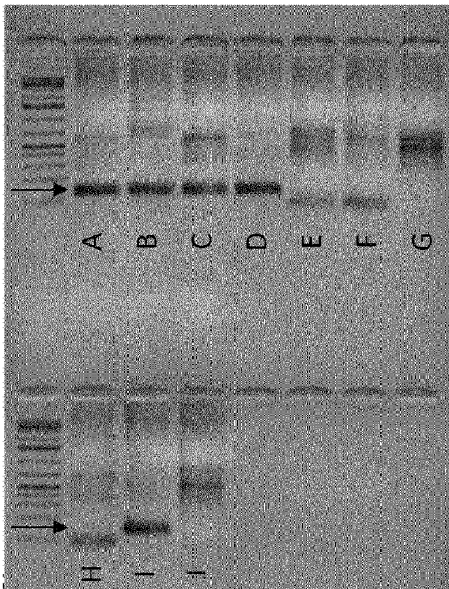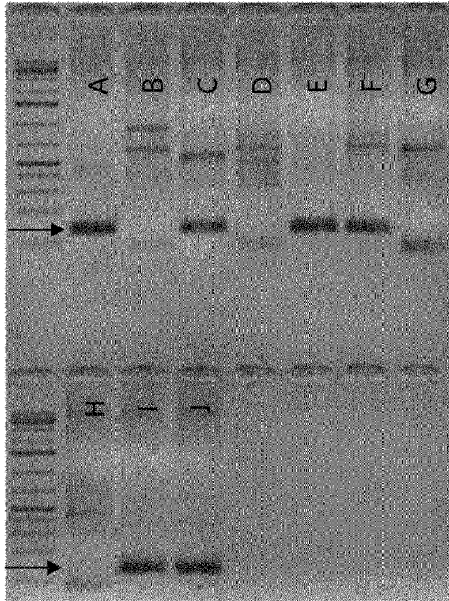
10

DRAWINGS

Figure 1
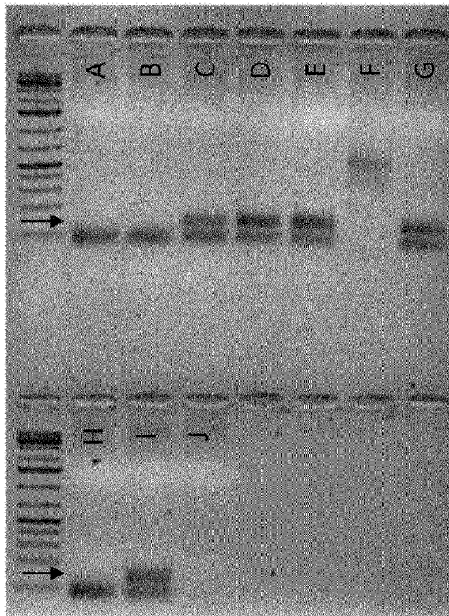

Figure 2

Figure 3

Figure 4

Figure 5

Start Mix
1:1972209

	Targets/µl (in start mix)	ng/µl	Targets / µg
Before Phi (w. 3,5 µl added)	0,02857142	187	0,15
Targets (added to each tube)	0,1		
ng (added)	654,5		
Targets / µg	0,15		

(3,5 µl to each reaction)

Phi | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10

(3,5 µl to 50 µl reaction)

Reamplification

(1.0 µl to 20 µl reaction)

Phi | #1 | #2 | #3 | #4 | #5 (1/3) | #6 (1/3) | #7 (1/3) | #8 (3/3) | #9 (3/3) | #10 (3/3)

Pooling of samples #8 + #10

(additional PCR)
PCR after Ph (8/13 positive)

Phi-quantification			
Targets (by PCR)/µl	ng/µl	Targets / µg	Gain
0,62	208	2,96	19,36

(additional PCR)
PCR (6/10 positive)
(Done twice - same result!)

Phi-quantification			
Targets/µl	ng/µl	Targets / µg	Gain
0,6	222	2,70	17,69

(additional PCR)
PCR (9/10 positive)
Undiluted -->

Phi-quantification			
Targets/µl	ng/µl	Targets / µg	Gain
0,9	207	4,35	1,61

(additional PCR)
PCR (5/10 positive)
x5 dilution -->

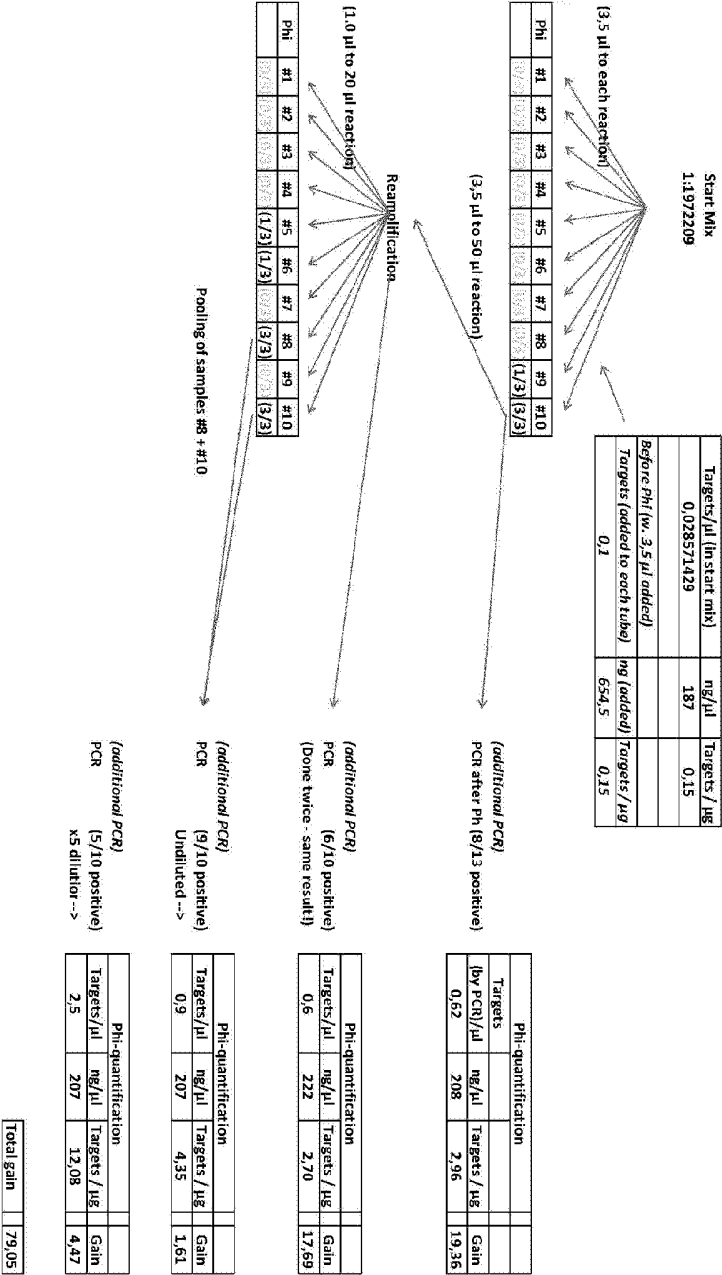| Phi-quantification			
Targets/µl	ng/µl	Targets / µg	Gain
2,5	207	12,08	4,47

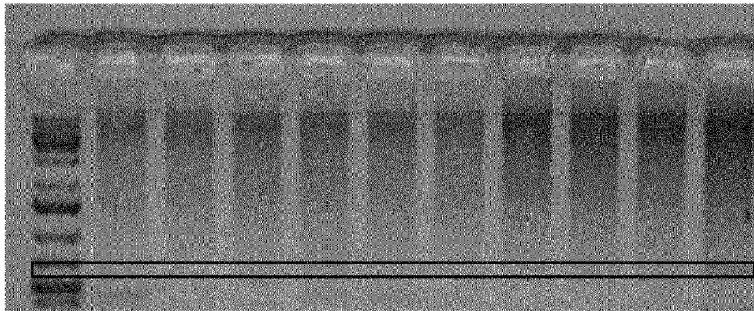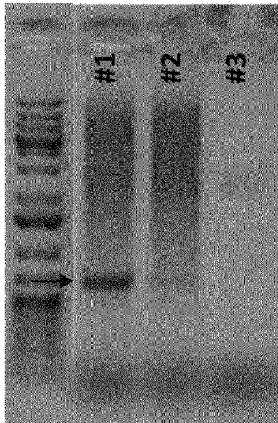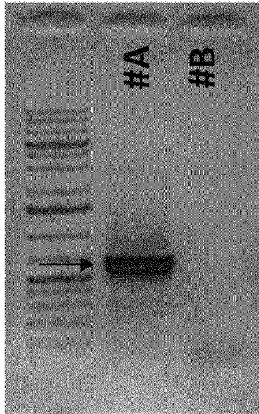| Total gain	79,05

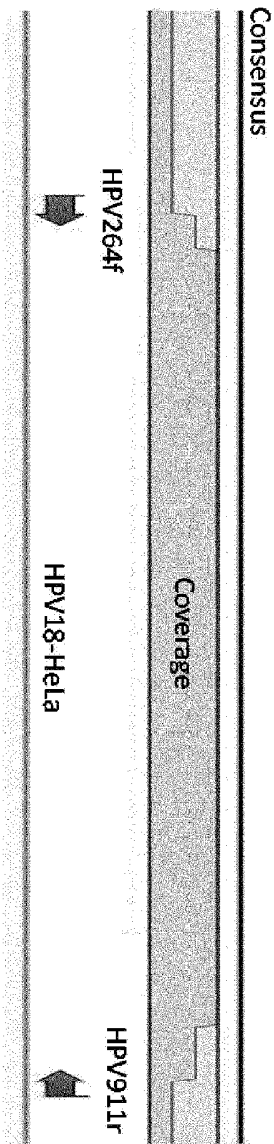Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

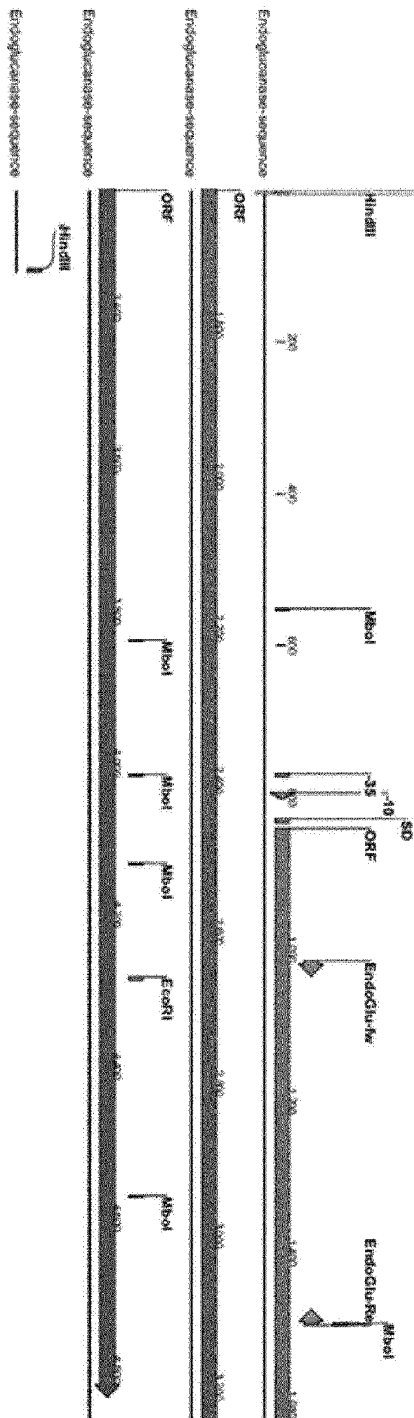| EndoGlu-Fw 5' – GCACATTGCTGGTACAGAGATA – 3' [SEQ ID NO: 5]	EndoGlu-Re 5' – GATCTGCTTCCATTATGCTCTG – 3' [SEQ ID NO: 6] (Reverse Complement to illustration)
ACGTTAA TAGGAACCAA TCAC▓▓▓▓▓▓▓▓▓▓▓▓▓▓GACT TGAT. ACGTTAA TAGGAACCAA TCACGCACAT TGCTGGTACA GAGATAGACT TGAT. ACATTAA TAGGAACAAA TCACGCACAT TGCTGGTACA GAGATAAACT TGAG. ACATTAA TAGGAACAAA TCACGCACAT TGCTGGTACA GAGATAAACT TGAG. ACNTTAA TAGGAACNAA TCACGCACAT TGCTGGTACA GAGATANACT TGAN. 	GAGAGATAAT GCC▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓CGCTG GAGAGATAAT GCCCAGAGCA TAATGGAAGC AGATCCGCTG GAGAGACAAT GCCCAGAGCA TAATGGAAGC AGATCCGCTG GAGAGACAAT GCCCAGAGCA TAATGGAAGC AGATCCGCTG GAGAGANAAT GCCCAGAGCA TAATGGAAGC AGATCCGCTG
B-all-27f – bacterial 16S rRNA forward primer	Eub338r – bacterial 16S rRNA reverse primer
5' - GAGTTTGATCCTGGCTCAG – 3' [SEQ ID NO: 7]	5' - GCTGCCTCCCGTAGGAGT - 3' [SEQ ID NO: 8]

FIGURE 11