



(51) International Patent Classification:
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/EP2017/076644

(22) International Filing Date:
18 October 2017 (18.10.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1617643.0 18 October 2016 (18.10.2016) GB

(71) Applicant: **OXFORD UNIVERSITY INNOVATION LIMITED** [GB/GB]; Buxton Court, 3 West Way, Oxford OX2 0JB (GB).

(72) Inventors: **ZHANG, Gang**; c/o The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford Oxfordshire OX3 7BN (GB). **DONNELLY, Peter**; c/o The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford Oxfordshire OX3 7BN (GB). **BOWDEN, Rory**; c/o The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford Oxfordshire OX3 7BN (GB). **HATTON, Edouard**; 36 quai Saint Laurent, 45000 Orleans (FR).

(74) Agent: **DEHNS**; St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHOD FOR COMPLETE, UNIFORM AND SPECIFIC AMPLIFICATION OF ULTRA-LOW AMOUNTS OF INPUT DNA

(57) Abstract: The present invention provides a method for the amplification of ultra-low amounts of input DNA, in particular the amount of genomic DNA which may be obtained from a single animal, plant or microbial cell, using random RNA primers and a Klenow fragment or an equivalent thereof, to obtain predominantly double stranded DNA amplification products. Further provided is a method of optimising nucleic acid sequence analysis of ultra-low amounts of DNA, said method comprising performing the DNA amplification method of the invention on an ultra-low amount of DNA prior to the nucleic acid sequence analysis of said DNA. Still further provided is a method of nucleic acid sequence analysis, said method comprising a step of sample preparation prior to the analysis step(s) in which the DNA amplification method of the invention is used to amplify an ultra-low amount of DNA in said sample.



- 1 -

Method for complete, uniform and specific amplification of
ultra-low amounts of input DNA

The present invention provides a method for the amplification of ultra-low amounts of input DNA, in particular the amount of genomic DNA which may be obtained from a single animal, plant or microbial cell. The method of the invention provides an amplification product that may, in certain embodiments, be significantly superior to that provided by existing whole genome, low input, amplification techniques in terms of its sequence coverage, the uniformity of that coverage and its specificity, and so the downstream use of the amplified product of the invention in whole-genome sequencing, copy-number profiling and other analytical techniques is associated with a reduced chance of introducing experimental artefacts. The inventors have found that a particular sequence of specific enzymatic DNA manipulations, which has been designed to effect a particular and unique amplification rationale, can provide an amplification product which is a population of small DNA fragments that randomly and extensively cover the nucleotide sequence of the starting material without appreciable bias and with only negligible amounts of off target products or sequence errors. This amplification product may therefore reflect the nucleotide sequence of the ultra-low amounts of starting DNA material sufficiently faithfully in terms of its sequence coverage, the uniformity of that coverage and its specificity such that the chances of experimental artefacts arising from the amplification of the starting material becoming apparent in downstream techniques are reduced as compared to the use of existing whole genome, low input, amplification techniques. This thereby may allow the recovery of greater amounts of useful sequence information from the starting material, as compared to existing whole genome, low input, amplification techniques.

The importance of complete and uniform coverage of the genome has been recognised in the description of techniques for the amplification of DNA from single cells. Most documented are the techniques termed MALBAC (Multiple Annealing and Looping Based Amplification Cycles) and MDA (Multiple Displacement Amplification).

MALBAC is disclosed, inter alia, in Zong et. al, 2012, Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human

- 2 -

Cell, Science, Vol 338,1622-1626; Lu et. al., 2012, Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing, Science, Vol 338, 1627-1630; and US 2014/0200146.

MDA is disclosed, inter alia, in Spits et. al., 2006, Whole-genome multiple displacement amplification from single cells, Nat Protoc, Vol 1(4), 1965-70; and WO 02/057487.

PCR-based techniques have also been devised: PCR-Based Whole Genome Amplification, in PCR, 2007, Chapter 18 (eds. Hughes and Moody).

These techniques have certain drawbacks. MDA uses random hexamer DNA primers which have a tendency to interact and amplify one another. MDA also shows allelic preference (where one allele at a specific genetic locus is amplified preferentially and at random in a sample) and even allelic dropout (where non-amplification of an allele occurs, usually randomly). This may be a consequence of the biochemical properties of Phi-29 DNA polymerase (i.e. high processivity and prolonged and continuous polymerisation from its initiation sites) used in this technique, in which randomly chosen contiguous areas of the genome that happen to be amplified in large runs, early in the amplification process, are thought to be favoured by the exponential nature of the MDA copying process and may produce substantially more (multiples of) product DNA than other regions. In addition, any DNA sequence or base content effect in which random primers are more likely to anneal to some parts of the genome will bias amplification towards those sequences, an effect that is multiplied by the highly exponential (many rounds of copying of copies) nature of the MDA process. Prolonged incubation time used in MDA may also contribute to amplification biases.

MALBAC relies on degenerate DNA primers and Bst I large fragment polymerase or a mixture of Phi-29 and Bst I large fragment to amplify the target DNA template.

The major drawback for MALBAC is the design of the degenerate primers, which consists of a mixture of 5 random nucleotides flanked by 27 and 3 nucleotides with fixed sequence at 5' and 3'sites, respectively, i.e. 5'-GT GAG TGA TGG TTG AGG TAG TGT GGA GNNNNNGGG-3' (SEQ ID NO: 1) and 5'-GT GAG TGA TGG TTG AGG TAG TGT GGA GNNNNNTTT-3' (SEQ ID NO: 2). This design substantially restricts the randomness of primer annealing onto the target DNA template, promoting preferential amplification.

A further drawback is that MALBAC requires multiple rounds of amplification, i.e. at least two, to generate amplified material, but further rounds may also be necessary to generate enough amplified material that is suitable for further processing, e.g. further PCR amplification. This proliferation in the number of rounds of amplification increases the probability of amplification biases becoming apparent.

Moreover, Bst I polymerase does not have 3'→5' proof reading activity. This leads to error prone amplification. In combination with multiple rounds, i.e. more than 2 rounds, of MALBAC amplification, the amplification fidelity may be diminished further.

PCR-based methods, e.g. GenomePlex and Picoplex, can induce significant amplification bias across different loci in the genome which results in significant amplification loci dropout. With these methods, only 5-30% of the genome is often sequenced with reasonable sequencing depth from amplified single cells.

Thus, while methods for whole genome amplification have been proposed, such methods can generally be inefficient, complex, error-prone and expensive. Therefore, a need exists for alternative methods of amplifying extremely small amounts of DNA, e.g. the genomic DNA from a single cell or small group of cells, and more preferably methods which are superior, ideally significantly superior, to those currently available.

In contrast to the prior art methods, the present inventors have adopted a different rationale to achieve amplification of ultra-low amounts of input DNA with acceptable coverage of the starting template, with acceptable uniformity and with acceptable amplification specificity. In certain embodiments the method of the invention can provide amplification of ultra-low amounts of input DNA with greater coverage of the starting template, with greater uniformity and with greater amplification specificity, as compared to existing whole genome, low input, amplification techniques. The inventors have found that by limiting the size of the nucleic acid fragments which are synthesised initially, and the yield thereof, but at the same time increasing the randomness of the priming reaction, the resulting amplification product contains a spectrum of fragments that can potentially amount to a more even, complete and specific representation of the starting material than the amplification products provided by existing whole genome, low input, amplification techniques, e.g. MALBAC, MDA, GenomePlex and Picoplex. In other words, through the use of overlapping polymerisation events which are more

random, there is the potential for a greater proportion of the entire nucleotide sequence content of the starting material to be represented in the amplification product with substantially less amplification bias and, ideally, fewer errors and less off target products, as compared to existing techniques. The inventors have also found that RNA primers may be used without significantly affecting the evenness or completeness of coverage of the starting material but whilst potentially limiting, e.g. preventing significantly or entirely, the formation of non-specific amplification products from random primer-interdependent polymerisation. In other words, the amplification method of the invention has the potential to be specific in that, essentially, it only produces copies of the nucleotide sequences in the starting material.

In order to meet this design rationale the inventors have devised a method having a particular sequence of steps involving particular enzymatic manipulations, each carefully chosen to provide a particular technical effect or effects that together, in the correct order, results in a DNA amplification product with acceptable coverage, uniformity within that coverage, and specificity, i.e. a DNA product that represents the nucleotide sequence of the starting DNA material sufficiently faithfully in terms of its sequence coverage, the uniformity of that coverage and its specificity to allow the recovery of acceptable amounts of useful sequence information from the starting material. In certain embodiments the method of the invention can potentially result in a DNA amplification product with superior coverage and superior uniformity within that coverage and which is specific, i.e. a DNA product that represents the nucleotide sequence of the starting DNA material sufficiently faithfully in terms of its sequence coverage, the uniformity of that coverage and its specificity to allow the recovery of greater amounts of useful sequence information from the starting material, as compared to the amplification products of existing whole genome, low input, amplification techniques.

The first key aspect of the method is the use of RNA random primers, specifically 10mer primers, with a DNA polymerase having the catalytic features of a Klenow fragment.

Experimental artefacts can occur in DNA amplification techniques when the primers required to initiate DNA synthesis act as both template and primer in the presence or absence of added target DNA template material. Such non-specific amplification product can itself be copied into greater amounts of product and this can quickly become a significant population of nucleic acid in the final reaction

product, leading to sources of error in the analysis thereof. Most DNA polymerases (including the Klenow fragment) cannot use RNA as a template efficiently, if at all, and so the use of RNA primers might in principle remove a major source of non-specific product during amplification if the polymerase can efficiently catalyse a DNA polymerisation reaction using RNA oligonucleotides as primers. The inventors have found that the Klenow fragment can use RNA primers to prime DNA polymerisation from very low input amounts of DNA template efficiently and thus the formation of problematic amounts of non-specific reaction product in the first stage of the method of the invention can be avoided to some extent, e.g. avoided significantly or even entirely. This can be contrasted to MDA in which illegitimate amplification of DNA primers and annealed primers can lead to significant output DNA product in the absence of added target material or when ultra-low amounts of target DNA material are used.

The Klenow fragment, a derivative of DNA polymerase I in which the portion of the protein carrying the 5'→3' exonuclease activity has been removed, is selected in part because its low temperature optimum for polymerisation is compatible with the temperature required for amplification of the widest range of annealed primers. In contrast, polymerases requiring higher temperatures for optimum polymerase activity demand reaction temperatures above that of low melting-temperature primer-target hybrids. This can result in less even amplification, higher percentage of GC content in amplified products, and/or incomplete coverage of the starting material.

Secondly, unlike Phi-29 DNA polymerase (used in MDA), the Klenow fragment tends to interact with the template and catalyse a polymerisation reaction in a transient and intermittent fashion (typically by adding approximately 20 nucleotides during each polymerisation reaction event, departing from the templates and randomly reinitiating polymerisation using annealed RNA primers or extended DNA chains from RNA primers on target templates as primers). This leads to more complete and even coverage of template DNA material.

Thirdly, the Klenow fragment has no 5'→3' nuclease activity, but it does have strand displacement activity. This allows the amplification of segments of template to extend past adjacent priming sites into segments that have already been copied, such that overlap occurs between successive product molecules on the same template molecule. This is a key beneficial property because it reduces the chance that stretches of template will be missed completely by the amplification

reagents during amplification or will be less represented in the amplified product, problems that could lead to “gaps” in the sequence data derived from the amplified DNA. In other words, such an activity has the potential to provide more complete and even coverage of the template material in the amplified product overall, in particular as compared to existing PCR-based whole genome, low input, amplification techniques.

The inventors have recognised that the Klenow fragment has a repertoire of technical features which are advantageous in the context of the present invention, e.g. whole genome amplification from a single cell. This repertoire of properties together with the use of RNA primers allows the development of a first amplification step which amplifies the input DNA in a manner which potentially provides extensive, uniform and accurate coverage of the template DNA, in particular as compared to existing whole genome, low input, amplification techniques.

At this stage the amplification reaction product is a complex mixture of single stranded and double stranded RNA-containing DNA molecules, much of which remains hybridised to the input DNA material in a complex branching structure, which require further manipulation to form them into more useable molecules, i.e. essentially free double stranded DNA molecules. This stage of the method of the invention is in effect a stage of nucleic acid repair and it is of paramount importance that this stage must be as accurate and efficient as possible whilst keeping to practical timescales. This part of the process can therefore be viewed as another limiting factor in terms of overall quality and quantity of the final amplification product and in terms of its usability. The invention achieves these ends over three simple steps.

In the first of these steps a DNA polymerase having the catalytic features of DNA polymerase I is used to initiate second strand synthesis. Pol I has the same polymerase properties as the Klenow fragment mentioned above, but it further has 5'→3' nuclease activity. This avoids displacement of downstream DNA strands and the formation of single stranded 5'-overhanging DNA flaps and thus results in a population of partially double-stranded nucleic acid molecules with small single-stranded gaps or nicks and overhanging and annealed sections of RNA sequences. Subsequently, a DNA polymerase having the catalytic features of Taq DNA polymerase may be used to further complete the double-stranded regions further so as to improve yield and sequence representation. Finally, a nuclease, e.g. having the catalytic features of S1 nuclease, a single-strand-specific DNA/RNA nuclease,

or BAL 31 nuclease, an exonuclease, is used under certain reaction conditions to efficiently remove the RNA primer sequences remaining in the amplification product, thereby conveniently forming a double stranded DNA product which is an amplified but faithful representation of the nucleotide sequence of the starting DNA material in terms of its sequence coverage, the uniformity of that coverage and its specificity, in particular as compared to existing whole genome, low input, amplification techniques.

The high quality low yield double stranded DNA amplification product which may be obtained directly from this method may be fed into downstream analytical techniques either directly or via further steps of amplification. Being potentially of superior fidelity and completeness and evenness of coverage, the risk of downstream amplification steps magnifying amplification errors in the original amplification of the starting material is reduced, in particular as compared to existing whole genome, low input, amplification techniques. Likewise, being potentially of superior fidelity and completeness and evenness of coverage, the amplification product may be used to prepare sequencing libraries which are more accurate and complete, in particular as compared to those which may be prepared using existing whole genome, low input, amplification techniques.

Thus in a first aspect there is provided a method for amplification of ultra-low amounts of input DNA, said method comprising

- (i) providing a sample of DNA,
- (ii) contacting said DNA with a population of RNA oligonucleotide primers consisting of essentially equal proportions of A, U, G and C bases, but wherein each primer of said population has a randomised base sequence, and a Klenow fragment, or a DNA polymerase having a DNA-dependent DNA polymerase activity, a nucleic acid strand displacement activity, a lack of 5'→3' nuclease activity, and an ability to prime DNA polymerisation from an RNA primer which are equivalent to said Klenow fragment, under conditions which permit at least one round of primer binding to said DNA and subsequent DNA polymerisation from said bound primers catalysed by said Klenow fragment, or said equivalent thereof, thereby forming a first reaction product,
- (iii) contacting said first reaction product with a DNA polymerase I, or a DNA polymerase having the DNA-dependent DNA polymerase activity, 5'→3' nuclease activity and an ability to prime DNA polymerisation from an RNA primer which are equivalent to said DNA polymerase I, under conditions which permit at

- 8 -

least one round of primer binding and subsequent DNA polymerisation from said bound primers catalysed by said DNA polymerase I, or said equivalent thereof, thereby forming a second reaction product,

(iv) optionally contacting said second reaction product with a Taq DNA polymerase, or a DNA polymerase having the DNA polymerase activity equivalent to a Taq DNA polymerase, under conditions which permit DNA polymerisation to be catalysed by said Taq DNA polymerase, or said equivalent thereof, thereby forming a third reaction product, and

(v) contacting said second reaction product, or said third reaction product if step (iv) is performed, with an endonuclease or an exonuclease or combination thereof capable of nucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, under conditions which permit said nucleolytic degradation, thereby forming an predominantly double stranded DNA amplification product.

In a preferred embodiment step (v) is a step of contacting said second reaction product, or said third reaction product if step (iv) is performed, with at least an S1 nuclease, or a RNA endonuclease having the RNA endonuclease activity of said S1 nuclease, under conditions which permit the endonucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, thereby forming an predominantly double stranded DNA amplification product.

In another preferred embodiment step (v) is a step of contacting said second reaction product, or said third reaction product if step (iv) is performed, with at least a BAL 31 nuclease, or an nuclease having the exonuclease activity of said BAL 31 nuclease, under conditions which permit the exonucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, thereby forming an predominantly double stranded DNA amplification product.

In other embodiments at least both an S1 nuclease and a BAL 31 nuclease (or the above described functional equivalents) are used in step (v).

The RNA oligonucleotide primers of randomised sequence of use in the invention are randomised in the sense that individual primers consist of a sequence of A, U, G and C bases in which each base is drawn at random from a pool of said four bases in which there is essentially equal amounts of said four bases.

Numerically this may be expressed as a pool consisting of about 25% A bases, about 25% U bases, about 25% G bases and about 25% C bases, and which together total 100%, e.g. about 23-27% A bases, about 23-27% U bases, about 23-27% G bases and about 23-27% C bases and which together total 100%. Because each base of the primer sequence is chosen at random, individual primers within the population may have a composition of A, U, G and C bases which is not essentially equal, but the primer population as a whole will consist of essentially equal amounts of A, U, G and C bases. Numerically this may be expressed as a population consisting of about 25% A bases, about 25% U bases, about 25% G bases and about 25% C bases, and which together total 100%, e.g. about 23-27% A bases, about 23-27% U bases, about 23-27% G bases and about 23-27% C bases and which together total 100%.

The primers of use in the invention may be of any length capable of priming effective Klenow fragment catalysed polymerisation from essentially random sites across the starting DNA material (i.e. priming polymerisation from non-specific sites with minimal sequence bias). Preferably the primers of use in the invention will consist of about 5 to 25 ribonucleotides, e.g. about 5 to 22, 5 to 20, 5 to 18, 5 to 16, 5 to 15, 5 to 14, 5 to 13, 5 to 12, 5 to 11, 5 to 10, 5 to 9, 5 to 8, 5 to 7, 5 to 6, 6 to 25, 7 to 25, 8 to 25, 9 to 25, 10 to 25, 11 to 25, 12 to 25, 13 to 25, 14 to 25, 15 to 25, 16 to 25, 18 to 25, 20 to 25 or 22 to 25 ribonucleotides. Any ranges which may be constructed from the above range end-points are expressly contemplated. In more preferred embodiments the primer of use in the invention will consist of about 6 to about 16 ribonucleotides, e.g. about 7 to about 13, about 8 to about 12 or about 9 to about 11 ribonucleotides. It may be most preferable for the primers to be nonamers (9-mer), decamers (10-mer), undecamers (11-mer) or dodecamers (12-mer), i.e. to consist of 9, 10, 11 or 12 ribonucleotides, or mixtures thereof.

Modified forms of RNA maybe used, e.g. phosphorothioate protected RNA or 2'-O-methyl-substituted RNA.

In certain embodiments the population of RNA oligonucleotide primers of randomised sequence and essentially equal proportions of A, U, G and C bases of use in the invention may be contacted with the DNA sample together with additional populations of RNA oligonucleotide primers, e.g. a population of RNA oligonucleotide primers consisting of A, U, G and C bases in varying proportions but wherein each primer of said population has a randomised base sequence. For

instance the additional primer population may be enriched for 1, 2 or 3 of A, U, G or C bases, i.e. may contain more of 1, 2 or 3 of A, U, G or C bases than the other(s).

In certain embodiments the additional primer population may be enriched for A and U bases or G and C bases, i.e. the population has more than 50% of the enriched pair of bases, e.g. more than about 60%, about 65%, about 70%, about 75%, about 80%, about 90%, or about 95% of the enriched pair of bases. In these embodiments the primer populations of enriched sequence may help the amplification of regions of the DNA sample which is enriched with the complementary pair of bases

Other additional primer populations may contain non-standard bases.

The additional primer populations, e.g. those of enriched A and U bases or G and C bases, may be contacted with the DNA sample together with the population of randomised RNA oligonucleotide primers consisting of essentially equal proportions of A, U, G and C bases in any appropriate ratio. Typically the population of randomised RNA oligonucleotide primers of essentially equal proportions of A, U, G and C will be the greatest proportion of any mixture used, but this may not be the case in all embodiments. In such embodiments the population of randomised RNA oligonucleotide primers of essentially equal proportions of A, U, G and C will account for no more than about 30% of the primers contacted with the DNA sample, e.g. no more than about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 90%, or about 95% of the primers contacted with the DNA sample.

In other embodiments, an additional primer population containing primers comprising non-randomised sequences, e.g. sequences specifically complementary to predetermined sequences in the template and/or predetermined sequence tags or other functional sequences, is not used in step (ii) and/or in any of the other steps of the method. In still further embodiments of the method of the invention, in step (ii) only the population of randomised RNA oligonucleotide primers consisting of essentially equal proportions of A, U, G and C bases is contacted with the sample of DNA. In still further embodiments of the method of the invention, the population of randomised RNA oligonucleotide primers consisting of essentially equal proportions of A, U, G and C bases is the only population of oligonucleotide primers used in the method. In still further embodiments DNA oligonucleotide primers are not used in the method of the invention. In still further embodiments primers which are not RNA primers are not used in the method of the invention.

In preferred embodiments the Klenow fragment is the large protein fragment produced when DNA polymerase I from *E. coli* is enzymatically cleaved by the protease subtilisin. This fragment has a DNA-dependent 5'→3' DNA polymerase activity which can be primed by an RNA primer, a nucleic acid strand displacement activity (and a 3'→5' exonuclease activity for removal of precoding nucleotides and proofreading), but not a 5'→3' exonuclease activity of practical significance at about 37°C in an appropriate polymerase buffer (e.g. as described herein). A functional equivalent of this enzyme of use in the invention will have essentially the same (e.g. at least about 70%, 80%, 90%, 95% or 99% of the) DNA-dependent DNA polymerase activity which can be primed by an RNA primer, nucleic acid strand displacement activity, and optionally 3'→5' exonuclease activity, at 37°C as described above. Functional equivalents will also have essentially no 5'→3' exonuclease activity (i.e. any 5'→3' exonuclease activity which may be present will not be of practical significance in the context of the invention). This may be a DNA polymerase I from *E. coli* which has been expressed without the 5'→3' exonuclease subunit or with a non-functional mutant form of the 5'→3' exonuclease subunit. Appropriately modified DNA polymerase I homologs from other prokaryotic organisms, e.g. bacteria, or eukaryotic organisms, e.g. fungi, may be used as well as mutants of the above mentioned DNA polymerase I (e.g. mutants in the 3'→5' exonuclease activity or mutants with silent mutations).

In preferred embodiments, the polymerase used in this step also has essentially the same 3'→5' exonuclease activity as the Klenow fragment.

General references to a particular activity of an enzyme are general references to a particular catalytic property of that enzyme, usually under a particular set of physical and chemical conditions. Such properties may be considered in more specific terms to encompass the specificity of the enzyme for the relevant reactants, the rate at which the enzyme catalyses the relevant reaction and/or the yield of that enzyme catalysed reaction. In preferred embodiments all three properties will be considered when assessing an activity of an enzyme. When referring to enzymes having the "same" or "equivalent" activity, this is a comparison of activities which have been measured under essentially identical conditions.

As used herein terms expressing the direction of nucleic acid processing, as 5'→3' or 3'→5', may be alternatively expressed as 5' to 3' or 3' to 5', or 5'-3' or 3'-5', respectively, and each form of the term is interchangeable with one another.

Conditions which permit primer binding to the DNA template and subsequent DNA polymerisation from said bound primers catalysed by a Klenow fragment include any buffer routinely used with Mg^{2+} dependent polymerases (e.g. about 1 mM to about 20mM, preferably about 10 mM, Tris-HCl (pH 7.0-9.0, preferably 8.0 @ 25°C) containing about 1 mM to about 20mM, preferably about 10 mM, $MgCl_2$, and optionally 10 mM to 100mM, preferably about 50mM KCl), an appropriate amount of primers (e.g. about 0.1 nmol to 1.0 nmol, preferably about 0.4nmol), an appropriate amount of dNTPs (e.g. about 0.05 mM to about 0.5mM, preferably about 0.12 mM) and an appropriate amount enzyme (e.g. about 10U to about 50U, preferably about 23U) with a primer annealing incubation temperature of about 0 °C to about 40 °C, e.g. about 20 °C to about 35 °C or about 28 °C to about 32 °C, preferably about 30 °C for about 1 to about 20 minutes, e.g. about 10 to about 20 minutes, preferably about 15 minutes followed by a polymerisation incubation of about 10 °C to about 45 °C, e.g. about 20 °C to about 39°C or about 35 °C to about 39 °C, preferably about 37 °C for about 30 to about 120 minutes, preferably about 60 minutes. Optionally, a polymerisation step having a temperature of about 32 °C to about 45 °C, preferably about 35 °C to about 40 °C, preferably about 37 °C, may be preceded by an incubation of about 28 °C to about 32 °C, preferably about 30 °C for about 10 to about 20 minutes, preferably about 15 minutes. This optional step may be helpful by providing further opportunity for primers to find their target nucleotides sequences, and/or by extending the annealed primers slightly, hence stabilising primer-template complexes that may not be stable at 37 °C.

In certain embodiments the only DNA polymerase, or the only nucleic acid polymerase, present in step (ii) in an amount effective to contribute materially to the first reaction product is said Klenow fragment or functional equivalent thereof. In other embodiments the only DNA polymerase, or the only nucleic acid polymerase, present in step (ii) is said Klenow fragment or equivalent thereof.

In preferred embodiments the DNA polymerase I is the DNA polymerase I of *E. coli*. This enzyme has a DNA-dependent 5'→3' DNA polymerase activity which can be primed by an RNA primer and a 5'→3' exonuclease activity (and a 3'→5' exonuclease activity for removal of precoding nucleotides and proofreading) at about 37°C in an appropriate polymerase buffer (e.g. as described herein). A functional equivalent of this enzyme of use in the invention will have essentially the same (e.g. at least about 70%, 80%, 90%, 95% or 99% of the) DNA-dependent

DNA polymerase activity which can be primed by an RNA primer and 5'→3' exonuclease activity, and optionally 3'→5' exonuclease activity, at about 37°C as described above. Functional equivalents of use in the invention may include DNA polymerase I homologues from other prokaryotic organisms, e.g. bacteria, or eukaryotic organisms, e.g. fungi, and mutants of DNA polymerase I (e.g. mutants in the 3'→5' exonuclease activity or mutants with silent mutations).

Conditions which permit primer binding to the first reaction product (specifically the newly synthesised DNA strands) and subsequent DNA polymerisation from said bound primers catalysed by a DNA polymerase I include any buffer routinely used with DNA polymerases (e.g. about 1 mM to about 20mM, preferably about 10 mM, Tris-HCl (pH 7.0-9.0, preferably 8.0 @ 25°C) containing about 1 mM to about 20mM, preferably about 10 mM MgCl₂ and 10 mM to 100mM, preferably about 50mM KCl), an appropriate amount of primers (e.g. about 0.1 nmol to 1.0 nmol, preferably about 0.4nmol), an appropriate amount of dNTPs (e.g. about 0.05 mM to about 0.5mM, preferably about 0.12 mM) and an appropriate amount enzyme (e.g. about 1U to about 20U, about 1U to about 15U, or about 1U to about 10U, preferably about 5U per reaction) with a primer annealing incubation temperature of about 0 °C to about 40 °C, e.g. about 20 °C to about 35 °C or about 28 °C to about 32 °C, preferably about 30 °C for about 1 to about 20 minutes, e.g. about 10 to about 20 minutes, preferably about 15 minutes followed by a polymerisation incubation of about 10 °C to about 45 °C, e.g. about 20 °C to about 39°C or about 35 °C to about 39 °C, preferably about 37 °C for about 1 to about 60 minutes, preferably about 20 minutes. Optionally, a polymerisation step having a temperature of about 32 °C to about 45 °C, preferably about 35 °C to about 40 °C, preferably about 37 °C, may be preceded by an incubation of about 28 °C to about 32 °C, preferably about 30 °C for about 10 to about 20 minutes, preferably about 15 minutes. This optional step may be helpful by providing further opportunity for primers to find their target nucleotides sequences, and/or by extending the annealed primers slightly, hence stabilising primer-template complexes that may not be stable at 37 °C.

In certain embodiments the only DNA polymerase, or the only nucleic acid polymerase, present in step (iii) in an amount effective to contribute materially to the second reaction product is said DNA polymerase I or functional equivalent thereof. In other embodiments the only DNA polymerase, or the only nucleic acid polymerase, present in step (ii) is said DNA polymerase I or equivalent thereof.

In preferred embodiments the Taq polymerase is the thermostable DNA polymerase of *Thermus aquaticus*. This enzyme has a 5'→3' DNA polymerase activity at about 68 °C (and a 5'→3' exonuclease activity) in an appropriate polymerase buffer (e.g. as described herein). A functional equivalent of this enzyme of use in the invention will have essentially the same (e.g. at least about 70%, 80%, 90%, 95% or 99% of the) DNA polymerase activity, and optionally 5'→3' exonuclease activity, at about 68 °C as described above. Functional equivalents of use in the invention may include Taq homologues from other prokaryotic organisms, e.g. bacteria, or eukaryotic organisms, e.g. fungi, and mutants of Taq polymerase (e.g. mutants in the 5'→3' exonuclease activity or mutants with silent mutations).

Conditions which permit DNA polymerisation to be catalysed by said Taq DNA polymerase upon contact of the second reaction product with Taq DNA polymerase include any buffer routinely used with DNA polymerases (e.g. about 1 mM to about 20mM, preferably about 10 mM, Tris-HCl (pH 7.0-9.0, preferably 8.0 @ 25°C) containing about 0.1 mM to about 10mM, preferably about 1.6 mM MgCl₂ and 10 mM to 100mM, preferably about 50mM KCl), an appropriate amount of dNTPs (e.g. about 0.5 mM to about 1.5mM, preferably about 0.8 mM) and an appropriate amount enzyme (e.g. about 0.1U to about 5U, preferably about 2U per reaction) with a polymerisation incubation of about 50 °C to about 75 °C, about 60 °C to about 75 °C, or about 65 °C to about 75 °C, preferably about 68 °C for about 10 seconds to about 40 minutes, about 1 minute to about 40 minutes, about 10 minutes to about 40 minutes, preferably about 20 minutes.

In certain embodiments the only DNA polymerase, or the only nucleic acid polymerase, present in step (iv) in an amount effective to contribute materially to the third reaction product is said Taq DNA polymerase or functional equivalent thereof. In other embodiment the only DNA polymerase, or the only nucleic acid polymerase, present in step (iv) is said Taq DNA polymerase.

The endonuclease, exonuclease or combination thereof of use in step (v) of the method of the invention is not limited so long as under suitable chemical and physical conditions the enzymes or combination thereof are capable of nucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, thereby forming an predominantly double stranded DNA amplification product (e.g. an amplification product the nucleic acid component of which contains

50% or more double stranded DNA, e.g. 60%, 70%, 80%, 85%, 90%, 95%, 98% or 99% or more double stranded DNA). Preferably said degradation of at least a portion of any remaining single stranded or annealed RNA is achieved without significant degradation of double stranded DNA. Suitable examples are S1 nuclease and nuclease BAL-31. The skilled person would be able to determine the quantity of enzyme, e.g. S1 and/or BAL-31, to use and also appropriate conditions, e.g. buffers, temperatures and reaction times, suitable for the selected enzymes which prevent excessive degradation of the amplification product.

S1 nuclease (EC 3.1.30.1, formerly EC 3.1.4.21) is also known as *Aspergillus* nuclease S1, deoxyribonuclease S1, endonuclease S1 and single-stranded-nucleate endonuclease. This enzyme catalyses the endonucleolytic cleavage of single stranded RNA (and single stranded DNA) at 25°C in an appropriate polymerase buffer (e.g. as described herein). A functional equivalent of this enzyme of use in the invention will have essentially the same (e.g. at least about 70%, 80%, 90%, 95% or 99% of the) RNA (e.g. single stranded RNA), and optionally the DNA, endonuclease activity, at about 25 °C as described above. Functional equivalents of use in the invention may include S1 homologues from other fungal or plant organisms and more specifically *Neurospora crassa* nuclease, mung bean nuclease, *Penicillium citrinum* nuclease P1 and RNaseH and mutants thereof (e.g. mutants in with silent mutations or mutants in the DNA endonuclease activity).

Conditions which permit the endonucleolytic degradation of any single stranded RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, (e.g. as a component of RNA, DNA or RNA-DNA duplex molecules) include a buffer containing about 0.01M to about 0.1M, preferably about 0.05M, sodium acetate (pH 4.0-5.0, e.g. 4.5, at 25°C), about 0.1 to about 0.5 M, preferably about 0.3M, NaCl, and 1 mM to about 10mM, preferably about 0.45mM ZnSO₄, an appropriate amount enzyme (e.g. about 1U to about 100U, preferably about 45U per reaction) and an incubation temperature of about 0°C to about 50°C, e.g. about 10°C to about 50°C, about 25°C to about 50°C or about 37°C to about 50°C or about 45°C for about 2 minutes to about 30 mins, e.g. about 5 mins to about 15 mins or about 5 mins to about 10 mins .

As shown in the Examples it has been surprisingly found that incubation of the S1 nuclease with the third reaction product at about 45°C (e.g. about 42°C to 48°C, 43°C to 47°C or 44°C to 46°C) for about 5 mins (e.g. about 1 min to about 10

mins, about 2 mins to 8 mins, 3 mins to 7 mins, or 4 mins to 6 mins) results in improved yields of amplification product in the shortest time compared to other temperatures and incubations time, especially when about 45U (e.g. about 30U to 60U, 35U to 55U, 40U to 50U or 42U to 48U) of nuclease was used in each reaction. This is significantly different to the standard conditions in which S1 nuclease is used, i.e. at room temperature with an incubation time of around 30mins. Without wishing to be bound by theory, the advantageous conditions described above result in the efficient degradation of the RNA in the second reaction product, or said third reaction product if step (iv) is performed, without significant collateral degradation of the DNA.

In these embodiments a functional equivalent of S1 nuclease of use in the invention will have essentially the same (e.g. at least about 70%, 80%, 90%, 95% or 99% of the) single stranded RNA, and optionally DNA, endonuclease activity (e.g. reaction rate, specificity and/or yield), at about 45 °C, as described above.

Steps (ii) and (iii) may independently comprise more than one round of primer annealing and subsequent DNA polymerisation, e.g. at least 2, 3, 4, 5, 8, 10, 15 or 20 rounds of primer annealing and subsequent DNA polymerisation. In preferred embodiments specificity of amplification and uniformity of coverage is maximised by limiting the number of rounds of primer annealing and subsequent DNA polymerisation to no more than 5, e.g. no more than 4, 3, 2 or 1 rounds of primer annealing and subsequent DNA polymerisation.

Steps (iii), (iv) and (v) require that the reaction product of the preceding step undergo further enzymatic manipulation as defined in said steps. Although it may be preferable that the direct product of the preceding step is the product to which the described enzymatic manipulation is applied, this should not be considered to exclude intervening processing steps, e.g. the addition of further components or the further modification or processing of certain of the components of the reaction product of the preceding step, so long as essentially none of the components of the reaction product, or at least none of the nucleic acid components, have been removed or substantially modified. The reaction product may therefore be described as a nucleic acid reaction product.

The DNA sample may come from any source, e.g. prokaryotic, eukaryotic or viral. Preferably the DNA is genomic DNA, e.g. from a eukaryote (animal, plant or fungus). Preferably it comprises the entire genome of a eukaryote. The eukaryotic genome may be haploid (e.g. comprising a gamete or a cell from another haploid

stage of an organism), or multiploid (e.g. diploid). It may be essentially pure or partially purified. In other embodiments the DNA is low copy number environmental DNA or DNA circulating in the fluids of an organism.

In certain embodiments the genomic DNA sample has a mass of less than about 1ng, e.g. less than about 500pg, 200pg, 100pg, 75pg, 50pg, 40pg, 30pg, 20pg, 10pg, 9pg, 8pg, 7pg, 6pg, 5pg, 4pg, 3pg, 2pg or 1pg. As appropriate in these embodiments the genomic DNA sample has a mass of greater than about 1pg, e.g. greater than about 2pg, 3pg, 4pg, 5pg, 6pg, 7pg, 8pg, 9pg, 10pg, 20pg, 30pg, 40pg, 50pg, 75pg, 100pg, 200pg, or 500pg. Any ranges which may be constructed from the above values are expressly contemplated.

In other embodiments the DNA sample, in particular the genomic DNA sample, contains less than 10000 DNA molecules, e.g. less than 5000, 1000, 500, 100, 50, 25 or 10 DNA molecules.

In the above embodiments the mean size of the DNA molecules in the sample may be at least 0.5 Mbp, e.g. at least 1, 2, 5, 10, 15 or 20 Mbp.

The above amounts of DNA sample may be taken as “ultra-low” amounts of input DNA.

In preferred embodiments the DNA sample is the genomic DNA from a single cell, i.e. a whole genome. Eukaryotic cells, including animal, plant or fungal cells are notable embodiments. In these instances the DNA sample may be the product of the lysis of a single cell or a partial purified product in which cell fragments and other components have been removed from the products of a lysis step. Techniques for lysing cells are routine in the art and any which do not deplete genomic DNA may be used in these embodiments. Thus, in certain embodiments the method of the invention comprises a preceding step of cell lysis with the optional step of removing a portion of the resulting cell fragments, e.g. by centrifugation. Typically a step of removing nucleic acid bound protein will also be included

As discussed above, the method of the invention provides for the amplification of ultra-low amounts of DNA with a high degree of coverage (the proportion of the nucleotide sequence of the starting DNA material present in the amplification product) and uniformity (the extent to which the proportions of discrete regions of the nucleotide sequence of the starting DNA material are maintained in the amplification product), e.g. across different loci in a genome. This may be evaluated in terms of the genomic coverage and read uniformity of a parallel

sequencing technique performed on the amplification product to a given sequencing depth. "Specificity of amplification" or the "specificity of the amplification product" refers to the extent by which off target products are avoided during the amplification of the starting DNA material and/or the sequence fidelity of the amplification product to the nucleotide sequence of the starting DNA material insofar as the nucleotide sequence of the starting DNA material is present in the amplification product. Put differently, amplification is specific if it essentially only produces copies of the starting DNA material. Specificity of amplification is therefore not affected by the extent of coverage or any lack thereof.

In this regard, in accordance with the invention, an amplification product that is superior to that provided by existing whole genome, low input, amplification techniques in terms of sequence coverage and the uniformity of that coverage is an amplification product which, when subjected to a parallel sequencing technique to a given sequencing depth, will provide a level of genomic coverage and/or read uniformity which will be superior to the level of genomic coverage and/or read uniformity which would be provided by the same parallel sequencing technique at the same sequencing depth using an amplification product generated by MALBAC, MDA, GenomePlex or Picoplex using the same starting DNA material and amounts thereof.

In the context of nucleic acid sequencing "sequencing depth" refers to the size of a sequencing dataset in relation to a reference genome, i.e. the amount of bases sequenced is divided by the predicted number of bases in the reference genome. Also in these contexts "genome coverage" defines the percentage of the predicted genomic information within a reference genome which is present in a sequencing dataset. It should be noted that this is independent of whether or not 100% of the genomic information within a reference genome is present in the starting DNA material.

Thus, in the context of the invention, the genome coverage of a parallel sequencing technique performed on a test amplification product may, in certain embodiments, be assessed (measured) by comparing the amount of sequence information obtained from a test amplification product at a given sequencing depth with the amount of sequence information obtained from a purified bulk sample of the starting DNA material at the same or greater sequencing depth using the same sequencing technique. In other words, a purified bulk sample of the starting DNA

material is used as a convenient reference genome. Other means of assessment may be used.

In the context of the invention the read uniformity of a parallel sequencing technique performed on a test amplification product may be assessed (measured) by plotting a Lorenz curve showing the distribution of the total sequencing reads over the available fraction of the genome for a sequenced test amplification product and comparing that curve to a corresponding curve obtained from the sequence data from a purified bulk sample of the starting DNA material. The closer the test and bulk sample curves match, the greater the read uniformity of the parallel sequencing technique performed on the test amplification product. Alternatively, the Lorenz curve of the test amplification product may be compared to the mathematical ideal of absolute uniformity, wherein the closer the test curve matches the ideal, the greater the read uniformity. Other means of assessment may be used.

Expressed numerically, the methods described herein provide an amplification product containing DNA molecules carrying between them an amount of the nucleotide sequence of the starting DNA material which may provide at least 40% of the sequence information provided by a purified bulk sample of the starting DNA material when the amplification product is subjected to a parallel sequencing technique with an approximate sequencing depth of at least 3 and the isolated bulk sample of the starting DNA material is subjected to a parallel sequencing technique with an approximate sequencing depth of equal to or greater than the sequencing depth to which the amplification product is sequenced.

In more specific embodiments at least 45%, e.g. at least 50%, 55%, 60%, 65%, 70%, 75% or 80%, of the nucleotide sequence information provided by a purified bulk sample of the starting DNA material may be provided by the amplification product.

In more specific embodiments the amplification product and, independently, the purified bulk sample of the starting DNA material is sequenced to an approximate sequencing depth of at least 4, e.g. at least 5, 6, 7, 8, 9, 10, 15 or 20, wherein the purified bulk sample of the starting DNA material is sequenced to an approximate sequencing depth of equal to or greater than the sequencing depth to which the amplification product is sequenced.

In still more specific embodiments at least 55%, e.g. at least 60%, 65%, 70%, 75% or 80% of the nucleotide sequence information provided by a purified

bulk sample of the starting DNA material may be provided by the amplification product when the amplification product is sequenced to an approximate sequencing depth of at least 5, e.g. at least 6, 7, 8, 9, 10, 15 or 20, wherein the purified bulk sample of the starting DNA material is sequenced to an approximate sequencing depth of equal to or greater than the sequencing depth to which the amplification product is sequenced.

In still more specific embodiments at least 65%, e.g. at least 70%, 75% or 80% of the nucleotide sequence information provided by a purified bulk sample of the starting DNA material may be provided by the amplification product when the amplification product is sequenced to an approximate sequencing depth of at least 7, e.g. at least 8, 9, 10, 15 or 20, wherein the purified bulk sample of the starting DNA material is sequenced to an approximate sequencing depth of equal to or greater than the sequencing depth to which the amplification product is sequenced.

In still more specific embodiments at least 70%, e.g. at least 75% or 80% of the nucleotide sequence information provided by a purified bulk sample of the starting DNA material may be provided by the amplification product when the amplification product is sequenced to an approximate sequencing depth of at least 8, e.g. at least 9, 10, 15 or 20, wherein the purified bulk sample of the starting DNA material is sequenced to an approximate sequencing depth of equal to or greater than the sequencing depth to which the amplification product is sequenced.

In further embodiments, the method of the present invention may also show consistent amplification performance across a plurality of separate input DNA samples which have been prepared in the same way, in particular as compared to the consistency of existing whole genome, low input, amplification techniques. This consistency of amplification performance can be assessed by comparing the genomic coverage of a parallel sequencing technique performed on the amplification products of a plurality of separate input DNA samples which have been prepared in the same way, as measured above. The more consistent the readings across the plurality of replicates, the more consistent the method. In accordance with the invention the genomic coverage of a parallel sequencing technique performed on the amplification products of a plurality of separate input DNA samples which have been prepared in the same way preferably differs by less than +/- about 40%, e.g. less than +/- about 35%, +/- about 30%, +/- about 25%, +/- about 20%, +/- about 15%, or +/- about 10%,

The parallel sequencing technique is preferably a high throughput technique based on a sequencing by synthesis approach, e.g. pyrosequencing (a DNA sequencing technique that relies on detection of pyrophosphate release upon nucleotide incorporation; Nyrén, P., 2007, *Methods Mol Biology*, 373: 1–14) or Illumina dye sequencing (a DNA sequencing method that relies on clonal amplification and reversible dye terminators; Canard, B. and Sarfati, R. S., 1994, *Gene*, 148 (1): 1–6, and WO 1998/044151). The parallel sequencing technique is preferably a technique which produces short reads, e.g. Illumina dye sequencing.

In these contexts a purified bulk sample of the starting DNA material is a sample of the starting DNA material that is sufficiently refined and intact to allow effective sequencing with the chosen parallel sequencing technique and which contains at least about 1 ng of DNA, preferably at least about 2 ng, e.g. at least about 5ng, 10 ng, 20 ng, 30ng 40 ng, or 50 ng. For the purposes of the invention the features of a method of the invention may be evaluated using sperm cells – with the input DNA material being that obtained from a single sperm cell and the bulk sample being that obtained from a number of sperm cells – as shown in the Examples. For the purposes of the invention the features of a method of the invention may be evaluated using GM19382 human lymphoblastoid cells – with the input DNA material being that obtained from a single GM19382 cell and the bulk sample being that obtained from a number of GM19382 cells – as shown in the Examples.

In the above embodiments the amplification product preferably comprises little, substantially few, or no, chimeric or non-specific sequences.

The double stranded DNA amplification product obtainable in accordance with the method of the invention may be used in whole-genome sequencing, copy-number profiling and other analytical techniques, e.g. targeted analysis such as detection, sequencing or genotyping, and for purposes such as the study of cellular heterogeneity and other single-cell research, forensics, the study of recombination, analysis of polar bodies, and pre-implantation or in utero genetic diagnosis. The amplification product of the invention may be used directly in such techniques or may undergo further processing prior to inputting into these downstream methods. As outlined in the Examples such further processing may include, but not be limited to, purification, end-repair, size selection, dA-tailing, adaptor ligation, targeted loci enrichment and further amplification. Such processing is routine in the art and will be familiar to the skilled man. Solid phase reversible immobilization (SPRI)

techniques to purify DNA are further processing techniques of note as such techniques have been found to provide good yield of DNA recovery and productivity and have the ability to separate DNA fragment based on size. Ampure XP is mentioned specifically.

Thus, in further embodiments the method of the invention comprises one or more further steps in which the double stranded DNA amplification product is exposed to one or more steps of purification, end-repair, size selection, dA-tailing, adaptor ligation and/or further amplification.

One of the advantages of the present invention is the provision of a double stranded DNA amplification product that reflects the nucleotide sequence of the ultra-low amount of starting DNA material sufficiently faithfully in terms of its sequence coverage, the uniformity of that coverage and its specificity such that the chances of experimental artefacts arising from the amplification of the starting material becoming apparent in downstream techniques, in particular nucleic acid sequencing techniques and other forms of nucleic acid sequence analysis, are kept to a minimum, in particular as compared to existing whole genome, low input, amplification techniques.

Thus, in a further aspect of the invention there is provided a method of optimising the nucleic acid sequence analysis of ultra-low amounts of DNA, said method comprising using the DNA amplification method as defined herein to accurately and uniformly amplify the ultra-low amount of DNA prior to the analysis of the nucleic acid sequence of said DNA.

The term "optimisation" encompasses an improvement in the accuracy and/or sensitivity of the sequence analysis of the ultra-low amount of DNA.

In a further aspect of the invention there is provided a method of nucleic acid sequence analysis, said method comprising a step of sample preparation prior to the analysis step(s) in which the DNA amplification method as defined herein is used to amplify an ultra-low amount of DNA in said sample.

In these latter aspects the features of the method of the invention as described above apply mutatis mutandis and may be recited in full, in any appropriate permutation, in place of the reference to using the method.

Preferably in these latter aspects the nucleic acid sequence analysis is a sequencing technique, e.g. the Sanger dideoxynucleotide sequencing method or a "next generation" or "second generation" sequencing approach (for instance, those involving pyrosequencing, reversible terminator sequencing, cleavable probe

sequencing by ligation, non-cleavable probe sequencing by ligation, DNA nanoballs, and real-time single molecule sequencing, e.g. nanopore based sequencing) or an oligonucleotide hybridisation probe based approach in which the presence of a target nucleotide sequence is confirmed by detecting a specific hybridisation event between a probe and its target.

The nucleic acid sequence analysis may provide information useful in the genotyping of an organism, e.g. for classification, identification, quantification, prognostic, diagnostic and/or forensic applications.

The invention will now be described by way of non-limiting Examples with reference to the following figures in which:

Figure 1 shows the profiles of amplified genomic DNA products obtained from the method of the invention at the stage of S1 nuclease digestion as seen on an E-gel® EX 1% agarose gel. 2 primer pools with different ratios of randomised RNA primers (equal AUGC:GC rich:AU rich) were tested in this experiment: A – 13:2:5 and B 10:3:7. DNA ladder loading: 50ng NEB 2-logs DNA ladder were loaded per each lane of marker (M)

Figure 2 shows the profiles of Illumina sequencing libraries made from shorter DNA fractions obtained from the method of the invention as seen on an E-gel® EX 1% agarose gel. 2 primer pools with different ratios of randomised RNA primers (equal AUGC:GC rich:AU rich) were tested in this experiment: A – 13:2:5 and B 10:3:7. Sample loading: 4 µl (20%) of unpurified PCR products from each corresponding PCR reaction (16 cycles) were subjected to the gel analysis. Ladder loading: 50ng NEB 2-logs DNA ladder were loaded per each lane of marker (M).

Figure 3 shows the impact of S1 nuclease digestion conditions on the yield of shorter DNA fractions obtained from the method of the invention. Different amounts of S1 nuclease (10 and 45 units) under different incubation conditions were tested (A - 37 °C for 10 minutes followed by 45 °C for 2.5 minutes; B - 45 °C for 5 minutes; C - 50 °C for 5 minutes).

Figure 4 shows the impact of S1 nuclease digestion conditions on production of Illumina sequencing libraries made from the shorter DNA fractions obtained from the method of the invention. Different amounts of S1 nuclease (10 and 45 units)

under different incubation conditions were tested (A - 37 °C for 10 minutes followed by 45 °C for 2.5 minutes; B - 45 °C for 5 minutes; C - 50 °C for 5 minutes).

Figure 5 shows the impact of S1 nuclease digestion conditions on production of PCR amplified (10 cycles) longer DNA fractions obtained from the method of the invention. Different amounts of S1 nuclease (10 and 45 units) under different incubation conditions were tested (A - 37 °C for 10 minutes followed by 45 °C for 2.5 minutes; B - 45 °C for 5 minutes; C - 50 °C for 5 minutes).

Figure 6 shows the percentage of the genome represented by one or more sequencing reads within the sequenced material from 69 separate sequencing experiments: bulk DNA were purified from a proportion of a pool of mouse sperm cells, from which 67 individual sperm cells were also isolated. The sequencing data of the 2 bulk DNA samples had a calculated sequencing depth of ~4 (the two higher traces), whereas the shorter fractions of the amplification products of the 67 single sperm samples were sequenced to an approximate sequencing depth of 3.6-8.5. All data have been normalised by removing regions not covered by any reads within the sequencing data obtained from the 2 ng bulk DNA sample that was sequenced to an approximate sequencing depth of 20, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means. When the single sperm amplification samples were sequenced, the reads obtained represent approximately 45% to 72% (median 62%) of the genome that is accessible to Illumina sequencing, as judged by setting the coverage of a bulk mouse sperm DNA sample, sequenced to an approximate sequencing depth of 20, to 100%.

Figure 7 shows a Lorenz curve presenting the distribution of total sequencing reads over the available fraction of genome for the 69 separate sequencing experiments described in Figure 6. All data have been normalised by removing regions not covered by any reads within the sequencing data obtained from the 2 ng bulk DNA sample that was sequenced to an approximate sequencing depth of 20, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means. The straight dotted line represents the idealised non-random completely uniform sequencing coverage

of the genome. The two traces closest to the dotted line represent the results from the bulk DNA sample.

Figure 8 shows the percentage of the genome represented by one or more sequencing reads within the sequenced material from 7 separate sequencing experiments: bulk DNA was purified from a proportion of a pool of human GM19382 cells, from which 6 individual cells were also isolated. The sequencing data of the bulk DNA sample had a calculated sequencing depth of ~60 (higher trace) and the combined shorter and longer fractions of the amplification products of the 6 single cell samples were also sequenced to an approximate sequencing depth of 60. All data have been normalised by removing regions not covered by any reads within the sequencing data obtained from the bulk DNA sample that was sequenced to an approximate sequencing depth of 60, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means. When the single-cell amplification samples were sequenced, the reads obtained represent approximately 89% to 93% (median 92%) of the genome that is accessible to Illumina sequencing, as judged by setting the coverage of the bulk DNA sample, sequenced to an approximate sequencing depth of 60, to 100%.

Figure 9 shows a Lorenz curve presenting the distribution of total sequencing reads over the available fraction of genome for the 7 separate sequencing experiments described in Figure 8. All data have been normalised by removing regions not covered by any reads within the sequencing data obtained from the bulk DNA sample that was sequenced to an approximate sequencing depth of 60, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means. The straight dotted line represents the idealised non-random completely uniform sequencing coverage of the genome. The trace closest to the dotted line represents the results from the bulk DNA sample.

EXAMPLES**Example 1 – Single Cell Whole Genome Amplification (SCWGA) Protocol****Reagents:**

Primers and oligonucleotides:

Unprotected random RNA decamers

Name	A:U:G:C %	Stock Concentration
Random RNA Decamer (RRD)	25:25:25:25	500 µM
Random RNA Decamer _GC (RRD_GC)	10:10:40:40	500 µM
Random RNA Decamer _AU (RRD_AU)	40:40:10:10	500 µM

The primer pool (R10mer) is used in SCWGA, and the primer ratio is

$$\text{RRD:RRD_GC:RRD_AU} = 46:1:3$$

PDON adaptor and primers

PDON-R21-adaptor I: **/5Phos/GTG TCT CAA AAT CTC TGA TGT/3BioTEG/** (SEQ ID NO: 3)

PDON-R21-adaptor II: **ACA TCA GAG ATT TTG AGA CAC *T** (SEQ ID NO: 4)

PDON-R21-PCRprimer: **ACA TCA GAG ATT TTG AGA CAC** (SEQ ID NO: 5)

Compounds and compositions:

DEPC Treated water	Invitrogen
Polyvinylpyrrolidone	PVP40, Sigma
DTT	Molecular biology grade (Sigma, stored at 4°C)
KOH	Molecular biology grade (Sigma)
1 M MgCl ₂	Sigma, M1028 (stored at -20°C)
3 M KCl	Fluka, 60135 (stored at room temperature)

dNTP	Mix of dA/T/G/CTP at concentration of 25 mM each (BioLine, stored at -20°C)
dATP	100 mM (NEB, N0440S, stored at -20°C)
1M Tris-HCl, pH 8.0	Molecular biology grade (Sigma)
Lysis solution (2x)	0.4 N KOH and 100 mM DTT
Neutralisation solution	0.2N HCl ₂ and 0.15M Tris-HCl (pH 8.0) (stored at -20°C)
10x Klenow Buffer	100 mM Tris-HCl (pH8.0 @ 25°C) and 100 mM MgCl ₂ (stored at -20°C)
10x Mg-free Pol I Buffer	100 mM Tris-HCl (pH8.0 @ 25°C), 500 mM KCl and 10 mM DTT (stored at -20°C)
10x RNA ligase Buffer	500 mM Tris-HCl, 100 mM MgCl ₂ and 10 mM DTT (pH 7.5 @ 25°C) (stored at -20°C)
100x TE buffer	1 M Tris-HCl and 0.1 M EDTA (pH 8.0) (Sigma)
Proteinase K solution	20 mg/ml (Invitrogen, stored at 20°C)
Klenow Fragment (KF)	77 u/μl, P7060H, Enzymatics (stored at -20°C) 5u/μl, P7060, Enzymatics (stored at -20°C)
E.coli DNA Polymerase I (Pol I)	10 u/μl, P7050, Enzymatics (stored at -20°C)
Taq DNA polymerase	5 u/μl (NEB M0320S), supplied with 10 x Standard Mg-free Buffer (100 mM Tris-HCl, 500 mM KCl, pH 8.3 @ 25°C) (stored at -20°C)
S1 nuclease	89 u/μl, (Promega M5761), supplied with 10x S1 nuclease buffer (0.5M sodium acetate (pH 4.5 at 25°C), 2.8M NaCl, 45mM ZnSO ₄ .) (stored at -20°C)
NEBNext® End-repair module	New England Biolab, E6050 (stored at -20°C)
NEBNext® dA-tailing module	New England Biolab, E6053 (stored at -20°C)
Quick Ligation™ Kit	New England Biolab, M2200 (stored at -20°C)
Q5® High-Fidelity 2X Master	New England Biolab M0492
DNA purification reagent	Ampure XP
Library analysis	High Sensitivity D1000 TapeStation kit (Agilent)
DNA quantification	QuantiFluor (Promega) 384 low volume black plate (Greiner, 784076)

Equipment:

FLUOstar OPTIMA (BMG Labtech)

Experimental procedures:**DNA preparation**Lysis of cells

- Individual cells are isolated in 1µl 0.25% PVP40 solution.
- Add 1µl lysis solution to each isolated cell, mixed briefly and then recollect the mixture by a brief centrifugation
- Incubate the samples on ice for 10 minutes

Neutralisation

- Add 2 µl neutralisation solution to each sample and mix gently
- Recollect the solution by a brief centrifugation

Proteinase K digestion:

- Add 0.5 µl of proteinase K (1:5 diluted in water) to each sample
- Mix gently and recollect the solution by a brief centrifugation
- Incubate the samples at 50°C overnight on a PCR machine
- Continue to incubate at 65°C for 10 minutes and then 80°C for 10 minutes on the PCR machine

Klenow fragment amplificationR10mer primer mix (µl)

Components in reaction	Volume per reaction	Final concentration in reaction
10x KF Buffer	0.6	1x
R10mer	0.8	50 µM
H ₂ O	0.6	
Total	2.0	

Klenow fragment mix (μ l)

Components in reaction	Volume per reaction	Final concentration in reaction
10x KF Buffer	0.20	1x
dNTP (5 mM total)	0.25	0.12 mM
Klenow Fragment (77 u/ μ l)	0.30	~3 u/ μ l
H ₂ O	1.25	
Total	2.00	

Method

- Add 2 μ l **R10mer primer mix** to each sample
- Mix gently and recollect the solution by a brief centrifugation
- Incubate at 94 °C for 3 minutes on a PCR machine and chill the samples in an icy water bath immediately
- Add 2 μ l **Klenow fragment mix** to each sample
- Mix gently and recollect the solution by a brief centrifugation
- Incubate at 30 °C for 15 minutes, 37 °C for 60 minutes, followed by 65 °C for 5 minutes, and then hold at 4 °C

Double-strand formation with Pol I and Taq DNA polymerasePol I reaction mix (μ l)

Components in reaction	Volume per reaction	Final concentration in reaction
10x Mg-free Pol I Buffer	0.2	1x*
dNTP (5 mM total)	0.1	~0.12 mM
H ₂ O	1.2	
DNA Pol I (10 u/ μ l)	0.5	~0.5 u/ μ l
Total	2.0	

* Final Mg⁺⁺ concentration in reaction is ~8 mM

- 30 -

Taq polymerase mix (μ l)

Components in reaction	Volume per reaction	Final concentration in reaction
10x Taq standard Mg-free Buffer	4.0	1x**
dNTP (40 mM total)	0.8	~0.8 mM
H ₂ O	34.8	
Taq DNA polymerase (5 u/ μ l)	0.4	0.04 u/ μ l
Total	40.0	

** Final Mg⁺⁺ concentration in reaction is ~1.6 mM

Method

- Heat the Klenow fragment reaction samples at 94°C on a PCR machine for 3 minutes, and immediately chill in an icy water bath.
- Add 2 μ l **Pol I reaction mix** to each sample.
- Mix gently and recollect the solution by a brief centrifugation.
- Incubate at 30 °C for 15 minutes, followed by 37 °C for 20 minutes.
- Add 40 μ l **Taq polymerase mix** to each sample, mix well, collect the solution by a brief centrifugation and incubate at 68 °C for 20 minutes.

The samples can be stored at -20 °C.

S1 nuclease digestionS1 mix (μ l)

Components in reaction	Volume per reaction	Concentration in reaction
10x S1 Buffer	7.0	1.2x
S1 nuclease (89 u/ μ l)	0.5	~0.8 u/ μ l
Total	7.5	

Method

- Add 7.5 µl **S1 mix** to each sample, following double strand formation
- Mix gently and recollect the solution by a brief centrifugation
- Incubation the sample at 45°C for 10 minutes
- Add 8 µl 100x **TE buffer** to each sample to stop the reaction
- Mix gently and recollect the solution by a brief centrifugation

Purification of DNA_using Ampure XP kit

The following protocol is based on the manufacturer's instructions.

- Warm the beads to room temperature and mix thoroughly before use.
- Add 65 µl mixed beads to each S1 nuclease treated sample, in order to achieve the DNA:beads ratio $\approx$ 1: 1
- Close the tubes, mix the suspension gently and incubate at room temperature for 5 minutes
- Place the samples on a magnetic stand, when the beads have been collected to the wall of the tube and the solution is clear, remove and discard the liquid. Keep the tubes on the stand until the washing step completed.
- Wash the beads by adding ~200 µl of freshly prepared 70% ethanol to the beads and incubating for 2 minutes at room temperature. Remove and discard the ethanol.
- Repeat the washing once.
- Keeping the tubes on the magnet and the caps open, dry the beads at room temperature for 3 minutes and remove any trace of ethanol from the bottom of the tube.
- Remove the tubes from the magnetic stand. Check and make sure no larger size droplets present in the tube.
- Add 16 µl of water to the dried beads and resuspend the beads thoroughly.
- Place the tubes on the magnetic stand until the solution is clear. Transfer the liquid to a fresh tube. The liquid contains purified DNA.

End-repair using NEBNext® End-repair moduleEnd-repair pre-mix (µl)

Components in reaction	Volume per reaction
10x End-repair Buffer	2.0
End-repair enzyme mix	0.5
Klenow Fragment (5 u/µl)	0.2
H ₂ O	1.3
Total	4.0

Method

- Mix 4 µl of **End-repair pre-mix** with the 16 µl purified DNA
- Incubate at 30 °C for 30 minutes

Purification/size selection of longer & shorter DNA fragments using Ampure XP kitConcentrating of the XP beads for use in the purification of shorter DNA fragments

For each sample, transfer 80 µl mixed XP beads to a fresh 1.5 or 2 ml tube, as appropriate, and place the tube on a magnetic stand. When the beads have been collected to the wall of the tube and the solution is clear, remove and discard 40 µl of the clear solution per sample. Remove the tube and resuspend the beads in the remaining solution. Add 40 µl of the concentrated XP beads to a set of fresh 0.2 ml tubes and the beads are ready to be used in the purification of shorter DNA fragments.

Purification of longer DNA fragments

- Mix 60 µl of water with 40 µl mixed XP beads per sample, and then add to each sample, following end-repair (the DNA:beads ratio = 1: 0.5).
- Close the tubes, mix the suspension gently and incubate at room temperature for 5 minutes.

- 33 -

- Place the samples on a magnetic stand, until the beads have been collected to the wall of the tube and the solution is clear.
- Transfer the **entire liquid fraction** from individual samples to the corresponding tubes **containing concentrated XP beads (above)** and mix well to form a **DNA/beads suspension for purification of shorter DNA fragments** (see below for the method for purifying shorter DNA fragments which begins from this mixture).
- Wash the **bead pellets** in the tubes on the magnetic stand by adding ~200 µl of freshly prepared 70% ethanol to the beads and incubating for 2 minutes at room temperature. Remove and discard the ethanol.
- Repeat the washing once.
- Keeping the tubes on the magnet and the caps open, dry the beads at room temperature for 3 minutes and remove any trace of ethanol from the bottom of the tube
- Remove the tubes from the magnetic stand. Check and make sure no larger size droplets present in the tube.
- Add 7 µl of water to the dried beads and resuspend the beads thoroughly.
- Place the tubes on the magnetic stand until the solution is clear. Transfer the liquid to a fresh tube. **The liquid contains purified longer DNA fragments (~500-3000 bp).**

Purification of shorter DNA fragments

- Incubate the **DNA/beads suspension for purification of shorter DNA fragments** as prepared above at room temperature for 5 minutes or longer.
- Place the samples on a magnetic stand, when the beads have been collected to the wall of the tube and the solution is clear, remove and discard the entire liquid fraction.
- Wash the beads by adding ~200 µl of freshly prepared 70% ethanol to the beads and incubating for 2 minutes at room temperature. Remove and discard the ethanol.
- Repeat the washing once.
- Keeping the tubes on the magnet and the caps open, dry the beads at room temperature for 3 minutes and remove any trace of ethanol from the bottom of the tube.

- 34 -

- Remove the tubes from the magnetic stand. Check and make sure no larger size droplets present in the tube
- Add 7 µl of water to the dried beads and resuspend the beads thoroughly.
- Place the tubes on the magnetic stand until the solution is clear. Transfer the liquid to a fresh tube. **The liquid contains purified shorter DNA fragments (~250-700 bp).**

Further manipulation of the purified longer DNA fragments:

dA-tailing

dA-tailing pre-mix (µl)

Components in reaction	Volume per reaction
10x RNA ligase Buffer	0.9
Klenow Fragment exo ⁻ (NEBNext dA-tailing module)	0.2
2 mM dATP	0.9
Total	2.0

Method

- Mix 2 µl of the pre-mix with each sample of the purified longer DNA fragments, following size-selection
- Incubate at 37 °C for 30 minutes, followed by 65 °C for 10 minutes

Adaptor ligation using Quick Ligation™ Kit

Ligation pre-mix (µl)

Components in reaction	Volume per reaction
2x Quick ligase Buffer	8
PDON adaptor (20 µM)	0.1
T4 Quick DNA ligase (NEB)	0.5

- 35 -

Total	8.6
-------	-----

- Mix 8.6 µl of the pre-mix with each of the dA-tailed DNA samples
- Incubate at 20 °C for 30 minutes

DNA purification using Ampure XP kit

- Mix 63 µl of water with 40 µl of the beads to each PDON adaptor ligated sample, in order to achieve the DNA:beads ratio $\approx$ 1: 0.5
- Close the tubes, mix the suspension gently and incubate at room temperature for 5 minutes.
- Place the samples on a magnetic stand, when the beads have been collected to the wall of the tube and the solution is clear, remove and discard the liquid.
- Wash the beads by adding ~200 µl of freshly prepared 70% ethanol to the beads and incubating for 3 minutes at room. Remove and discard the ethanol.
- Repeat the washing once.
- Keeping the tubes on the magnet and the caps open, dry the beads at room temperature for 3 minutes and remove any trace of ethanol from the bottom of the tube. Remove the tubes from the magnetic stand. Check and make sure no larger size droplets present in the tube.
- Add 10 µl of water to the dried beads and resuspend the beads thoroughly.
- Place the tubes on the magnetic stand until the solution is clear. Transfer the liquid to a fresh tube. **The liquid contains purified PCR-ready longer DNA fragments.**

Amplification

Amplification pre-mix (µl)

Components in reaction	Volume per reaction
Q5® High-Fidelity 2X PCR Master Mix	10.0
PDON-R21-PCR primer (25µM)	0.1

Total	10.1
-------	------

- Mix 10.1 µl of the pre-mix with each of the Adaptor-linked DNA samples
- Run the following programme on a PCR machine
 - 94 °C 2 min
 - 94 °C 30 sec
 - 50 °C 30 sec } 10 cycles
 - 68 °C 2.5 min
 - 68 °C 2 min
 - 4°C hold

Purification of DNA post PCR

- Mix 60 µl of water with 40 µl mixed XP beads per sample, and then add to each sample, following the PCR (the DNA:beads ratio = 1: 0.5).
- Close the tubes, mix the suspension gently and incubate at room temperature for 5 minutes.
- Place the samples on a magnetic stand, until the beads have been collected to the wall of the tube and the solution is clear.
- Remove and discard the entire liquid fraction.
- Wash the **bead pellets** in the tubes on the magnetic stand by adding ~200 µl of freshly prepared 70% ethanol to the beads and incubating for 2 minutes at room temperature. Remove and discard the ethanol.
- Repeat the washing once.
- Keeping the tubes on the magnet and the caps open, dry the beads at room temperature for 3 minutes and remove any trace of ethanol from the bottom of the tube
- Remove the tubes from the magnetic stand. Check and make sure no larger size droplets present in the tube.
- Add 57 µl of water to the dried beads and resuspend the beads thoroughly.
- Place the tubes on the magnetic stand until the solution is clear. Transfer the liquid to a fresh tube. **The liquid contains purified amplified longer DNA fragments (~500-3000 bp).**

Determination of DNA yield

- Warm up QuantiFluor to room temperature, and dilute the dye in 1x TE buffer (1:100 dilution)
- Transfer 2.5 µl of diluted dye to individual wells of a 384 low volume black plate (Greiner, 784076)
- Transfer 2.5 µl of each serially diluted DNA solution with adequate and known concentrations to corresponding wells
- Transfer 2.5 µl of each DNA samples to corresponding wells
- Measure the DNA concentration on FLUOstar OPTIMA, according to manufacturer's instruction
- The DNA yield is calculated as

$$\text{Yield (ng)} = \text{FLUOstar reading (pg/}\mu\text{l)} \times 2 \text{ (dilution factor)} \times 57 \text{ (}\mu\text{l)} / 1000$$

Example 2 – Whole genome amplified DNA profiles after S1 nuclease digestion

The profile of an amplified genomic DNA product of the above described method at the stage of S1 nuclease digestion was assessed. Along with no template controls, 35 pg mouse genomic DNA were used as amplification template. The amplification was performed, according to the protocol of Example 1. Following S1 digestion and subsequent DNA purification, entire amplified products from 2 individual reactions under corresponding experimental conditions were subjected to agarose gel electrophoresis on an E-gel® EX 1% agarose gel. 2 primer pools with different ratios of randomised RNA primers (equal AUGC:GC rich:AU rich) were tested in this experiment: A – 13:2:5 and B 10:3:7.

The results are shown in Figure 1. It can be seen that 1) there is no detectable amplification was observed in no template controls, demonstrating template-dependent manner of amplification; 2) the amplified DNA size ranged approximately between 0.3 to 3 kbp, with majority around 400 -1200 bp; and 3) no significant difference between the two primer pools was identified.

Example 3 - DNA profiles of Illumina sequencing libraries made from shorter DNA fractions

In order to test the consistency and sensitivity of the method of the invention, samples of purified mouse genomic DNA were used as amplification templates in a series of reactions. Amounts of DNA applied in individual reactions were 3.5 or 35 pg, equivalent to 1 or 10-fold that of a haploid genome, respectively, and the samples were grouped in duplicate. 2 primer pools with different ratios of randomised RNA primers (equal AUGC:GC rich:AU rich) were tested in this experiment: A – 13:2:5 and B 10:3:7. Amplification of input genomic DNA and purification of short fragments was performed according to the protocol of Example 1. Illumina sequencing libraries were made using the shorter DNA fraction of the entire amplification product from 3.5 pg inputs or 10% of the product from the 35 pg input samples and by applying 16 cycles of PCR. 20% of unpurified PCR product was run on an E-gel® EX 1% agarose gel.

The results are shown in Figure 2. Measured as the amount of amplification product, the amplification of the input genomic DNA by the method of the invention is similarly efficient for different input concentrations / amounts of DNA and for different primer mixes. No significant amount of product, if any, was detectable for zero-input samples.

Example 4 - Optimisation of the S1 nuclease digestion step

Investigations were performed to determine the most favourable conditions for the S1 nuclease digestion step of the method of the invention, e.g. those under which S1 nuclease could efficiently remove the RNA sequences incorporated at the 5' end of the amplified DNA molecules without significant damage to the DNA fragments.

Different amounts of S1 nuclease (10 and 45 units) under different incubation conditions were tested (A - 37 °C for 10 minutes followed by 45 °C for 2.5 minutes; B - 45 °C for 5 minutes; C - 50 °C for 5 minutes).

- 39 -

In a first experiment, 4 pg, of purified genomic DNA (approximately equivalent to 1 haploid genome), along with no DNA controls, were used as initial input DNA template and the method as described in Example 1 was carried out up to the end of "Double-strand formation with Pol I and Taq DNA polymerase". Conditions of S1 nuclease digestion were then tested, followed by downstream process as described in Example 1, until DNA purification post the end-repair stage. The DNA concentration of the shorter DNA fraction from individual samples was measured using QuantiFluor (Promega) on FLUOstar OPTIMA (BMG Labtech), and then DNA yields were calculated, accordingly.

The results are shown in Figure 3. As can be seen there was no significant difference in DNA yields when 10 and 45 units of S1 nuclease were used in combination with 37 °C for 10 minutes followed by 45 °C for 2.5 minutes incubation, or when the DNA amplification products were treated with 45 units of S1 nuclease at 45 °C for 5 minutes. However, DNA yields were lowered when the DNA amplification products were incubated with 45 units of S1 nuclease at 50 °C for 5 minutes, indicating undesirable degradation of DNA fragments by the S1 nuclease under such conditions.

Again, the results also show consistency and template-dependence of the amplification reaction of the method of the invention.

In a second experiment, 4 pg of purified genomic DNA (approximately equivalent to 1 haploid genome), along with no DNA controls, were used as initial input DNA template and the shorter DNA fraction, generated under varying S1 nuclease digestion conditions, was obtained using the method as described in Example 1. Illumina sequencing libraries were then made using the entire shorter DNA fraction of the purified amplification product from individual samples, and 10 cycles of PCR were applied. Following purification, the DNA concentration from individual PCR reactions were measured using QuantiFluor (Promega) on FLUOstar OPTIMA (BMG Labtech), and then DNA yields were calculated, accordingly.

The results are shown in Figure 4. As can be seen, the highest DNA yield was observed in the libraries prepared from amplified DNA treated with 45 units of S1 nuclease at 45 °C for 5 minutes, whereas the lowest DNA yield was seen in the

- 40 -

libraries prepared from amplified DNA treated with 45 units of S1 nuclease at 50 °C for 5 minutes. The DNA yield from the other treatment groups fell in between the two. These results suggest that, among the tested conditions, S1 nuclease digestion using 45 units of S1 nuclease in combination with an incubation of 45 °C for 5 minutes is the most optimal condition to remove the RNA sequences incorporated at the 5' end of the amplified DNA molecules without significant damage to the DNA fragments.

Once again, the results also show consistency and template- dependence of the amplification reaction of the method of the invention.

In a third experiment 4 pg of purified genomic DNA (approximately equivalent to 1 haploid genome), along with no DNA controls, were used as initial input DNA template and the longer DNA fraction, generated under varying S1 nuclease digestion conditions, was obtained using the method as described in Example 1. The entire longer DNA fraction of the purified amplification product from individual samples were further manipulated using the method described in Example 1, and 10 cycles of PCR were applied. After purification, DNA concentration from individual library samples were measured using QuantiFluor (Promega) on FLUOstar OPTIMA (BMG Labtech), and then DNA yields were calculated, accordingly.

The results are shown in Figure 5. As can be seen, the highest DNA yield was observed in the libraries prepared from amplified DNA treated with 45 units of S1 nuclease at 45 °C for 5 minutes or at 37 °C for 10 minutes followed by 45 °C for 2.5 minutes, whereas the lowest DNA yield was seen in the libraries prepared from amplified DNA treated with 45 units of S1 nuclease at 50 °C for 5 minutes. The DNA yield from the group treated with 10 units of S1 nuclease was in the middle

Taken together with the other results, these results suggest that, among the conditions tested, S1 nuclease digestion using 45 units of S1 nuclease in combination with an incubation of 45 °C for 5 minutes is the most optimal condition to remove the RNA sequences incorporated at the 5' end of the amplified DNA molecules without significant damage to the DNA fragments.

Once again, the results also show consistency and template- dependence of the amplification reaction of the method of the invention.

Example 5 – Analysis of Illumina sequencing libraries made from shorter DNA fractions of haploid cells

Illumina sequencing libraries were made from shorter DNA fractions of 67 individual mouse sperm cells by the method as described in Example 1. These 67 libraries and 2 bulk DNA samples isolated from the pool of mouse sperm cells from which the individual cells were isolated were subjected to Illumina sequencing. The libraries made from the shorter DNA fractions were sequenced to an approximate sequencing depth of 3.6 to 8.5, and the bulk samples were sequenced to an approximate sequencing depth of 4. The results are shown in Figures 6 and 7 following normalisation to remove regions which are not covered by any reads within the sequencing data obtained from the 2 ng bulk DNA sample that was sequenced to an approximate sequencing depth of 20, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means.”

Figure 6 shows that when the single sperm amplification samples were sequenced, the reads obtained represent approximately 45% to 72% (median 62%) of the genome that is accessible to Illumina sequencing, as judged by setting the coverage of a bulk mouse sperm DNA sample, sequenced to an approximate sequencing depth of 20, to 100%.

Figure 7 shows a Lorenz curve presenting the distribution of total sequencing reads over the available fraction of genome for the 69 separate sequencing experiments. The distribution within the test samples (sequencing libraries made from amplification products of the 67 individual mouse sperm cells by the method described in Example 1) is not markedly different to that seen with bulk samples using genomic DNA as starting material for preparation of sequencing libraries. This degree of similarity indicates that the method described in Example 1 is capable of amplifying the haploid genome of individual sperm cells with sufficient uniformity to represent an surprisingly effective whole genome amplification technique.

Example 6 – Analysis of Illumina sequencing libraries made from DNA amplified from diploid cells

Illumina sequencing libraries were made separately from shorter DNA fractions and longer DNA fractions of (6) individual cells from the GM19382 human lymphoblastoid cell line by the method as described in Example 1. These (12) libraries and (1) library from a bulk DNA sample isolated from the pool of GM19382 cells from which individual cells were isolated were subjected to Illumina sequencing. The single-cell libraries made from the shorter DNA fractions and from the longer fractions were all sequenced separately to an approximate sequencing depth of 30, and the bulk samples were sequenced to an approximate sequencing depth of 60. The data from the short and long fractions from each cell were combined and the results are shown in Figures 8 and 9 following normalisation to remove regions which are not covered by any reads within the sequencing data obtained from the bulk DNA sample, as these regions within the genome may not be accessible to sequencing with Illumina technology and/or mapping by standard bioinformatics means.

Figure 8 shows that when the single-diploid-cell amplification samples were sequenced, the reads obtained represent approximately 89% to 93% (median 92%) of the genome that is accessible to Illumina sequencing, as judged by setting the coverage of a bulk GM19382 DNA sample sequenced to an approximate sequencing depth of 60 to 100%.

Figure 9 shows a Lorenz curve presenting the distribution of total sequencing reads over the available fraction of genome for the 7 separate sequencing experiments. The distribution within the test samples (sequencing libraries made from amplification products of the 6 individual GM19382 cells by the method described in Example 6) is not markedly different to that seen with bulk samples using genomic DNA as starting material for preparation of sequencing libraries. This degree of similarity indicates that the method described in Example 6 is capable of amplifying the diploid genome of individual GM19382 cells with sufficient uniformity to represent a surprisingly effective whole genome amplification technique.

CLAIMS

1. A method for amplification of ultra-low amounts of input DNA, said method comprising
- (i) providing a sample of DNA,
 - (ii) contacting said DNA with a population of RNA oligonucleotide primers consisting of essentially equal proportions of A, U, G and C bases, but wherein each primer of said population has a randomised base sequence, and a Klenow fragment, or a DNA polymerase having a DNA-dependent DNA polymerase activity, a nucleic acid strand displacement activity, a lack of 5'→3' nuclease activity, and an ability to prime DNA polymerisation from an RNA primer which are equivalent to said Klenow fragment, under conditions which permit at least one round of primer binding to said DNA and subsequent DNA polymerisation from said bound primers catalysed by said Klenow fragment, or said equivalent thereof, thereby forming a first reaction product,
 - (iii) contacting said first reaction product with a DNA polymerase I, or a DNA polymerase having the DNA-dependent DNA polymerase activity, 5'→3' nuclease activity and an ability to prime DNA polymerisation from an RNA primer which are equivalent to said DNA polymerase I, under conditions which permit at least one round of primer binding and subsequent DNA polymerisation from said bound primers catalysed by said DNA polymerase I, or said equivalent thereof, thereby forming a second reaction product,
 - (iv) optionally contacting said second reaction product with a Taq DNA polymerase, or a DNA polymerase having the DNA polymerase activity equivalent to a Taq DNA polymerase, under conditions which permit DNA polymerisation to be catalysed by said Taq DNA polymerase, or said equivalent thereof, thereby forming a third reaction product, and
 - (v) contacting said second reaction product, or said third reaction product if step (iv) is performed, with an endonuclease or an exonuclease or combination thereof capable of nucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, under conditions which permit said nucleolytic degradation, thereby forming an predominantly double stranded DNA amplification product.

- 44 -

2. The method of claim 1, wherein step (iv) is performed.
3. The method of claim 1 or claim 2, wherein step (v) is a step of contacting said second reaction product, or said third reaction product if step (iv) is performed, with an S1 nuclease, or an RNA endonuclease having the endonuclease activity of said S1 nuclease, under conditions which permit the endonucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, thereby forming an predominantly double stranded DNA amplification product.
4. The method of any one of claims 1 to 3, wherein step (v) is a step of contacting said second reaction product, or said third reaction product if step (iv) is performed, with at least a BAL 31 nuclease, or an nuclease having the exonuclease activity of said BAL 31 nuclease, under conditions which permit the exonucleolytic degradation of at least a portion of any single stranded or annealed RNA remaining in the second reaction product, or said third reaction product if step (iv) is performed, thereby forming an predominantly double stranded DNA amplification product.
5. The method of any one of claims 1 to 4, wherein said RNA oligonucleotide primers are contacted with the DNA sample together with at least one additional population of RNA oligonucleotide primers consisting of A, U, G and C bases in varying proportions but wherein each primer of said population has a randomised base sequence.
6. The method of claim 5, wherein the additional population of RNA oligonucleotide primers is enriched for A and U bases or G and C bases.
7. The method of claim 6, wherein, the additional population of RNA oligonucleotide primers has more than about 80% A and U bases or G and C bases.
8. The method of any one of claims 1 to 7, wherein the population of randomised RNA oligonucleotide primers of essentially equal proportions of A, U, G

- 45 -

and C will account for more than about 50%, of the primers contacted with the DNA sample.

9. The method of any one of claims 1 to 8, wherein said RNA oligonucleotide primers consist of 6 to 16 ribonucleotides.

10. The method of claim 9, wherein said RNA oligonucleotide primers consist of 9, 10, 11 or 12 ribonucleotides, or a mixture thereof.

11. The method of claim 10, wherein said RNA oligonucleotide primers consist of 10 ribonucleotides.

12. The method of any one of claims 1 to 11, wherein in step (v) said second reaction product, or said third reaction product if step (iv) is performed, is contacted with an S1 nuclease, or functional equivalent thereof, for about 1 mins to 10 mins, at about 42°C to 48°C.

13. The method of claim 12, wherein in step (v) said second reaction product, or said third reaction product if step (iv) is performed is contacted with an S1 nuclease, or functional equivalent thereof, for about 5 mins, at about 45°C.

14. The method of claim 12 or 13, wherein about 30U to 60U of said S1 nuclease, or functional equivalent thereof, is present.

15. The method of claim 14, wherein about 45U of said S1 nuclease, or functional equivalent thereof, is present.

16. The method of any one of claims 1 to 15, wherein said DNA sample is genomic DNA.

17. The method of claim 16, wherein said DNA sample is the entire genome of a single eukaryotic cell.

18. The method of any one of claims 1 to 17, wherein said DNA sample contains less than 100 DNA molecules

19. The method of any one of claims 1 to 18, wherein said method comprises a preceding step of cell lysis, wherein the conditions of said lysis step does not deplete genomic DNA.

20. The method of any one of claims 1 to 19, wherein said method comprises one or more further steps in which the double stranded DNA amplification product is exposed to one of more steps of purification, end-repair, size selection, dA-tailing, adaptor ligation and/or further amplification.

21. A method of optimising nucleic acid sequence analysis of ultra-low amounts of DNA, said method comprising performing the DNA amplification method as defined in any one of claims 1 to 19 on an ultra-low amount of DNA prior to the nucleic acid sequence analysis of said DNA.

22. A method of nucleic acid sequence analysis, said method comprising a step of sample preparation prior to the analysis step(s) in which the DNA amplification method as defined in any one of claims 1 to 19 is used to amplify an ultra-low amount of DNA in said sample.

23. The method of claim 21 or claim 22, wherein the nucleic acid sequence analysis is a sequencing technique or an oligonucleotide hybridisation probe based technique.

FIGURE 1

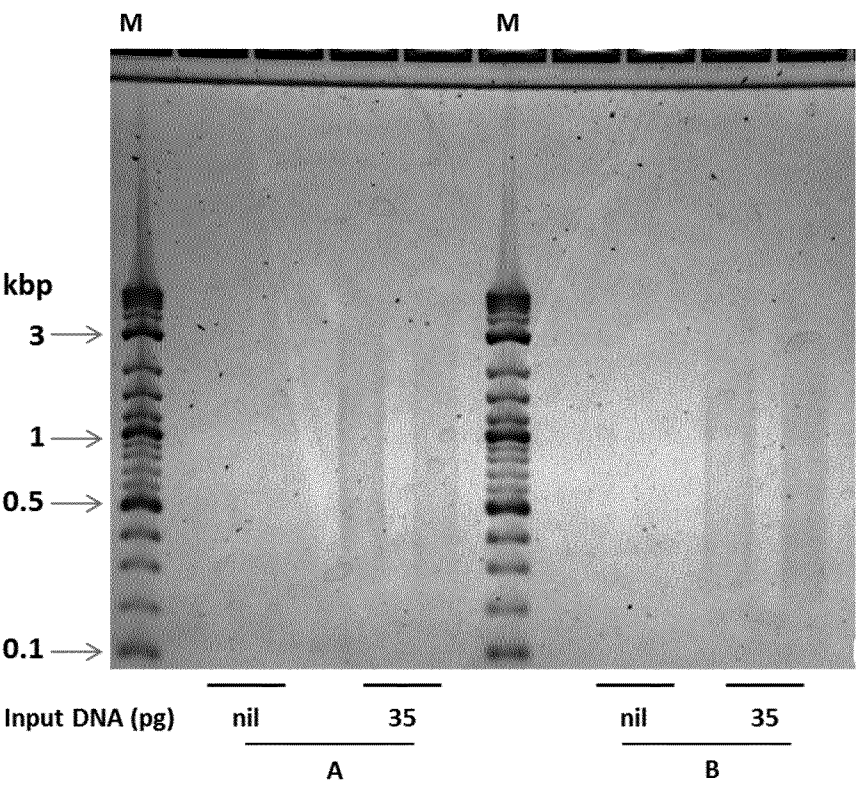


FIGURE 2

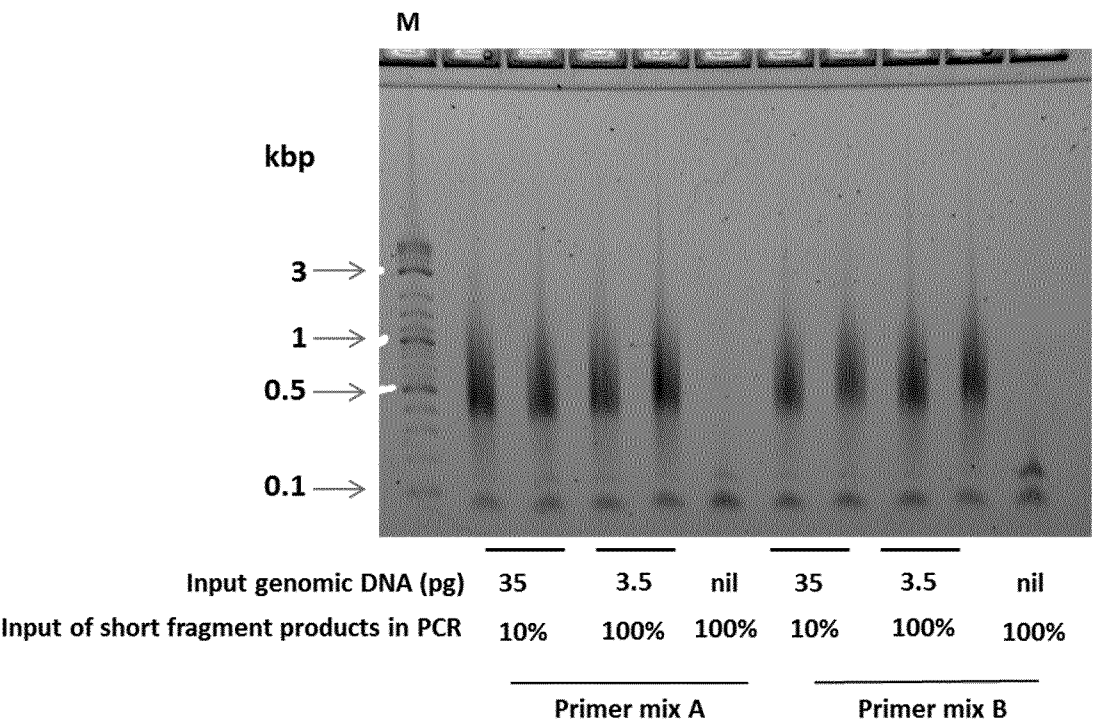


FIGURE 3

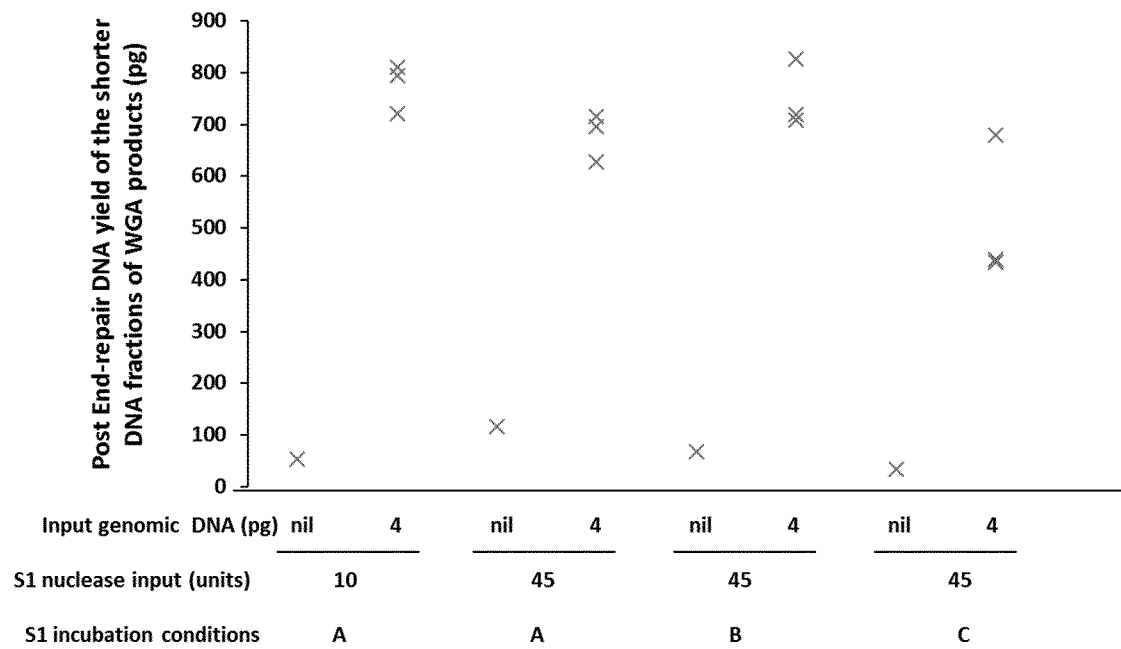


FIGURE 4

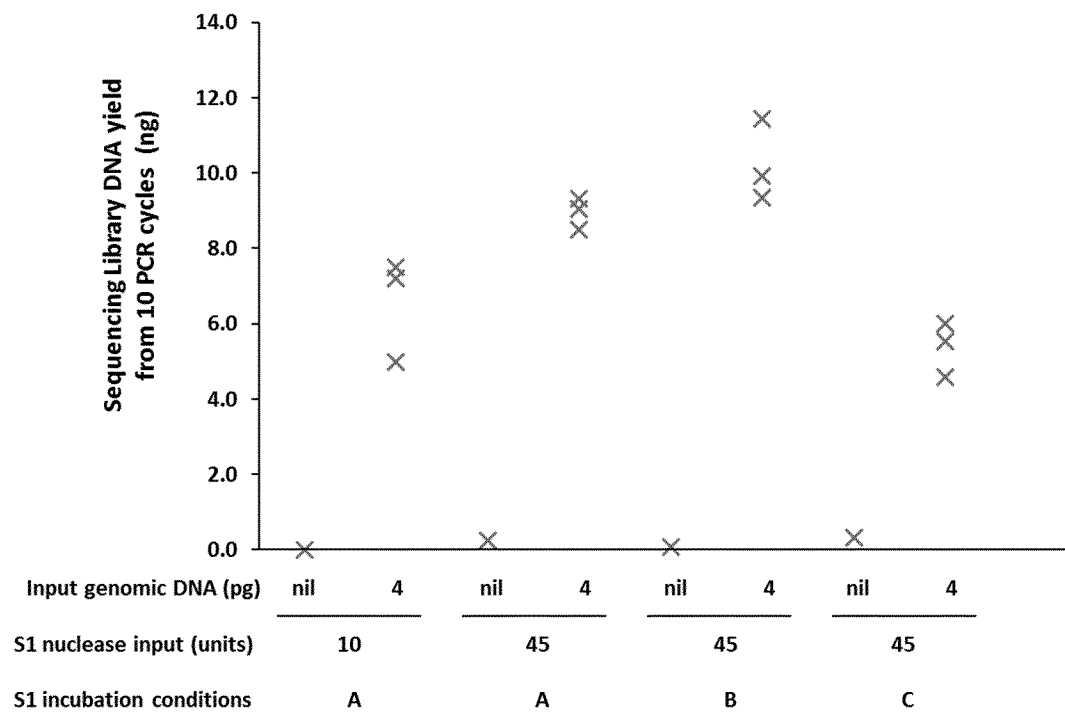


FIGURE 5

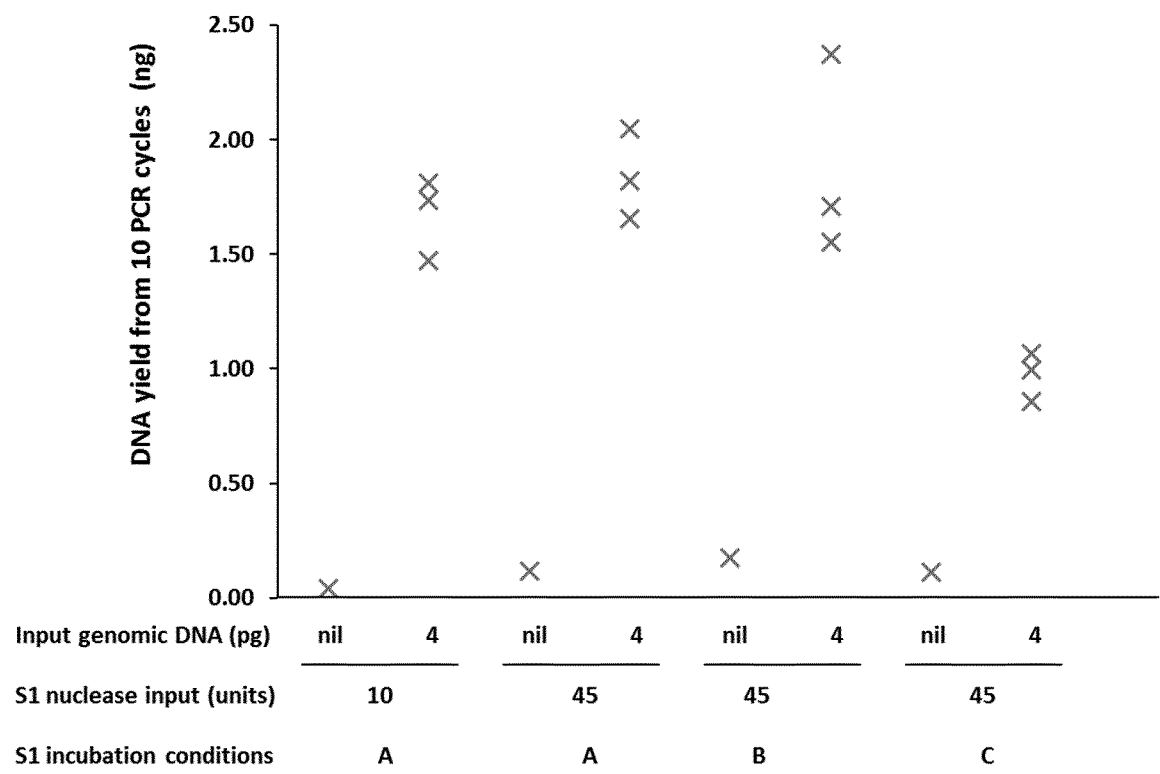
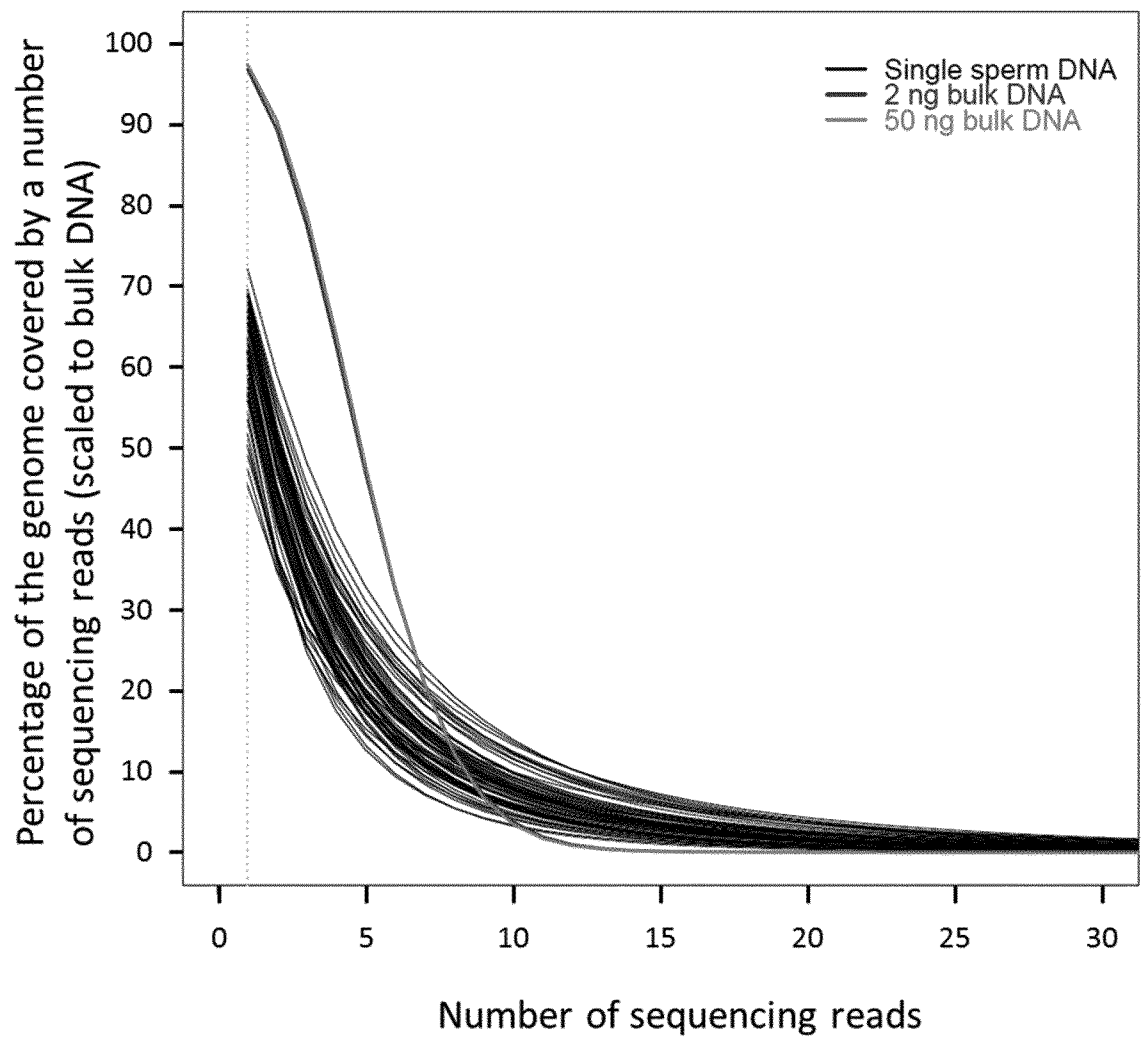
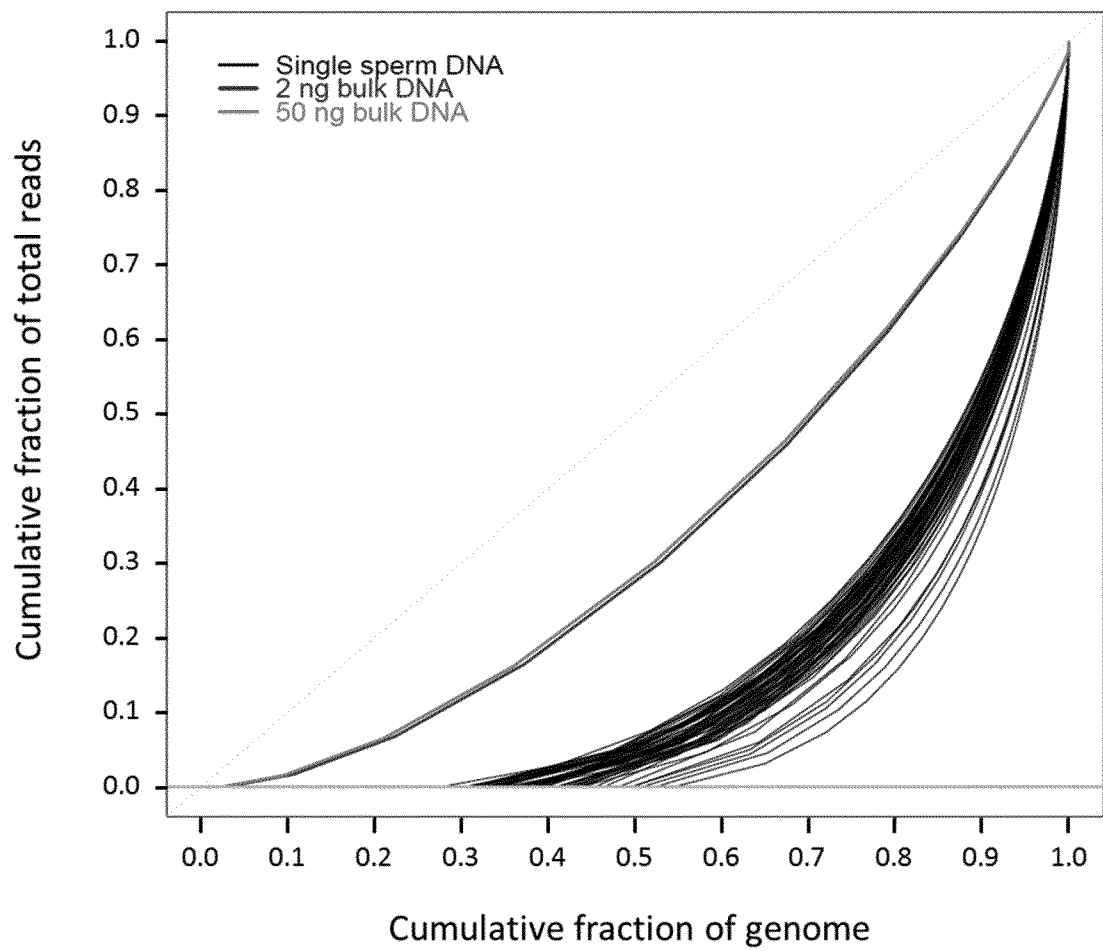


FIGURE 6

5/7

FIGURE 7

6/7

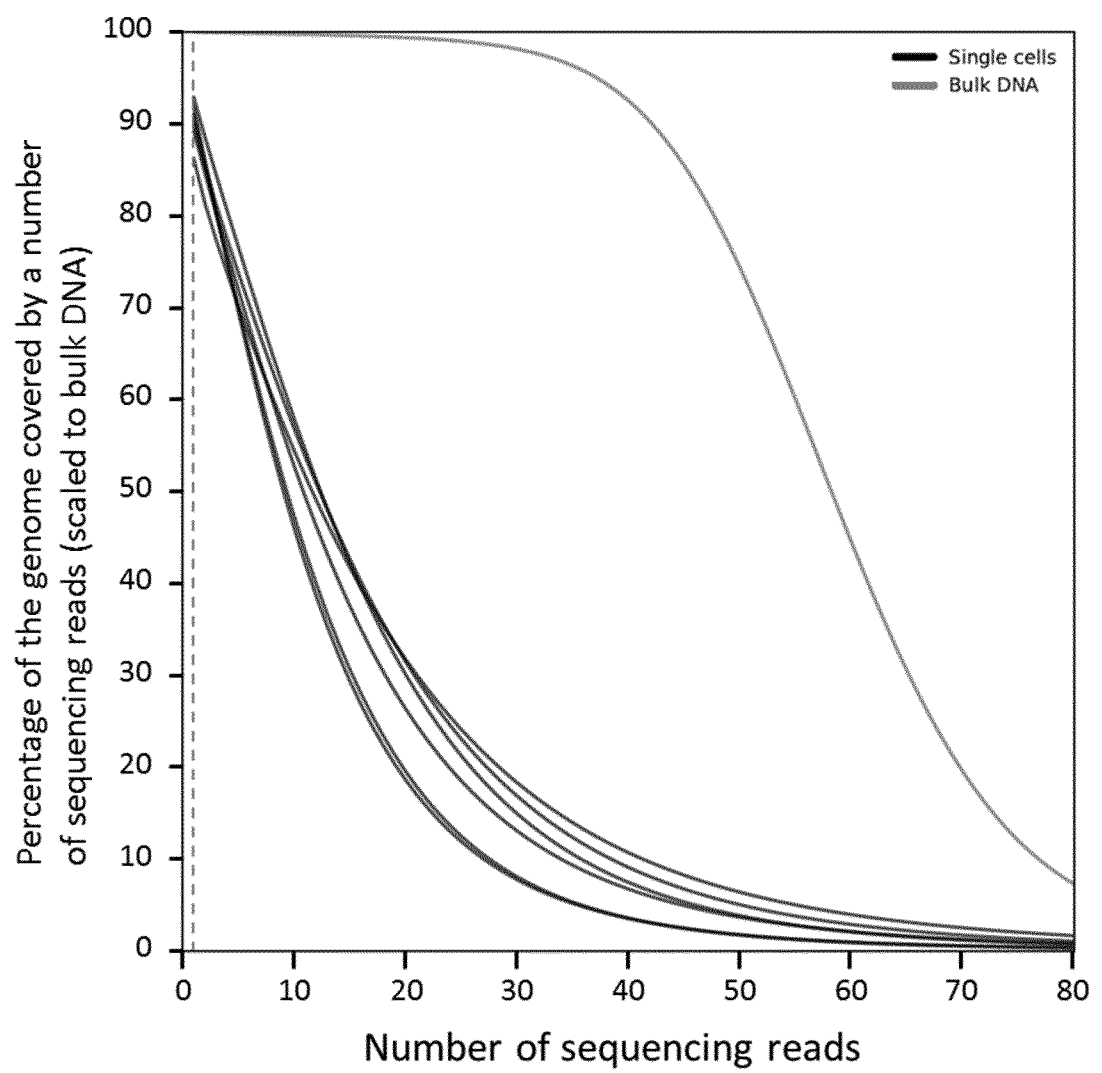
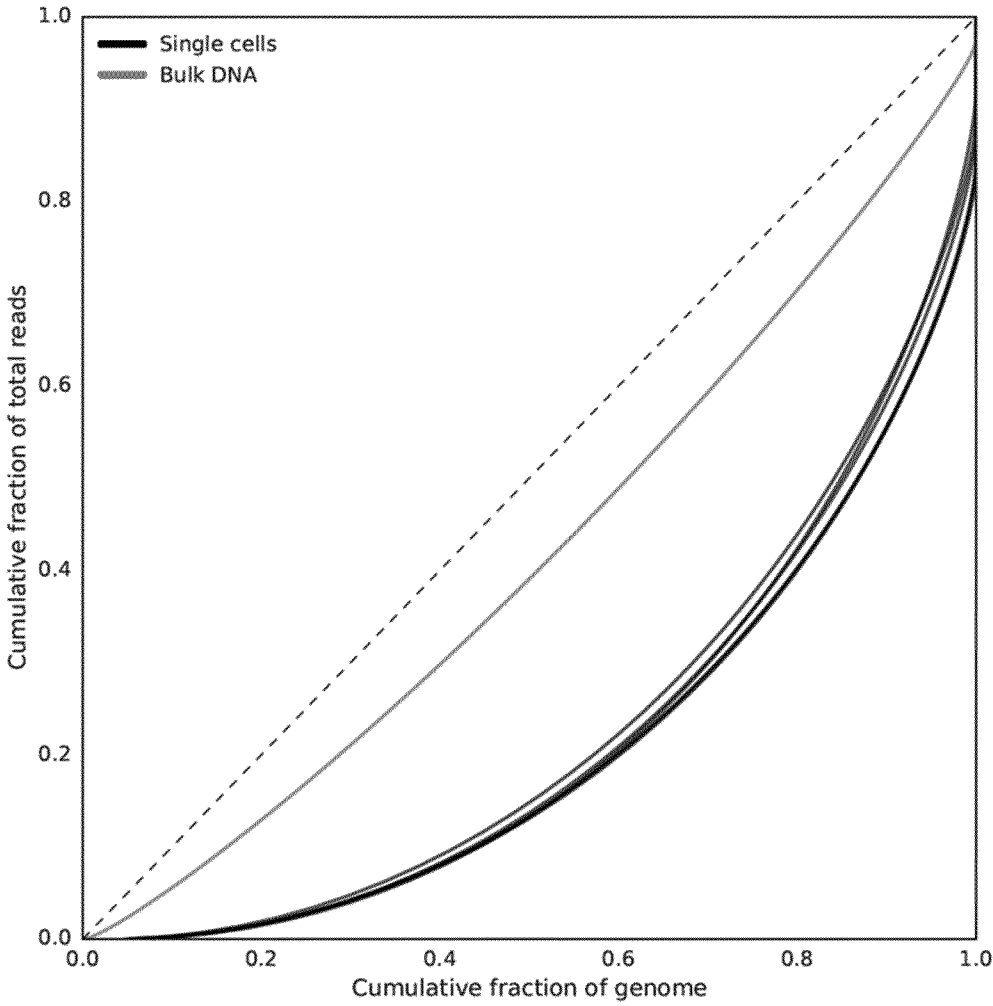
FIGURE 8

FIGURE 9



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/076644

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

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 practicable, search terms used)

EPO-Internal, 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	JP 2010 094091 A (NAT AGRICULTURE & FOOD RES ORG) 30 April 2010 (2010-04-30)	1,16,18, 21-23
Y	the whole document para. 10, 14, 19, 20, 23, 35, 58, 70, 75; Fig. 1	2-15,17, 19,20
Y	----- JP 2006 141357 A (NAT INST OF AGROBIO SCIENCES) 8 June 2006 (2006-06-08) the whole document para. 1-3, 9-13, 18, 24-25	1-23
Y	----- US 2015/275285 A1 (ZHANG YILIN [US]) 1 October 2015 (2015-10-01) the whole document para. 13-14, 52, 90, 93, 94 ----- -/-	1-23



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 December 2017

Date of mailing of the international search report

04/01/2018

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Sauer, Tincuta

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/076644

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 824 517 A (CLEUZIAT PHILIPPE [FR] ET AL) 20 October 1998 (1998-10-20) the whole document col. 15, ll. 5-19, Fig. 6 -----	1-23
Y	US 2015/360193 A1 (FAN JIAN-BING [US] ET AL) 17 December 2015 (2015-12-17) the whole document para. 53-54, 69, 75, 80, 85, 95, 129, 150, Fig. 7 -----	1-23

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/076644

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2010094091 A	30-04-2010	JP 5652843 B2 JP 2010094091 A	14-01-2015 30-04-2010
JP 2006141357 A	08-06-2006	JP 4793842 B2 JP 2006141357 A	12-10-2011 08-06-2006
US 2015275285 A1	01-10-2015	CN 105189780 A EP 2925893 A1 HK 1217518 A1 US 2015275285 A1 WO 2014088979 A1	23-12-2015 07-10-2015 13-01-2017 01-10-2015 12-06-2014
US 5824517 A	20-10-1998	AT 214741 T CA 2200627 A1 DE 69619959 D1 DE 69619959 T2 EP 0787209 A1 FR 2737223 A1 US 5824517 A WO 9704126 A1	15-04-2002 06-02-1997 25-04-2002 07-11-2002 06-08-1997 31-01-1997 20-10-1998 06-02-1997
US 2015360193 A1	17-12-2015	EP 2877593 A1 US 2015360193 A1 WO 2014018093 A1	03-06-2015 17-12-2015 30-01-2014