

METHODS FOR DETECTING GANODERMA

TECHNICAL FIELD

[0001] The present disclosure generally relates to markers that can be used to detect
5 Ganoderma. More specifically, the present disclosure relates to markers that can be used to
detect pathogenic Ganoderma species. The present disclosure also relates to methods of
screening and determining Ganoderma pathogenicity in the *Ganoderma boninense* (*G.*
boninense) species as well as other species. The disclosure further relates to methods for
analyzing a sample such as, but not limited to a soil sample, to detect the presence of *G.*
10 *boninense* and methods for predicting the pathogenicity of the *G. boninense* detected using
one or more markers of the present disclosure.

BACKGROUND

[0002] *G. boninense* is a type of white rot fungus which is threatening sustainable oil palm
15 production in South East Asia, especially Malaysia and Indonesia. *G. boninense* breaks down
the lignin in wood, leaving the lighter-colored cellulose behind, hence leading to the name
“white rot fungus”. This fungus is a threat to the oil palm industry as it causes the basal stem
rot (BSR) disease in oil palm. BSR is characterized by a decay of the bole, visible symptoms
on the diseased palm such as multiple unopened spears and production of *G. boninense*
20 fruiting bodies on the base of the trunk.

[0003] Colonization of *G. boninense* at the basal stem of the oil palm will cause disruption
to the plant vascular system, resulting in reduced fruit yield. As oil palm is a food crop, BSR
disease can cause severe economic losses through the direct reduction in oil palm bunch
numbers and weight of fruit bunches from standing diseased palms. A survey conducted by
25 Idris et al. in 2011 found that 3.71% of 1.59 million hectares of oil palm plantation were
infected with *Ganoderma*, and this figure is increasing, with the reported disease incidence
higher in replanting areas and in younger palms. Besides the economic loss caused by the
disease, the disease also threatens the safety of workers in the field as often an infected oil
palm tree looks sturdy before they topple over and this can lead to injuries.

30 [0004] Studies have been conducted for the past 40 years to understand the disease and
spread of the disease. However, to date, the understanding of the fungus and the disease
caused is still insufficient to efficiently manage the spread of the disease.

[0005] Visual observation of the symptoms is the current detection method of the disease. However, little can be done once the physical symptoms are observed, as the damage to the plant has already been done. It should be noted that though BSR is often associated with *G. boninense*, it may be possible for other *Ganoderma* species to also have pathogenic ability in oil palm, and it is also possible for different varieties of *G. boninense* to have varying degrees of pathogenicity.

[0006] Therefore, there is a need for a means of detecting the presence of pathogenic *Ganoderma*, in particular pathogenic strains of *G. boninense*, at an early stage of infection to enable early treatment of an affected plant, and to restrict spread of the disease. This enables planters to better manage the spread of disease and consequently reduce loss of fresh fruit bunch yield and plants before their full commercial value is realized.

[0007] There is also a need to detect the presence of pathogenic *Ganoderma* species, in particular pathogenic strains of *G. boninense*, directly from an environmental sample, such as but not limited to, a soil sample.

[0008] However, in the current state of the art, there are only methods to identify the *Ganoderma* isolate or strain but there are no known methods for identifying and differentiating between pathogenic and non-pathogenic *Ganoderma* isolates using any available biological markers.

[0009] Therefore, there is further need to provide molecular markers that are capable of differentiating between pathogenic and non-pathogenic *Ganoderma* species to enable detecting and identifying a pathogenic *Ganoderma* species.

SUMMARY

[0010] Disclosed is a method for determining the pathogenicity of a *Ganoderma* species, wherein the method comprises the step of detecting the presence or absence of one or more sequences selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is non-pathogenic.

[0010a] In one aspect, there is provided a method for determining the pathogenicity of Ganoderma species, wherein the method comprises the step of detecting the presence or absence of one or more sequences selected from a group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a single nucleotide polymorphism (SNP) variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof, indicates that the Ganoderma species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the Ganoderma species is not pathogenic.

[0011] Advantageously, the method allows for direct identification of Ganoderma, and its pathogenicity from a sample, such as a soil sample. This translates into a shorter time for analyzing a sample, such as a soil sample, as the method eliminates the need for isolation of Ganoderma strains before analysis can be performed.

[0012] Disclosed is a method for detecting presence of a pathogenic Ganoderma species in a sample, comprising the steps of: (a) extracting DNA from the sample; (b) subjecting the DNA to sequencing to determine sequences of the extracted DNA; (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof; and (d) determining the absence or presence of one or more sequences selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the Ganoderma species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the Ganoderma species is non-pathogenic.

[0012a] In another aspect, there is provided a method for detecting presence of pathogenic Ganoderma in a sample, comprising the steps of: (a) extracting DNA from the

sample; (b) subjecting the DNA to sequencing to determine the sequences of the extracted DNA; (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from a group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof; and (d) determining the absence or presence of one or more sequences selected from a group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the *Ganoderma* species is not pathogenic.

[0013] In another embodiment, there is provided a method for identifying a nucleic acid sequence containing a marker for *Ganoderma* pathogenesis comprising the steps of: (a) extracting DNA from a group of pathogenic and non-pathogenic *Ganoderma* strains; (b) preparing whole-genome sequencing libraries of each of the pathogenic and non-pathogenic *Ganoderma* strain using the extracted DNA of step (a); (c) sequencing the libraries from step (b) to form a plurality of sequencing reads; (d) assembling the sequencing reads from step (c) to generate a backbone sequence for the pathogenic *Ganoderma* strain and a backbone sequence for the non-pathogenic *Ganoderma* strain; (e) mapping the sequencing reads from step (c) from each strain to the backbone sequence of step (d) to perform comparative genomics to identify variant sequence(s); (f) designing and preparing primers from the variant sequence(s) from step (e); (g) applying the primers prepared from step (f) to DNA extracted from the pathogenic and non-pathogenic *Ganoderma* strains for DNA amplification and sequencing; and (h) classifying the variant sequence(s) into those associated with the

pathogenic Ganoderma strains and those associated with the non-pathogenic Ganoderma strains.

[0014] Disclosed is an oligonucleotide primer comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 124 to 369.

5

10

15

20

25

30

[0014a] Disclosed is an oligonucleotide primer comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 288-289, 178-179, 336-337, 198-199, 214-215, 234-235, 256-257, 312-313, 318-319, 344-345, 172-173, 250-251, 284-285, and 308-309.

5 [0015] Advantageously, the methods of the present disclosure are able to determine the presence of pathogenic *Ganoderma* in a sample, such as a soil sample, with confidence level of up to about 100%. For example, using one of the markers (SEQ ID NO: 83) isolated by the methods of the present disclosure, it is possible to determine the presence of pathogenic *Ganoderma* at about 100% confidence level.

10 [0016] Advantageously, the methods of the present disclosure are also highly sensitive in detecting DNA from pathogenic *Ganoderma* in a sample. For example, one of the markers identified by the present disclosure (SEQ ID NO: 25) is capable of detecting DNA from pathogenic *Ganoderma* in a soil sample that is present in an amount as low as 10^{-2} ng.

[0017] Disclosed is a kit for use in the methods defined herein, wherein the kit comprises:
15 (a) one or more primer(s) selected from the group consisting of SEQ ID NOs: 124 to 369 for determining the presence or absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123; (b) one or more container(s) suitable for sample collection; (c) one or more container(s) suitable for sample processing; (d) one or more reagents for sample processing; (e) one or more buffer(s); (f) one or more PCR reagent(s);
20 and (g) instructions for use in accordance with the methods described herein.

[0017a] Also disclosed is a kit for use in the methods defined herein, wherein the kit comprises: (a) one or more primer(s) selected from the group consisting of SEQ ID NOs: 288-289, 178-179, 336-337, 198-199, 214-215, 234-235, 256-257, 312-313, 318-319, 344-345, 172-173, 250-251, 284-285, and 308-309 for determining the presence or absence of one
25 or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93; (b) one or more container(s) suitable for sample collection; (c) one or more container(s) suitable for sample processing; (d) one or more reagents for sample
30 processing; (e) one or more buffer(s); (f) one or more PCR reagent(s); and (g) instructions for use in accordance with the methods described herein.

DEFINITION OF TERMS

[0018] The following words and terms used herein shall have the meaning indicated:

[0019] The term “molecular marker” or “marker”, as used therein, refers to a sequence of genetic information that is associated with a location in the genome, a trait associated with the marker or a DNA-based feature, such as a repeat region or identifiable and conserved DNA region. A molecular marker may refer to the entirety of the sequence, or fragments of that sequence, or a region of flanking genetic region, that will allow for the identification of the molecular marker. Ideally, a molecular marker is polymorphic in nature so that it can be used to differentiate between the different features that they are related to. In the present disclosure, the feature related to the marker is pathogenicity of a *Ganoderma* isolate.

[0020] In one example, molecular markers may comprise but are not limited to single nucleotide polymorphisms (SNPs), simple sequence repeats (SSR), microsatellites, insertion-deletion of bases (INDELs), amplified fragment length polymorphism (AFLP), random amplification of polymorphic DNA (RAPD), restriction fragment length polymorphism

(RFLP), unique-event polymorphism (UEP), transposable element position, copy number variation, conserved DNA regions and such.

[0021] The term “single nucleotide polymorphism” or “SNP”, as defined herein, is a DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) in the genome sequence is altered or differs between members of a biological species. Thus, as used herein, a SNP is any polymorphism characterized by a different single nucleotide at a particular physical position in at least one allele. Each individual in a given population has many single nucleotide polymorphisms that together create a unique DNA pattern for that individual.

[0022] As used herein, “sequence identity” refers to the residues in two sequences that are the same when aligned for maximum correspondence over a specified window of comparison by means of computer programs known in the art, such as Burrows-Wheeler Alignment (BWA) version 0.5.9. and GAP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1996, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711) (Needleman, S.B. and Wunsch, C.D., (1970), Journal of Molecular Biology, 48, 443-453).

[0023] Percentage of mappable reads is defined as the percentage of alignment of the trimmed and filtered reads to a *G. boninense* backbone sequence.

[0024] The term “variant” as used herein includes a reference to substantially similar sequences. These sequence variants may have at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to the “non-variant” nucleic acid, or to a section within the “non-variant” nucleic acid sequence. For example, a variant of any one of SEQ IN NOs: 1 to 123 may have at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98% or at least about 99%. In one example, where the variant sequence includes a molecular marker, the differences that make up the variants may occur outside of the section where the molecular marker is located. That is to say that the differences that constitute the variant sequences can be found outside the location of the molecular marker. By way of an example, if a sequence comprises a SNP at base pair 267, the differences that make up the variants then can be found at any location within the variant sequence other than base pair 267.

[0025] The term "nucleic acid" includes a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. The terms "nucleic acid", "nucleic acid molecule", "nucleic acid sequence" and polynucleotide etc. may be used interchangeably herein unless the context indicates otherwise.

[0026] The term "fragment" as used herein includes a reference to a nucleic acid molecule that is a constituent of a particular nucleic acid or variant thereof. Fragments of a nucleic acid sequence do not necessarily need to encode polypeptides which retain biological activity, for example, hybridisation probes or PCR primers.

[0027] The term "oligonucleotide" as used herein refers to short nucleic acid molecules useful, for example, as hybridizing probes, nucleotide array elements or amplification primers. Oligonucleotide molecules are comprised of two or more nucleotides, i.e. deoxyribonucleotides or ribonucleotides, preferably more than five and up to 30 or more. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. Oligonucleotides can comprise ligated natural nucleic acid molecules or synthesized nucleic acid molecules and comprise between 10 to 150 nucleotides or between about 12 and about 100 nucleotides which have a nucleotide sequence which can hybridize to a strand of polymorphic DNA, e.g. to permit detection of a polymorphism. Such oligonucleotides may be nucleic acid elements for use on solid arrays (e.g. synthesized or spotted). In some examples, such oligonucleotides can comprise as few as 12 hybridizing nucleotides, e.g. for assays where the oligonucleotide also comprises a detectable label. In other examples, the oligonucleotide can comprise as few as about 15 hybridizing nucleotides, e.g. for single base extension assays. Such oligonucleotides may also be primers for use in polymerase chain reaction (PCR) or other reactions.

[0028] The term "primer" as used herein refers to a nucleic acid molecule, such as an oligonucleotide, whether derived from a naturally occurring molecule such as one isolated from a restriction digest or one produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single stranded for maximum

efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer may be an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors according to the particular application, including temperature and source of primer. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains at least 18, at least 19, at least 20, at least 21, at least 22, at least 23 or at least 24 nucleotides, which are identical or complementary to the template and optionally a tail of variable length which need not match the template. It is understood the longer the nucleic acid binding sequence of the primer, the more specific the binding of the primer to the intended target sequence. However, the length of the tail should not be so long that it interferes with the recognition of the template. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

[0029] The primers herein are selected to be “substantially” complementary to the different strands of each specific sequence to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer. Computer generated searches using programs such as Primer3 (www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi), STSPipeline (www-genome.wi.mit.edu/cgi-bin/www-STSPipeline), or GeneUp, for example, can be used to identify potential PCR primers. Exemplary primers include primers that are 18 to 24 bases long, where at least between 18 bases are identical or complementary to at least 18 bases of a segment of the template sequence. In some examples described herein, primers that are 20 bases long are used. In other examples, primers that are 19, 21, 22, 23, or 24 bases long are used.

[0030] The term “backbone sequence” refers to an artificial sequence generated from assembly of sequencing reads from multiple strains of an organism (such as Ganoderma in this case), which may share a similar characteristic (such as being pathogenic or non-pathogenic). The assembly which generates such a “backbone sequence” may be performed using CLC Bio Assembly Cell Version 4.06beta.67189.

[0031] This disclosure also contemplates and provides primer pairs for amplification of nucleic acid molecules in order to detect, for example, polymorphisms or single nucleotide polymorphisms (SNPs). As used herein “primer pair” refers to a set of two oligonucleotide primers based on two separated sequence segments of a target nucleic acid sequence. One primer of the pair is a “forward primer” or “5’ primer” having a sequence which is identical to the more 5’ of the separated sequence segments (sometimes also denoted as the “+” strand). The other primer of the pair is a “reverse primer” or “3’ primer” having a sequence which is the reverse complement of the more 3’ of the separated sequence segments (sometimes denoted as the “-” strand). A primer pair allows for amplification of the nucleic acid sequence located between the binding sites of the two primers on the double-stranded template nucleic acid sequence. Optionally, each primer pair can comprise additional sequences, e.g. universal primer sequences or restriction endonuclease sites, at the 5’ end of each primer, e.g. to facilitate cloning, DNA sequencing, or re-amplification of the target nucleic acid sequence.

[0032] The term “isolate” when used in reference to a microorganism isolate, such as but not limited to, Ganoderma isolates, refers to a culture of microorganism removed from its original environment.

[0033] Unless specified otherwise, the terms “comprising” and “comprise”, and grammatical variants thereof, are intended to represent “open” or “inclusive” language such that they include recited elements but also permit inclusion of additional, unrecited elements.

[0034] As used herein, the term “about”, in the context of, but not limited to, concentrations of DNA, chemicals, chemical solutions, enzymes or components of a buffer, typically means +/- 5% of the stated value, more typically +/- 4% of the stated value, more typically +/- 3% of the stated value, more typically, +/- 2% of the stated value, even more typically +/- 1% of the stated value, and even more typically +/- 0.5% of the stated value.

[0035] Throughout this disclosure, certain embodiments may be disclosed in a range format. It should be understood that the description in range format is merely for convenience

and brevity and should not be construed as an inflexible limitation on the scope of the disclosed ranges. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges as well as individual numerical values within that range. For example, description of a range such as from one to six should be considered to have specifically disclosed sub-ranges such as from one to three, from one to four, from one to five, from two to four, from two to six, from three to six etc., as well as individual numbers within that range, for example, one, two, three, four, five, and six. This applies regardless of the breadth of the range.

[0036] Certain embodiments may also be described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the disclosure. This includes the generic description of the embodiments with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0037] The disclosure illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the disclosure claimed. Thus, it should be understood that although the present disclosure has been specifically disclosed by preferred embodiments and optional features, modification and variation of the disclosures embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this disclosure.

DETAILED DISCLOSURE OF THE EMBODIMENTS

[0038] The present disclosure provides methods for identifying markers indicative of pathogenicity of a *Ganoderma* species. These markers may be identified by comparing genomic DNA sequences between pathogenic and non-pathogenic species of *Ganoderma*. For example, when a sequence is only present in the DNA of a pathogenic *Ganoderma* species and not in the DNA of a non-pathogenic *Ganoderma* species, the sequence may be selected for further use as a marker indicative of pathogenicity of a *Ganoderma* strain or isolate. For

example, the present disclosure provides methods for the identification of single nucleotide polymorphic (SNP) markers through amplification of DNA that are able to identify and differentiate between pathogenic or non-pathogenic *Ganoderma* species in varying confidence intervals using the strategy as set out in Figure 2.

5 [0039] The present disclosure also provides uses of the identified markers for identifying a pathogenic *Ganoderma* species, differentiating a pathogenic *Ganoderma* species from a non-pathogenic *Ganoderma* species, and detecting a pathogenic *Ganoderma* species in a given sample. For example, the present disclosure provides means for screening a panel of *Ganoderma* isolates with known pathogenicity (pathogenic or non-pathogenic) using the
10 MiSeq genotyping system.

[0040] A pathogenic *Ganoderma* species is one that is capable of causing disease in plants, such as but not limited to, stem rot disease. The diseased plants may present one or more of the following symptoms such as, but not limited to white mass of mycelia, leaf chlorosis, drying up of the plant, and plant death. Table 1 provides the symptoms visible on
15 plants, according to the severity of the infection. These symptoms may also be present in one or more of the following parts of a plant infected with a pathogenic *Ganoderma* species such as, but not limited to, root, shoot, stem, bark, leaves, fruits and seed. A non-pathogenic *Ganoderma* species, on the other hand, is one that does not cause disease in plants, typically with no visible symptoms as described above.

20 [0041] The inventors have identified 123 markers (comprising 123 SNP markers) that are able to differentiate between pathogenic and non-pathogenic *Ganoderma* species, with up to 100% accuracy, as shown in Table 4. For example, the markers comprising SEQ ID NO: 36, SEQ ID NO: 39, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 78, SEQ ID NO: 84, SEQ ID NO: 96, and SEQ ID NO: 116 are able
25 to achieve the differentiation at about 50% accuracy. In another example, the markers comprising SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 45, SEQ ID NO: 48, SEQ ID NO: 51, SEQ ID NO: 54, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 62, SEQ ID NO: 66, SEQ ID NO: 70, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 87, SEQ ID NO: 94, SEQ ID NO: 99, SEQ ID NO: 101, SEQ ID NO: 103, SEQ ID NO: 112, SEQ ID NO: 119, and
30 SEQ ID NO: 122 are able to achieve the differentiation at about 60% accuracy. In another example, the markers comprising SEQ ID NOs: 25 to 27, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 53, SEQ ID NO: 64, SEQ ID NO: 74, SEQ ID NO: 81, SEQ ID NO: 85, SEQ

ID NO: 86 SEQ ID NO: 93, SEQ ID NO: 106, SEQ ID NO: 117, and SEQ ID NO: 120 are able to achieve the differentiation at about 70% accuracy. In another example, the markers comprising SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 49, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 63, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 90, SEQ ID NO: 95, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 102, SEQ ID NO: 104, SEQ ID NO: 109, SEQ ID NO: 111, and SEQ ID NO: 113 are able to achieve the differentiation at about 80% accuracy. Markers such as those comprising SEQ ID NO: 19, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 58, SEQ ID NO: 92, SEQ ID NO: 100, SEQ ID NO: 105, SEQ ID NO: 107, SEQ ID NO: 108, and SEQ ID NO: 110 are able to achieve the differentiation at about 90% accuracy. In another example, the marker comprising SEQ ID NO: 83 is able to achieve the differentiation at about 100% accuracy.

[0042] The present disclosure also provides the determination of pathogenic and non-pathogenic *Ganoderma* species by performing artificial infection in oil palm seedlings. The present disclosure further provides the use of *Ganoderma* isolates to test the pathogenicity (pathogenic or non-pathogenic) detection ability of the SNP-linked PCR primers. Exemplary isolates which may be useful include, but are not limited to, the *Ganoderma* isolates as set out in Table 2 of the present disclosure. The skilled person can further determine other *Ganoderma* strains/isolates suitable for testing pathogenicity using methods disclosed in the present disclosure. The flanking regions containing the SNP markers are provided in Table 7.

[0043] These SNP marker related primer sets can be used as a tool for early detection of *Ganoderma* in the field and in the laboratory and do not rely on the exhibition of physical symptoms of the *Ganoderma* wood rot disease (which are typically only exhibited at a late stage of infection). The sequences for these SNP marker related PCR primer sets are provided in Table 8 which may be used to amplify a DNA region from a *Ganoderma* isolate that is flanked by sequences complementary with the PCR primer sets, to thereby identify the SNP sequence linked to the pathogenicity (pathogenic or non-pathogenic) prediction and make a pathogenicity (pathogenic or non-pathogenic) prediction for the isolate. With the disclosed early detection method, planters are able to better manage the spread of the disease and subsequently reduce the loss of fresh fruit bunch yield and loss of plants before their full commercial value is realized.

[0044] In one aspect, there is provided a method for determining the pathogenicity of a *Ganoderma* species, wherein the method comprises the step of detecting the presence or

absence of one or more sequences selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is non-pathogenic.

[0045] In the context of detecting marker sequences, “absence” can refer to when the marker cannot be detected using a particular detection method or if the level of marker detected is below a pre-determined threshold value. For example, a marker may be considered to be absent if its detected level in a sample containing a *Ganoderma* DNA is below a pre-determined threshold value, such as a concentration of less than 10^{-2} ng.

[0046] In the context of detecting marker sequences, “presence” can refer to when the marker can be detected using a particular detection method or if the level of marker detected is above a pre-determined threshold value. For example, a marker may be considered to be present if its detectable level in a sample containing *Ganoderma* DNA is above the pre-determined threshold value, such as concentration of more than 10^{-2} ng. Symptoms of infection (for example, by a *Ganoderma* species), such as those that are listed in Table 1, may or may not be visible on the plant when one or more of the markers is detected and determined to be present.

[0047] Variants of the markers of the present disclosure may have at least 80% identity, at least 85% identity, at least 90% identity, at least 95% identity or at least 99% identical to any one of SEQ ID NOs 1 to 123. The variants may have about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98% or about 99% identical to any one of SEQ ID NOs 1 to 123. In one embodiment, the variant has at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of SEQ ID NOs: 1 to 123.

[0048] In one embodiment, the one or more marker sequences are selected from the group consisting of SEQ ID NO: 19, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 56,

SEQ ID NO: 58, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 74, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO 102, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 113; SEQ ID NO: 117, and SEQ ID NO: 120, or a variant thereof. In one embodiment, the sequences show a confidence interval of at least 70%.

[0049] In another embodiment, the one or more sequences are selected from the group consisting of SEQ ID NO: 25, SEQ ID NO: 28, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 107, and SEQ ID NO: 111, or a variant thereof. Advantageously, these sequences have a percentage of mappable reads of at least 60% and are detectable in a soil sample when present in an amount as low as 10^{-2} ng using the methods of the present disclosure.

[0050] The one or more sequences, or a variant thereof, as described above may comprise a single nucleotide polymorphism (SNP). For example, a sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 123, or a variant thereof, comprises a SNP. The nucleotide sequences flanking and containing SNP markers that can be used to identify and differentiate between pathogenic or non-pathogenic *Ganoderma* species with varying confidence levels using MiSeq are shown in Tables 4 and 7 of the present disclosure (in Example 3).

[0051] Exemplary methods which may be used for detecting SNPs include, but are not limited to, sequencing, hybridization-based methods (such as, but not limited to, dynamic allele-specific hybridization, molecular beacons, and SNP microarrays), enzyme-based methods such as, but not limited to, restriction fragment length polymorphism, Flap endonuclease, primer extensions, 5'-nuclease, oligonucleotide ligation assay, and other PCR-based methods), and other post-amplification methods which are based on physical properties of the SNP-containing amplified DNA (such as, but not limited to, single strand conformation polymorphism, temperature gradient gel electrophoresis, denaturing high performance liquid chromatography, high-resolution melting of the entire amplicon, use of DNA mismatch-binding proteins, SNPLEX and surveyor nuclease assay).

[0052] The detection using the methods of the disclosure may be carried out on any suitable sample. For example, the detection may be conducted on a sample that potentially contains or is suspected of containing a pathogenic *Ganoderma* strain. Alternatively, the detection may also be conducted on a sample where there is no suspicion of pathogenic *Ganoderma* strain infection or presence, for example in a routine check for infection.

[0053] Samples on which the methods of the disclosure may be applied include, but are not limited to, environmental samples such as soil samples, and plant parts selected from root, shoot, stem, bark, leaves, fruits, seed, and laboratory samples such as an isolated microbial strain. The “isolated microbial strain” may be one that has been isolated from environmental samples such as soil samples, and plant parts selected from root, shoot, stem, bark, leaves, fruits and seed, and may be present, for example, as a cell culture.

[0054] The detection in the methods of the disclosure may be achieved using techniques available in the art, such as but not limited to sequencing (such as MiSeq sequencing and HiSeq sequencing), hybridization-based methods, enzyme-based methods, PCR-based methods, and post-amplification methods. In one embodiment, the PCR in a PCR-based method is conducted using at least one set of primers selected from the group consisting of SEQ ID NOs: 124 to 369.

[0055] In one embodiment, the method of the disclosure comprises the steps of: (a) extracting DNA from a sample; (b) subjecting the DNA to sequencing to determine sequences of the extracted DNA; (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof; and (d) determining the absence or presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the *Ganoderma* species is non-pathogenic. Exemplary methods which may be used for extracting DNA from *Ganoderma*, include, but are not limited to, the rapid preparation protocol based on Reader and Broda (1985), CTAB extraction method and Mericon extraction method. Other protocols suitable for extracting DNA from a fungal sample may also be used. Exemplary methods which may be used for extracting DNA from a

soil sample include, but are not limited to, the use of the MOBIO Powerlyzer PowerSoil DNA Isolation Kit. Other protocols suitable for extracting DNA from soil may also be used.

[0056] The methods and uses described herein may be used for detection of pathogenic Ganoderma infection in a wide variety of plants such as, but not limited to, plants or trees susceptible to Ganoderma infection. Exemplary plants which may be susceptible to Ganoderma infection, include, but are not limited to crops (such as oil palm, *Mucuna bracteata*, coconut, betel nut, sugar palm, sago palm, rubber tree, *Acacia mangium*), forest trees, and ornamental plants (such as ornamental palm *Chrysalidocarpus lutescens*).

[0057] A pathogenic Ganoderma strain that may be detected or identified using the methods of the disclosure includes, but is not limited to, a strain from *Ganoderma boninense* species. Additional exemplary Ganoderma species which may also be detected include, but are not limited to, *G. steyaertanum*, *G. weberianum*, *G. lucidum*, *G. orbiforme*, *G. pseudoferreum*, *G. fornicatum*, *G. williamsianum* and *G. applanatum*. Exemplary pathogenic species which may be detected are such as, but not limited to *G. boninense*, *G. steyaertanum* and *G. orbiforme*. Exemplary non-pathogenic species which may be detected are such as, but not limited to *G. weberianum*, *G. lucidum*, *G. fornicatum*, *G. williamsianum*, *G. pseudorerreum* and *G. applanatum*.

[0058] In another aspect, there is provided a method for detecting presence of a pathogenic strain of Ganoderma in a sample, comprising the steps of: (a) extracting DNA from the sample; (b) subjecting the DNA to sequencing to determine sequences of the extracted DNA; (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof; and (d) determining the absence or presence of one or more sequences selected from a group consisting of SEQ ID NOs: 1 to 123 or a variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the Ganoderma strain is a pathogenic strain, and the absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123 or a variant thereof indicates that the Ganoderma strain is a non-pathogenic strain. Exemplary methods which may be used for extracting DNA from Ganoderma, include, but are not limited to the rapid preparation protocol based on Reader and Broda (1985), CTAB extraction method and Mericon extraction method. Other protocols suitable for extracting DNA from a fungal sample may also be used. Exemplary methods which may be used for extracting DNA from a

soil sample, include, but are not limited to the use of the MOBIO Powerlyzer PowerSoil DNA Isolation Kit. Other protocols suitable for extracting DNA from soil may also be used.

[0059] In another aspect, there is provided a method for identifying a nucleic acid sequence containing a marker for Ganoderma pathogenesis comprising the steps of: (a)

- 5 extracting DNA from a group of pathogenic and non-pathogenic Ganoderma strains; (b) preparing whole-genome sequencing libraries of each of the pathogenic and non-pathogenic Ganoderma strain using the extracted DNA of step (a); (c) sequencing the libraries from step (b) to form a plurality of sequencing reads; (d) assembling the sequencing reads from step (b) to generate a backbone sequence for the pathogenic Ganoderma strain and a backbone
10 sequence for the non-pathogenic Ganoderma strain; (e) mapping the sequencing reads from step (c) from each strain to the backbone sequence of step (d) to perform comparative genomics to identify variant sequence(s); (f) designing and preparing primers from the variant sequence(s) from step (e); (g) applying the primers prepared from step (f) to DNA extracted from the pathogenic and non-pathogenic Ganoderma strains for DNA amplification
15 and sequencing; and (h) classifying the variant sequence(s) into those associated with the pathogenic Ganoderma strains and those associated with the non-pathogenic Ganoderma strains. Comparative genomics is a technique which enables comparison of a genomic feature (such as a DNA sequence) from different organisms in order to identify similar or different features. In the methods of the present disclosure, comparative genomics may be used to
20 identify differences in DNA sequence between pathogenic and non-pathogenic species of Ganoderma which may subsequently be used as molecular markers to determine pathogenicity of any given strain of Ganoderma.

[0060] The extraction of DNA from a sample may be conducted using methods known in the art, such as, but not limited to the rapid preparation protocol based on Reader and Broda
25 (1985), CTAB extraction method and Mericon extraction method; or using commercial kits such as, but not limited to DNeasy Qiagen Plant Mini Kit and MOBIO Powerlyzer PowerSoil DNA Isolation Kit. Sequencing may be performed using methods known in the art, such as, but not limited to next-generation sequencing (such as MiSeq and HiSeq), chain-termination methods and *de novo* sequencing.

- 30 [0061] In another aspect, there is provided an oligonucleotide primer comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 124 to 369.

[0062] In another aspect, there is provided a kit for use in the methods defined herein, wherein the kit comprises: (a) one or more primer(s) selected from the group consisting of SEQ ID NOs: 124 to 369 for determining the presence or absence of one or more sequences selected from the group consisting of SEQ ID NOs: 1 to 123; (b) one or more container(s) suitable for sample collection; (c) one or more container(s) suitable for sample processing; (d) one or more reagents for sample processing; (e) one or more buffer(s); (f) one or more PCR reagent(s) and (g) instructions for use in accordance with the methods described herein. The kit may further comprise other components such as, but not limited to, filtration columns for sample purification. PCR reagents may include, but are not limited to, nucleotides, template DNA (to be used as negative or positive controls in a reaction), PCR buffers, and enzymes (such as DNA polymerase).

[0063] Establishing a platform and method for the detection of Ganoderma in environmental sample such as soil is challenging due to the complexity of soil DNA as compared to plant samples or pure pathogen. The SNP markers of the present disclosure allow for the early prediction of presence of a pathogenic Ganoderma strain from any plant or environmental sample and will be useful for the planning of how to address the infection status. The markers of the present disclosure can be used in different formats, kits and protocols as determined by a person skilled in the art, for determining the presence of a pathogenic Ganoderma strain.

[0064] Unlike other known methods that detect Ganoderma in diseased palms, the development of SNP markers in the present disclosure allows the differentiation of pathogenic Ganoderma strains from non-pathogenic Ganoderma strains. Known methods for detecting Ganoderma involve detection of ergosterol (found in cell membrane of fungi), detection of internal transcribed spacer (ITS) region, protein profiling (WO 2014/109629 A1) and fragment length analysis (WO 2013/066144 A1). In contrast, the present disclosure utilizes molecular markers (particularly SNP markers) that can differentiate between, and thereby detect pathogenic and non-pathogenic Ganoderma strains. SNP markers are more sensitive and specific compared to biochemical markers such as ergosterol and protein profiling. A SNP marker which is based on a particular sequence is also more specific than fragment length analysis which is based on detection of the length of a DNA fragment. Advantageously, a SNP marker which is present in a specific location in the genome is

unique and thus more specific than an ITS marker which may be present as multiple copies in a genome due to its high copy number.

[0065] Other known methods also use gel electrophoresis system (which are of a smaller scale), whereas the present disclosure utilizes high-throughput sequencing with the ability to screen 384 samples within 40 hours. In addition, known methods are restricted to pure isolates, while the present disclosure provides methods that can be applied directly onto environmental samples, such as soil samples. The methods of the present disclosure circumvent the additional steps needed for growing and isolating microbes from samples prior to identification of Ganoderma, which results in shorter detection time.

BRIEF DESCRIPTION OF THE DRAWINGS

[0066] The disclosure will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

[0067] **Fig. 1.** The figure shows visual symptoms of Ganoderma infection as described in Table 1 for categorizing the disease into one of the 5 disease classes.

[0068] **Fig. 2.** The flow chart describes the process of identification of the markers for determining the presence of pathogenic or non-pathogenic Ganoderma directly from soil sample as described in Example 3.

TABLES

[0069] Table 1: Disease severity index for Ganoderma disease in oil palm.

[0070] Table 2: Test list of pathogenic and non-pathogenic Ganoderma strains.

[0071] Table 3: List of pathogenic and non-pathogenic Ganoderma strains used for amplicon screening.

[0072] Table 4: Percentage confidence level for each SNP marker in differentiating pathogenic and non-pathogenic Ganoderma.

[0073] Table 5: List of markers showing the associated mappable reads (criterion 1) and their detectability at 10^{-2} ng in soil samples (criterion 2). The markers which were ultimately selected were those which have mappable reads of at least 60% and are able to consistently detect Ganoderma DNA in soil present in an amount as low as 10^{-2} ng.

[0074] Table 6: Shortlist of markers showing the associated mappable reads (criterion 1) and their detectability at 10^{-2} ng in soil samples (criterion 2).

[0075] Table 7: Flanking sequences containing SNPs associated with pathogenicity of Ganoderma.

[0076] Table 8: Primer sequences for amplicon sequencing.

5

EXAMPLES

Example 1: Materials and Methods

Method for identifying a nucleic acid sequence containing a marker for Ganoderma pathogenesis

[0077] In order to mine for candidate markers for distinguishing between pathogenic and non-pathogenic Ganoderma, comparative genomics was performed between both groups. The genome size of Ganoderma species is estimated to be approximately 50 Mbps. For the marker discovery, a total of 16 pathogenic Ganoderma isolates (i.e. strains) and six non-pathogenic Ganoderma isolates were sequenced. Whole genome sequencing of at least 100x coverage for each isolate was performed on Illumina HiSeq2500. After the whole genome sequencing, all the reads generated were pulled together to be assembled as a combined assembly using CLC Bio Assembly Cell version 4.06beta.67189. The assembled sequence serves as the backbone for the comparative genome for the SNP mining. Upon establishing the backbone sequence, all the reads from each isolate were mapped using Burrows-Wheeler Alignment (BWA) version 0.5.9 to the backbone and variant called using SAMtools (version 0.1.16) with base quality cut off of $\geq Q25$ and base coverage of ≥ 4 reads. Thereafter, polymorphic SNPs between pathogenic and non-pathogenic Ganoderma were identified. A total of 123 SNPs were shortlisted to be used for amplicon screening using the 86 isolates from Table 3.

Whole-genome Sequencing (WGS)

[0078] WGS Illumina libraries were prepared using DNA isolated from 16 pathogenic and six non-pathogenic Ganoderma isolates (listed in Table 2). DNA extraction was performed according to the rapid preparation protocol based on Reader and Broda (1985). Genomic DNA (2 μ g) was sheared using the Covaris S220 to generate insert size of approximately 550 bp. The sheared DNA was end repaired, size selected, added with A-tail and ligated with Illumina adaptors using TruSeq DNA PCR-free Library Prep Kit (from Illumina). The prepared libraries were verified using Agilent Technologies 2100 Bioanalyzer for qualitative purposes on the library size distribution. A SYBR green quantitative PCR (qPCR) assay with

primers specific to the Illumina adapters (KAPA Library Quantification Kits) was used to determine the concentration of the libraries. The final library was mixed with Illumina-generated PhiX control libraries and denatured using sodium hydroxide. The detailed protocol with index sequences is provided by Illumina protocol and the supplemental material (which can be found on the Illumina website at https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqdnpcrfree/truseq-dna-pcr-free-library-prep-guide-15036187-d.pdf, https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseqkits/hiseq-ga-denaturing-diluting-libraries-reference-guide-15050107-03.pdf and https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseq2500/hiseq-2500-system-guide-15035786-01.pdf). The denatured libraries were loaded into cBot for cluster generation on HiSeq flowcell (TruSeq Cluster Kit v3) and were sequenced on the Illumina HiSeq 2500 using TruSeq SBS Kit v3, paired-end 2x100 cycles.

Ganoderma Genotyping-By-Sequencing (GBS) Analysis

[0079] Raw reads generated from Next-Generation Sequencing (NGS) of 16 pathogenic and six non-pathogenic *Ganoderma* isolates (listed in Table 2) using HiSeq2000 2 x 100bp pair-ended approach was filtered at base quality cut off of $\geq Q25$ and trimmed at sequence length ≥ 72 bp using trim_fastq.pl. Filtered and trimmed reads were mapped to the *G. boninense* backbone sequence using Burrows-Wheeler Alignment (BWA) version 0.5.9. Single nucleotide polymorphism (SNP) variant calling was performed using SAMtools version 0.1.16 with base quality cut off of $\geq Q25$ and base coverage of ≥ 4 reads.

Shortlisting of the Candidate markers from GBS results

[0080] SNPs markers were selected based on variation that shows homozygous / heterozygous bases differentiating at least 50% of each of the pathogenic and non-pathogenic *Ganoderma* isolates. The SNP markers differentiating pathogenic and non-pathogenic *Ganoderma* isolates were shortlisted for MiSeq sequencing analysis.

Sample preparation for sequencing

[0081] A total of 71 libraries were constructed using the pool of 123 amplicons of the regions of interest from 71 *Ganoderma* samples listed in Table 3. The DNA used for

constructing the libraries was isolated from *Ganoderma* using the rapid preparation protocol based on Reader and Broda (1985). Amplicons were generated using a high-fidelity polymerase (KAPA HiFi HotStart ReadyMix) and then were purified using a magnetic bead capture kit (Agencourt AMPure XP) and quantified using a fluorometric kit (QuantIT PicoGreen; Invitrogen) and Agilent Technologies 2100 Bioanalyzer (Agilent High Sensitivity DNA Kit). The purified amplicons were pooled according to the sample and indexed using Nextera XT index kit. The libraries were then purified using magnetic bead capture kit and verified using Agilent Technologies 2100 Bioanalyzer (Agilent High Sensitivity DNA Kit) on the library size distribution. Samples from all libraries were pooled using equal molar quantities of DNA into one final library and the final concentration of the pooled library was determined using qPCR assay as described above in the WGS protocol using Illumina HiSeq 2500. The final pooled library was mixed with Illumina-generated PhiX control libraries and denatured using sodium hydroxide. The detailed protocol with index sequences is provided by Illumina protocol for 16S metagenomics sequencing library preparation (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf and https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-reagent-kit-v3-reagent-prep-guide-15044983-b.pdf) and the supplemental material. The sequencing runs were performed on MiSeq using MiSeq Reagent V3 kit (paired-end 2x300 cycles) with Illumina Real Time Analysis (RTA) and HiSeq Control Software (HCS).

[0082] Subsequent screenings of amplicon sequencing using MiSeq system to further shortlist the markers were carried out using similar methods as described above using the appropriate number of markers and samples.

MiSeq Sequencing Analysis

[0083] Sequence base calling of MiSeq flow cell was performed using MiSeq reporter with zero base mismatch allowed for de-multiplexing based on sample index sequences. For the MiSeq run with *Ganoderma* DNA, de-multiplexed reads were mapped to the targeted region of the SNP location without secondary de-multiplexing step. For MiSeq run with soil DNA, de-multiplexed reads based on samples were further secondly de-multiplexed based on amplicons primer sequences with minimal three base mismatches using `fastx_barcode_splitter.pl` script in `fastx` toolkit version 0.0.13.2. Raw sequences were filtered

with base quality cut off of $\geq Q25$ and trimmed at sequence length ≥ 72 bp using trim_fastq.pl. Filtered and trimmed sequences were mapped to the backbone sequence prepared as described above using Burrows-Wheeler Alignment (BWA) version 0.5.9 with default parameter settings ($\leq 4\%$ mismatch of the read length). Single nucleotide polymorphism (SNP) variant calling was performed using SAMtools version 0.1.16 with based quality cut off of $\geq Q25$ and base coverage of ≥ 4 reads.

Example 2: Determination of Ganoderma Pathogenicity Level

[0084] To measure and observe the effect that isolated Ganoderma strains have on oil palm, the oil palm seeds were germinated and then infected with the Ganoderma strains through the use of rubber wood blocks that were inoculated with the Ganoderma strains. The rubber wood blocks were prepared using the following method:

[0085] Step 1: Rubber wood blocks measuring 2.5" x 2.5" x 5" were placed into an autoclavable plastic bag with media suitable for Ganoderma growth, sealed and then autoclave sterilized.

[0086] Step 2: The prepared rubber wood blocks containing media were then inoculated with Ganoderma by introducing media with cultivated Ganoderma on it.

[0087] Step 3: Using this method, each of the strains of Ganoderma were localized onto rubber wood blocks which were then used to challenge the oil palm seedlings as part of a pathogenicity trial so that the pathogenicity (pathogenic or non-pathogenic) of the Ganoderma strains can be determined.

[0088] A total of 22 strains of Ganoderma (listed in Table 2) were tested using this method.

[0089] The experimental design of the pathogenicity trial consisted of randomized complete block design with four replicates and 12 plots per replicate for each of the Ganoderma strain. After the germinated oil palm seeds were introduced to the Ganoderma impregnated rubber wood blocks, bi-weekly observations were carried out to observe the development of the disease symptoms until 36 weeks post-infection. A disease severity index was created as shown in Table 1 and the Ganoderma strains were assessed to be either pathogenic or non-pathogenic based on visual inspection of the symptoms shown in Figure 1. In cases where the oil palm seedlings did not exhibit symptoms of Ganoderma disease, the Ganoderma strain used were categorized as non-pathogenic.

[0090] Table 1: Disease severity index for Ganoderma disease in oil palm.

Disease Class	Description
0	Healthy plant
1	Without white mass of mycelia on any part of plant, with leaf chlorosis
2	With white mass of mycelia on any part of plant, without leaf chlorosis
3	With white mass of mycelia on any part of plant, with leaf chlorosis
4	With Ganoderma fruiting body, at least 50% of total leaf show chlorosis, plant drying up, dying or already dead

Example 3: Pathogenicity Testing5 **Marker Identification**

[0091] A total of 22 Ganoderma isolates (listed in Table 2) which are either non-pathogenic or pathogenic were selected for genome sequencing. On average, each genome has a genome coverage of 100X to provide a good assembly. The genomes were individually mapped to the Ganoderma backbone sequence.

10 [0092] Table 2: Test list of pathogenic and non-pathogenic Ganoderma strains.

Isolate No.	Isolate ID	Identification	Pathogenic / Non-pathogenic
1	JOP01-G2-B	<i>Ganoderma sp. MT-1</i>	Pathogenic
2	S06-G1-A	<i>G. boninense</i>	Pathogenic
3	ATCC	<i>G. boninense</i>	Pathogenic
4	SG09-G2-B	<i>G. boninense</i>	Pathogenic
5	TJ02-G1-B	<i>G. boninense</i>	Pathogenic
6	BS05-G1-C	<i>G. orbiforme</i>	Pathogenic
7	KB12-G1-A	<i>Ganoderma sp. BL-9 18S</i>	Pathogenic
8	T26-G1-B	<i>G. boninense</i>	Pathogenic
9	SP02-G1-C	<i>G. boninense</i>	Pathogenic
10	SL07-G1-A	<i>Ganoderma sp. C16452</i>	Pathogenic
11	SG12-G1-A	<i>G. boninense</i>	Pathogenic
12	SR05-G1-A	<i>G. boninense</i>	Pathogenic
13	SR09-G1-A	<i>G. boninense</i>	Pathogenic
14	C03-G1-A	<i>G. boninense</i>	Pathogenic
15	T12-G3-B	<i>G. steyaertanum</i>	Pathogenic
16	SG02-G1-A	<i>Ganoderma sp. BL-94 18S</i>	Pathogenic
17	TG05	<i>G. weberianum</i>	Non-pathogenic
18	GL-B	<i>G. lucidum</i>	Non-pathogenic

Isolate No.	Isolate ID	Identification	Pathogenic / Non-pathogenic
19	IMI 389522	<i>G. pseudoferreum</i>	Non-pathogenic
20	LW49	<i>Ganoderma sp. G31</i>	Non-pathogenic
21	SL65	<i>G. fornicatum</i>	Non-pathogenic
22	LY59	<i>G.williamsianum</i>	Non-pathogenic

[0093] Comparative genomics for each of the 16 pathogenic and 6 non-pathogenic genomes (from the 22 strains in Table 2) to the backbone sequences found that the amount of SNP varies from isolate to isolate. The number of SNP discovered in each genome ranges from 8,841 to 958,490. *In silico* comparison of SNPs that are able to differentiate the pathogenic and non-pathogenic group was done and 123 unique SNPs were identified (SEQ ID NOs: 1 to 123 as set out in Table 7). These SNPs have no flanking SNPs within 300 bp. The criterion for selection is that each SNP has to be able to predict the pathogenicity or non-pathogenicity of an isolate for at least 50% of a test list of pathogenic and non-pathogenic isolates. Each SNP is compared against the list and only those that satisfied the criterion of being able to correctly predict pathogenicity or non-pathogenicity 50% of the time is selected.

[0094] Primers designed for the 123 SNP markers (as set out in Table 8) were screened on 71 *Ganoderma* isolates using MiSeq which amplified 800bp amplicons. Mapping of the sequences based on amplicon found that the 123 markers have different percentage confidence level of differentiating pathogenic and non-pathogenic *Ganoderma*. The fragments of the sequences generated from the sequencing were mapped to the expected SNP position on the backbone sequence to identify the SNP call on the sequenced fragment (of each sample).

[0095] Table 3: List of pathogenic and non-pathogenic *Ganoderma* strains used for amplicon screening.

No	Isolate ID	Pathogenic/ Non-pathogenic	Identification	Accession	Similarity identification measured as high scoring pair (HSP) percentage (%)
1	SG13-G1-A	P	<i>Ganoderma boninense</i> isolate G8	KR093029.1	90.18
2	KB06-G1-A	P	<i>Ganoderma boninense</i> isolate GB001	KX092000.1	94.36

No	Isolate ID	Pathogenic/ Non-pathogenic	Identification	Accession	Similarity identification measured as high scoring pair (HSP) percentage (%)
3	S02-G2-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.68
4	S02-G5-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.68
5	S03-G1-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.84
6	SP01-G1-A	P	Ganoderma boninense isolate GB001	KX092000.1	97.71
7	SR05-G1-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.84
8	T10-G2-B	P	Ganoderma boninense isolate GB001	KX092000.1	99.84
9	TM03-G2-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.52
10	TM04-G2-B	P	Ganoderma boninense isolate GB001	KX092000.1	99.68
11	TM08-G2-A	P	Ganoderma boninense isolate GB001	KX092000.1	99.68
12	TM09-G1-A	P	Ganoderma boninense isolate GB001	KX092000.1	98.56
13	TM11-G1-A	P	Ganoderma boninense isolate GB001	KX092000.1	90.97
14	JOP03-G1-A	P	Ganoderma boninense isolate UPMGB001	KM220584.1	100
15	T26-G1-B	P	Ganoderma boninense isolate UPMGB001	KM220584.1	99.82
16	TJ02-G1-A	P	Ganoderma boninense isolate UPMGB001	KM220584.1	99.47
17	TJ02-G2-A	P	Ganoderma boninense isolate UPMGB001	KM220584.1	99.29
18	TJ03-G2-A	P	Ganoderma boninense isolate UPMGB001	KM220584.1	99.82
19	T11-G3-A	P	Ganoderma boninense isolate UPMGB003	KM271998.1	90.48
20	SL05-G1-B	P	Ganoderma boninense strain PER71	KM015454.1	100
21	T25-G1-A	P	Ganoderma boninense strain PER71	KM015454.1	99.65
22	BS05-G1-C	P	Ganoderma orbiforme isolate BCC22325	KX421867.1	100
23	BS07-	P	Ganoderma orbiforme	KX421867.1	99.65

No	Isolate ID	Pathogenic/ Non-pathogenic	Identification	Accession	Similarity identification measured as high scoring pair (HSP) percentage (%)
	G1-A		isolate BCC22325		
24	C08-G2-A	P	Ganoderma orbiforme isolate BCC22325	KX421867.1	99.4
25	SG05-G1-A	P	Ganoderma orbiforme isolate BCC22325	KX421867.1	99.35
26	TM12-G1-A	P	Ganoderma orbiforme isolate BCC22325	KX421867.1	98.4
27	KB02-G1-A	P	Ganoderma sp. BL-6 18S	JN400509.1	90.59
28	KB05-G2-A	P	Ganoderma sp. BL-6 18S	JN400509.1	98.73
29	T15-G2-A	P	Ganoderma sp. BL-6 18S	JN400509.1	92.43
30	KB05-G1-A	P	Ganoderma sp. BL-9 18S	JN400510.1	97.13
31	KB12-G1-A	P	Ganoderma sp. BL-9 18S	JN400510.1	99.84
32	TM06-G1-A	P	Ganoderma sp. BL-9 18S	JN400510.1	91.52
33	TM13-G2-A	P	Ganoderma sp. BL-9 18S	JN400510.1	99.65
34	SG02-G1-A	P	Ganoderma sp. BL-94 18S	JN400511.1	99.52
35	SG06-G2-B	P	Ganoderma sp. BL-94 18S	JN400511.1	99.52
36	SG03-G1-A	P	Ganoderma sp. BL-94 18S	JN400511.1	98.1
37	TM01-G1-A	P	Ganoderma sp. BL-94 18S	JN400511.1	99.52
38	C06-G1-A	P	Ganoderma sp. BP-16 18S	JN400513.1	99.52
39	KB10-G1-A	P	Ganoderma sp. BP-16 18S	JN400513.1	99.02
40	KB11-G1-A	P	Ganoderma sp. BP-16 18S	JN400513.1	95.27
41	SR02-G1-B	P	Ganoderma sp. BP-16 18S	JN400513.1	99.21
42	T27-G1-A	P	Ganoderma sp. BRIUMSa	JN234427.1	99.65
43	T14-G2-A	P	Ganoderma sp. BRIUMSb	JN234428.1	96.16

No	Isolate ID	Pathogenic/ Non-pathogenic	Identification	Accession	Similarity identification measured as high scoring pair (HSP) percentage (%)
44	C02-G1-B	P	Ganoderma sp. BRIUMSc	JN234429.1	90.81
45	SG04-G2-A	P	Ganoderma sp. BRIUMSc	JN234429.1	94.72
46	KB13-G1-A	P	Ganoderma sp. C16452	EU239386.1	99.14
47	SL07-G1-A	P	Ganoderma sp. C16452	EU239386.1	99.82
48	SL09-G1-A	P	Ganoderma sp. C16452	EU239386.1	99.15
49	BS06-G1-B	P	Ganoderma sp. MT-1	AY220543.1	100
50	JOP01-G1-A	P	Ganoderma sp. MT-1	AY220543.1	100
51	JOP01-G2-B	P	Ganoderma sp. MT-1	AY220543.1	100
52	SG14-G1-C	P	Ganoderma sp. MT-1	AY220543.1	99.82
53	SL21	P	Ganoderma boninense isolate GB001	KX092000.1	99.68
54	SL22	P	Ganoderma orbiforme strain BCC22324	JX997990.1	99.68
55	SL65	NP	Ganoderma fornicatum	JX840347.1	99.68
56	GL-A	NP	Ganoderma lucidum strain GI-20	GU213481.1	99
57	LW17	NP	Ganoderma sp. G31	KR093030.1	99.82
58	LW47	NP	Ganoderma sp. G31	KR093030.1	99.82
59	TM49	NP	Ganoderma sp. G31	KR093030.1	99.36
60	SL132	NP	Ganoderma sp. G31	KR093030.1	99.64
61	LW08	NP	Ganoderma sp. G31	KR093030.1	99.64
62	LW16	NP	Ganoderma sp. G31	KR093030.1	99.64
63	SL36	NP	Ganoderma sp. G31	KR093030.1	99.16
64	LW15	NP	Ganoderma sp. G31	KR093030.1	99.64
65	LW49	NP	Ganoderma sp. G31	KR093030.1	99.64
66	TG05	NP	Ganoderma weberianum strain CBS	JQ520219.1	100
67	LY59	NP	Ganoderma williamsianum	KU219995.1	100

No	Isolate ID	Pathogenic/ Non-pathogenic	Identification	Accession	Similarity identification measured as high scoring pair (HSP) percentage (%)
68	LY11	NP	Ganoderma sp. MEL 2382607	KP012934.1	99.84
69	LY57	NP	Ganoderma sp. MEL 2382607	KP012934.1	99.84
70	LY12	NP	Ganoderma applanatum	KR867655.1	99.84
71	LY10	NP	Ganoderma sp. MEL 2382607	KP012934.1	99.68

[0096] To determine the confidence level of a SNP marker, if for example 50 out of 56 of the pathogenic isolates tested have the expected SNP call associated with pathogenicity, then the confidence level is 89%. The percentage confidence level for each marker based on this calculation method is as tabulated in Table 4.

[0097] There is one SNP marker that showed 100% confidence level in differentiating pathogenic and non-pathogenic Ganoderma, 11 markers at 90% confidence level, 17 SNP markers at 80% confidence level, followed by 15 markers at 70% confidence level, 21 SNP markers at 60% confidence level, 11 SNP markers at 50% confidence level and 47 SNP markers that is less than 50% confidence level. The assay was repeated and the results were reproducible.

[0098] Table 4: Percentage confidence level for each SNP marker in differentiating pathogenic and non-pathogenic Ganoderma. Each SNP marker was screened on 71 pure Ganoderma samples. SNP call [P] refers to the SNP observed in pathogenic Ganoderma.

SEQ ID NO.	Marker ID	SNP call	% confidence level
1	GMS001	T	< 50
2	GMS002	G	< 50
3	GMS003	T	< 50
4	GMS004	T	< 50
5	GMS005	C	< 50
6	GMS006	A	< 50
7	GMS007	T	< 50
8	GMS008	A	< 50
9	GMS009	T	< 50

SEQ ID NO.	Marker ID	SNP call	% confidence level
10	GMS010	C	60
11	GMS011	G	< 50
12	GMS012	A	< 50
13	GMS013	G	60
14	GMS014	T	< 50
15	GMS015	C	< 50
16	GMS016	T	< 50
17	GMS017	T	< 50
18	GMS018	T	< 50
19	GMS019	T	90
20	GMS020	T	< 50
21	GMS021	T	< 50
22	GMS022	G	< 50
23	GMS023	T	< 50
24	GMS024	T	< 50
25	GMS025	T	70
26	GMS026	T	70
27	GMS027	A	70
28	GMS028	G	90
29	GMS029	G	< 50
30	GMS030	A	< 50
31	GMS031	G	< 50
32	GMS032	T	< 50
33	GMS033	T	70
34	GMS034	T	< 50
35	GMS035	C	< 50
36	GMS036	C	50
37	GMS037	C	70
38	GMS038	T	80
39	GMS039	T	50
40	GMS040	C	< 50
41	GMS041	C	< 50
42	GMS042	C	< 50
43	GMS043	C	< 50
44	GMS044	T	< 50
45	GMS045	T	60
46	GMS046	T	80
47	GMS047	A	< 50
48	GMS048	T	60
49	GMS050	G	80

SEQ ID NO.	Marker ID	SNP call	% confidence level
50	GMS051	G	90
51	GMS052	C	60
52	GMS053	G	90
53	GMS054	C	70
54	GMS055	A	60
55	GMS056	T	80
56	GMS057	C	80
57	GMS058	T	60
58	GMS059	T	90
59	GMS060	T	60
60	GMS061	A	50
61	GMS062	C	50
62	GMS063	A	60
63	GMS064	A	80
64	GMS065	A	70
65	GMS066	G	50
66	GMS067	T	60
67	GMS068	T	80
68	GMS069	C	< 50
69	GMS070	A	80
70	GMS071	A	60
71	GMS073	T	50
72	GMS074	T	50
73	GMS075	T	< 50
74	GMS076	T	70
75	GMS077	G	< 50
76	GMS078	T	60
77	GMS079	C	60
78	GMS080	G	50
79	GMS081	T	< 50
80	GMS082	G	< 50
81	GMS083	A	70
82	GMS084	T	< 50
83	GMS085	T	100
84	GMS087	T	50
85	GMS088	T	70
86	GMS090	T	70
87	GMS092	A	60
88	GMS093	G	< 50
89	GMS094	C	< 50

SEQ ID NO.	Marker ID	SNP call	% confidence level
90	GMS096	A	80
91	GMS097	G	< 50
92	GMS098	A	90
93	GMS099	T	70
94	GMS100	A	60
95	GMS101	G	80
96	GMS102	C	50
97	GMS103	T	80
98	GMS104	C	80
99	GMS105	C	60
100	GMS106	A	90
101	GMS107	C	60
102	GMS108	A	80
103	GMS109	G	60
104	GMS110	G	80
105	GMS111	G	90
106	GMS112	T	70
107	GMS114	A	90
108	GMS115	G	90
109	GMS116	A	80
110	GMS117	A	90
111	GMS118	T	80
112	GMS119	C	60
113	GMS120	T	80
114	GMS122	T	< 50
115	GMS123	C	< 50
116	GMS124	T	50
117	GMS125	A	70
118	GMS126	A	< 50
119	GMS127	C	60
120	GMS128	G	70
121	GMS129	G	< 50
122	GMS130	T	60
123	GMS132	T	< 50

[0099] Based on the screening results of 123 markers on 71 pure *Ganoderma* samples, 44 markers with at least 70% confidence level (as listed in Table 5) were selected to be tested on environmental samples such as soil to assess the sensitivity of the SNP markers in detecting pathogenic *Ganoderma* in the soil using MiSeq.

[00100] Six soil samples were collected from oil palm estate (three of which are from Ganoderma infected areas and another three from non-infected areas).

[00101] In order to test the sensitivity limit of the 44 markers, five sets of the same soil samples were spiked with different concentrations of DNA isolated from a pathogenic strain of Ganoderma (Isolate ID: SG12-G1-A): 1ng, 10^{-2} ng, 10^{-4} ng, 10^{-6} ng, and 10^{-8} ng. DNA was extracted from Ganoderma using the rapid preparation protocol based on Reader and Broda (1985). DNA was extracted from these six sets of soil samples (one set original soil and five sets spiked with Ganoderma soil) for the marker screening using MiSeq using the MOBIO Powerlyzer PowerSoil DNA Isolation Kit according to the manufacturer's recommendation.

Similar protocols and parameters were adopted for the DNA sequencing and bioinformatics analysis. The results showed that detection of Ganoderma in the soil could be achieved even at concentration as low as 10^{-2} ng. These 44 markers were further shortlisted to 14 markers which can effectively detect Ganoderma in soil samples.

[00102] The criterion for SNP marker selection was that there must be more than 60% of the total reads mappable to the targeted regions in both spiked and non-spiked samples, and also be able to detect Ganoderma DNA at concentration as low as 10^{-2} ng.

[00103] Table 5 summarizes the results of screening 44 markers on soil samples. A total of 14 SNP markers in the table were shortlisted based on the two criteria described above.

[00104] Table 5: List of markers showing the associated mappable reads (criterion 1) and their detectability at 10^{-2} ng in soil samples (criterion 2). The markers which were ultimately selected were those which have mappable reads of at least 60% and are able to consistently detect Ganoderma DNA in soil present in an amount as low as 10^{-2} ng.

SEQ ID NO.	Marker ID	Criterion 1: Percentage mappable reads of at least 60%	Criterion 2: Detection of expected SNP in at least 10^{-2} ng spiked samples
19	GMS019	0.00%	No
25	GMS025	73.16%	Yes
26	GMS026	91.14%	No
27	GMS027	35.85%	No
28	GMS028	92.71%	Yes
33	GMS033	88.04%	No
37	GMS037	23.89%	No
38	GMS038	91.23%	Yes
46	GMS046	91.91%	Yes

SEQ ID NO.	Marker ID	Criterion 1: Percentage mappable reads of at least 60%	Criterion 2: Detection of expected SNP in at least 10⁻² ng spiked samples
49	GMS050	62.93%	No
50	GMS051	82.18%	No
52	GMS053	92.04%	No
53	GMS054	22.97%	No
55	GMS056	68.65%	No
56	GMS057	86.27%	Yes
58	GMS059	75.33%	No
63	GMS064	0.00%	No
64	GMS065	74.15%	Yes
67	GMS068	83.25%	Yes
69	GMS070	73.22%	No
74	GMS076	0.00%	No
81	GMS083	66.06%	Yes
83	GMS085	70.41%	Yes
85	GMS088	16.30%	No
86	GMS090	0.00%	No
90	GMS096	0.00%	No
92	GMS098	0.00%	No
93	GMS099	79.70%	Yes
95	GMS101	86.99%	Yes
97	GMS103	0.00%	No
98	GMS104	87.30%	Yes
100	GMS106	0.00%	No
102	GMS108	0.00%	No
104	GMS110	58.35%	No
105	GMS111	0.00%	No
106	GMS112	79.65%	No
107	GMS114	63.40%	Yes
108	GMS115	77.59%	No
109	GMS116	73.95%	No
110	GMS117	0.00%	No
111	GMS118	81.01%	Yes
113	GMS120	60.54%	No
117	GMS125	38.70%	No
120	GMS128	45.91%	No

[00105] Table 6: Shortlist of markers showing the associated mappable reads (criterion 1) and their detectability at 10^{-2} ng in soil samples (criterion 2).

SEQ ID NO.	Marker ID
25	GMS025
28	GMS028
38	GMS038
46	GMS046
56	GMS057
64	GMS065
67	GMS068
81	GMS083
83	GMS085
93	GMS099
95	GMS101
98	GMS104
107	GMS114
111	GMS118

5 [00106] Table 7: Flanking sequences containing SNPs associated with pathogenicity of Ganoderma.

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
1	GMS001	CTCATTCTCCATCCTTATTGGCCTAGACCTTCGCGATG CTAGGGACGCCCACGTCGTCTT[T]GAAGCCGTTCGAC TGGGCGTGCTTCCCTTAATCACTCGTCGTCTCACC GC CAACGAGCGT
2	GMS002	CCAGTTGTGGCCTGACGCCTCGAGGCGGCCGTCCCCG ATCGCCACACCTCCGTGCATAG[G]GACTCCGAGAA GTCCCGGTTGTGGGTGATGACGAGCACACCTCCCTCG AACTCCTTGAGC
3	GMS003	TGACACGCTCACACGCGTCACCGAATATGTCCCCGAG ATCGTATCGTTTCGTGGAGGGGAT[T]GTCAAGAACGG GTACGGGTACGAGGTCGAGGGCAGTGTCTACTTCGA CACGCGGGCATT
4	GMS004	ATGATGCCTCCACCCGCTTACCCCGCCAAACCCAAAT GCATCCAAGGTTCTCCTCCTCGG[T]TCCGGAGACTCG GGCAAGTCCACCATCCTCAAGCAGATGCGTCTAATCC ACCGCGTCCCC
5	GMS005	GTCGCGCCCCCTGGGCCAGGATGATGATCGAGGACC GAGCTTCCAGGCCACAGGGCCGC[C]TTCTTCTTCAA GCTTGAGCGAGAGCTGGAGAAGGTGCGCGCACATGC

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		GCTCCGATAGGTC
6	GMS006	AGGCTTCAGAACGCCCCCGGATGGAGCCCCTCCTGTC CACGTCAGGAGCGTTATTGTCTG[A]GTGTGAGATTGG ATCGTCAGCGCTTTCATTTTGGTTGTGAGCGGAGGCA TTTTATGGGAA
7	GMS007	GAAGTTGCATGACACAAGGAGCTTGTCGAGTTCGAC GGCACGCGCAACGGACTTGACGAA[T]ATGACGACCT TAACCGAGAACGGTTATGCCATGAGAAATCAGGGGA CGGGCGTTCGAGAC
8	GMS008	CCAAGTCACTGATCGCACGTAGGAGCTCCGGCTTCAA CAGGAAATCCCTACACGAAAGTG[A]CACGCGTAAAT ACGCGATCTCACGAGAACGAGGACGCGTCTGACATA CCTGAATCCTGTA
9	GMS009	AGTCCTTCAATCTCCTTTCCATTTATGCGTCTGTTTAC GCCTTCATCAGGTCGTCTTCGG[T]GAGGTTCTTGAGG GTATGGACATCGTGCACGCTATCGGTACGTCTACTCT CCGTGATGAA
10	GMS010	GCGCCAGTTACGTGACGGAGAAATGATCGTGGACGG CAACGAGATGGTCCAGCGGTCTGA[C]GCTGTCGGCG GTGTCCAATCATCCACCGACATGAAGGGCAAGGGGC GCGCAGTGGATCTC
11	GMS011	TCGCTCGGCATGGTTCTCGTCTTCTTCATCCTCCAGTA CCCCAAGGGCGGCTTCACGCTC[G]ACTGGTGGGGCA ACGACGTCTGGCAGACCACTGCCGACGCGATGGGTC TGCCCCGCCTCG
12	GMS012	CACTTGGGTCTGTGTGACGTACATGCTGACCCAGACC GCCTTCCAACCTCTCTATGGGAA[A]CTCTCGGATCTC ATAGGTTCGGACGGTCGGTGCTTCTCCATAGACCGACA CGAGTACGCCG
13	GMS013	GCAGTGCCTGCGCATTGGAGCAATCACGAACCTCAA GCATGCGGCCAATGGCATTGAACA[G]CCATCGGTGG GAGTTGATCTTCTTTTGATTCTTCGCCCTTGATGACCAA AATAATCTGGAC
14	GMS014	ATTCAAACCAACCCCCATGCAGTTCAAGGATCTGCGT GCGGCACGCTACCTTTTGTACCA[T]AACTCTCGGTAT CAGGCCAAGATCCAGAAGCGGGACGCGCGCATCGCG GACGTCGAGCGC
15	GMS015	GGCGACGAACGGCACGGCGTTGTGTTTTGGAAGCAG AAGCACGAGTGCGTCGTATCACAC[C]GCAAGAAGTG TGGGGGTACACACCGACGAATCGGAGGAAAGCTGAG GTGAGCGTTGCTTT
16	GMS016	GTA TCCGCGTTCTCATTGTGTTTGTGTAACGATTCCG ACAAGATGAGGAACGTGAAGCG[T]GTCACTTCCAGG AGTTCTGAGCCAGCAGACTCGCCCCGCGATGTCTCCT GCGAGAAAGTC

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
17	GMS017	CTCAGGGTCCACATAGTGTTC AATCTCGGCCATGGTG AACTCGCGGACGCGAAGAAGCCC[T]GCACGAGGCGA GATTTTCATTACGGAACGAGCGACCGATCTGAGCGGA AGCGAAGGGAATG
18	GMS018	GTGTGAGAGCTCCGTGAGGGT GCGCCATCCCGTTGAT CCGACACTGACCTGCGCTAGAAT[T]GTCTCATACTCC TTAACCTGTTCGTGAGCCAGTTTGACGGCGACAGCCT TCGCCTTGTC
19	GMS019	GCGTCTCAAGTCTTCGTGCGTCCTCGACCTGAAGATG GGCACGCGGCAGTACGGGATGGA[T]GCAACGCTCGC GAAGAAGAAGAGCCAGCGGAAGAAGTGCGACCGGA CGACGAGTAGGGCT
20	GMS020	GTTCCGTGCCGCCTTCCCCACCGCCAATGAGGAAGCC GAGAAGGGCGAAACCACCTGGAT[T]AAGCTCTCTTC GACACCTCCGGCGCGAACAAGTCGGGCAAGGCTCGC TTCGCCGGCACT
21	GMS021	CGTCAGGAGCTCCTTGAAC TCGGCGTCCTCCAGCGAC TTCTCCGACTTGACGAGCTCGGG[T]TTGTACCATAACC GGTGGGGAGGTGATCATCGCTGGGCCACCGGGAGG AGTGCGATCGTA
22	GMS022	CATCGCCCTTTGCACCATATGTGTGTCCTATCCGCTCT CACTCTACCGCGATATCCACAA[G]CTCTCGCGTGCAT CCGGGCTTGCGCTCGTCGGCATGCTCATCATCGTCAC CTCGGTCCTT
23	GMS023	CGAGGACCAGTCCTTCCGCAAGATCAAGCTCCGGGTT GACGAGGTCCAGGGCAAGAACTG[T]CTCACCAACTT CCACGGCATGTCCTTCACATCGGACAAGCTTCGGTCC CTCGTCCGCAAG
24	GMS024	CGGTAGGGGAGAGCGAGAGGACCGGGGTATGCGAG GATAATCGGGATGAAAGACACCGGA[T]GATAGGACT CACTTCGTGGAGGGCGTGTCCACACTCGGGAACGAT ATGAGGCTTCTCGCC
25	GMS025	GTTCTGCCGGACGATGCACGGGGTGATGAAGAACAT CGCGCACCTGTGCAAGCGTGACCG[T]AGCAAGACAT GGGGGAAGGAGGGCTGGAAGAAGGTGGTGGTGTGTA TCGTGTCGGATGGC
26	GMS026	CACGACGAACAGGCAAGGAGAAGGCTCCCCGGCCAA GATCGGGAGCTCGTCCAAGGTGAA[T]GAGAACGGGA TGCGCGTTGTGTCGAGCACGACGATTCAGACAGCG GGAGCGAGGCTGGC
27	GMS027	AGGTACTACATACATGCTACTTGATCTTGCGCAGCTG CTTGATCGCCGCGTGCATCTCCA[A]CTCCTTGGTCTTG CCCAGCTGCCTGCTCTCCCGGAAGAAGTACTTCGGGC CCCGGTACTC
28	GMS028	GGAACACTCGTACCCGAAGTGAAAGACGCCGTGGGA TTTCGCGCGCCTCGAAGTCTCGTC[G]TAGCGGTGTTT

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		CAGGTAGATGGGGTGCGAGCCATACCTAAGCGACAG GTACACGTGAGTT
29	GMS029	CTGGCTGAACACCGCCGAGTCGACGAGCTGGCTGAG CGCGGGCACCCCCGCCTTCTTGAT[G]GTGTACGCCCC GCTGAACACGCTCTTGCGCAGCGTCCCGCTCGCGTTA ATCTTCGCCATG
30	GMS030	GGATGCACTCACGTATGCGACGGTAACAGGCTGGAT GATCGTGGACAGAAGATCAATCAT[A]ACTATGAACC GCATGGAGAAACAACAGAAACCGCAAAGCTGGTCCA AGAAGATGAGTTCC
31	GMS031	GTCGTCGGCGTGGACGTCGAACACGTACTTGACGGTC AGCGCAGCACATAGGAGGTAGCC[G]CCAGTCGTGTG GACGACACCCTTGGGCTTGCCCGTCGAGCCGGAGGTC TACAAACATCGG
32	GMS032	ACAGCCCGTAATGTTGTAGGTAAATCCGTTGATAACG TTCGCCTTTCCGATGACGTCAGA[T]GTGAGCGTCGGT ACGATGTAGAGCCTCCCATTTGTCGACGAACGAGTTGT TCGGCGAGTTG
33	GMS033	CCGGTGCACGTCCCAGCACACCACCTTACCGCTTTTT GTCGCGCACAGGCTGTAGCTGGA[T]GTGATGGGCGA TAGCCAGTCCACCTCTTCCTGCGGCGGCGCGCTGATC ACCTTGTACCAC
34	GMS034	CCACTTTCAATAATATAGGTACCACCCACGAGTCCTG TACATTGACATCGATATCCACCA[T]GGAGACGGCGTC GAGTTTGCCTTCTACCACACGGACCGTGTGATGACCG TCTCTTTCCAC
35	GMS035	CGTCGCCGGAGACAGTTTTTCTGCCGCAATAAGCGAC AAGGGCGAACTTCGTGCTTGGGG[C]GCTTTCTCGGTA CGCATAACATTCACCTCTTTTTATCGTACGTTGTTGTA ATTGAGTATG
36	GMS036	CCAGCGCCTGGTGGACCGTCTGCGCGTCCCAGCGGCC CTTGGCTACCTCCTTGCCCTCGA[C]GTGGAGCGCGGG GATCCAGTAGACATACTTGCGCTTCCAGCGCTCCTGG CCCGGGTCCTG
37	GMS037	TGTGGACGCCCCCTGGGCTCGAGGCGGACTGGGGTTG GCACGTATCACGATGCTGTTTCCA[C]GTCTTGATGGC GGTGGCATTGTCATCCGTCAGGATCGGGATCCATTTC TCGCAAGAGCGA
38	GMS038	TGCTGCTGCCCCGGAGGTTCCGATGACACCTATCATC ATGAACGCACCTGGCCTCATGCA[T]GAGGACAAGCA CAACAGGTACCTCACGGACGAAGAGCTCGACGCCAT TCTCCCCTCCTCT
39	GMS039	TGTCACGCCCCCTCCTGGTTATGCTCCTCTTGTCCGAC CTCCCAAGCTCTCGGGTCCTTC[T]GTGACCGAAGTTG GAGGTTTCCAAATCCAGGAAGGCTCAGATGCAGCAG CAGCAGCAGCT

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
40	GMS040	TTTGTGTCGGATGCACACTCAAAGACAAAAGGGAGA CGTACCATGTGCGCGCTTCGGCGG[C]AGCCTGGGTAT CTACTGCGACGCCGCGAGGGAAAGTCACAACCTTTCTG AACGCTTTTTGA
41	GMS041	GAATCTGCACGGGGCTACGTCACGGATTTAGACTTCA AAGAGCCCAGATGATAGCAAGCA[C]GCTTCCCGTGG CGAGCGCGATAGCCCCGCCACAGACAGGCACGACA GCTCCGCGGATCG
42	GMS042	CAAATGGACCTCGAGTGGGTCCAACATGCCCTTCCGG GTGACGATGACCTGTTTCGGCTG[C]TCTTCGACCGGT AGGGACCGGACCAGTGCGGCAACTTCTTCAGCGGTCT TCCACTTGGCC
43	GMS043	GGGGGCCCAAAGCATCCAATGCTTACTGCCAAAGGA GACGGAGCGACCCATGCGTACGTG[C]AGGAGGGTCA ACCCCGAGATGACGAGAAGGCAGGTCTGGCGGAGAC TCATGGCGAACC
44	GMS044	GACGATCGTAATGTCCTCCCGCGTCAGGTTCTTGATC TTCCTGTCCTTCACTTCCCACAG[T]TCCTCGGCGACTT GAGAAATCAGCCCTGTAAGGAGAAGGCGCCCGCAA CGCGAGACGCG
45	GMS045	TACCCCTTCCCAGCGGTCCTTCGAGGTCGTAAATCGG CCTACCCGTCAACGTCGGGCGAG[T]GGTAACAGTCTC GCCGCGTTCTCAGACCCGCGTCCTTCCCATCGCCAC GCAACCTCAGT
46	GMS046	GTCTCGTGGCTGCAAGGCGCTGAACGCCGGCGGAGG TGTCACGACTGTTCTCACCGCCGA[T]GGCATGACGCG CGGGCCTGCAATTGACTTCCCCTCGATAGTGCGCGCG GCAGAAGCGAAG
47	GMS047	GGAGGCTTTGGGACCTCGCCGAGGGCCTCGCGCTTGC GCTTGACCTGAGTTGCGGGATCG[A]CCTTGGAGGAG GCAGCCACAGCCACGGACTTTGCCCTGGTGGCCGCG GTGGAGGCGGTAG
48	GMS048	GTGTGAGAGCTCCGTGAGGGTGCGCCATCCCGTTGAT CCGACACTGACCTGCGCTAGAAT[T]GTCTCATACTCC TTAACCTGTTGTCAGCCAGTTTGACGGCGACAGCCT TCGCCTTGTC
49	GMS050	CGCGACCGTTTTTCGACTTCATGTAGCGCTCCTCGCAT TTCGTGATGTCTCGGGAGAGAC[G]CGGCGTTTTGAA AGGTCGATATAGCCTGCATGCGGAGAGATCAAGGGA AGGGAGGTTTCGC
50	GMS051	ACGGGTGCGACGCACTCTTGGTGAATTGCGGGCACCC TTGATACCGTGCCAGATGGTACC[G]CCGACGGCACCC ATGGCGAAGGCGCTACCAAAGTCAGAGAGGATAACC CATGGGCTGGGG
51	GMS052	ATCCGCTTCTGTGCGGGTCATCCGTGCGCGCAGGGCCC ACTGCCCTCGGGGCTATCGTCCA[C]TACTAACGACCC

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		AGATGGTCTCGCACCCCTTCCTCCAGCGACGCGCTCTC ACAGAGCCCGA
52	GMS053	GCCCTTCGGGCGCGACGAGGAGAAGAACGCGTTCTG TTGCTCGTTGTGCGCCTTCGAGTCC[G]TGGAGGAAGGG GAAGAGGACATGGTCCGGGGCATGGGTGGTAAGGTG CTTCAGATGGAGG
53	GMS054	CGGCATCGCGAGCACCTCGCCAAACTCGTGCGCGAA CTTGACGGCGTCCTCGGTGCGCGA[C]GCGTTGAACGC GGGGTAGAAGTACGTCTGGCCAGAGGTGTAGTGGGG TAGGCACGCTGTC
54	GMS055	AGGTCGGACGTCAAGCTTCTTGCGGAAGGCCGAGGA TTCAGATTGGAGCTTCTGACCCTC[A]GCGTAGAGCTC CCAGCCCTTTCCAAGGACGTCTTCAACACGCTTCATG TACGTCTGTAGC
55	GMS056	TTGGTTCTGTACAGCTCCCCTCCCTCCCGCCTGGTGCT CCCTCCTCTCCGCTCGTCCTCT[T]ATACACGGATCTTT GACATGGGTGTCGCCTCCAGTTGCTGCGAGTCTTGTT GTGCGTGGC
56	GMS057	AGGAATTGCAGGGTTGAAGCCTCGAGGCGGCTGGCG GTAGGGTATGTTGGTGGCCACGTT[C]GGTGGACGCCC GTGGACGAATGCCGGGGGAGCGCCGCCTGGAGGTAC GGGCACGTTGGGG
57	GMS058	GCCAAGGGATTGGACGAATGATGCGAGAAGGCCAAT AGTGATGCCGATGGGTATAGAGAT[T]CTGGGGTCGG ACGAGTCGACGGCCATAAGGAGAGTGCGGGGTGTGG GCCTGTATAGGATG
58	GMS059	CGACGACTGGCGGCGCATGAGTAACCCGCGAAGCAT ACCCCCGGTGGTTCGAACCAAACGA[T]GGCGACTGCG TCTGCGAAGGAATGGGCAGCCCGCGCTTGTCGACCTG CACCAATACCAAT
59	GMS060	GAGCCAGCACCCGAACATCGTCAACTTCCTCGACGCT TACCTCGTTAAGAACAACGAGCT[T]TGGGTCGTGATG GAGTACATGGAGGGCGGTGCGCTCACGGACATCATT GAGAACAACACG
60	GMS061	CACTACCGCGAGCACTGCCGCTACGAGCGCGAGTGG GACGGGGAGTGCGGAGGAGGCGGC[A]TCGGCGTCAG GGACGTCGAGCAAGGACCCACGGCATGAAGAACTGT AGAATAGATTATAC
61	GMS062	CCATGTTCTGTTTCGCCCAGACATCTTCATCGGCCTCA ACGTGGTCCGCGCCCTCAGCAT[C]ATCTCACTTGTGC TTCTCTTCGCAAGCAGCATAGTCACCATGGTCCATGA CATTGAGGCT
62	GMS063	CCGGCCACGCGCACCTTCTTCCCCTCCCGACGCGCCA GCTCCAGCACGAGCTCGCATTGC[A]CCTCGTTCTCCG GCTCGAACACGCACAGCGGCCGGCACGTGTACGTCA GTCCCCAGTTCG

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
63	GMS064	GAAGTCCATGATGGCGGCCTCGATGCATCGTTCTGGC GTGAAGTTCGGGTTGTGGGTCAT[A]ACATCCTGGAGG AGGCCCGCAGCATAGCGTTCTGAACACTCCAGATGCT GCGAGATGAAC
64	GMS065	GCAAGGGGGTTCAGGAACAAGAACAAGGTTGAGGAG GGTCGCGGTCCAAGCTCGGTGCGCG[A]TGCCGGCGC CAACGGCGGAGACGCGCGCGACGTCGCGCGTCGCGT GAACCATGCACTGAT
65	GMS066	GGGCCAGGGACGTGCGGGCGGGGAAGCATCGACAGA AGCTTACGTGAAGCCAACGAGGGC[G]CGCTCCAACG CCATAATCAATGCATCCATCATCTTCAACTGCCAGTG CCACCGGTCCGTC
66	GMS067	CCTGCCGCTCTTCGCGGTCTCGCCGCGAGCTGCTTC TCCCTGCTCGCGTTCATGAACGT[T]AGCGCGGGCGCG GAGACGGTGTTCAACTGGTTCGTGAACCTGTGACCG TCGGCGGGTTC
67	GMS068	GCACGACAGCTCGATGTCGAGCACCTTGGCTGTCGCC TTAATCTGGGCCTGCAGCCGGTG[T]GTCGGCCCGCCA AACATCATCATCGCGCGTGCGAGTTTGAGGATGAACT CCTGCCGACAT
68	GMS069	GAAGCAGGCGTCCAAGGACGCCTCGAAGGAGATAAG GACGCTGCGCGAGGCTGCGCTCTC[C]ATGCTGCTCCA CCACCCGTACATCTGCGGGATGCGGGAGATGATCGTC CACCAGCATCAC
69	GMS070	GAAACGGATGGCGAACGGGCGGAACGGTCACTCTCG CTGGGGAGAGTCGTTTTCCAACCTC[A]TCGCTCGCGAT CTCGCACGACGGCAGCTTCTCTACGACGACTACGAAG GTTGACCAATCG
70	GMS071	CCTCACGTGTGTCGACAGCGTCGCTACTACTTCAACA AAACCACCGCATGCGATCAAGAC[A]GATATCAAGCG CGTCCTCGAGCGCATGCAAGTCCAGTACCGCGAGAC GAAGACCGGCTTC
71	GMS073	GCGGTTTCGCGCCGGTGTTGTAAAGCACGATCTGGAG GTCCTCGAGACGGCCGGTGACCTC[T]CGCATTGAGTC CGCGCGCTTGTCCGCATTACGTCAATGGGGTATAGG GTCGCACCCATC
72	GMS074	GATGATGTCCAGCATCTGCTTCATCTGGCAGAAAATC AGCACGCGGTGCTGCGAGAACGC[T]CCCGTGCCCGTC TCGTCCGCGTCCACGAGCTCGGACTTGCGGGCATCGC CCGTCGCGCCC
73	GMS075	CTACCAGTTCATCCAGGCGCGCGGGATGATTCTCATC GAGAGCGAGATCGGGCTCGTCGC[T]GTGCTCCCAATT GGCATGGCGCAGGTGTCGTCGGTCGTGCTGAGCGTG AAGAAGGGCGAT
74	GMS076	GAAGGGCGATTGGGTGGAGAAGGGCCAGGAGATCTC GTGCTTCCAGCTGGGTGGGTCGGA[T]ATCGTGATGGT

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		GTTCCAGAAGGACGCGCAGGTCCAGCTCGACCAGGA GAAGAACGTCCAC
75	GMS077	GTTGCATACCGCTACTCGATAGCACCCGGCTGCATGA TTCTGCTTCGCTGCCTGCAAATA[G]AGGTGGTACGCG CGTTCGTGGTCGACCTGCAATCCCAGCATACCCGTGC CATGACAGTTA
76	GMS078	GATTCTCAAGGTCATCCGCAAGAACCTCGTCAAGAA GTGCATGGACCTTTTCTCCGAGAT[T]GCGGAAGACAA GGACAACTTCAACAAGTTCTACGAGGCGTTCGGCAA GAACATCAAGCTT
77	GMS079	TGCCCCGTTGTACTCATTCGTCTATAGACTACATCACC CGCATGCCCCGAGGTGCAGAAGAA[C]ATTTACTACCTT ACCGGCGAGTCTCTGTCCGCAGTCAAGGACTCTCCCT TCCTTGAGGTC
78	GMS080	GTCGAAATTGCGCAACTTCTTCGGGCAGCGCCCGCCC AGCGAGCTCATCACGAACCACCT[G]CAGGAATACTTC CCGTTACAGGAGAAGCGCGTGCTTGAGCGGACACGG AGACAGAGCATC
79	GMS081	CAGCGAGCTCATCACGAACCACCTGCAGGAATACTTC CCGTTACAGGAGAAGCGCGTGCT[T]GAGCGGACACG GAGACAGAGCATCATGCGACAGAGCGGCGTACCTGG AAGGCGGGACAGC
80	GMS082	GTA TCTCGAGGAACTCTTTCATCTTCTCGGGGAGCTCG TCAGCAGTTTTGCAGCGGATGGC[G]TGGACGTTTCATC GCACGCGCGAGCGCAACGAAGTCGGGGTTGACCATG CGCGTGTGCGCG
81	GMS083	ACGACGTCAGTCGTACAAAGTATAGACACGCACCAC ACGATACAGAGACGGAGGGTTGAC[A]TTCCGGATAT AATCGAAGAGGTCACCATTTGAGCACAGCTCGCTGA CGATCATGACCGGC
82	GMS084	CTCCTTGCGAAGGTAGGCGCTGTGCTGGTCCTTGATG TATTTGACAATCGGGTCCCAGCT[T]CAGAATACATGG AAGACGCGTCAGGACGGCGTTACGACGCTTCAAGCA TGCATGAAGAGG
83	GMS085	TGCAATCTCGTCCTCCTTCAAGCCCTCCAAGCGTATC CGCTTACAGGCAACCATCTGGCC[T]GTGTTCAAGTTC AGCGCTCGGTACACGGCGCCGAACCTGGCCACGACCG ATGCAGTTCCCC
84	GMS087	CTCCTCGGACGCTCTTCCCTTGCCCTTTCCCGCTCCAT GCGAGAGCATCCCTGTCTCGTC[T]ACCGGCGCCGCAT GTTTCGCCCTCGACTTTGTGGCCGTCCGCAGACTCTG CTCCAGGGAG
85	GMS088	TTCCCACCGTGTGTGACGTGGACACAAGCCCGGAAA ACGCACCTGGTTATTTGGCTCGAA[T]GATAGACCTAG CTGGATGGCCTCCTTGGCCTCGTGGTACTTGTTGAGC TTCATGAGAGAG

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
86	GMS090	GGGCTCTGATCCATTTCGAGGGGGCCAGAGTAAGGAC GTTGATATTTTCCGAGTTGGATGG[T]TTGCGCCTGGA AGAACCAGATGGGGTCTTGCTACGTGAGGAAGCAGC ATCGGGCCCCGCGG
87	GMS092	GAGTCGCTGTTTCACCTTCTTGACAAACATGTTTCCGT CCGCCGCGAAGCCGTTGACCGC[A]AGCACGGCGCGA TCAGTTCTGTGTATGCGCGATAGCAGAGGCGGGTGA GCGGATGATCGC
88	GMS093	GTTCTTCATGCAATTTCGAAGGTCTGTTTCCGATTTTC CCTTCGCTCGCTATTCATTCAT[G]TTCTCCTAGATCTT GTTCAAGATGCCAGTACTGTCATGACATTTTACGCGA CGTCGGTGG
89	GMS094	AGCTGCCTGTTTGAAGACATTGGGGCGTTTCAAAGGA TTGGAAGCCTTCAACGGATTTCGT[C]GCAGGTTTGAGG GCATTGAGCTTGATGCCTGAGGAAGACGCTCCGGGCT TCGGCGTTAAG
90	GMS096	GTCTTCAGCTGGGAGTCCGCTTGGCCTGAGATCGCTG AGGGCCTCGTGAACGTCAGCGAC[A]CCAAGGATCAA GCCGGCCTCCAAGCGGCCCGAGCTGCGAACAAGGTC TACATGATGCGTA
91	GMS097	AGATAGCCTGGGGAAGCCGCAGGGTGCGGAGAACGC GCGCGACCAGATCAACATCGCTGT[G]CTCCTGCCGTC GAACACGACGGGGGTGACGATGCGTGTGACGAGCGG GGCAATGTCGTC
92	GMS098	CCCGCTGGTCCCACAGCTTCTTCTAGCATCTCGTCGC CCTTCCGACGCCCCGACTCCCCT[A]CTACTAACTTCC CGTCGTCAGCGCCACGCGAATCCAGGCGTTCCTTCTC TCAGTCGCGAC
93	GMS099	AGACCGATAATGGAGACGGCGATATCGAACAGCGAG GAAGCAGGACCGTCCAACGTTTCGG[T]AATGCTGATAT ACGGCGCCATAGGTCATTCGACAGTGGCAGGTGCTCC CCGGCGCAAGAC
94	GMS100	GCAGTTTGCTGAGTTACGGATACACCCCCACTCGCAT ACAGCGAATATGCAGGTCATCAC[A]TCTATTTACTTG AACTGCAGGCCCGACCTCCGTGATGAATGGCTTACAG GCATCGAAGTC
95	GMS101	GTACAGCTTCGCCACGAATGGTGACTCGGCCTGCTTC ATCAGGATCATGCGCTCGGCCTT[G]ACGTTTCGTGATC TGGTTCTTCGCTATCATGTCCGCCTTCTTCAGGACCTT GATCGCAAAG
96	GMS102	CTTTGTCCGCTGAACAGCAGTCTTGTACGGGCTCAGA ACCAGGAGATCAGAGTGGTACTT[C]AGGGATTGCGG GTCCTTTTGTGCCGCTCCTCCGGCCAGGGGTACGTC GCGTGCGTCTGG
97	GMS103	GTCTGCTCAGATCCGGGAAATTAGGAAAAAGATGTT CGAGATCATGACCCGGGAGGCGAC[T]ACTTGCGACTT

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		GAAGGAGCTGGTGCAGAAGTTCGTTCCGGAGGCGAT CGGTCGTGAGATT
98	GMS104	TGCGTTGCAGGATTTCTCAAGGCCCTTCAACAGCGT CACCATGATGAGCTGTTGGAGGA[C]GTCTTGTCCCGA CAATGATAGGCCAGAGCGTCAGATGTTAGCCGTTGTG CCGTTGTTAGA
99	GMS105	CTTGGCCTCCTCGTTCGACACCGATGGGTGGACGATG ACATCGTCGCCCTTGTTCAGTT[C]ACGGGAGTCGTA ATGCGGTACTTGTGCGCCGATCTGGAGGGCTGGAGAC AGGACAAACGAT
100	GMS106	CTCGACGACGTCATGGCGGACCATCTCGAAGTTAGG GTGGCCGACCCAGTGGCTGACGGT[A]GTTTTCGAGCC GGTGAAGAAGTTGTGCGAGTACGGTAACTTCGTGGCC GAGGAGCATAAGA
101	GMS107	GCCCTGCCC GCCGGCGATCGTCGCATTGCCGTCTCCG TTGTGCAACATTTTGTGGGGGAC[C]AGCGCGCCGCCA AGGCCGAGAAGTCGCAACCTGACGCCGCCGATGTCG AGGCAGCGCGTC
102	GMS108	GCACCCTCACCAGCAGGGATTCCAGGCCCCATACGTG AATGGTATTGGGGCACATTCTCT[A]GGCCACGGACAT CTCACGCAGCCACAATTTGGGGCCCATCTTCTGGGC CCGGTGCTGGA
103	GMS109	CTCCTTGATAACCCTCCGCTCCCAGAGCAAGGGTGCCG TACGGCAATCGGGCGGAGCCAGA[G]ACGCTCTTGAG CTCGGAGACCTTCTCGGACTCCTCACTAGAAGGGAAA ATCAGAGGTGGT
104	GMS110	TCCTGGCATCGGTACGTCTCCACCTCTGATCCCTCTTC CGACAAGCTTCTCCTGGACGCC[G]CCCTCCTCCCGCG CCTCTGGGAGTTGCGCCAGCGGACTCTTCTTGACCCC GCCGGTGTCT
105	GMS111	ACAGGCCGTTCTCCACCATGCCGGTCAGACATTGTC GCATTGTGGAACGATCCGGTCGT[G]CGACAGATCCTG CGAAGACGAAAGGTACGGCTAGAAGAGGGCCCTGGA TTGTGAGTATTT
106	GMS112	GCTGATGCGCCGGAGGAAGTGGATGGGGTTCGGGTA GCTGAGGTTCCAATCGATGGTCTT[T]AGGACATAACG CTCCGCTTGGAGAATCTCAGCCTCGGTGTACGAAGAG TCGGCGCAGTAC
107	GMS114	GATTCATCTTCGCAAGGAAGGCGGCTTCAAGACCTTC ACTGACGCTCAGGCTGGTAAGCT[A]GCTGGCGAGAA CCCGGACTATGGTATCCAGACCTTGTTTCGAGGATATC GAGGCCGGCAAG
108	GMS115	GTAGATAAGCACGCGGATGCCGCTCTCCACGAGCTC GGGGAGGAGCTTCGCACGGTTGCG[G]GCACCGTCGC CTTGGAACGCGAACGCCTGGTTCACCTGCATGTTGCA GCTCTCGAAGTCG

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
109	GMS116	GGCACGACATGACGACATGGGCTCACCGAATCTGAA CAGTCGCTCGTTGGGCACCTGCTT[A]CACCTCCGAT CATGTAGCGGGCATGGCTCTTGGCCATCTCGTCGTCG CTGATGTCGACC
110	GMS117	GGCCTACATCAAGACGGGTAACCTACCGCTGGAACGC GGGTATGTTTCGTGACGAAGGCGAC[A]TTCCTCATGCA GCTGCTCAAGGAGTACAAGCCGGAGCTGCACGCGGG GCTCGAGAAGATC
111	GMS118	CTTCCAGACCCGTAACCCCATGCACCGCGCTCACCGT GAGCTCACCGTCCGCGCCGCTCG[T]CAGCGCCAGGCC AACGTCCTCATCCACCCCGTCGTCGGTCTACCAAGC CCGGCGACGTC
112	GMS119	CGAGATCGTCAAGGAGTACAGCCAGTCCGTCGTGAC TACACTTCACGAGAAGAACCTGCT[C]TTGCTCGGGGA AGATCGCTTCCTCACCACGGTCCTCCTCCGCACATTC CCCAACCGCAAG
113	GMS120	TCCGCGCTTCATTAAATTAACCTCTTCGTCTATTAGGA TTCTCGTCCTGAAGGATGGGCG[T]ATCGTCGAGCAGG GCACCCACGCCGAGTTGCTCGCCAGGGCGGCGTTTT CGCGGAGATG
114	GMS122	TGAGCGTGGCAGTACGTACGAGGTGTAGCCAGCCAA GAAGTGGTAGCTCGCCTGCTCGGG[T]GAGAGCTTCGA GACAGGTGGGAGGACACCGAACGCGTCGCAGGTGAG CATGATGATGTTT
115	GMS123	ACCCACACAGATCTCGTCCTTGGAGGTGTCCTCGTAG ACGATACGCTTGTCTTTGGGGCT[C]CGTCCGGTCTTCT TGCCGGAAGTTGATGAGCGCTCCGGAGGAAGAGA CGACGGCACCT
116	GMS124	CTTGGGCACGAGCTCGAGCGCCTTCAAGTTCTCGAGC AGCTGCTCGGGCTTCGTGCGGCC[T]AGGATGACCGTC GACGTGTTCTGGGTTCGCGCGACCCACGCGAGCGCG AGCTGCGTCGAC
117	GMS125	GACCTCCCCGACGATGAAGTGGCACTGGAAGCCGAC GCCGTCGAGCGGCACGCCCGCGGC[A]AGGAGCTGCT TGACGAGGCTGAGCATCGCGGTCGCCTTCGCGCCCGA CTGCTCGATGTTG
118	GMS126	GAGAGGGGAGGATGCGCGATACGCACAGATTTGGCC CTTGTAGTGGCCGACGAGCGTGCC[A]CAGTGGTTCTG GATGACGGTCGTCAGGTCGGCGGCGGTGAACTGCCC GTTGGACACCCAG
119	GMS127	CAAGGCGCTCTACTTCGACTACGACACCGTGCAGGA GAAGCAGAGCCTGTACGAGAGTGT[C]ACCGCGTTTTG GACGCTGTGGGCGGGGTGCGCGAGCGAGGAGCAGTG CTGGAAACTCATG
120	GMS128	GTTCCCGTAAACTTTGTCGCCGCCGTAGTCAATTATC AGAGCGCTTCCTGCTGAGCGGGA[G]CCGTCCGAGGC

SEQ ID NO.	Marker ID	120 Flanking bases, [SNP]
		TTTGTCGTGGAGCAGCTCACCGACTTGGCGGGCAATC TTGAATGATGCA
121	GMS129	CCTCGTCAAGGTCGGCTCCGTCGACTTCCTCCAGGAC GACTCCCGCGGCAACCCCGTCGT[G]GAGTACCTCAAG CGCAACGGCAAGCCTCAGGGTCTCGTCAGCTCGCTCT CCAACGAGGGC
122	GMS130	GGCACATGGCCGTAACGGCGACTTCCAAGTACAAGC ACCACTCGTCCTCGGTCATGAAGC[C]GCTGGCATCGT CACCGCCGTCGGCCCTGGTGTAAGCACCTCGTCCCC GGCCAACGAGTA
123	GMS132	AGGTCAACATGTTCCGCGAGGACGGCAACGTCCTGC ACTTCAGCACGCCTAAAGGTACGC[T]GCGCGTCCATC CTTATCCCACCCATCTCCCCTATCCATTTCTCCCCCG CACCATCCCTT

[00107] Table 8: Primer sequences for the amplicon sequencing.

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
1	GMS001	23	SEQ ID NO: 124	GGTAGCGACCCAGGATAC GCTTCT	23	SEQ ID NO: 125	ACGTTGGACTTCAACCAG GTGAG
2	GMS002	22	SEQ ID NO: 126	CATGCGCTCCTTCTTGA GCTTC	23	SEQ ID NO: 127	CTCAAGGAGTTCGAGGGA GGTGT
3	GMS003	23	SEQ ID NO: 128	GTAAAAGCCGACACATCA GGTTTG	23	SEQ ID NO: 129	TGGCCAAAGAACTTCTGG TCTTG
4	GMS004	23	SEQ ID NO: 130	GTCTCGACGGTGGTGAA GTTTTG	23	SEQ ID NO: 131	CCACACGAAAAACATTTC CAGAC
5	GMS005	23	SEQ ID NO: 132	CATCTCTTCTCTTGCCGC AAATC	22	SEQ ID NO: 133	AATGGACGTGTGGGGGA CCTAT
6	GMS006	23	SEQ ID NO: 134	CGACGATCTTAAAGTCG GTCAGC	23	SEQ ID NO: 135	CGGATACGGCAACAAAG ACAATC
7	GMS007	21	SEQ ID NO: 136	CATTCAACCCACTCCAC CGTC	23	SEQ ID NO: 137	GTGAGCATGTCTTTCCCA GCGTA
8	GMS008	20	SEQ ID NO: 138	GCTCGAACACGCTCAGC ACT	23	SEQ ID NO: 139	CTCACCTCAGCTTTCCCTCC GATT
9	GMS009	23	SEQ ID NO: 140	GAGCCGAGACTGTAACC GAGGAA	22	SEQ ID NO: 141	GCATCTGCATTCTGTTCG CACT
10	GMS010	23	SEQ ID NO: 142	GAAGGAGGGACCAGAA GTACGGA	20	SEQ ID NO: 143	CTTCTTCCCCTCGGCCTCT C
11	GMS011	24	SEQ ID NO: 144	TGATCCATATCCATATC CAATGCG	23	SEQ ID NO: 145	TAGCCATTATTGTCGGC ACTTC
12	GMS012	20	SEQ ID NO: 146	AAGTGTGCACGCGGAA AAT	23	SEQ ID NO: 147	CAGGGCTACAAGAAGAT CCAGCC
13	GMS013	23	SEQ ID NO: 148	GACACACCTTGCGGTG TTCTTG	23	SEQ ID NO: 149	TCTTACCCCGATCTCAC AGTCG
14	GMS014	21	SEQ ID NO:	CAGTCGCCTTCAGTCTG	24	SEQ ID NO:	TCCATCTTGAGAACTGAG

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			150	CCAT		151	CAATCG
15	GMS015	23	SEQ ID NO: 152	ATTCAATGGCTTCAGGG AAGACG	22	SEQ ID NO: 153	ATACGTGCAGGCGACGA AAATC
16	GMS016	23	SEQ ID NO: 154	GGAAAGCTCATCGTCTGC TTCACT	22	SEQ ID NO: 155	CCGACTTTCTCGCAGGAG ACAT
17	GMS017	22	SEQ ID NO: 156	CTCGTTTGCCACAATCC CCTTA	22	SEQ ID NO: 157	CTAAACTTCTCCCGCCTC CTCG
18	GMS018	22	SEQ ID NO: 158	AAACAGGGTGCTGCC AGTAG	23	SEQ ID NO: 159	GAAAAAGAGGACAAAGG CGAAGG
19	GMS019	23	SEQ ID NO: 160	CTTGGGACCAACCCTCAA TCATAG	23	SEQ ID NO: 161	TTTGGTCATCAAGGCGAA GAATC
20	GMS020	23	SEQ ID NO: 162	CCGAGACGACGTCCTGT ATGTCT	23	SEQ ID NO: 163	GTCGGATCGCTGTAGATG CTCTG
21	GMS021	23	SEQ ID NO: 164	GTCGGACGCGTTAAGAT CCCTAC	20	SEQ ID NO: 165	GAGCCTTGCCCGACTTGT TC
22	GMS022	23	SEQ ID NO: 166	TGCGCATTCGGAAATCAT TATAGG	23	SEQ ID NO: 167	TAAGGACCGAGGTGACG ATGATG
23	GMS023	23	SEQ ID NO: 168	GTACGTTCTCAGTGATC CTGCCG	23	SEQ ID NO: 169	TACACGTGTGATAGGGTG TCGGA
24	GMS024	24	SEQ ID NO: 170	CCATATTTGAAACGTTTC GAGAGGC	23	SEQ ID NO: 171	AGAAGCCTCATATCGTTTC CCGAG
25	GMS025	20	SEQ ID NO: 172	ACTGGACATGTGTCCGC AGC	20	SEQ ID NO: 173	CGCTCACAGAAACACACGC AC
26	GMS026	19	SEQ ID NO: 174	TCAGTCCTGCTGCTCGC CT	20	SEQ ID NO: 175	CCTGCCTGATCGTGATGT CC
27	GMS027	20	SEQ ID NO: 176	ACAATGCGTGCGGTAGG TGA	18	SEQ ID NO: 177	GAGCCCCGAGGCGAAGGT C
28	GMS028	20	SEQ ID NO: 178	TGGGCCCCGCAAAAGAAAT AGA	20	SEQ ID NO: 179	GTTGGAAACCGTGCAAACC CT
29	GMS029	20	SEQ ID NO:	ACACGACTCGGTATCC	18	SEQ ID NO:	CAACTGCCCGGGGACAT

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			180	CGT		181	
30	GMS030	20	SEQ ID NO: 182	ATCCCATTTGCGACGAA GGA	20	SEQ ID NO: 183	ACAAGACCCAAATTCGTGC GG
31	GMS031	20	SEQ ID NO: 184	CGCCGAGTAGAACTGCG TGA	20	SEQ ID NO: 185	CCTCGCTATTGCCCTCCA GA
32	GMS032	20	SEQ ID NO: 186	CTCGCGACGACTTCCAC CTT	20	SEQ ID NO: 187	TCGACATGAGTGGCTTTG GG
33	GMS033	20	SEQ ID NO: 188	CTCCCGCGTGTGTGGTA GTG	20	SEQ ID NO: 189	ACGCCGTCAAATCAAGCGA CT
34	GMS034	20	SEQ ID NO: 190	GGAGCGGTGGTCTCCAT CAT	20	SEQ ID NO: 191	TCGTCGATGCCGTCTTTG AG
35	GMS035	20	SEQ ID NO: 192	CATGCGGAACGAACGAC AAC	20	SEQ ID NO: 193	GCGGAGACCCACTTCTCG TC
36	GMS036	20	SEQ ID NO: 194	CTGTCTGTGTGCGTCCTC CTC	20	SEQ ID NO: 195	TCCTCGGCGCATGACAAGC TA
37	GMS037	19	SEQ ID NO: 196	GGAGTGCACACGCGACT GC	20	SEQ ID NO: 197	AGTGCGAAGAGCAGCCA TCC
38	GMS038	20	SEQ ID NO: 198	GCGGAGCCTGAGGAGG ATAA	21	SEQ ID NO: 199	TGAGCCCTTCCTGGATTG GAA
39	GMS039	18	SEQ ID NO: 200	ACCAAGCCCCTGCTGCT G	20	SEQ ID NO: 201	GTCCACCGACAGCTCCGT CT
40	GMS040	20	SEQ ID NO: 202	GGGATGTTGCTGGAGA GGA	19	SEQ ID NO: 203	GACCCACTGCGCCCTCTT T
41	GMS041	20	SEQ ID NO: 204	CTGAAGGAGGACGTGG GTGG	20	SEQ ID NO: 205	CGTCACACTACGACACCG CC
42	GMS042	20	SEQ ID NO: 206	ATCTGCGGTTGGGTAGC ACG	20	SEQ ID NO: 207	TTCTGTGCAGACGGAGGCC AT
43	GMS043	20	SEQ ID NO: 208	ATGATGAGGGTCCCGAC GAG	20	SEQ ID NO: 209	TCATGTTGCGACCTCCTT GC
44	GMS044	20	SEQ ID NO:	CCCGAGGAGAACCGAA	18	SEQ ID NO:	ACGCACGGGGAAACACGG

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			210	CAGA		211	T
45	GMS045	18	SEQ ID NO: 212	GATCGTGCAGCGAGCCC T	20	SEQ ID NO: 213	ATGGGGCTGAGGGGATAC TC
46	GMS046	20	SEQ ID NO: 214	GACTACTCCCCGCGTGAT GGG	19	SEQ ID NO: 215	GAACCCGCACGAACAGCG TC
47	GMS047	20	SEQ ID NO: 216	CGAATCGACAACAGTGC GGA	20	SEQ ID NO: 217	ACAAGGAGAACGCCGACA GCA
48	GMS048	20	SEQ ID NO: 218	ACCAGGGTGCTGCCCCAG TAG	20	SEQ ID NO: 219	AACGTTCTGGAGGCCCGA TT
49	GMS050	20	SEQ ID NO: 220	CTCCTCGTGCCCCCTCCTT TT	20	SEQ ID NO: 221	GTCGTCGTCGTCCTCCGA GT
50	GMS051	21	SEQ ID NO: 222	TGACAGCACAGTCGAAG GTCG	20	SEQ ID NO: 223	ACGCCGACGACGACTAG ACA
51	GMS052	20	SEQ ID NO: 224	CATGCCCTTCCACATCCT CCC	20	SEQ ID NO: 225	CGCGTCGACCCACACTGAG AA
52	GMS053	20	SEQ ID NO: 226	TCCATGGCGACATCGTC TTC	20	SEQ ID NO: 227	CCTGTCCTATCCTCCCGC CT
53	GMS054	20	SEQ ID NO: 228	GTGCGGTGATGTTGTCC TCG	20	SEQ ID NO: 229	CGCTGCGTACCAGGATGT TG
54	GMS055	20	SEQ ID NO: 230	ATGCGGCACCTGGAAGT TCT	21	SEQ ID NO: 231	GCCCAAAGGTATATGCGC TGA
55	GMS056	20	SEQ ID NO: 232	ACGGAGGAGGGGGCTT ACTG	20	SEQ ID NO: 233	TGCGTTGGAGGTCGACAT TG
56	GMS057	21	SEQ ID NO: 234	CGTGCCATCGTAGGGTT ATCG	20	SEQ ID NO: 235	CCGCCTATGACCCCCAATC AA
57	GMS058	20	SEQ ID NO: 236	GGCACACATACCAAAGC CGA	20	SEQ ID NO: 237	TCAATGTGGAGCTTGGA CG
58	GMS059	20	SEQ ID NO: 238	TCCTGTTGCTGTCGCC ATA	20	SEQ ID NO: 239	CTCAAAAGGATCAGCGTC GG
59	GMS060	20	SEQ ID NO:	CGATGACCCGGAAGGCTC	20	SEQ ID NO:	CGCTTCGACTTCTGGTCC

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			240	TGT		241	GT
60	GMS061	20	SEQ ID NO: 242	TGGTCGAGAAGCAGA AGGG	20	SEQ ID NO: 243	TGGGAACAAACGGGGTTT CA
61	GMS062	23	SEQ ID NO: 244	TCTTGCTTTACGCCCTTCC TTCTC	20	SEQ ID NO: 245	AAGCGGTTGAAGAAGGC TGC
62	GMS063	20	SEQ ID NO: 246	CTCCGGATGACGCACAG ATG	20	SEQ ID NO: 247	GCTCCGATATGATGTCGC CC
63	GMS064	20	SEQ ID NO: 248	TCGCCAATGACAGTCCT GCT	20	SEQ ID NO: 249	CCCCCTTGCTTTCTTCCAC A
64	GMS065	20	SEQ ID NO: 250	GTTCAACCAGTTGTGGCC GGT	20	SEQ ID NO: 251	ACGCGTGCGTAAGTTGGT CA
65	GMS066	20	SEQ ID NO: 252	CGTGGGCGAGGTTTGGG TAT	20	SEQ ID NO: 253	GTACGTCGGACGGATCTC GG
66	GMS067	20	SEQ ID NO: 254	CAGGAATCAAGGTGCTC CCG	20	SEQ ID NO: 255	TGAAGAGCGGGACGTTG ATG
67	GMS068	20	SEQ ID NO: 256	TTAACGCCACAGCAAGT CGC	20	SEQ ID NO: 257	TCCAGCGGTGTTTCACAG TC
68	GMS069	21	SEQ ID NO: 258	GGAGCATGGGCAAGGT AAAGC	20	SEQ ID NO: 259	GGTCGGCAGGGTTGTACA GG
69	GMS070	20	SEQ ID NO: 260	CGGACTCCCACTCGTT CAT	20	SEQ ID NO: 261	AGTAACGGCATCGGGAC GAG
70	GMS071	20	SEQ ID NO: 262	GAGAACGGTGCGAGGG AAGA	20	SEQ ID NO: 263	AAGGTTCTCCACGCGTGCC AG
71	GMS073	20	SEQ ID NO: 264	CCAGTTGGATCTGGGCG ATG	20	SEQ ID NO: 265	AGGGCCATGAAAAACGG AGC
72	GMS074	18	SEQ ID NO: 266	ACGTGCGTCGTGAGCAG C	20	SEQ ID NO: 267	CAAGATGGAGGCCGAGG ATG
73	GMS075	20	SEQ ID NO: 268	GATAGCCCTGACAGCAC GCC	20	SEQ ID NO: 269	GGGCAAGGACGACCCAG ATA
74	GMS076	20	SEQ ID NO:	CGCGCTAACAAAGAAGGC	20	SEQ ID NO:	GCGTTCCGCGTCTTCCAGC

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			270	GAA		271	TA
75	GMS077	20	SEQ ID NO: 272	AGGGTTCGGTTGCTGCT TCA	20	SEQ ID NO: 273	CAACTGCCCTCCGTGAGCC TT
76	GMS078	20	SEQ ID NO: 274	ATCAAGCTCTACGTCCG CCG	20	SEQ ID NO: 275	GCGTACTCGTCGATGGGG TC
77	GMS079	20	SEQ ID NO: 276	AGGCGTTCGGCAAGAAC ATC	20	SEQ ID NO: 277	CCGTTCCATGTTTCGAGGA CC
78	GMS080	20	SEQ ID NO: 278	AGCGAGCTCGGTCAGCA TCT	20	SEQ ID NO: 279	CGTGGTCGCATTTCGAGAG GT
79	GMS081	20	SEQ ID NO: 280	TTAGGCGTGAAGATCCC CGA	20	SEQ ID NO: 281	GGGCTCAACGTCGACATG CT
80	GMS082	20	SEQ ID NO: 282	CCGGAACGTAGGGTGAA GGA	19	SEQ ID NO: 283	CTCGATGGTCACCTCGGG C
81	GMS083	20	SEQ ID NO: 284	CGACTGCTTGACCTTTG CCA	20	SEQ ID NO: 285	AGTGGCGGCTTCAAGGAT GT
82	GMS084	20	SEQ ID NO: 286	GACGGGAACAACGTTGA CGA	20	SEQ ID NO: 287	TTGCAGGCTCACTGACTC GG
83	GMS085	20	SEQ ID NO: 288	TCCGCGTACCTGCAATT CAA	20	SEQ ID NO: 289	TCCAGAGCGAGGTTGACG CT
84	GMS087	20	SEQ ID NO: 290	GGCGATCCTTGTCCCTTG TGG	20	SEQ ID NO: 291	TCCCTTCCGCCCGTCTTAT T
85	GMS088	23	SEQ ID NO: 292	TGTACACATAACCGTAGC ATACCA	20	SEQ ID NO: 293	GATGAGGGAAACGTGCG TCC
86	GMS090	20	SEQ ID NO: 294	TCGCGAGGGAAGTTGAG GTC	20	SEQ ID NO: 295	AGGTACGTCCTCGCCCT TC
87	GMS092	20	SEQ ID NO: 296	GGCGTAGAGCATCGTT TGG	20	SEQ ID NO: 297	TGCCCATCTCTCGTCTTCC C
88	GMS093	20	SEQ ID NO: 298	AAGGCCTCTCCATTGTC CCG	20	SEQ ID NO: 299	TCTGCATGGATTGCTCT GC
89	GMS094	20	SEQ ID NO:	CATTACGCAGCCGACTC	19	SEQ ID NO:	AGCGGAGCAAGGAACCG

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			300	ACG		301	AC
90	GMS096	20	SEQ ID NO: 302	TTGTCAGCACCTTTTGAC GGC	20	SEQ ID NO: 303	CGTGTGTCGTAGCCATT CG
91	GMS097	20	SEQ ID NO: 304	CGAATGGCTACGACCAC ACG	20	SEQ ID NO: 305	CGCATTCAAAATCCCTCTG CC
92	GMS098	20	SEQ ID NO: 306	AGGCAGTGCTCTTCCAC CCA	20	SEQ ID NO: 307	CGTGCTTGTCTCGACGTCCA TT
93	GMS099	20	SEQ ID NO: 308	AGGGTTTTCTGGGTTTT GCG	20	SEQ ID NO: 309	GGTGTATCATCCCCCACC CA
94	GMS100	20	SEQ ID NO: 310	CTGACAGGTGTCGGCGG TAA	20	SEQ ID NO: 311	GTCTGGACCGGACGAAAGC AT
95	GMS101	20	SEQ ID NO: 312	AGCCAAGAACCCCTCA GCG	20	SEQ ID NO: 313	CCGAAGCTGTTGATCACC CC
96	GMS102	20	SEQ ID NO: 314	GTGATAGTCGGCGCTGC CTT	20	SEQ ID NO: 315	CCACGTATGCCATCAAAG CG
97	GMS103	20	SEQ ID NO: 316	TCTCACCAACTTCCACG GCA	23	SEQ ID NO: 317	TGTGCAAGTGTGGAGATA CAAGG
98	GMS104	20	SEQ ID NO: 318	CAGTGTGGACCGAAGCG AGA	20	SEQ ID NO: 319	TGGAAGTGGGTGGAAGA CGC
99	GMS105	20	SEQ ID NO: 320	CGCACATGTGTCTTCGT CCA	20	SEQ ID NO: 321	TGCCGATCTCTCACACTC CG
100	GMS106	21	SEQ ID NO: 322	TTGATCGTCTTCACGGC ATTG	20	SEQ ID NO: 323	ACACCGTAGGGTTCCAGC GA
101	GMS107	20	SEQ ID NO: 324	TCCGCCTTAAGGACAAG CCC	20	SEQ ID NO: 325	ACCCCCTGAGACACTGAC GC
102	GMS108	20	SEQ ID NO: 326	GTTCGATTACCCGCCCTG TGC	20	SEQ ID NO: 327	ACGTCGACTCCTTGTGG GG
103	GMS109	20	SEQ ID NO: 328	ATGAGGCCCTTGCCCTC TTC	20	SEQ ID NO: 329	CATGGACGCATTCAACCTG GA
104	GMS110	20	SEQ ID NO:	CCTCCAGATCCCCAACCC	20	SEQ ID NO:	GATGTGCGGGCAGAAAG

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			330	TCC		331	AGG
105	GMS111	20	SEQ ID NO: 332	AAGCCACCAACCTCGGC TAT	20	SEQ ID NO: 333	GAATGCGTTCTGACCTGC GA
106	GMS112	20	SEQ ID NO: 334	CAGAAAGGCTCACCCAGT CCG	20	SEQ ID NO: 335	ATTTGAACCGATGTATG CG
107	GMS114	22	SEQ ID NO: 336	GGACTACCTCTCCCCAA ATCCG	20	SEQ ID NO: 337	GAGAAAGAGCCGCGACTG GAG
108	GMS115	20	SEQ ID NO: 338	CACCCAAACCGGAAAGGT CAC	20	SEQ ID NO: 339	GGTACCCACCTGTCAACG CC
109	GMS116	20	SEQ ID NO: 340	ACCTGCGCCGTGATTCC TAT	20	SEQ ID NO: 341	GTTCGTGCGAGAACGAGC AG
110	GMS117	20	SEQ ID NO: 342	AAGCCGTCGAAGTCGCA AAG	20	SEQ ID NO: 343	CAAGACCCGCTGTCAACG GA
111	GMS118	20	SEQ ID NO: 344	GGCTATCCAGCCCCCAA CTC	20	SEQ ID NO: 345	ACGTCACTCTGCTGGAAGG GC
112	GMS119	20	SEQ ID NO: 346	GCGACAGTCATGGGTTA CGG	20	SEQ ID NO: 347	CGATAAGGAGGACGAGG GGG
113	GMS120	20	SEQ ID NO: 348	AGGACATCCAGAAAGGC GCTG	20	SEQ ID NO: 349	GGGTCTTGTCTTGCACTC GT
114	GMS122	20	SEQ ID NO: 350	GATGAGGGGGTATGCGT GGA	20	SEQ ID NO: 351	CAGACCCAGGAGGAAG CTG
115	GMS123	20	SEQ ID NO: 352	GGTTGACCGGTACTCAC GCC	20	SEQ ID NO: 353	CTCCCATCCTGCAGCAAT CA
116	GMS124	20	SEQ ID NO: 354	GAACGCGGGGAGAAAGGA CAGA	21	SEQ ID NO: 355	GCGATACTGTCTGCAATG GGC
117	GMS125	20	SEQ ID NO: 356	TGATACCCACACAGCCCC TCG	20	SEQ ID NO: 357	ACGACCGTCATCCAGAAC CA
118	GMS126	20	SEQ ID NO: 358	GACGTGCCGAGCGTGT GTA	20	SEQ ID NO: 359	CACCCCTCTGGTTCTGGCC TC
119	GMS127	20	SEQ ID NO:	CGAGCTCGAGGAGGACT	20	SEQ ID NO:	GTGGAGGCCCTGTCTGAACG

Marker SEQ ID NO.	Marker ID	Forward Primer length	Forward primer SEQ ID NO:	Forward Primer sequence	Reverse Primer length	Reverse primer SEQ ID NO:	Reverse Primer sequence
			360	TCG		361	AT
120	GMS128	20	SEQ ID NO: 362	AAACATCCCGAGGAAAC GCA	20	SEQ ID NO: 363	AACCTTGACCGACGCTGT CC
121	GMS129	20	SEQ ID NO: 364	GGAGTGGTTCCGAGTGGG ACG	20	SEQ ID NO: 365	CGGGGAGAAATCATCCCAA CA
122	GMS130	20	SEQ ID NO: 366	ACACCAGGGTCAGGGCTC AGG	20	SEQ ID NO: 367	AGGGCTGCTTGCTCGAAG GT
123	GMS132	20	SEQ ID NO: 368	TGCACGATTACTTACGC GGG	20	SEQ ID NO: 369	CTGCTGGATGGCCTGGTA GG

[00108] References:

- 5 IDRIS, A S; MIOR, M H A Z; MAIZATUL, S M and KUSHAIRI, A (2011). Survey on status of Ganoderma disease of oil palm. Proc. of the PIPOC 2011 International Palm Oil Congress – Agriculture Conference. MPOB, Bangi. p. 235-238
- READER, U. and BRODA, P (1985) Rapid preparation of DNA from filamentous fungi. Letters in Applied Microbiology, 1:17-20

CLAIMS

1. A method for determining the pathogenicity of *Ganoderma* species, wherein the method comprises the step of detecting the presence or absence of one or more sequences
5 selected from a group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93 or a single nucleotide polymorphism (SNP) variant thereof, wherein the presence of one or more sequences selected from the group consisting of
10 SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof, indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID
15 NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the *Ganoderma* species is not pathogenic.
- 20 2. The method of any one of the preceding claims, wherein the one or more sequences, or a SNP variant thereof, comprises a SNP.
3. The method of claim 2, wherein the one or more sequences, or a SNP variant thereof, comprises a SNP selected from the group consisting of SEQ ID NO: 83, SEQ ID NO:
25 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93.
4. The method of any one of the preceding claims, wherein the SNP variant has at least
30 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO:

95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93.

- 5 5. The method of any one of the preceding claims, wherein the detecting is conducted on a sample potentially containing or suspected of containing said *Ganoderma* species.
6. The method of the method of claim 5, wherein the sample is selected from the group consisting of soil, root, shoot, stem, bark, leaves, fruits, seed and an isolated microbial strain.
- 10 7. The method of any one of the preceding claims, wherein the detecting is conducted using a technique selected from the group consisting of sequencing, hybridization-based methods, enzyme-based methods, PCR-based methods, and post-amplification methods.
- 15 8. The method of claim 7, wherein the PCR is conducted using at least one set of primers selected from the group consisting of SEQ ID NOs: 288-289, 178-179, 336-337, 198-199, 214-215, 234-235, 256-257, 312-313, 318-319, 344-345, 172-173, 250-251, 284-285, and 308-309.
- 20 9. The method of any one of claims 5 to 8, comprising the steps of:
 - (a) extracting DNA from the sample;
 - (b) subjecting the DNA to sequencing to determine sequences of the extracted DNA;
 - (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof;
 - (d) determining the absence or presence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof.
- 25 sequence selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof;
- 30 (d) determining the absence or presence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof.

10. The method of any one of the preceding claims, wherein the *Ganoderma* species is pathogenic in a plant selected from the group consisting of crops, wherein the crop is optionally selected from a group consisting of oil palm, *Mucuna bracteata*, coconut, betel nut, sugar palm, sago palm, rubber tree, and *Acacia mangium*; forest trees; and ornamental plants, optionally selected from *Chrysalidocarpus lutescens*.
11. The method of any one of the preceding claims, wherein the *Ganoderma* is selected from the group consisting of a *Ganoderma boninense*, *Ganoderma orbiforme*, *Ganoderma sp.*, *Ganoderma orbiforme*, *Ganoderma fornicatum*, *Ganoderma lucidum*, *Ganoderma weberianum*, *Ganoderma williamsianum*, *Ganoderma steyaertanum* and *Ganoderma applanatum*.
12. The method of claim 11, wherein the *Ganoderma sp.* is selected from the group consisting of *Ganoderma sp.* BL-6 18S (accession no. JN400509.1), *Ganoderma sp.* BL-9 18S (accession no. JN400510.1), *Ganoderma sp.* BL-94 18S (accession no. JN400511.1), *Ganoderma sp.* BP-16 18S (accession no. JN400513.1), *Ganoderma sp.* BRIUMSa (accession no. JN234427.1), *Ganoderma sp.* BRIUMSb (accession no. JN234428.1), *Ganoderma sp.* BRIUMSc (accession no. JN234429.1), *Ganoderma sp.* C16452 (accession no. EU239386.1), *Ganoderma sp.* MT-1 (accession no. AY220543.1), *Ganoderma sp.* G31 (accession no. KR093030.1), and *Ganoderma sp.* MEL 2382607 (accession no. KP012934.1)
13. The method of claim 11, wherein the *Ganoderma* is a *Ganoderma boninense*, *Ganoderma steyaertanum* or *Ganoderma orbiforme* species.
14. A method for detecting presence of pathogenic *Ganoderma* in a sample, comprising the steps of:
 - (a) extracting DNA from the sample;
 - (b) subjecting the DNA to sequencing to determine the sequences of the extracted DNA;
 - (c) comparing the DNA sequences determined from step (b) to a reference DNA sequence selected from a group consisting of SEQ ID NOs: SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID

NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof; and

(d) determining the absence or presence of one or more sequences selected from a group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof, wherein the presence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the *Ganoderma* species is pathogenic, and the absence of one or more sequences selected from the group consisting of SEQ ID NO: 83, SEQ ID NO: 28, SEQ ID NO: 107, SEQ ID NO: 38, SEQ ID NO: 46, SEQ ID NO: 56, SEQ ID NO: 67, SEQ ID NO: 95, SEQ ID NO: 98, SEQ ID NO: 111, SEQ ID NO: 25, SEQ ID NO: 64, SEQ ID NO: 81, and SEQ ID NO: 93, or a SNP variant thereof indicates that the *Ganoderma* species is not pathogenic.

15. The method of claim 14, wherein the sample is selected from the group consisting of root, shoot, stem, bark, leaves, fruits, seed and an isolated microbial strain.