



(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/NL2016/050626

(22) International Filing Date:
9 September 2016 (09.09.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15184943.7 11 September 2015 (11.09.2015) EP

(71) Applicant: **NEDERLANDSE ORGANISATIE VOOR TOEGEPAST-NATUURWETENSCHAPPELIJK ONDERZOEK TNO** [NL/NL]; Anna van Buerenplein 1, 2595 DA 's-Gravenhage (NL).

(72) Inventor: **SCHUREN, Frank Henri Johan**; c/o TNO / IP & Contracting, Anna van Buerenplein 1, 2595 DA 's-Gravenhage (NL).

(74) Agent: **JANSEN, C.M.**; V.O., P.O. Box 87930, 2508 DH Den Haag (NL).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,

MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *of inventorship (Rule 4.17(iv))*

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: DETECTION OF QUALITY OF SURFACE WATER

(57) Abstract: The present invention relates to a method for developing a classifier for classifying the trophic state of surface water comprising the steps of: a. taking samples of surface water of different trophic states; b. determining the nucleotide sequence of 18S RNA of micro-organisms in said sample; c. selecting for those sequences that are unique for a specific trophic state. Further, the present invention relates to a method for classifying the trophic state of surface water comprising the steps of: a. taking a surface water sample; b. extracting the nucleotide sequence of the 18S RNA of the microorganisms in said sample; c. checking for the presence of one or more of the nucleotide sequences depicted in any of Tables 1 - 6 in the extracted nucleotide sequences of step b; d. classifying the sample on basis of the presence or absence of said sequences. Further, a kit comprising an array for hybridization forms part of the invention.



Detection of quality of surface water

The invention relates to the field of biological testing, more particularly the testing of biodiversity, especially the testing of the biodiversity in order to detect the quality of surface water.

Surface water makes up around 4% or 368,000 m² of the European Union and its quality significantly impacts public health and commercial enterprises as well as flora and fauna. As this water is under continuous pressure from contamination by heavy metals and nutrient overload from fertilisers and pesticides and remainders of all kinds of chemical substances from human waste, it has to be regularly monitored to enable timely interventions.

For this purpose, surface water is generally divided into three main categories, dependent upon the level of nutrients that is provided in the water, also called the trophic state of the water. The three main classes are eutrophic, mesotrophic and oligotrophic. The classification of water into these categories is basically provided by Carlson's trophic state index (as e.g. used by the United States Environmental Agency for determining the trophic state of lakes). This index is defined as the total weight of algal biomass in a given water body. An oligotrophic lake is a lake with low nutrient content. These lakes have low algal production, and consequently, often have very clear waters, with high drinking-water quality. The bottom waters of such lakes typically have ample oxygen; thus, such lakes often support many fish species, like lake trout, which require cold, well-oxygenated waters. Ecologists use the term oligotrophic to distinguish unproductive lakes, characterized by nutrient deficiency, from productive, eutrophic lakes, with an ample or excessive nutrient supply. A eutrophic body of water, commonly a lake or pond, has high biological productivity. Due to excessive nutrients, especially nitrogen and phosphorus, these water bodies are able to support an abundance of aquatic plants. Usually the

water body will be dominated either by aquatic plants or algae. When aquatic plants dominate the water tends to be clear. When algae dominate the water tends to be darker. The algae engage in photosynthesis which supplies oxygen to the fish and biota which inhabit these waters.

- 5 Occasionally an excessive algae bloom will occur and can ultimately result in fish kills due to respiration by algae and bottom living bacteria.

Accordingly, the difference in the trophic state of surface water is not related to the (drinking) quality of the water, but more or less reflects the amount of water life that is present in the water.

- 10 However, changes in the trophic state may be indicator of ecological changes, not seldom caused by environmental changes, such as pollution. Accordingly, monitoring of the trophic state of surface water and changes therein is of utmost importance.

- 15 The current way of assessing the trophic state is cumbersome. An estimate of the state can be provided by measuring either the amount of chlorophyll pigments, the total phosphorus content or the so-called Secchi depth. Of these three the chlorophyll pigments will most accurately indicate the trophic state, but this changes with the season. The Secchi depth measurement is considered to be the least accurate, but is nowadays the
20 cheapest and most readily available method. This parameter measures water transparency and by that would indicate the concentration of dissolved and particulate matter in the water, which in turn can be used to derive the biomass.

- 25 All of these parameters have the drawbacks that they are very general measurements, which may be influenced by other phenomena than the amount of nutrients (e.g. the season, sudden bursts of algae growth, and so on).

- 30 Currently the most reliable way to assay the trophic state of surface water is to take a sample and to look at the microbial lifeforms in the water, especially the diatom populations.

Diatomeae are unicellular algae of the phylum Bacillariophyta that are characterized by an exocellular silica skeleton. The wealth of species is enormous: it is estimated that there are more than 100,000 different species. They are water-bound and can be found in both deep and shallow waters, where they make up most of the phytoplankton. As such the population of diatoms in a water sample reflects the nutritional, i.e. trophic state of the water and thus diatoms can be used for the classification of the trophic state of surface water. However, currently this classification is performed by microscopic determination of the diatoms in the water sample and this method, although it is very reliable, is very labour intensive and thus costly.

Accordingly, an improved method for classification of the trophic state of surface water is needed.

SUMMARY OF THE INVENTION

The present invention relates to a method for developing a classifier for classifying the trophic state of surface water comprising the steps of:

- a. taking samples of surface water of different trophic states;
- b. determining the nucleotide sequence of 18S RNA of micro-organisms in said sample;
- c. selecting for those sequences that are unique for a specific trophic state.

Further, the present invention relates to a method for classifying the trophic state of surface water comprising the steps of:

- a. taking a surface water sample;
- b. extracting the nucleotide sequence of the 18S RNA of the microorganisms in said sample;
- c. checking for the presence of one or more of the nucleotide sequences depicted in any of Tables 1 - 6 in the extracted nucleotide sequences of step b;

d. classifying the sample on basis of the presence or absence of said sequences.

Further, a kit comprising an array for hybridization forms part of the invention.

5

DETAILED DESCRIPTION

The term "nucleic acid" as used herein, is interchangeable with the term "polynucleotide", and refers to a nucleotide multimer or polymeric form of nucleotides having any number of nucleotides, e.g.,
10 deoxyribonucleotides or ribonucleotides, or compounds produced synthetically (e.g. PNA as described in U.S. Pat. No. 5,948,902 and the references cited therein) and can be either double- or single-stranded. A polynucleotide can hybridize with other polynucleotides in a sequence
15 specific manner, e.g. can participate in Watson-Crick base pairing interactions. The term also includes modified, for example by methylation and/or by capping, and unmodified forms of the polynucleotide

The terms "deoxyribonucleic acid" and "DNA" as used herein mean a polymer composed of deoxyribonucleotides.

20 The term "oligonucleotide" refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alkyl and aryl-phosphate, phosphorothioate), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases
25 (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl ribosyl, 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains

capable of forming secondary structures (e.g., stem-loop, pseudo knots and kissing loop structures).

The term "nucleotide sequence homology" as used herein denotes the presence of homology between two polynucleotides. Polynucleotides have

5 "homologous" sequences if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence. Sequence comparison between two or more polynucleotides is generally performed by comparing portions of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison

10 window is generally from about 20 to 200 contiguous nucleotides. The "percentage of sequence homology" for polynucleotides, such as 50, 60, 70, 80, 90, 95, 98, 99 or 100 percent sequence homology may be determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison

15 window may include additions or deletions (i.e. gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by: (a) determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions; (b)

20 dividing the number of matched positions by the total number of positions in the window of comparison; and (c) multiplying the result by 100 to yield the percentage of sequence homology. Optimal alignment of sequences for comparison may be conducted by computerized implementations of known algorithms, or by visual inspection. Readily available sequence comparison

25 and multiple sequence alignment algorithms are, respectively, the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990; Altschul et al., 1997) and ClustalW programs, both available on the internet. Other suitable programs include, but are not limited to, GAP, BestFit, PlotSimilarity, and FASTA in the Wisconsin Genetics Software Package

30 (Genetics Computer Group (GCG), Madison, WI, USA) (Devereux et al.,

1984). Especially the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489) should be mentioned.

5 As used herein, "substantially complementary" means that two nucleic acid sequences have at least about 70%, preferably about 80%, more preferably about 90%, even more preferably 95%, and most preferably about 98%, sequence complementarity to each other. This means that primers and probes must exhibit sufficient complementarity to their template and target
10 nucleic acid, respectively, to hybridise under stringent conditions. Therefore, the primer and probe sequences need not reflect the exact complementary sequence of the binding region on the template and degenerate primers can be used. For example, a non-complementary nucleotide fragment may be attached to the 5'-end of the primer, with the remainder of the primer
15 sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer has sufficient complementarity with the sequence of one of the strands to be amplified to hybridize therewith, and to thereby form a duplex structure which can be extended by the polymerising
20 means. The non-complementary nucleotide sequences of the primers may include restriction enzyme sites. Appending a restriction enzyme site to the end(s) of the target sequence would be particularly helpful for cloning of the target sequence. A substantially complementary primer sequence is one that has sufficient sequence complementarity to the amplification template to
25 result in primer binding and second-strand synthesis. The skilled person is familiar with the requirements of primers to have sufficient sequence complementarity to the amplification template.

 The term "hybrid" in the context of nucleic acids refers to a double-stranded nucleic acid molecule, or duplex, formed by hydrogen
30 bonding between complementary nucleotide bases. The terms "hybridise" or

"anneal" refer to the process by which single strands of nucleic acid sequences form double-helical segments through hydrogen bonding between complementary bases.

The term "probe" refers to a single-stranded oligonucleotide sequence that will recognize and form a hydrogen-bonded duplex with a complementary sequence in a target nucleic acid sequence or its cDNA derivative.

The term "primer" as used herein refers to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach, thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product is induced, i.e., in the presence of nucleotides and an agent for polymerisation such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerisation. The exact lengths of the primers will depend on many factors, including temperature and composition (A/T en G/C content) of primer. A pair of bi-directional primers consists of one forward and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

It will be understood that "primer", as used herein, may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the target region to be amplified. Hence, a "primer" includes a collection of primer oligonucleotides containing sequences representing the possible variations in the sequence or includes nucleotides which allow a typical base pairing.

Oligonucleotides, such as primers or the sequences as defined in any of SEQ ID NO: 1 - 97 may be prepared by any suitable method. Methods for preparing oligonucleotides of specific sequence are known in the art, and include, for example, cloning and restriction of appropriate sequences, and
5 direct chemical synthesis. Chemical synthesis methods may include, for example, the phospho di- or tri-ester method, the diethylphosphoramidate method and the solid support method disclosed in e.g. U.S. Pat. No. 4,458,066. The oligonucleotides may be labelled, if desired, by incorporating means detectable by for instance spectroscopic, fluorescence, photochemical,
10 biochemical, immunochemical, or chemical means.

Template-dependent extension of the oligonucleotide primer(s) is catalysed by a polymerising agent in the presence of adequate amounts of the four deoxyribonucleotide triphosphates (dATP, dGTP, dCTP and dTTP, i.e. dNTPs) or analogues, in a reaction medium which is comprised of the
15 appropriate salts, metal cations, and pH buffering system. Suitable polymerizing agents are enzymes known to catalyse primer- and template-dependent DNA synthesis. Known DNA polymerases include, for example, E. coli DNA polymerase I or its Klenow fragment, T4 DNA polymerase, and Taq DNA polymerase. The reaction conditions for catalysing DNA synthesis
20 with these DNA polymerases are known in the art.

The products of the synthesis are duplex molecules consisting of the template strands and the primer extension strands, which include the target sequence. These products, in turn, serve as template for another round of replication. In the second round of replication, the primer extension
25 strand of the first cycle is annealed with its complementary primer; synthesis yields a "short" product which is bound on both the 5'- and the 3'-ends by primer sequences or their complements. Repeated cycles of denaturation, primer annealing and extension result in the exponential accumulation of the target region defined by the primers. Sufficient cycles
30 are run to achieve the desired amount of polynucleotide containing the

target region of nucleic acid. The desired amount may vary, and is determined by the function which the product polynucleotide is to serve.

The PCR method is well described in handbooks and known to the skilled person.

5 After amplification by PCR, the (presence of) nucleic acids in a sample may be detected by hybridisation with a polynucleotide (designated as 'marker' or 'probe') which forms a stable hybrid with that of the target sequence under stringent to moderately stringent hybridisation and wash conditions. If it is expected that the probes will be essentially completely
10 complementary (i.e., about 99% or greater) to the target sequence, stringent conditions will be used. If some mismatching is expected, for example if variant strains are expected with the result that the probe will not be completely complementary, the stringency of hybridisation may be lessened. However, conditions are chosen which rule out non-specific/adventitious
15 binding. Conditions which affect hybridisation, and which select against non-specific binding are known in the art, and are described in, for example, Sambrook et al., (2001). Generally, lower salt concentration and higher temperature increase the stringency of binding. For example, it is usually considered that stringent conditions are incubations in solutions which
20 contain approximately 0.1xSSC, 0.1% SDS, at about 65°C incubation/wash temperature, and moderately stringent conditions are incubations in solutions which contain approximately 1-2xSSC, 0.1% SDS and about 50°-65°C incubation/wash temperature. Low stringency conditions are 2xSSC and about 30°-50°C.

25 The terms "stringency" or "stringent hybridisation conditions" refer to hybridisation conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimised to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid
30 sequence. The terms as used include reference to conditions under which a

probe or primer will hybridise to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridise specifically at higher

5 temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridises to a perfectly matched probe or primer. Typically, stringent

10 conditions will be those in which the salt concentration is less than about 1.0 M Na^+ ion, typically about 0.01 to 1.0 M Na^+ ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions

15 may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or "conditions of reduced stringency" include hybridisation with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 2x SSC at 40°C. Exemplary high stringency conditions include hybridisation in 50% formamide, 1 M NaCl,

20 1% SDS at 37°C, and a wash in 0.1x SSC at 60°C. Hybridisation procedures are well known in the art and are described by e.g. Ausubel et al., 1998 and Sambrook et al., 2001.

The term "hybridization pattern" refers to the list of measurements of spot intensities obtained after hybridizing the array with a

25 sample comprising a variety of nucleic acids.

The term "nucleotide" is used to denote a deoxyribonucleoside, a ribonucleoside, or a 2'-O-substituted ribonucleoside residue.

The term "array" refers to an array of individual fragments of DNA or oligonucleotides that are bound to a substrate such as a microscope

30 slide. The purpose of an array experiment is to determine the amount of

matching nucleic acid fragments in a particular sample. The target nucleic acid fragments are extracted from the cells or tissues of interest, optionally converted to DNA, and labeled. They are then hybridized to all the DNA or oligonucleotide spots on the array. Matches (i.e. the spots where

5 hybridization has taken place) are identified using antibody detection, optionally combined with precipitation, chromatography, colorimetry, and/or phosphorescent or fluorescent imaging. The data resulting from an array experiment are a list of measurements of spot intensities. A preferred array of the invention is an array which has less than 1000, preferably less than

10 500, more preferably less than 200, and preferably more than 50, more preferably more than 70, more preferably more than 80, more preferably more than 90, more preferably more than 95 spots, and which contains at least 50 of the DNA sequences of SEQ ID NO:1 – 97. Such an array may, next to the sequences that are actually designed for the detection or

15 quantification of nucleotides in a sample, also contain sequences that are used as controls for the overall performance of the test.

For instance, arrays can be formed on non-porous surfaces (such as glass) by robotic micropipetting of nanoliter quantities of DNA to predetermined positions on a non-porous glass surface (as in Schena et al,

20 Science 270 : 467-470, 1995, and WO 95/35505). This is a "spotting" technique. Generally, in a spotting technique, the target molecules are delivered by directly depositing (rather than flowing) relatively small quantities of them in selected regions. For instance, a dispenser can move from address to address, depositing only as much target as necessary at

25 each stop. Typical dispensers include an ink-jet printer or a micropipette to deliver the target in solution to the substrate, and a robotic system to control the position of the micropipette with respect to the substrate. In other embodiments, the dispenser includes a series of tubes, a manifold, an array of pipettes, or the like so that the target polypeptides can be delivered

30 to the reaction regions simultaneously.

In addition the oligonucleotide array can be synthesized by a light-directed chemical coupling, (see Pirrung et al, (1992) U.S. Pat. No. 5,143, 854; Fodor et al, (1998) U.S. Pat. No. 5,800,992; Chee et al, (1998) U.S. Pat. No. 5,837,832. In brief, the light-directed combinatorial synthesis of oligonucleotide arrays on a glass surface proceeds using automated phosphoramidite chemistry and chip masking techniques. In one specific implementation, a glass surface is derivatized with a silane reagent containing a functional group, e.g., a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithographic mask is used selectively to expose functional groups which are then ready to react with incoming 5' photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface. Combinatorial synthesis of different oligonucleotide analogues at different locations on the array is determined by the pattern of illumination during synthesis and the order of addition of coupling reagents.

The data generated by assaying an array can be analyzed using known computerized systems. For instance, the array can be read by a computerized "reader" or scanner and quantification of the binding of probe to individual addresses on the array carried out using computer algorithms. Likewise, where a control probe has been used, computer algorithms can be used to normalize the hybridization signals in the different spots of the array. Such analyses of an array can be referred to as "automated detection" in that the data is being gathered by an automated reader system. In the case of labels that emit detectable electromagnetic wave or particles, the emitted light (e. g., fluorescence or luminescence) or radioactivity can be

detected by very sensitive cameras, confocal scanners, image analysis devices, radioactive film or a Phosphoimager, which capture the signals (such as a color image) from the array. A computer with image analysis software detects this image, and analyzes the intensity of the signal for each
5 probe location in the array. Signals can be compared between spots on a single array, or between arrays (such as a single array that is sequentially tested with multiple different samples), or between the labels of different probes on a single array. Computer algorithms can also be used for comparison between spots on a single array or on multiple arrays. In
10 addition, the data from an array can be stored in a computer readable form. Certain examples of automated array readers (scanners) will be controlled by a computer and software programmed to direct the individual components of the reader (e. g., mechanical components such as motors, analysis components such as signal interpretation and background
15 subtraction). Optionally software may also be provided to control a graphic user interface and one or more systems for sorting, categorizing, storing, analyzing, or otherwise processing the data output of the reader. To "read" an array, an array that has been assayed with a detectable molecule to produce binding (e. g., a binding pattern) can be placed into (or onto, or
20 below, etc., depending on the location of the detector system) the reader and a detectable signal indicative of probe binding detected by the reader. Those addresses at which the molecule has bound to an immobilized nucleic acid probe provide a detectable signal, e. g., in the form of electromagnetic radiation. These detectable signals could be associated with an address
25 identifier signal, identifying the site of the "positive" hybridized spot. The reader gathers information from each of the addresses, associates it with the address identifier signal, and recognizes addresses with a detectable signal as distinct from those not producing such a signal. Certain readers are also capable of detecting intermediate levels of signal, between no signal at all
30 and a high signal, such that quantification of signals at individual addresses

is enabled. Certain readers that can be used to collect data from the arrays, especially those that have been tested with using a fluorescently tagged molecule, will include a light source for optical radiation emission. The wavelength of the excitation light will usually be in the UV or visible range, but in some situations may be extended into the infra-red range. A beam splitter can direct the reader emitted excitation beam into the object lens, which for instance may be mounted such that it can move in the x, y and z directions in relation to the surface of the array substrate. The objective lens focuses the excitation light onto the array, and more particularly onto the (polypeptide) targets on the array. Light at longer wavelengths than the excitation light is emitted from addresses on the array that contain fluorescently-labeled probe molecules (i. e., those addresses containing a nucleic acid molecule within a spot containing a nucleic acid molecule to which the probe binds). In certain embodiments, the array may be movably disposed within the reader as it is being read, such that the array itself moves (for instance, rotates) while the reader detects information from each address. Alternatively, the array may be stationary within the reader while the reader detection system moves across or above or around the array to detect information from the addresses of the array (see 2001 WO 01/73134).

20

The term “trophic state” when relating to surface water, is herein defined as a measure for the indication of the level of nutrients that is present in the water. Although there is a classification into oligotrophic (little nutrients), mesotrophic (mediocre amount of nutrients) and eutrophic (high amount) and although there are even classes in between those (oligo-mesotrophic, meso-eutrophic) and even beyond those (hypereutrophic), it should be understood that classification is based on a sliding scale of an increase in the amount of nutrients. This scale does not recognize any decisive points where there is a clear distinction between two neighbouring

classes. The borders of the classes will thus not be clearly defined and accordingly also classification of a sample into a class may be fuzzy.

The current invention provides a classification method for the
5 trophic state of surface water.

It was found that it was possible to develop a set of
biomarkers/probes that was able to distinguish between the different
trophic states of surface water. This was achieved by taking samples of
surface water with known characteristics (i.e. a confirmed trophic state),
10 isolating the 18S nucleic acid from the micro-organisms in the sample and
comparing the sequences found with sequences from other samples with the
same trophic state. Doing this for the three trophic states (but also
recognizing intermediate states like oligo-mesotrophic and meso-eutrophic)
yielded a total of a few hundred sequences, some of which were uniquely
15 found in a specific state, others that occurred in more than one state. From
these many sequences now 97 sequences have been defined as being suitable
for providing such a classification of surface water into one of the three, or
preferably 5 or 6 trophic states.

The general method for finding these sequences can be described as a
20 method for developing a classifier for classifying the trophic state of surface
water comprising the steps of:

- a. taking samples of surface water of different trophic states;
 - b. determining the nucleotide sequence of 18S RNA of micro-organisms in
said sample;
 - 25 c. selecting for those sequences that are unique for a specific trophic state.
- For such a method samples of water having a known trophic state need to be
available.

To ensure that only the eukaryotic organisms are considered in the sample
it has been chosen to look at the 18S RNA sequences, since these are
30 characteristic for eukaryotic organisms. If needed, the water specimens first

can be filtered to remove larger structures such as debris and larger eukaryotic organisms (seaweed, little molluscs and crustaceae). The 18S RNA is very suited for discovering the variation in a sample, because it contains highly conserved parts for which it is possible to prepare consensus
5 primers for amplification using PCR or similar techniques. The amplicons that are produced in this way will next to the consensus parts also contain highly variable sequences that are specific on a family, genus, species or even variety level. It is this variety that enables the correlation the 18S sequences of the micro-organism population with the trophic state of the
10 surface water.

The invention accordingly is also related to a method for the classification of the trophic state of surface water comprising the steps of:

- a. taking a surface water sample;
- b. extracting the nucleotide sequence of the 18S RNA of the microorganisms
15 in said sample;
- c. checking for the presence of one or more of the nucleotide sequences depicted in any of Tables 1 - 6 in the extracted nucleotide sequences of step b;
- d. classifying the sample on basis of the presence or absence of said
20 sequences.

As has been discussed above, extraction of the nucleotide sequence encoding the 18S rRNA subunit can be achieved by using an amplification of said nucleotide sequence with the use of consensus primers. In the current
25 invention the use of the primers having the sequences **[5'-TGCAGTTAAAAAGCTCGTAG-3' and 5'-TAATCATCTTCGATCCCCTA-3']** has been found very advantageous, since these primers anneal to a consensus sequence in the 18S rRNA coding sequence. Amplification with these primers should be performed under less stringent conditions to allow
30 mismatches to be present.

After amplification the nucleic acids in the sample will be denaturised to make them single stranded. Thereafter will be applied to an array on which the probes (also indicated as biomarkers or marker sequences) have been provided. Amplicons that are complementary to the marker sequences will
5 be bound to them and can be detected.

The sequences of markers given in the sequence listing are about 120 – 130 nucleotides long. However, this does not necessarily mean that the complete sequence as shown in any of the SEQ ID's needs to be present on the array: the assay can be performed with sequences, which are
10 parts thereof. It should, however, be understood, that a minimal length is necessary for obtaining a correct and distinctive hybridization with any sample nucleic acid. Thus, the length of these part(s) of the sequences should range from about 30 to about full length nucleotides, i.e. the sequences can be shorter than those presented in the sequence listing, but
15 for a sufficient hybridization to occur they should be at least 30 nucleotides long and preferably contain the variable region which discriminates the 18S RNA sequence from other 18S RNA sequences. Especially the 5' and 3' ends may be truncated since these are consensus sequences that served as anchoring point for the primers.

Further, it will depend on the length of the probes and the hybridization conditions if and to what extent mismatches in the sequences would still lead to hybridization to the probe. Preferably, however, the hybridization assay is performed under stringent conditions to be certain that the probes do not aspecifically bind to sequences in the sample thereby
25 decreasing the reliability of the assay. Nevertheless, it is assumed that any nucleotide sequence or part thereof (as defined above) which has a homology of at least 90%, preferably at least 95%, more preferably at least 97%, even more preferably at least 98% with the sequences of SEQ ID NO: 1 through SEQ ID NO: 97 is applicable in and intended within the scope of the
30 present invention.

In the current invention the term "homologues" is used to indicate both (parts of) the sequences of the invention of at least 30, but preferably at least 50, more preferably at least 75, more preferably at least 100 and most preferably at least about 120 nucleotides, and sequences
5 which have a degree of homology of at least 90%, preferably at least 95%, more preferably at least 97%, even more preferably at least 98% with the sequences of SEQ ID NO: 1 through SEQ ID NO: 97.

The above discussed markers, or fragments thereof may be spotted on a surface to provide for a DNA micro-array. In order to facilitate
10 coupling of the fragments, the surface of the array (e.g. the slide, the surface of which may i.a. be glass, gold, etc.) may be modified. Spotting may occur by any method available, for instance by using ElectroSpray Ionization (ESI) micro-array printing. After spotting of the markers or fragments thereof, the slide surfaces may be blocked to prevent further attachment of
15 nucleic acids, e.g. by treatment with boro-anhydride in case of formaldehyde modified glass-slide surfaces.

To facilitate detection of successful hybridization, the sample DNA is suitably labeled, preferably with a compound which is uniquely detectable by an antibody. Such a compound can for instance be biotin.
20 Labeling of nucleotides with biotin and subsequent detection through antibodies specific for biotin, has been sufficiently described in the literature and will be routine experimentation for a person skilled in the art. Alternatively, the nucleotides are fluorescently labeled (e.g. by using CyTM labels [Amersham Pharmacia Biotech]). Fluorescent labeling kits are
25 commercially available from various manufacturers. In order to be able to judge the signals caused by the hybridization of the marker sequences with sample nucleic acid, preferably the array also comprises reference nucleotide sequences, which can serve as positive and negative controls. As negative controls, sequences should be used of about the same length as the
30 average length of the markers, but which will not hybridize with any

nucleotide sequence in the sample, i.e. sequences, which do not occur in eukaryotic 18S rRNA coding sequences. As positive control, a sequence should be used which will be present in (nearly) all samples to be tested, for instance through adding a known amount of an artificial sequence to the sample just before testing.

The average size of sample nucleic acid, which is defined by the primers that are chosen for the amplification procedure, has an effect on the signal distribution on the array. Larger sample molecules comprise more information and are thus more likely to find a suitable hybridization partner in more of the spots. Reducing the average size of the sample nucleic acid can reduce this phenomenon. On the other hand, when the sample nucleic acid is too small, the nucleic acid fragments in the sample contain too little genetic information and also find suitable hybridization partners in many spots. The average size of the fragments in the sample nucleic acid is preferably between about 30 and 3000 nucleotides. More preferably, the average size of the fragments in the sample nucleic acid comprises a size of between about 50 and 1000 nucleotides, more preferably between about 100 and 500 nucleotides.

After performance of the hybridization assay the pattern that shows up from the (labeled) sequences that are detected shows the trophic class in which the sample can be classified. Below are six Tables over which in total 97 sequences (SEQ ID NO: 1 – 97) are listed (in most cases there is overlap between consecutive Tables) showing which sequences would be typically found for each of the six trophic states. If a classification into three classes is performed then only Tables 1, 3 and 5 may be considered, although also in this case it would be advisable to additionally use a large part, if not all markers of Tables 2, 4 and 6. If an unknown sample is tested and all 97 sequences would be available on the array for hybridizing, it simply results in counting from which Table the most sequences are recognized; the sample then will be classified according to the class of the

Table where the most hits are found. It has, however, been found that an acceptable classification may be performed with about 30 of the 97 sequences of Tables 1 – 6, which means that 5 sequences from every Table of the Tables 1 – 6 would be sufficient for a classification that still would be acceptable from a commercial perspective. Preferably however, because a larger amount of probes increases the performance of the assay, at least 40, more preferably at least 50, more preferably at least 70, more preferably at least 80, more preferably at least 90, more preferably at least 95 of the sequences of Table 1 – 6 should be used in a hybridisation assay for the classification of the trophic state of surface water. However, since the trophic state presumable has a Gaussian distribution over the classes, it would be advisable to take more samples from the middle classes (Tables 2 – 5) than from the extreme classes (Table 1 and 6).

It will, of course, be possible to add more sequences to the 97 sequences that are presented in the Table below. It is expected that performing a further search for additional sequences that can be used in the classification will provide further useful sequences, especially if samples are taken from climatologic different origin. The origin of the present set of probes is from a variety of surface water sources in the Netherlands , and thus would represent at least a variety that is useful for classification of a surface water sample derived from Northern-Europe. Although it is believed that the present set of markers would also be adequate for classifying surface water samples from the rest of the world, it is expected that when the current panel is completed with markers that were found using a test panel of surface water from other parts of the world and/or specific climatologic conditions, the performance of the classification can be improved.

Next to the *per se* classification of unknown samples, the present assay can also be advantageously used in monitoring the water quality by looking at changes over time in samples from the same origin. Although the trophic state of water may be subject to seasonal variation (see introduction) monitoring for

detecting any positive or negative changes in the quality of the surface water. In such a case a full array as described above may be used, but it would also be possible to use just those markers that are predictive for the class in which the sample was originally classified (and maybe its neighbouring classes) to see if it would still fall within that class. For such applications arrays may be produced which consist of the markers of only one or a few classes.

Although not specifically designed for quantitative measurements, the assay of the present invention can be used to at least get an impression of the total amount of micro-organisms that are present in the sample. This can be derived from the intensity of the label at the spot(s) where a particular biomarker is present. If, for instance, the label is a fluorescent label, the emission of fluorescent radiation will be higher if more label (i.e. more target sequences from the sample) is hybridized to the marker sequence. Of course, this semi-quantitative measurement can be turned into a reliable quantitative measurement if there is a control measurement which has a known concentration and if the signal provided by the label is linear with respect to the concentration of the label at the spot. Further, it should be ensured that the spot is provided with sufficient marker molecules to enable binding to multiple target sequences from the sample.

20

Also part of the invention is a kit for the classification of the trophic state of surface water, wherein said kit comprises an array for hybridization optionally together with the necessary reagents for performing the hybridization assay, such as buffers, etc. , Said array is provided with at least 30 spots, each spot containing a sufficient number of molecules comprising one of the sequences as depicted in any of Tables 1 – 6. The nucleic acids may be bound to non-nucleic acid moieties for immobilization on the surface of the array plate as has been defined above. In one spot there should be sufficient identical molecules to enable differentiation in intensity when many identical target molecules are present in the sample to be tested. Next to the spots with sequences derived from the below Tables, the array may further contain positive and negative control sequences and

30

optionally also additional areas for measurement of e.g. pH, oxygen concentration and the like.

5

Table 1 Diatomeae 18S marker sequences for oligotrophic surface water

Class	SEQ ID NO	sequence
uniqueoligotroof_66	1	TTGGATTTCTGGTTTGAGCGTCCGGTCGACCTTGTGTCTGTA CTATGGATGTTTCGAGCCATATTTTGTGAGGATGCCTTTCTGC CATTGAGTTGGTGGTTGGGTGAACTTGCATCGTTTACTGTG
uniqueoligotroof_26	2	TCGGATTTTCGGGTGCTGCCCCGGTGTAGCCAAGGCGAACGCC GGGTGGTGCCTTCCTTCTGGGATTCTCTGCACTCTTCACTGA GTCGGATGGCGTTCCAGGACGTTTACTGTG
uniqueoligotroof_233	3	TTGGATCTGGGGTTCGGGTCGGTGGTTTGCCGCTCTGCGGC TACTGCCGTGCTCGACTCTTACTAGCCGGTGCCTCTTGGT GCTCTTCATTGAGTGCCACGAGAGCCTGGCATGTTTACTTTG A
oligotroof_9	4	TTGAATTTCTGGCAGGGGCCATCAAATTTCCACATGGTGGAT TGTGTGGTGGCCCCTGCCTTCTTTGGGTGAAGTTCGGTGGC ATTAACCTGTTGCTGTTACGACATCCATCGTTTACTGTGA
oligotroof_10	5	TTGGATTTGTGGTTGAGCTGTCCGGTCCGACACTTTTGTGTG GGTACTTGGACGGGTCGTCCATCCTTGGGTGGAATCTGTGT GGCATTAGGTTGTCGTGCAGGGGATGCCCATCTTTTACTGTG
oligotroof_11	6	TTGGATTTCTGGTCGCGTCCGCCGGTCCGGCCCTCACGGGTT GTGTACCTGGCGTGTGCAGCCATCCTCGGGGAGAGCACATC TGGGATTAAGTTGTCGGGTGTGGTATCCTCGTCATTTACTGT GA
oligotroof_13	7	TTGAATTTCTGGCTCCGAATCGCGGTCTGGCCCAAGGGTTAT GTACCAGCGAGTTGGAGCCATCCTCGAGGAGAACTTATCTG GCATTCATTTGTTGGGTAAGAGACCCTCGTCATTTACTGTGA
oligotroof_14	8	TTGAACCTTGGGCCTGGCTGGCCGGTCCGCCTCACC GCGTGC ACTGGTCCGGCCGGGTCTTTCCTTCTGAGGAGCCGCATGCCC TTCACTGGGTGTGTGCGGGGAATCAGGACTTTTACTTTG
oligotroof_15	9	TTGAATTTTGGACCTGGTTTAATGGTTAGGCCTTTGGCTTTTA CTATTGATTGGGTCTTTCCTTCTAGGGAACCTGTGTGCACTTT ATTGTGTGTGCAGGGGATCTAGGACTTTTACTTTG
oligotroof_20	10	TTGAACCTTGGGCCTGGCTGGCCGGTCCGCCTCACC GCGTGC ACTGGCTCGGCCGGGTCTTTCCTCCTGGGGAGCCGCATGCCC TTCACTGGGTGTGTGCGGGGAACCAGGACTTTTACTTTG

oligotroof_23	11	TTGAATCTTGGGCCTGGTTTGGCGGTCGCGCTTTAGGGCGT GTACTGTGCGCCGGGTCTTTTACTTGCGGGGAACGTGCGTG CCTCTCGGGGTGTGTACGGGAACCGTGGCTATTACTTTG
oligotroof_37	12	TTGGATTTCTGGTTTGAGCGTCCGGTCGAGTTTATCTCTGTAC TATGGATGTTTGAGCCATATTTTGTGAGGATGCCTTTCTGCC ATTCAGTTGGTGGTTGGGTAGACTTGATCGTTTACTGTG
oligotroof;oligo- mesotroof_48	13	TTGGATTTGTGGCAGGTGCGTGCGGCCCGACACTGTGTGTG GGTGCTTGTGCGCGCCGCCATCCTTGGGTGGAACCTGTGTG GCATTAGGTTGTCGTGCAGGGGATGCCATCTTTTACTGTG
oligotroof;oligo- mesotroof_53	14	TTGGATTTCTGGTTGTGGACAGCTTTCAGCACTTCATTGTGTC TGACTTGTGTCTGCGTCCATATTCGGGGGAAATTTATCTGG CATTAAAGTTGTCGGGTAAAGGATCCTCGTCATTTACTGTGA
oligotroof;oligo- mesotroof_62	15	TTGGATTTCTGCTGAGGACGATCGGTCCGCCCTCTGGGTGA GTATCTGACTCGGCCTTGGCATCTTCTTGGAGAACGTTACTG CACTTGACTGTGTGGTTCGGTATCCAGGACTTTTACTTTGA
oligotroof;oligo- mesotroof_64	16	TTGAACCTTGGGCCTGGCTGGCCGGTCCGCCTACCGCGTGC ACTGGTCCGGCCGGGCCTTTCCTTCTGGGGAGCCGCATGGC CTTCACTGGCCGTGTCGGGGAACAGGACTTTTACTTTG
oligotroof;oligo- mesotroof_70	17	TTGGATTTCTGGTTTGAGCGTCCGGTCGACCTTGTGTCTGTA CTATGGATGTTTGGGCCATATTTTGTGAGGATGCCTTTCTGC CATTGAGTTGGTGGTTGGGTAGACTTGATCGTTTACTGTG

Table 2 Diatomeae 18S marker sequences for meso-oligotrophic surface water

Class	SEQ ID NO	sequence
oligotroof;oligo- mesotroof_48	13	TTGGATTTGTGGCAGGTGCGTGCGGCCCGACACTGTGTGTGGGT GCTTGTGCGCGCCGCCATCCTTGGGTGGAACCTGTGTGGCATTAG GTTGTCGTGCAGGGGATGCCATCTTTTACTGTG
oligotroof;oligo- mesotroof_53	14	TTGGATTTCTGGTTGTGGACAGCTTTCAGCACTTCATTGTGTCTGA CTTGTTGTCTGCGTCCATATTCGGGGGAAATTTATCTGGCATTAAAG TTGTCGGGTAAAGGATCCTCGTCATTTACTGTGA
oligotroof;oligo- mesotroof_62	15	TTGGATTTCTGCTGAGGACGATCGGTCCGCCCTCTGGGTGAGTAT CTGACTCGGCCTTGGCATCTTCTTGGAGAACGTTACTGCACTTGAC TGTGTGGTTCGGTATCCAGGACTTTTACTTTGA
oligotroof;oligo- mesotroof_64	16	TTGAACCTTGGGCCTGGCTGGCCGGTCCGCCTACCGCGTGCACT GGTCCGGCCGGGCCTTTCCTTCTGGGGAGCCGCATGGCCTTCACT GGCCGTGTCGGGGAACAGGACTTTTACTTTG
oligotroof;oligo- mesotroof_70	17	TTGGATTTCTGGTTTGAGCGTCCGGTCGACCTTGTGTCTGTACTAT GGATGTTTGGGCCATATTTTGTGAGGATGCCTTTCTGCCATTGAGT TGGTGGTTGGGTAGACTTGATCGTTTACTGTG
uniqueoligo- mesotroof_117	18	TTGGATCTCAGTTCTGGACGGGCGGTTTCATGTAGCAATGATACCG TTCGTACCGGACCCATTGCTGGTTCTACCTAGGTGTCCTTCACTGG

		GTGCCCTGGGAGACCAGAAAGTTTACTTTG
uniqueoligo- mesotroof_477	19	TTGAACCTTGGGCCTGGCTGGCCGGTCCGCCTCACCGCGTGCACT GGTTCGGCCGGGCCTTTCCCTCTGGGGAGCCTCATGTCCTTCACTG GGTGTGATGGGGAACCAGGACTTTTACTGTG
uniqueoligo- mesotroof_109	20	TTGGATTTCTGTGTTGAATGTCCGGTCGATTCCCTTTGGGAGTGTG TACTTGTGACATTTGATGCCTTTTTGTGAGGATGTTTTCTGTCAT TAAGTTGATGGTTGAATGGACTTGCATCGTTTACTGTG
uniqueoligo- mesotroof_83	21	TTGGATTTCTGGTCGTGGTAATCGGTCCGCCCTTAACGGGGTGTG TACCAGGTTCTGCGTCCATCCTCGAGGAGAACGTATCTGGCATT AAGTTGTCGGGTATGATATCCTCGTCATTTACTGTGA
uniqueoligo- mesotroof_330	22	TTGAATTTCTGGCCGTGGGTCCGGTCCGTCGCCGAAGGGTGTG TACCAGGAATTCGCGGCCATCCTCGAGGAGAGTATATCTGGCATT CAGTTGTTGGGTATAGGACCCTCGTCATTTACTGTGA
oligo- mesotroof_81	23	TTGGATCTCGGCTCTCAGGGGGCGGTCCGTTATTTGACGCGACTG CCCCATTTGAGAGTGTTTCTCGGCGGAACCTCCGGGGTGCTCTTC GCCGAGTGTCGGCTTGGGGGCCGCCAGTTTACTTTG
oligo- mesotroof_83	24	TTGAATTTTGGACCTGGTTGGGCGGTCTTCTCATTGAGATGCACTG CTCTGACTGGGTCTTTCCTTCTGGCGAACTCTTGCTATTCATTTA GTGCGATTGGGAACCAGGACTTTTACTTTGA
oligo- mesotroof_85	25	TTGAATTTTGGGCCTGGTTGGTTGGTCTGCCTTTGGTATGAACTGA CCGACTGGGCCTTTCCTTCTGGCGAACTATTGTGCTATTCATTTAG TGCGATAGGGAACCAGGACTTTTACTTTG
oligo- mesotroof_86	26	TTGGATTTCTGGAAATAGCGAGTGAAATATCCTTTACGGGAGATG TTGCTCGCATTATCCATCTTTGAAGGATAGAAGTCTGGCATTGGT TGTCGGGCTACTGTGCTTCGTCATTTACTGTGA
oligo-mesotroof; mesotroof_87	27	TTGAATTTCTGGTCTCGGACAGCGGTCTGCCTCAACCGAGGTATG TACTATCTGTCTAAGTCCATCCTCGGGGAAAACCTATCTCTGATTA AGTTTCAGGGGTATGTGATCCTCGTCATTTACTGTGA
oligo-mesotroof; mesotroof_90	28	TTGGATTTCTGGTAGCGACATACGGTCCGCCCCTCACGGGGTGCG CACTAGTATGTTGCATCCATCCTCGAGGAGAACGTGTCTGGCATT CAGTTGTGCGGATACGGTATCCTCGTCATTTACTGTGA
oligo-mesotroof; mesotroof_96	29	TTGAATTTCTGGTTTTCGGAGTGCGGTCTGCCCCGTGGGGTAAGTA TCAGTGCTTCGGAGCCATCCTCGAGGAGAACATATCTGTGATTGA GTTTCATGGGTATGGGAGCCTCGTCATTTACTGTGA

Table 3 Diatomeae 18S marker sequences for mesotrophic surface water

Class	SEQ ID NO	sequence
oligo-mesotroof; mesotroof_87	27	TTGAATTTCTGGTCTCGGACAGCGGTCTGCCTCAACCGAGGTATG TACTATCTGTCTAAGTCCATCCTCGGGGAAAACCTATCTCTGATTA AGTTTCAGGGGTATGTGATCCTCGTCATTTACTGTGA
oligo-	28	TTGGATTTCTGGTAGCGACATACGGTCCGCCCCTCACGGGGTGCG

mesotroof;mesot roof_90		CACTAGTATGTTGCATCCATCCTCGAGGAGAACGTGTCTGGCATT CAGTTGTCGGATACGGTATCCTCGTCATTTACTGTGA
oligo-mesotroof; mesotroof_96	29	TTGAATTTCTGGTTTCGGAGTGCGGTCTGCCCCGTGGGGTAAGTA TCAGTGCTTCGGAGCCATCCTCGAGGAGAACATATCTGTGATTGA GTTTCATGGGTATGGGAGCCTCGTCATTTACTGTGA
mesotroof_104	30	TTGGATTTTCGGATGGGAGTTGCCGGTCCGTCGTTACGATGTGCAC TGGCCGCTCCTATCTTGTTGTCGGGGACTAGCTCCTTGGATTCACT TCCTGGGACTAGGAGCTGACGAGGTTACTTTGA
mesotroof_107	31	TTGGATTTTGGGCTGGAGGATTCGGTCTGCCTTACGGGTTGATAC TGTTTCCTCCATCCTTCTGCCGGGGACGCGCTCTGGCCTTCACTG GCTGGGGCGCGGAGTCGGCGATGTTACTTTG
mesotroof_115	32	TTGGATTTGTGGGTGGAGGTGCTGGCCCAACGCTTTCGCTGGGT GCTGGCTACCTCTGTCCATCCTTGGGTGGAACCTGTGTGGCATTGA AGTTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof_117	33	TTGAATTTCTGGTCTCGGACAGCGGTCTGCCTCAACCGAGGTATG TACTATCTGTCTAAGTCCATCCTCGGGGAAAACCTATCTCTGATTC AGTTCAGGGGTATGTGATCCTCGTCATTTACTGTGA
mesotroof;meso- eutroof_123	34	TTGGATTTTCGGGTGGGTACCGCCGGTCTCCATTGGGTATGTACT GGCGTGGCCCGCCTTGCTGTCGGGGACGATCTCCTGGGCTTCACT GTCCGGGAATCGGAGTCGACGAGGTTACTTTGA
mesotroof;meso- eutroof_131	35	TTGGATTTGTGGATGGGCCGTCCGGTGCGCTCTTTTGAGCGTTTC ATGGGTAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso- eutroof_136	36	TTGGATCTCGGGTCCAGGCTTGC GGTTTCGCTCGCGCGATTACT GCTCGTCCTGACCTACCTCCCGGTTTTCCCTTGGTGCTCTTCGCTG AGTGCTTCGGGTGGCCGGAACGTTTACCTTG
mesotroof;meso- eutroof_144	37	TTGGACCTTGGGCCGGGTGCGCCGGTCCGCCTCACGGCGAGCAC CGACCTACTCGACTCTATTGCCGGCGATGCGCTCCTAGCCTTAATT GGCCGGGTGCTGCCTCCGGCGCCGTTACTTTGA
mesotroof;meso- eutroof_147	38	TTGGATTTTCGGGTGGGTACCGCCGGTCTCCATTGGGTAGTACTG GCGTGGCCCGCCTTGCTGTCGGGGACGATCTCCTGGGCTTACTG TCCGGGAATCGGAGTCGACGATGTTACTTTGA
mesotroof;meso- eutroof_148	39	TTGGATTTGTGGATGGGCCGTCCGGTGCGCTCTTTTGAGCGTTAC ATGGGTAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso- eutroof_150	40	TTGGATTTGTGGATGGGCCGTCCGGTGCTCTTTTGAGCGTTAC TTGGGCAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso- eutroof_156	41	TTGGATTTGTGGCAGTCGCATGCGGCCCGGCATTCTGTGTCGGTGT TTGCTCGCGTCGCCATCCTTGGGTGGAACCTGTGTGGCATTAGGT TGTCGTGCAGGGGATGCCATCATTACTGTG
mesotroof;meso- eutroof_167	42	TTGGACTTTAGGCAAGGCGGTCTGGTCGCCCACTGGGTGTACTGG GCTCGCCGCGCCTCACCTTCGGGGAAGTGTGTCTTCGGTGCG CTAGGGAACCGAGCCTTACTTTGA

Table 4 Diatomeae 18S marker sequences for meso-eutrophic surface water

Class	SEQ ID NO	sequence
mesotroof;meso-eutroof_123	34	TTGGATTTGCGGTGGGTACCGCCGGTCTCCATTGGGTATGTACT GGCGTGGCCCGCCTTGCTGTCGGGGACGATCTCCTGGGCTTCACT GTCCGGGAATCGGAGTCGACGAGGTTACTTTGA
mesotroof;meso-eutroof_131	35	TTGGATTTGTGGATGGGCGGTCCGGTGCGCTCTTTGAGCGTTTC ATGGGTAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso-eutroof_136	36	TTGGATCTCGGGTCCAGGCTTGCGGTTGCGCTCGCGGCGATTACT GCTCGTCCTGACCTACCTCCCGGTTTTCCCTTGGTGCTCTTCGCTG AGTGCTTCGGGTGGCCGGAACGTTTACCTTG
mesotroof;meso-eutroof_144	37	TTGGACCTTGGGCCGGGTGCGCCGGTCCGCCTCACGGCGAGCAC CGACCTACTCGACTCTATTGCCGGCGATGCGCTCCTAGCCTTAATT GGCCGGGTCTGCTCCGGCGCCGTTACTTTGA
mesotroof;meso-eutroof_147	38	TTGGATTTGCGGTGGGTACCGCCGGTCTCCATTGGGTAGTACTG GCGTGGCCCGCCTTGCTGTCGGGGACGATCTCCTGGGCTTACTG TCCGGGAATCGGAGTCGACGATGTTACTTTGA
mesotroof;meso-eutroof_148	39	TTGGATTTGTGGATGGGCGGTCCGGTGCGCTCTTTGAGCGTTAC ATGGGTAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso-eutroof_150	40	TTGGATTTGTGGATGGGCGGTCCGGTGCTCTTTGAGCGTTAC TTGGGCAGCTCGTCCATCCTTGGGTGGAATCTGTGTGGCATTAAAG TTGTCGTGCAGGGGATGCCATCTTTACTGTG
mesotroof;meso-eutroof_156	41	TTGGATTTGTGGCAGTCGCATGCGGCCCGGCATTGCTGTCGGTGT TTGCTCGCGTCGCCATCCTTGGGTGGAACCTGTGTGGCATTAGGT TGTCGTGCAGGGGATGCCATCATTACTGTG
mesotroof;meso-eutroof_167	42	TTGGACTTTAGGCAAGGCGGTCTGGTCGCCCCTGGGTGTACTGG GCTCGCCGCGCCTCACCTTCGGGGAAGTAGTTGTGCTTCGGTGCG CTAGGGAACCGAGCCTTACTTTGA
uniquemeso-eutroof_189	43	TCGACATATCTCGATACCAGACATCTTTCTGCACTTCTACCTCTGGT AGTTGTCGTAGAACCTTTACTTTGA
uniquemeso-eutroof_190	44	TTGAACCTTGGGTTTGGTTGGTCCGCCTACCGCGTGCACT GATCCGACCGAATCTTCTTCTGGGGAACCGTATGCCCTTCACTG GGTGTGCTGGGGAACCAAGGACTTTACTTTG
meso-eutroof_175	45	TTGGATCTGAGGACTCGAATTGGTGGTGCAGCAAACGCTGTTACT GCTGATTCGGCCTACCTAGTCGGTGCTCGCTCTGGTGTTCTTAAT TGAATGCCCTGAGTGCCCGACATGTTACTTTGA
meso-eutroof_176	46	TTGAACCTTGGGCCTGGCTGGCCGGTCTGCTTCACCGCATGTACT GGTCTGGCTGGGTCTTCTTCTTGGGGAACCGTATGCCCTTACTG GGTGTACTGGGGAACCAAGGCTTTACTTTG
meso-	47	TTGGATCTCGGGTCCAGGCTTGCGGTTGCGCTCGCGGCGATTACT

eutroof_178		GCTCGTCCTGACCTACCTCCCGGTTTTCCCTTGGTGCTCTTCGTTGA GTGCTTCGGGTGGCCGGAACGTTTACCTTG
meso- eutroof_180	48	TTGGATTTCTGTTCAGGACGACCGGTCCGCCCTCTGGGTGTGTATC TGGCTCGGCCTGGGCATCTTCTTGGAGAACGTCGCTGCACTTGAC TGTGTGGCACGGTATCCAGGACTTTTACTTTGA
meso- eutroof_182	49	TTGGATTCCGGGAGTGAGCCGCCGGTCGCCGCCGCGGTGCGATA TCGGCGCTCTTCCTTACCAGCCTCGGCCGCCGCTGTGCTTCGCTG CTCGGGCGCGCGCCGCGGGCGATGTTACCTTGA
meso- eutroof_212	50	TTGAACCTTGGGCCTGGCTGGTCGGTCCGCCTCACCGCGTGTACT GATCCGGCCGGGCCTTTCTTCTGGCTAGCCGCATGCCCTTCACTG GGTGTGTTGGTGAACCAGGACTTTTACTTTG
meso- eutroof_218	51	TTGGATTTCTGCTGAGGATGACTGGTCCGCCCTCTGGGTGAGTAT CTGGCTCAGCCTTGGCATCTTCTTGGAGAACACAGCTGCACTTCAC TGTGTGGTGTGGAATCCAGGACCTTTACTTTGA
meso- eutroof_220	52	TTGGATCTCGGGTACTAGCATATGGTACGCTTATTGCTGTTACTAT ATGTCTGGTCCTAATTCATCGTTGGCTGTTGATGTCCTTTATTGG GTGTTTTAGTTGCCGATACGTTTACTTTGA
meso- eutroof_227	53	TTGGATTTGTGGATGATCTTGGTGTCCACATTATTGTTGGTACTAT TGAGATTGTCCATCTTTGGGTGGAATCTGTGTGGCATTAGGTTGT CGTGCAGGGGATGCCATCGTTTACTGTG
meso- eutroof_232	54	TTGAACCTTGGGCCTGGGCGGGCGGTCTACCTATGGTTTGTACTG CTGCGGCCGGGTCTTTCCTTCTGGCTACCCCTCCGGGGGGGAACCA GGACGATTACTTTGA
meso- eutroof_235	55	TTGGATCTCGGGAAGTACATGTGGTACGCTTATTGCTGTTACTA CATGTCTGGTCCTAATCAATCGGTTGGCTGTTAATGCCCTTACTG GGTGTTCAGTAGCCGGTACGTTTACTTTGA
meso-eutroof; eutroof_258	56	TTGGATCTGCGCCTTCGGACTCGGTCCGCCAACGGGTGTGAACT GGGATCCAAGGCTTATACTGCTGGTTTTCCCTTGATGCTCTTACT GGGTGTCTTGGGTGGCTAGCGAGTTTACTTTG
meso-eutroof; eutroof_260	57	TTGGATTTGTGGTGTGTGGTGTGTTCCGGGCACTTGTGTCTGAG TAACACGCTGCTGCCATCCTTGGGTGGAACCTGTATGGCATTAGG TTGTCGTGCAGGGGATGCCATCTTTTACTGTG
meso-eutroof; eutroof_296	58	TTGGATTTTGGGTGGGCCGACCGGTCCGCCTCTGGGTGTGCACC GGTGTCTCGTCCCTTCTACCGCGATGCGTTCCTGGCCTTAACTG GCCGGTCTGCTCCTCCGTGCTGTTACTTTG
meso-eutroof; eutroof_331	59	TTGAACCTTGGGCCTAGCCGGCCGGTCCGCCTCACCGCGTGCAC GGCTCGGTTGGGCCTTTCTTCTGGAGAACCCATGTCCTTCACTG GGTGTGGTGGGGAACCAGGACTTTTACCGTGA
meso-eutroof; eutroof_336	60	TTGAATTTTGGGCCTGGTCGGGCGGTCTTGCTGCTTATTGTGGTA AGCACTGCTTGGTCGGGTCTTTCCTTCTGGGGAAGTGTGCTGCCT TTTACTGGGACTGGCCTAGCGATCCAGGACTTTTACTTTGA
meso-eutroof; eutroof_352	61	TTGAAGTTTGGTCTCGGACGCTGGGTCTGCTTAATTGCATGTACTT GACGGTCCGAGACTTCTTCTTGGTGAACGGCCGCTTCGGGTGG TCCGGAACCAGGACTATTACTTTG
meso-eutroof; eutroof_358	62	TTGGATTTCTGGCAGGAGCGTCCGGTCTTGCACTCTGTGCGGGAA CTTGGATGTCTCTGGCCATCTTGGGGATATCCTGTTTGGCATTAA GTTGTGCGGGCAGGGGATACCCATCGTTTACTGTG

meso-eutroof; eutroof_372	63	TTGGATTTCTGGGACGGAGACGTCGGTCCTCCCTTGTGGTCGATAC TGACTCTCTTCTTAATTGCCTCGAGCGTCGCCTAGTCTTCATTGCC TGGGCGCGCTACGCGGCGCCGTTACCTTGA
meso-eutroof; eutroof_381	64	TTGGACTTTGGGTTGGGTTGATCGGTCCGCCTTCTGGTGTGTACCT ATCGTCTCGTCCCTTCTGTGCGCGATACGTTCTTGCCTTAATTGG CCGGGTCGTGCCTCCGACGCTGTTACTTTGA
meso-eutroof; eutroof_406	65	TCGAATTTCTGGGACGGATGCTATGGTCCACCTCTTAGGTGGGACC AATGTGTCTAGTCCTATTTCTAGGGACGCTATAGCCTTAACGGGT GTGTAGCCGTGTCTAGGACTTTTACTTTGA
meso-eutroof; eutroof_421	66	TTGGATTTGTGGCGTTCGCGTGC GGCCCGGCATTATTGTGGTGT TTGCTAGCGTCGCCATCCTTGGGTGGAACCAAGTGTGGCATTAAAGT TGTCGTGCTGGGGATGCCCATCGTTTACTGTG
meso-eutroof; eutroof_431	67	TTGAACCTCGGGCCTGGCTGGCCGGTCCGCCTCACCGCGTGCACCT GGTCCGGCCGGGTCTTTCCTCCTGTGGAACCTCGGGTCCTTCACTG GGCTCGGCGGGAAAACAGGACATTTACTTTG
meso-eutroof; eutroof_502	68	TTGAACCTTAGGCAAGGCGGCCTGGTCACCCACTGGGTGTACTGG GCTCGCCGCGCCTCACCTTCGGGGAAGTGTGCTGCTTCGGTGC CTAGGGAACCGAGCCTTACTTTGA
meso-eutroof; eutroof_518	69	TTGGATTTCTGGGTGGGTACCGCCGGTCTCCCTTGGGTATGTACT GGCGTGGCCCGCCTTGTGTCGGGGACGGTCTCCTGGGCTTCACT GTCCGGGAATCGGAGTCGACGAGGTTACTTTGA

Table 5 Diatomeae 18S marker sequences for eutrophic surface water

Class	SEQ ID NO	sequence
meso- eutroof;eutroof_258	56	TTGGATCTGCGCCTTCGGACTCGGTCCGCCAACGGGTGTGAA CTGGGATCCAAGGCTTATACTGCTGGTTTTCCCTTGATGCTCTT TACTGGGTGTCTTGGGTGGCTAGCGAGTTTACTTTG
meso- eutroof;eutroof_260	57	TTGGATTTGTGGTGTGTGGTGTGTTCCGGGCACTTGTGTCTG AGTAACACGCTGCTGCCATCCTTGGGTGGAACCTGTATGGCAT TAGGTTGTCGTGCAGGGGATGCCCATCTTTACTGTG
meso- eutroof;eutroof_296	58	TTGGATTTTGGGTTGGGCCGACCGGTCCGCCTCTGGGTGTGCA CCGGTTGTCTCGTCCCTTCTACCGCGATGCGTTCCTGGCCTTA ACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTG
meso- eutroof;eutroof_331	59	TTGAACCTTGGGCCTAGCCGGCCGGTCCGCCTCACCGCGTGCA CTGGCTCGGTTGGGCCTTTCTTCTGGAGAACCCCATGTCCTTC ACTGGGTGTGGTGGGGAACCAAGGACTTTTACCGTGA
meso- eutroof;eutroof_336	60	TTGAATTTTGGCCTGGTCGGGCGGTCTGCTGCTTTATTGTG GTAAGCACTGCTTGGTCGGGTCTTCTTCTGGGGAAGTGTG CTGCCCTTTACTGGGACTGGCCTAGCGATCCAGGACTTTTACTT TGA
meso-	61	TTGAAGTTTGGTCTCGGACGCTGGGTCTGCTTAATTGCATGTA

eutroof;eutroof_352		CTTGACGGTCCGAGACTTCCTTCTTGGTGAACGGCCGCTTCG GGTGGTCCGGAACCAGGACTATTACTTTG
meso- eutroof;eutroof_358	62	TTGGATTCTGGCAGGAGCGTCCGGTCTTGCACTCTGTGCGGG AACTTGATGTCTCTGGCCATCTTTGGGGATATCCTGTTTGGC ATTAAGTTGTGCGGGCAGGGGATACCCATCGTTTACTGTG
meso- eutroof;eutroof_372	63	TTGGATTTCGGGACGGAGACGTCGGTCCTCCCTTGTGGTCGAT ACTGACTCTCTTCCTTAATTGCCTCGAGCGTCGCCTAGTCTTCA TTGCCTGGGCGCGCTACGCGGCGCCGTTACCTTGA
meso- eutroof;eutroof_381	64	TTGGACTTTGGGTTGGGTTGATCGGTCCGCCTTCTGGTGTGTA CCTATCGTCTCGTCCCTTCTGTGCGCGATACGTTCTTGCCTTA ATTGGCCGGGTCGTGCCTCCGACGCTGTTACTTTGA
meso- eutroof;eutroof_406	65	TCGAATTCGCGGCAGGATGCTATGGTCCACCTCTAGGTGGGA CCAATGTGTCTAGTCTATTTTCTAGGGACGCTATAGCCTTAAC GGGTGTGTAGCCGTGTCTAGGACTTTTACTTTGA
meso- eutroof;eutroof_421	66	TTGGATTGTGGCGTTCGCGTGCGGCCCGGCATTATTGTCGGT GTTTGCTAGCGTCGCCATCCTTGGGTGGAACCAAGTGTGGCATT AAGTTGTCGTGCTGGGGATGCCATCGTTTACTGTG
meso- eutroof;eutroof_431	67	TTGAACCTCGGGCCTGGCTGGCCGGTCCGCCTACCGCGTGCA CTGGTCCGGCCGGGTCTTCTCCTGTGGAACCTCGGGTCCTT CACTGGGCTCGGCGGGAAAACAGGACATTTACTTTG
meso- eutroof;eutroof_502	68	TTGAACTTTAGGCAAGGCGGCCTGGTCACCACTGGGTGTACT GGGCTCGCCGCGCCTCACCTTCGGGGAAGTAGTCGTGCTTCG GTGCGCTAGGGAACCGAGCCTTACTTTGA
meso- eutroof;eutroof_518	69	TTGGATTTCGGGTGGGTACCGCCGGTCTCCCTTTGGGTATGTA CTGGCGTGGCCCGCCTTGCTGTGCGGGACGGTCTCCTGGGCTT CACTGTCCGGGAATCGGAGTCGACGAGGTTACTTTGA
uniqueeutroof_355	70	TTGGATTCTGACAGGGGCGCATGGTGGGGCTCGCAAGGGTC CCGACCGATGTCAGTCTTTGTCATCCTTGGGGAGAGCCGTATT GCCATTAACCTGGTAGTGCGGGGATCCTCATCGTTTACTGTGA
uniqueeutroof_269	71	TTGACCAATTGGCCTGGCTGGCAGGTCCGCCTACCGCGTGCA CTTGTCGGGCCGGGCCTTTCTTTCTGGAGAACCGCATGCCCTTC ACTGGGCGTGTTGGGGACCAGGACTTTTACTTTG
uniqueeutroof_357	72	TTGAACCTTGGGCCTGGCTGGCTGGTCCGCCTACCGCGTGTA CTGGTTGGCCGGGCCTTCTTCTGGAGAACCCCATGCCCTTC ACTGGGTGTGCGGGGGAACCAGGACTTTTACTGTG
uniqueeutroof_187	73	TTGGACTTCTGGCAGGAGCGTCTGGTCTCACACTCTGTGCGGG AACTTGATGTCTCTGGCCATCTTTGGGGATATCCTGTTTGGC ATTAAGTTGTGCGGCAGGGGATACCCATCGTTTACTGTG
uniqueeutroof_289	74	TTGACCAATTGGCCTGGCTGGCGGGTCCGCCTACCGCGTGCA CTTGTCGGGCCGGGCCTTTCTTTCTGGAGAACCGCATGCCCTTC ACTGGGCGTGTTGGGGACCAGGACTTTTACTTTG
eutroof_527	75	TTGGATTGTGGTTGCTCGCTGGGTCTCTTTTGTAGGCATCCG GTTGGGTGACCATCCTTGGGTGGATGTTGCGTGGCATTGCGTT GTCGCGCAGGGGATGCCATCGTTTACTGTG
eutroof_529	76	TTGGACCTTGGGCCGGGTAGGTCGGTCCGCCTCTCGGGTGTG AACCAGCCTCCTCGTCCCTTCTGCCGGCGATGCGCTCCTGTCTT TAATTGGCCGGGTGCTGCCTCCGGCGCTGTTACTTTGA

eutroof_535	77	TTGGATTTCTGGAGCGGGCCCCGTCCGGTCCACCGCAAGGTGAGC TACTGGCGGGTCTGTTCTTCTTCGAAAGACTGCAGGTGCACT TCGCTGTGTGCTTGTGCGGATTTGTGACGTTTACTTTG
eutroof_542	78	TTGGATTTCTGGATACGGCAGCCGGTCTCCCTTAACGGGGTT GTGTACTAAGTTGCTGTGTCCATCCTCGAGGAGAACATGCCTG GCATTTATTTGTGCGGTATGGTATCCTCGTCATTTACTGTG
eutroof_543	79	TTGAATTTCTGGTCCCGGAACACGGTCAGCTCCAAGGAGTTCTG TACCTGTGGTCTAGGACCATCCTCGGGGAGAACATATCTGCCA TTAAGTTGGTGGGTATGGGACCCTCGTCATTTACTGTGA
eutroof_544	80	TTGGATTTCTGGGTGGGTTCTAGCGGTCCGCCTCTGGTGTGTAC TGCTATGGCTCACCTTTCTGTGCGGGGACCAGCTCTTGGGCTTA ACTGTCCGGGACTGGGAATCGGCGAGGTACCCGTGA
eutroof_548	81	TTGGACTTGTGCATGTCCCTGTATGTCCATATTGTTTGGTGCCT TCAAGGGCGTGCATCCTTGGGTGAAACCTGTGTGGCATTAGG TTGTCGTGTAGGGAATGCTCATCGTTTACTGTG
eutroof_572	82	TTGGATTTCTGGGTGGGGCCTGCCGGTCCGCCGTTTCGGTGTGC ACTGGCAGGGCCACCTTGTTGCCGGGGACGGGCTCCTGGGC TTCCTGTCCGGGACTCGGAGTCGGCGCTGTTACTTTGA
eutroof_577	83	TTGAATTTCTGGTCTCGGACATCGGTCTGCCTCGAACGAGGTA CGTACCAGTTGTCTGAATCCATCCTCGGGGAAAACGTGTCTGT CATTCAAGTTAATGGGCACGGGATCCTCGTCATTTACTGTGA
eutroof_589	84	TTGGACTTTGGGTTTGGCCGGCCGGTCCGCCCTCGTGGTGTGC ACCGGTCGTCTCATCCCTTCTACCGGCGATACGCTCCTGTCCTT AATTGGCCGGGTCGTGCCTCCGGTGTGTTACTTTGA
eutroof_600	85	TTGAAATTGGGGAATTATTTGATTGGAGACTTATGGTCGACA ATTGATTTATCCTTCGTGTCATGTATATTGGGAAGGTTCTTAA TTGGATTTTCTTGATAACTGACAAGTTTACTTTG
eutroof_612	86	TTGGATTTCTGGGCGGGGCCTGCCGGTCCGCCCTTGGTGTGCAC TGGCAGGGTCCGCCCTTGCTGTGCGGGACGGGTTCTGGGCTT TACTGTCTGGGACTCGGAGTCGACGTGGTTACTTTGA
eutroof_640	87	TTGGATCTCGGCACCCCTCTGCCGGTCCGCTGTTCTGGGCGGAC TGGCACGATCGGGGTGTGACGCTTTGGCGGAGGCCCGGGGG TGCTCTTGACCGAGTGTCGCCGCGGTGCCGTACGTTTACTTT G
eutroof_654	88	TTGGATTTCTGGTAGGATAACTTTGCCCTCTTGATGAGTGCGC AGAGTGATCCTGCCATCCTTCTGTTAACCTTCAGGCATTCTTT TGTTTGTTGGGGGCTCAGATATTTTACCTTGA
eutroof_659	89	TTGGATTTCTGGCAGGAGTGACCGGCCATGCACTCTGTGCGTG AGCTGTGTCATTCTGGCCATCCTTGGTGAGATCCTGTTTGGCA TTAAGTTGTGCGGGCAGGGGATAACCATCGTTTACTGTGA
eutroof_665	90	TTGGATTTCTGTTTTGAACGTCCGGTCCGCTTCTTAGAAGTGTG TTACTAGGATGTTTGAGACATTTTTGTGAGGATGTTCTTCTGC CATTCAAGTTGGTGGTTGAATAGACTTGATCGTTTACTGTG
eutroof_672	91	TTGAAATTGGGCACGGCCGGGCGGTCCACCTTAGGGTGCCT ACTGCTCAGTGGCGTTTCGGCGTCACGAGGCTGAGCCTTGGC CGCGTCTCGTGCAGGGGCAACCTTGACGCGGGACAGGCAG TTACCTTGA

eutroof_688	92	TTGGATTTTGGGTTGGGCCGACCGGTCCGCCTCTGGGTGTGCA CCGGTTGTCTCGTCCCTTTTACCGGCGATGCGTTCCTGGCCTTA ACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGA
eutroof; hypereutroof_750	93	TTGAATTTCTGGTCTCGGAACACGGTCAGCTCCTAGGAGTTCG TACCTGTGAGCCGGGGCCATCCTCGGGGAGAACATATCTGCC ATTAAGTTGGTGGGTATGGGACCCTCGTCATTTACTGTGA
eutroof; hypereutroof_763	94	TTGGATTTCTGCCGAGGACGACCGGTCCGCCCTCTGGGTGAG CATCTGGCTCGGCCTGGGCATCTTCTTGAGAACGTATCTGCA CTTGACTGTGTGGTGCGGTATTCAGGACTTTTACTTTGA
eutroof; hypereutroof_787	95	TTGAATTTTGGGCCTAGCTGGGTGGTCTAGCTTAACGGTTAGT ACTACTTTGGCTGGGTCTTTCCTTCTGGCTAACCAAGTGTGCCT TAACTGGGTGCGCTGGGGCTCCAGGACTTTTACTGTG
eutroof; hypereutroof_793	96	TTGAACCTTGGGCCTGGCTGGCAGGTCCGCCTACCGCGTGTA CTTGTCGGGCCGGGTCTTCTTCTGGAGAACCTCATGCCCTTCA CTGGGCGTGTTGGGGACCAGGACTTTTACTTTG
hypereutroof_817	97	TTGGATTTCTGGTAGGAATATCGATGTCAGATTGTTCTGTGCA GACATATCTTACCATCCTTCTGTAACTTGGTTTGCTCTTCAAA GGCGGCCAAGGGCTCAGATATTTTACCTTGA

Table 6 Diatomeae 18S marker sequences for hypereutrophic surface water

5

Class	SEQ ID NO	sequence
eutroof; hypereutroof_750	93	TTGAATTTCTGGTCTCGGAACACGGTCAGCTCCTAGGAGTTCG TACCTGTGAGCCGGGGCCATCCTCGGGGAGAACATATCTGCC ATTAAGTTGGTGGGTATGGGACCCTCGTCATTTACTGTGA
eutroof; hypereutroof_763	94	TTGGATTTCTGCCGAGGACGACCGGTCCGCCCTCTGGGTGAG CATCTGGCTCGGCCTGGGCATCTTCTTGAGAACGTATCTGCA CTTGACTGTGTGGTGCGGTATTCAGGACTTTTACTTTGA
eutroof; hypereutroof_787	95	TTGAATTTTGGGCCTAGCTGGGTGGTCTAGCTTAACGGTTAGT ACTACTTTGGCTGGGTCTTTCCTTCTGGCTAACCAAGTGTGCCT TAACTGGGTGCGCTGGGGCTCCAGGACTTTTACTGTG
eutroof; hypereutroof_793	96	TTGAACCTTGGGCCTGGCTGGCAGGTCCGCCTACCGCGTGTA CTTGTCGGGCCGGGTCTTCTTCTGGAGAACCTCATGCCCTTCA CTGGGCGTGTTGGGGACCAGGACTTTTACTTTG
Hypereutroof_817	97	TTGGATTTCTGGTAGGAATATCGATGTCAGATTGTTCTGTGCA GACATATCTTACCATCCTTCTGTAACTTGGTTTGCTCTTCAAA GGCGGCCAAGGGCTCAGATATTTTACCTTGA

Examples

Example 1: DNA isolation from fresh water samples and sequencing

5

DNA isolation from biological samples

The method consists of concentrating biological material through sedimenting by careful centrifugation for collecting eukaryotic material (18s rDNA fraction) followed by high speed centrifugation for collecting bacterial material (16s rDNA fraction).

Eukaryotic material is collected by overnight sedimentation of water samples of half a liter volume. Liquid is removed by pipetting until 50 ml is left with the sediment. This material is then centrifuged for 15 min at 300 g. The supernatant is removed by pipetting until 1 ml is left. The remaining sediment is considered the 18s rDNA fraction.

DNA is isolated from these samples with the following protocol:

1. In a 2 ml screw cap eppendorf tube the following components are mixed:
 - a. 400µl 18s rDNA fraction
 - 20 b. 400µl 0.5 mm zirconium beads (Biospec.), washed in lysisbuffer (Agowa mag Mini DNA isolation kit art. nr. 40401)
 - c. 500µl Phenol, water saturated with 10 mM Tris pH8.0.
 - d. 300µl lysisbuffer (Agowa).
2. 2 minutes beadbeating, cooling down on ice
- 25 3. 10 minutes centrifugation at 10000 rpm
4. Transfer 500 µl supernatant to new 2 ml eppendorf tube filled with 1 ml binding buffer and 20 µl magnetic beads (Agowa mag Mini DNA isolation kit art. nr. 40401). Mix by pipetting.
5. Incubate 10 minutes at rT while shaking (1000 rpm).
- 30 6. Place tubes in magnetic separator, incubate 1 minute

7. Remove supernatant followed by removing magnet.
8. Add 400µl washing buffer 1, mix by pipetting
9. Incubate 5 minutes at rT while shaking (1000 rpm)
10. Place tubes in magnetic separator, incubate 1 minute
- 5 11. Remove supernatant followed by removing magnet
12. Add 200µl washing buffer 2, mix by pipetting
13. Incubate 5 minutes at rT while shaking (1000 rpm)
14. Place tubes in magnetic separator, incubate 1 minute
15. Remove supernatant followed by removing magnet
- 10 16. Dry magnetic beads at 55°C
17. Resuspend magnetic beads in 63µl elution buffer
18. Incubate 10 minutes at 55°C
19. Place tubes in magnetic separator, incubate 1 minute
20. Remove supernatant and bring to new cup.
- 15 This DNA is now ready for PCR amplification

PCR amplification

The amplification mix contained 1 unit of Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific, Pittsburgh, PA, USA) and 1x Phusion HF buffer (Thermo Scientific), 200 µM dNTP and 1 µM of each primer (18s F 5'-TGCAGTTAAAAAGCTCGTAG-3' and 18s R 5'-TAATCATCTTCGATCCCCTA-3'). After denaturation (98°C; 30 sec), 35 cycles were performed that consisted of denaturation (98°C; 10 sec), annealing (50°C; 30 sec), and extension (72°C; 30 sec) followed by a final extension step (72°C; 5 min). As a next step PCR products were analyzed with the Caliper Labchip GX (Perkin Elmer, Santa Clara, CA, USA) for exact sizing and quantitation. Based on the results of the Caliper Labchip GX analysis, the amplicons of the individual samples were pooled in equimolar amounts in one library. This pool was gel purified and amplicons of the expected size were eluted by means of the QIAquick PCR Purification

Kit (Qiagen, Hilden, Germany) followed by AMpure bead purification (Beckman Coulter, Brea, CA, USA). This purified amplicon pool was then used for 454 pyrosequencing.

5 Sequencing data were processed as follows: data were trimmed by removing primer sequences and low-quality data, overlapping paired-ends were stitched together, sequences that did not have an exact match to the barcode, sequences that had an ambiguous base call (N) in the sequence. Sequences were grouped according to the barcode. We then aligned the
10 sequences with the MOTHUR align function (Schloss et al., 2009. Appl Environ Microbiol 75: 7537-7541) using kmer search. Chimera.uchime (available in MOTHUR) was used to remove sequence chimera's from the dataset.

15 Unique sequences were aligned using the "align.seqs" command and the MOTHUR-compatible Eukaryotic SILVA SEED database. Sequences were taxonomically classified by the RDP-II Naïve Bayesian Classifier using a 60% confidence threshold. Sequences were assigned to taxa according to the RDP training set using kmer search. In cases where
20 reads were equidistant to multiple 18S reference sequences, and/or where identical 18S sequences were derived from longer sequences mapping to different taxa, reads were assigned to the lowest common taxon of at least 60% of the sequences.

25 **Example 2 Classification of fresh water samples**

On basis of a hybridization assay with the 97 sequences from Tables 1 – 6 a number of samples of known trophic state was classified. The following results were obtained:

Table 7 Results of classification of 100 samples on basis of a hybridization assay with all 97 probes of Tables 1 - 6

	oligo	oligo-meso	meso	meso-eu	eutrophic	hyper	%
Assigned	23	13	6	26	38	3	70
2 groups	0	3	5	9	6	3	17
Unassigned	1	3	3	6	5	2	13

In this Table it shows that at least 70% of the samples are classified in the correct class when a six-class system was used. A further relatively large number of samples (17 out of 100) was assigned to two neighbouring classes, most probably because the sample was somewhat ambiguous because of the gradual changes from class to class. When looking at 5 classes (i.e. the classes eutrophic and hypereutrophic combined) the percentages for this joint class were: assigned 45 of 57 and unassigned 7/57. The remaining samples again would be classified as belonging to two groups.

Claims

1. A method for developing a classifier for classifying the trophic state of surface water comprising the steps of:
 - a. taking samples of surface water of different trophic states;
 - b. determining the nucleotide sequence of 18S RNA of micro-organisms in
5 said sample;
 - c. selecting for those sequences that are unique for a specific trophic state.
2. A method for classifying the trophic state of surface water comprising the steps of:
 - 10 a. taking a surface water sample;
 - b. extracting the nucleotide sequence of the 18S RNA of the microorganisms in said sample;
 - c. checking for the presence of one or more of the nucleotide sequences depicted in any of Tables 1 - 6 in the extracted nucleotide sequences of step
15 b;
 - d. classifying the sample on basis of the presence or absence of said sequences.
3. A method according to claim 2, wherein the classification is
20 provided into three classes: eutrophic, mesotrophic and oligotrophic.
4. A method according to claim 2, wherein the classification is provided into 5 classes: eutrophic, meso-eutrophic, mesotrophic, oligo-mesotrophic and oligotrophic, specifically wherein the class of eutrophic is
25 subdivided into two classes eutrophic and hypereutrophic.

5. A method according to any of claims 1 – 4, wherein the micro-organisms are diatomaceous micro-organisms.
6. A method according to any of claims 2 – 5, wherein the presence of
5 one or more of the sequences of Table 1 is indicative for the oligotrophic state of the surface water.
7. A method according to any of claims 2 – 5, wherein the presence of
one or more of the sequences of Table 3 is indicative for the mesotrophic
10 state of the surface water.
8. A method according to any of claims 2 – 5, wherein the presence of
one or more of the sequences of Table 5 is indicative for the eutrophic state
of the surface water.
15
9. A method according to any of claims 4 – 5, wherein the presence of
one or more of the sequences of Table 2 is indicative for the oligo-
mesotrophic state of the surface water.
- 20 10. A method according to any of claims 4 – 5, wherein the presence of
one or more of the sequences of Table 4 is indicative for the meso-eutrophic state of the surface water.
11. A method according to any of claims 4 – 5, wherein the presence of
25 one or more of the sequences of Table 6 is indicative for the hypereutrophic state of the surface water.
12. A method according to any of claims 2 – 5, wherein said sequences
of any of Tables 1 – 6 are available on an array for hybridization with the

sequences of the sample.

13. A method according to claim 12, where the array comprises from 50 to all of the sequences of Tables 1 – 6, preferably wherein the array
5 comprises more than 70, preferably more than 80, preferably more than 90 sequences.

14. Kit for the classification of the trophic state of surface water, comprising an array for hybridization, wherein said array is provided with
10 at least 30 spots, each spot containing molecules comprising one of the sequences as depicted in any of Tables 1 – 6.

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2016/050626

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

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012/021429 A1 (RUBLEE PARKE A [US] ET AL) 26 January 2012 (2012-01-26) paragraph [0131] - paragraph [0150]; claim 17; figure 1	1-13
Y	----- SFRISO ADRIANO ET AL: "Using phytoplankton and macrophytes to assess the trophic and ecological status of some Italian transitional systems", CONTINENTAL SHELF RESEARCH, vol. 81, 13 April 2014 (2014-04-13), pages 88-98, XP028653959, ISSN: 0278-4343, DOI: 10.1016/J.CSR.2014.03.013 abstract paragraph [03.2] - paragraph [0005] ----- -/--	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

19 December 2016

Date of mailing of the international search report

04/01/2017

Name and mailing address of the ISA/

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

Authorized officer

Cornelis, Karen

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2016/050626

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 101 824 481 A (OCEAN UNIV CHINA) 8 September 2010 (2010-09-08) the whole document -----	1-14
A	WO 2015/008011 A1 (MEDLIN LINDA [GB]; MONTRESOR MARINA [IT]; GRANELI EDNA [SE]; REGUERA B) 22 January 2015 (2015-01-22) page 3 - page 4 -----	14
A	MAIALEN GARMENDIA ET AL: "Phytoplankton composition indicators for the assessment of eutrophication in marine waters: Present state and challenges within the European directives", MARINE POLLUTION BULLETIN, vol. 66, no. 1-2, 1 January 2013 (2013-01-01), pages 7-16, XP055253173, GB ISSN: 0025-326X, DOI: 10.1016/j.marpolbul.2012.10.005 the whole document -----	1-14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/NL2016/050626

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012021429	A1	26-01-2012	NONE

CN 101824481	A	08-09-2010	NONE

WO 2015008011	A1	22-01-2015	EP 3022317 A1 25-05-2016
			US 2016281178 A1 29-09-2016
			WO 2015008011 A1 22-01-2015
