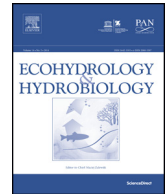




ELSEVIER

Contents lists available at ScienceDirect

Ecohydrology & Hydrobiology

journal homepage: www.elsevier.com/locate/ecohyd

Review Article

DNA barcoding – A new device in phycologist's toolbox

Zuzanna Kowalska^{*}, Filip Pniewski, Adam Latała

Institute of Oceanography, University of Gdańsk, Piłsudskiego 46 Ave., 81-378 Gdynia, Poland



ARTICLE INFO

Article history:

Received 4 July 2018

Accepted 13 January 2019

Available online 14 February 2019

Keywords:

DNA barcoding

Algae

Taxonomy

Biodiversity

ABSTRACT

Up to now only a fraction of all living organisms in the world has been identified. Organisms are identified using morphological characteristics, which is laborious and requires specialized knowledge. Barcoding DNA helps to solve the limitations of traditional species identification. The nucleotide sequence obtained using specific primers is presented in the form of a strip resembling a bar code which allows comparing and distinguishing species. The standard sequence used for the barcoding of almost all animal groups is derived from the mitochondrial cytochrome c oxidase gene (Cox1). For fungi Cox1 and ITS (internal transcribed spacers) were proposed as barcodes. In contrast, the world of plants does not have a single standard sequence for identifying species. The ideal marker should consist of a highly variable region enabling species discrimination, and be flanked by highly conserved regions allowing universal primer design for the DNA barcodes standardization. This paper describes the markers used to identify eukaryotic photosynthetic organisms, their properties, advantages and disadvantages, and the proposed standard DNA barcodes. This work focuses on microalgae and macroalgae, which are of great ecological importance and are widely used in food, pharmaceutical or energy industries where they fast, correct and reliable identification is crucial.

© 2019 European Regional Centre for Ecohydrology of the Polish Academy of Sciences.

Published by Elsevier B.V. All rights reserved.

1. Introduction

Diversity of life is the base of all biological research. In fact, up to now only a fraction of all living organisms (number of which is hotly debated; e.g. [Hammond, 1992](#); [Hebert et al., 2003](#); [Guiry, 2012](#); [Mann and Vanormelingen, 2013](#)) in the world has been identified. Identification of species has four important limitations. First, both the phenotypic plasticity and the genetic variation used to identify species may lead to incorrect identifications. Second, putatively cryptic taxa are common in many groups ([Knowlton, 1993](#)). Third, morphological keys are

often effective at a particular stage of life or sex only, thus many individuals cannot be identified. Finally, the use of keys often requires a high level of expertise and there are cases where they can still be misinterpreted.

Limitations associated with morphological species determination led to the need for a new approach in the identification of organisms. Problems with morphological taxonomy were enough to start developing DNA-based identification systems. The number of cases in which such systems are used is continually increasing ([Brown et al., 1999](#)). The genomic approach to taxonomic diagnosis uses diversity within DNA sequences to identify organisms ([Kurtzman, 1994](#); [Wilson, 1995](#)). This paper describes the markers used to identify plant organisms, their properties, advantages and disadvantages, and the proposed standard DNA barcodes for photosynthetic

* Corresponding author.

E-mail address: zuzannakowalska0@gmail.com (Z. Kowalska).

organisms. This work focuses on microalgae and macroalgae, which are of great ecological importance in aquatic environments and are widely used in food, pharmaceutical or energy industries.

2. Barcoding DNA as a tool for describing biodiversity

In 2003 Hebert proposed “barcoding DNA” as a way of identifying species. In barcoding a short genetic sequence from the standard part of the genome is used to mark an organism. Barcoding DNA helps to solve the limitations of traditional species identification. This does not mean that traditional taxonomy has become less important. On the contrary, DNA barcoding has become for taxonomists a new tool complementing their knowledge and an innovative device for non-experts who need to quickly identify the organism.

Species identification by barcode DNA starts from the specimen. Samples can be taken from various sources, e.g. strains from collections of various institutions, sediments or water samples. In the laboratory, small fragments of thallus or just a several cells of microorganisms are used to isolate DNA. The barcode region is amplified by polymerase chain reaction (PCR) and then sequenced. The obtained sequence is a representation of a unique species-specific series of four nucleotides, i.e. adenine, cytosine, guanine, thymine denoted by a series of letters ACGT, respectively (Fig. 1).

The use of DNA barcodes to describe life has not been without criticism. Mallet and Willmott (2003), for example, were concerned that differences in DNA sequences between closely related species were not sufficient to enable their identification. Other researchers worried that with barcoding DNA (potentially) replacing traditional taxonomic practice new species could never be formally

described (Robba et al., 2006). The DNA barcoding posed a number of other problems: a serious limitation of the usefulness of barcoding as a molecular diagnostics is the human error and uncertainty in the creation of reference libraries (Collins and Cruickshank, 2013). The idea of creating such libraries is the ability to identify unknown species using existing sequences of previously marked taxa.

Contrary to these arguments, DNA barcode designs have already achieved positive results and brought benefits to ecologists, conservationists, and diverse agencies committed to control the use of pests, food safety and monitor the occurrence of invasive species (Hebert and Gregory, 2005; Robba et al., 2006). Furthermore, Hebert et al. (2003) also emphasized that DNA barcoding in conjunction with traditional taxonomy is a valuable tool in revealing hidden diversity.

Only a few areas of the genome are considered to be suitable for the use as DNA barcodes. There are several requirements for selecting the appropriate DNA markers. Firstly, low intra-specific nucleotide variation (less than 2%) and high variability among species (Mirek et al., 2007). Secondly, and equally important, the marker should be universally reproducible for different taxa, which means that there should be a high success of amplification and sequencing, preferably with a universal set of PCR primers. Taking into account Sanger sequencing technique, it is also desirable that the DNA barcode tag should be short enough to be sequenced in a single reading (<700 bp). Longer fragments may also be sequenced, but this requires more financial effort, as well as time and work. Markers that may have microsatellite or regions rich in long repeats of or one or two base pairs are not beneficial because they cause problems with direct PCR product sequencing (Pečňikar

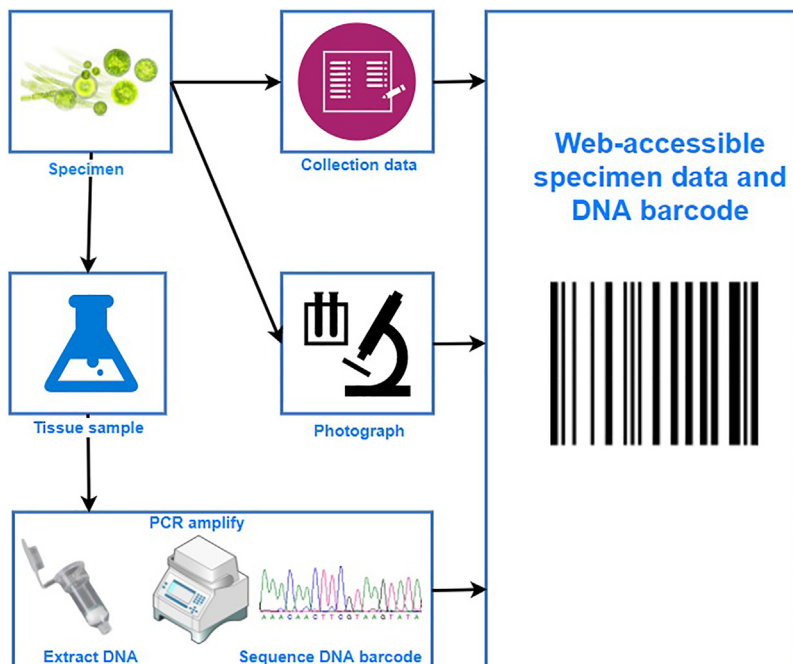


Fig. 1. Application of DNA barcoding technique for new species identification.

and Buzan, 2014). A good marker should contain a small number of insertions and/or deletions. These mutations impede the alignment of sequences obtained from different organisms and in genetic analysis are generally not regarded as diagnostic; nucleotide substitutions are preferred. However, Kress and Erickson (2007) rejected this requirement by stating that DNA barcodes are intended for species identification, not for phylogenetic analyses. Any mutation, whether it is insertion or deletion or substitution, is a feature of the organism, and is equally good if it allows for identification (Mirek et al., 2007). Furthermore, the ideal marker should consist of a highly variable region that provides enough information for species discrimination, and be flanked by highly conserved regions, which allow universal primer design as well as DNA barcodes standardization. It should also appear in the genome in a large number of copies which is of great importance when it comes to working with small thallus fragments or with fossil material. The greater the number of copies of the target sequence in the genome, the greater the chance that some of them have survived intact (Pečnikar and Buzan, 2014).

3. Markers used to identify algae

The standard sequence used for barcoding of almost all animal groups was derived from the gene 5' fragment of cytochrome c oxidase subunit I of mtDNA (Cox1) (Hebert et al., 2003). For fungi Cox1 and nuclear ITS sequences (internal transcribed spacers) were proposed as barcodes. In contrast, the world of photosynthetic organisms does not have a single standard DNA fragment for species identification. It is estimated that in the world there are over 72,500 species of algae, while only 44,000 of them have been described (Guiry, 2012) (Fig. 2). There are still many undiscovered species, so a new system of identification is needed to make the description of algae more efficient.

Currently, intensive research is conducted on the identification of algae and vascular plants. The work continues on the selection of molecular markers that will be successfully used in a wide spectrum of organisms for

DNA barcoding. Plant Working Group deals with the development of a standard for the description of plant DNA barcodes. The standard Cox1 sequence for animals shows insufficient variation in vascular plants (Kress et al., 2005), but can be used for some algal groups; it has been successfully applied in red algae research (Saunders, 2005; Robba et al., 2006; Le Gall and Saunders, 2010; Saunders and Kucera, 2010), while for green algae it has shown insufficient amplification outcome (Hall et al., 2010; Saunders and Kucera, 2010). The mitochondrial DNA (mtDNA) of plants is also not suitable source of barcodes because of frequent rearrangements and gene replacements with the nuclear genome (Hebert et al., 2003). The genes of nuclear DNA (nDNA), which are being present in very low copy number, may cause difficulties in sequencing of degraded DNA. Regarding the above mentioned reasons, researchers began to use chloroplast DNA (cpDNA), which is present in plant cells in multiple copies. This type of DNA consists of both low variability coding and much more variable non-coding parts. At the current state of knowledge, it is not possible to indicate the ideal DNA barcode sequence in any of the plant genomes (nDNA, cpDNA, mtDNA), thus it is necessary to use several less variable sequences forming a DNA barcode (Kress et al., 2005).

One universal sequence which could enable establishing of a single-sequence DNA barcode system seems to be not possible. It is also argued that the use of several target sequences may not be sufficient for the identification of all photosynthetic organisms. Therefore, it is expected that in the case of particular taxonomic groups occupying different phylogenetic positions, DNA barcoding will operate on different standard sequences (Mirek et al., 2007).

In 2006, Newmaster and his colleagues proposed to identify organisms based on two sequences. The first sequence would allow classifying an organism into a systematically higher unit, while the second one would enable species identification within the higher systematic unit to which an object has been included by means of the "primary" sequence. Kress and Erickson (2007) further developed this idea by evaluating sequences that could

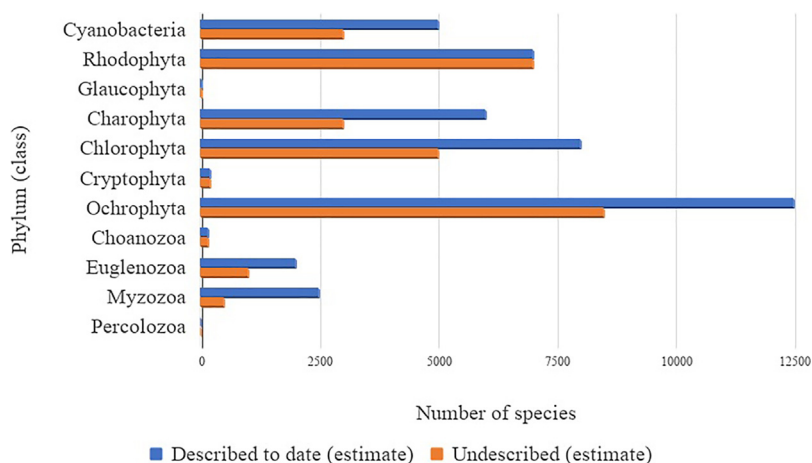


Fig. 2. The number of described species of algae and the estimated number of undescribed species. Based on Guiry (2012).

potentially be used as the barcode of plant DNA through the prism of two criteria (universality of amplification and the ability to differentiate genres). Regarding macroalgae Cox1 DNA region was proposed as the main “primary” marker for brown and red algae, whereas *tufA* (elongation factor) as the main marker of green algae (Kress and Erickson, 2012). As the “secondary” marker LSU D2/D3 (region of the large ribosomal subunit) was chosen as it has been proven to be effective in estimating the phylogenetic relationship of organisms at interspecific and intraspecific level (Kress and Erickson, 2007, 2012; Armeli Minicante et al., 2014).

So far, most of the molecular research on algae has used a fairly limited range of markers such as *rbcL* (plastid Rubisco large subunit), ITS, mitochondrial *Cox2-3* or ribosomal and nuclear DNA (Provan et al., 2004; Chase et al., 2007). The lack of available markers was particularly problematic in the study of species variations and overall there is still a need for universal markers that can be used for a wide range of species (Provan et al., 2004; Mirek et al., 2007).

4. Sequencing methods

Initially, in order to identify the organism using the DNA barcoding, sequencing of the obtained material was performed using the Sanger method (1977). In the last decade, various studies have been carried out to develop alternative approaches for biodiversity assessment using methods based on DNA sequences, such as molecular phylogenetics and DNA barcoding (Hajibabaei et al., 2011). Continuous development and linking of DNA reference libraries will provide a database for a DNA-based identification system. The application of new generation sequencing (NGS) methods in biodiversity science may further extend the use of DNA information for biomonitoring at an unprecedented scale. NGS technologies are characterized by enormous bandwidth, which significantly reduces the cost of reading the nucleotide sequence in comparison with the standard Sanger method (Shokralla et al., 2012; Heather and Chain, 2016). Analysis of eDNA (environmental, e-environment) using DNA barcodes is a key application for NGS sequencing in ecological research. Parallel access to huge amounts of sequences, as well as subsequent improvements in reading the length and bandwidth of various sequential platforms, have led to a wider representation biodiversity of the environmental samples tested. The latest achievements of nanotechnology allow direct reading of sequences from a single-molecule sequencing (SMS), without the need for its amplification – third generation sequencing (Mikheyev and Tin, 2014). Oxford NanoPore Minion is the first portable NGS platform. This pocket-sized device allows to obtain sequences anywhere, requiring only a connection to a laptop. Other benefits include low cost and real-time data production (Jain et al., 2016). Portable genomics have enormous potential and can allow DNA barcoding in the field (Hollingsworth et al., 2016). New technologies are growing rapidly and have the potential to radically extend environmental research and thus our knowledge of the environmental (Taberler et al., 2012).

5. Reference databases

One of the most important elements of the DNA barcoding initiative is to establish a public reference library containing taxa identifiers that can be used to assign unknown individuals to known species. Currently, there are two main databases containing DNA barcoding sequences that fulfill this role, i.e. BOLD and GenBank. However, regarding reference sequences for algae, and more specifically – diatoms, one of the most important databases is R-Syst:: diatom.

The GenBank reference database contains more than 200 million sequences from different organisms, collected mainly through the direct deposition/submission of sequences by individual laboratories originating from large-scale sequencing projects (including those related to whole genomes and those from nucleotide Archive and Japan Data Bank makes the content of the database available worldwide (Costa and Carvalho, 2007; Benson et al., 2018). GenBank can be accessed via the NCBI database (National environmental sampling) (Benson et al., 2018). Daily data exchange with the European NucleCenter for Biotechnology Information), which integrates data from all major DNA and protein databases, including taxonomy, genomes, mapping and protein structure information (Pruitt et al., 2008) as well as biomedical literature via PubMed. The BLAST algorithm provides search for sequence similarity in GenBank and other databases. NCBI also offers a wide range of search and analysis services based on GenBank data (Sharma et al., 2018). The GenBank database and associated resources are freely available on the NCBI home page (Benson et al., 2018). Every year, more and more reference sequences are coming up, as demonstrated by the Benson study (1994–2018). Data from 1994 indicates that the GenBank database contained 182,753 sequences, while the data from 2018 show a significant increase to more than 200 million sequences. The more information is gathered the faster and more credible the interpretation of the obtained results will be.

Information systems such as BOLD (Barcode of Life Database) support the acquisition, storage, analysis and publication of DNA barcode records (Strugnell and Lindgren, 2007). BOLD is mining DNA barcodes from GenBank. This program provides an integrated bioinformatic platform that supports all stages of the analytical path from sampling to a strictly verified barcode library (Stoeckle and Hebert, 2008). BOLD is freely available to any researcher interested in DNA barcoding and contains more than 6 million DNA barcode sequences. Providing specialized services it facilitates the compilation of records that meet the standards required to obtain a DNA barcode in global sequence databases (Costa and Carvalho, 2007). Owing to its online data delivery system, BOLD can support projects involving extensive research initiatives (Ratnasingham and Hebert, 2007).

Reference barcode database for diatoms R-Syst:: diatom developed within the framework of R-Syst, the network of systematic supported by INRA (French National Institute for Agricultural Research) (Rimet et al., 2016). R-Syst:: diatom is used to identify barcodes from natural

samples. The data comes from two sources – a culture collection of freshwater algae maintained in INRA, in which new strains are regularly deposited and codified, and from the NCBI nucleotide database. Two molecular markers were selected for database support: 18S and rbcL due to their effectiveness (Rimet et al., 2016; Vasselon et al., 2017). Data are developed using innovative and classic bioinformatic tools (BLAST, classic phylogenesis) and current taxonomy (Rimet et al., 2016).

6. DNA-qua-Net project

In the fast changing world and progressing destruction of the natural environment protection, preservation and restoration of aquatic ecosystems and their functions are of crucial importance. For European countries, those actions have become legally binding, mainly through the EU-Water Framework Directive (WFD, 2000/60/EC) and the Marine Strategy Framework Directive (2008/56/EC) (Leese et al., 2016). The current approach to the assessment and monitoring of biodiversity is based on morphological taxonomy. This approach has many disadvantages, but the key is the possibility of making an error by an analyst (Leese et al., 2018). New genomic tools may bypass imperfections of morphological identifications and be complementary to traditional biomonitoring methods. However, there are many various approaches that are independently developed in different institutions preventing their universal application and for that reason the DNA-qua-Net project was created. The aim of this project is to determine gold-standard genomic tools and new eco-genomic indicators that can be utilized in the assessment of biodiversity and in biomonitoring of European water reservoirs. Researchers have been working on the development and use of markers among else for algae, which are a very important group used in bioindication and biomonitoring of the aquatic environment. Furthermore, this project is also to provide a platform for training of future researchers. Together with water managers, politicians and other interested parties a conceptual framework for the use of eco-genomic tools is

being developed for their subsequent implementation as a part of legally binding assessments (Leese et al., 2016; Lefrançois et al., 2018).

7. Problems with algae identification

Microalgae include very diverse groups of organisms, from single-celled forms to complex multicellular forms, occurring commonly in the marine environment around the world. Microalgae are among the main primary producers in the marine environment (Robba et al., 2006). New species are being discovered every year, and molecular research has discovered many cryptic species and unexpected phylogenetic connections (Hall et al., 2010). Identification of microalgae is problematic due to the fact that many species do not have obvious structural features, and some traits may differ between the same taxa. Typically, taxonomic diagnosis was made by electron microscopy (Hall et al., 2010). However, many groups contain taxa that have not yet been described, and the exact number of species remains undetermined. This is because they are extremely difficult to identify and classify solely on the basis of morphological characteristics. Their morphology can be extremely variable within and between species, and characteristic features that allow for easy identification are often lacking (Provan et al., 2004). In addition, convergent morphology is common and often conceals mysterious species that have only just been revealed on the basis of molecular data (Mirek et al., 2007). In the case of microalgae, barcoding DNA has been tested on several molecular markers. Each of them has its own advantages and disadvantages (Table 1).

Due to the use of diatoms in the bioindication of the environment and biological research, or industry where it is necessary to determine their exact taxonomy, they are a very important group in the aquatic environment. In the case of diatoms, barcoding DNA has been tested on several molecular markers (e.g. Evans et al., 2007; Moniz and Kaczmarska, 2010; Pniewski et al., 2010; Zimmermann et al., 2011). There are three main problems here. First of

Table 1
Markers used for microalgae barcoding DNA and their properties.

Marker	Features of the marker	References
Cox1	Useful in many systematic and phytogeographic studies of red algae Ineffective in the amplification (green algae) Variable enough to reveal potentially cryptic species and works in some cases as a geographical marker (diatom) Problem of the universality	Provan et al. (2004) Moniz and Kaczmarska (2009) Evans et al. (2007); Mann et al. (2010)
LSU	Insufficiently variable to differentiate green algae Insufficiently variable to differentiate diatoms	Evans et al. (2007); Mann et al. (2010) Hall et al. (2010)
V4	Sufficient variability to differentiate diatoms in environmental trials High amplification success	Moniz and Kaczmarska (2009) Zimmermann et al. (2011)
UPA	Great success of amplification and sequencing Insufficiently variable (green algae)	Sherwood et al. (2010) Hall et al. (2010)
ITS	Possible discovery of the bacterial sequence Highly variable Chosen for intra-population diatom studies Direct sequencing of ITS impossible in several species	Del Campo et al. (2010) Mann et al. (2010)
rbcL	Sufficiently variable to differentiate the strains of green algae within species Not working for species, without functional plastids (obligatory heterotrophs) Sufficiently variable to differentiate the strains of green algae within species	Hall et al. (2010) Mann et al. (2010) Hall et al. (2010)

all, the diversity of diatoms is unknown (Mann and Vanormelingen, 2013; Rimet et al., 2016). Secondly, identification has become increasingly difficult because existing taxonomic aids are obsolete by new discoveries. Thirdly, contrary to the long-standing belief that all taxa are cosmopolitan (Vyverman et al., 2007; Kooistra et al., 2008), they may have biogeography. The traditional taxonomic diagnosis of diatoms was based on their morphological characteristics, but exclusive reliance on these traits may lead to erroneous determination of the taxon due to the widespread cryptic species in this group (Evans et al., 2007).

Identifying macroalgae is also very difficult due to the fact that macroalgae show a very simple morphology and a certain degree of phenotypic plasticity, which is influenced by environmental conditions, age of the thallus and life style (Armeli Minicante et al., 2014). Macroalgae were differentiated mainly based on gametangia positions and the mezosporous membrane ornamentation (Hall et al., 2010). Moreover, the occurrence of cryptic taxa makes identification difficult only on the basis of morphological features. Therefore, communities from the Mediterranean were largely underestimated due to the presence of genetically different but morphologically similar species. For this reason, taxonomists are increasingly using molecular tools to identify individuals, discover species and other related taxonomic tasks (Armeli Minicante et al., 2014). In case of macroalgae barcoding DNA has been tested on several molecular markers (Table 2).

8. Cox1 and Cox2-3

For algae, the mitochondrial intergenic spacer between the cytochrome oxidase subunits 2 and 3 (Cox2-3) amplified by primers designed by Zuccarello et al. (1999) is a potentially universal marker of non-coding DNA. The 300–350 bp sequence was found to be useful in many systematic and phytogeographic studies of red algae (Provan et al., 2004). However, the success of sequencing

and amplification of the Cox1 marker in comparison to the others used for the molecular identification of red algae turned out to be fairly small (Sherwood et al., 2014). In the case of biodiversity research, where the red algae strains must be successfully amplified and sequenced, the Cox1 marker should be supplemented with additional markers.

As for the usefulness of the Cox1 marker for the identification of green algae, its amplification has not been effective in most of taxa. The marker is difficult to amplify in green algae due to the fact that this gene contains several introns which exact number and positions are not known in many species (Moniz and Kaczmarska, 2009).

For diatoms Cox1 is variable enough to distinguish between previously identified taxa, to reveal potentially cryptic species and it may work in some cases as a geographical marker (Evans et al., 2007; Mann et al., 2010). For some species Cox1 can provide useful information on the structure of the population or racial differences. According to Evans et al. (2007), Cox1 has a higher degree of variation than other markers used in research on diatom biodiversity (rbcL, ITS, 18S) on the other hand, the evolution of the mitochondrial genomes of algae is unclear (Evans et al., 2007). It can also pose major problems in amplification and sequencing (Moniz and Kaczmarska, 2009; Mann et al., 2010). Therefore, it seems that Cox1 may be valuable only in special cases, such as for example *Sellaphora*, unless the problem of the universality of this marker is solved in the future.

The main disadvantage of Cox1 is the current absence of universal amplification primers. Original primers for algae are successful in some groups, but show mixed results in other lines, which may result from heterogeneity within the species at positions near the 3' end of the primers (Clarkston and Saunders, 2010). The sequence of Cox1 due to the lack of sufficient information for different groups of macroalgae in a data bank requires additional use of another marker to assign the organisms to the specific species (Manghisi et al., 2010). Studies have confirmed the effectiveness of the Cox1 marker in the differentiation of

Table 2
Markers used for macroalgae barcoding DNA and their properties.

Marker	Features of the marker	References
LSU	Poor amplification efficiency for green algae Good results in terms of universality, ease of amplification and discriminating power	Saunders and Kucera (2010) Kress and Erickson (2012)
tufA	Universality and ease of amplification Pose challenges in some lines where the gene is encoded in the nucleus Great power to differentiate closely related species, satisfactory success of the amplification	Hall et al. (2010) Hall et al. (2010) Armeli Minicante et al. (2014)
Cox1	Effective in differentiating among closely related species belonging to different families of brown and red algae Absence of universal amplification primers Satisfactory success of the amplification	McDevit and Saunders (2009); Saunders (2005) Clarkston and Saunders (2010) Armeli Minicante et al. (2014)
UPA	Non-differentiating closely related species Use to evaluate the distribution of macrophytes Great success of amplification and sequencing	Clarkston and Saunders (2010) Wallace et al. (2015) Sherwood et al. (2010)
ITS	Low success of sequencing and amplification for green algae Difficulties with alignment	Saunders and Kucera (2010) Hall et al. (2010)
rbcL	Presence of introns some marine macroalgae negatively affect the universality High degree of universality, low levels of contamination and high sequence quality Low level of sequence discrepancy between closely related taxa More useful in explaining phylogeny than in DNA barcoding	Saunders and Kucera (2010) Saunders and Kucera (2010) Freshwater et al. (1994) Poong et al. (2014)

closely related species belonging to different families of brown and red algae (McDevit and Saunders, 2009; Saunders, 2005).

9. *tufA*

The *Cox1* and *tufA* markers are often used to identify macroalgae, and they can be very successful in delineation of new species as shown by Armeli Minicante et al. (2014) who described two new species *Ulva ohnoi* and *Pyropia yezoensis* occurring on the Sicilian coast. Both markers showed great power to differentiate closely related species, and the success of the amplification was satisfactory.

The *tufA* marker is characterized by high universality and ease of amplification (Hall et al., 2010). It is moderately variable and it may pose challenges in some lines where the gene is encoded in the nucleus (Hall et al., 2010). The *tufA* gene was probably transferred to the nucleus after separation of the main classes of green algae, as they all seem to have at least the remainder of the gene in their cpDNA, but before the emergence of terrestrial plants which are lacking *tufA* in their chloroplast genome. If this is the case, then the transfer must have occurred in the line leading to terrestrial plants, Charophyceae (Baldauf et al., 1990). It may result in the appearance of paralogs. Paralogs are genes created by duplication within a genome and may evolve new functions which may complicate the identification of species in some lines (Hall et al., 2010). Nevertheless, the marker appears to be promising at the species level in chlorophytes (Saunders and Kucera, 2010).

10. Ribosomal RNA genes

There is no doubt that molecular techniques have radically changed the face of phycological research, especially in the field of biodiversity and systematics (Mirek et al., 2007). In the early years, the researchers concentrated strongly on two commonly used markers, i.e. *rbcl* (Freshwater et al., 1994; Chase et al., 2007) and a small rDNA subunit (Ragan et al., 1994; Harper and Saunders, 2001). Sequencing of ribosomal RNA genes (rDNA) has been used as a standard in microbial ecology to describe the structure and diversity of their population. The most commonly used gene is the small ribosomal subunit (SSU), 16S for bacteria and 18S for eukaryotes. Unfortunately, the slow evolution of 18S does not allow distinguishing of closely related species (Evans et al., 2007). However, the large ribosomal subunit (LSU) gene (23S for bacteria and 28S for eukaryotes) may also be used to determine phylogeny and taxonomy of studied taxa (Gutell et al., 1994; Harper and Saunders, 2001; Sherwood and Presting, 2007). LSU genes potentially encode more phylogenetic information due to greater length and sequence variability (Gutell et al., 1994). Another advantage of LSU sequences is its greater database representation (about 2:1 compared to SSU genes) (Gutell et al., 1994). The tested LSU markers have shown great success in amplification and sequencing in the case of red algae, but due to the more conservative nature they are not sufficient to distinguish closely related

species, which is crucial for barcoding DNA (Sherwood et al., 2010).

Hall in his research (2010) tested the utility of various markers in the identification of algae from the Chlorophyceae cluster (including the plastid 23S rDNA [UPA], also 18S, *rbcl*, ITS1, ITS2 and *Cox1*). These studies suggest that 18S is not the best choice when it comes to creating a DNA barcode because the region is not sufficiently variable. Similar conclusions were drawn by Moniz and Kaczmarska (2009) who investigated the possibility of using 18S as a barcode in diatoms. 18S is less variable than *rbcl* and *Cox1* and is relatively rarely used as a diatom DNA barcode (Mann et al., 2010). However, according to Zimmermann et al. (2011), the subregion of 18S rDNA-V4 turned out to be a good candidate for a diatom DNA barcoding. It exhibits sufficient variability to differentiate organisms in environmental trials and shows very high amplification success. However, in the case of cryptic species, it still shows insufficient variability, therefore it has been postulated that it should be supplemented with additional markers (Zimmermann et al., 2011).

In case of macroalgae the LSU D2/D3 region gives very good results in terms of universality, ease of amplification and discriminating power. The resulting sequence quality is satisfactory (Kress and Erickson, 2012). However, when it comes to green algae, PCR amplification caused problems (Saunders and Kucera, 2010). These problems could result from the high frequency of introns within the green algae. In addition, the presence and amplification of contaminations in the samples is an additional explanation of the encountered difficulties (Van Oppen et al., 1993; Bhattacharya et al., 1996; Saunders and Kucera, 2010). Further development of primers can solve the problem with poor amplification efficiency (Saunders and Kucera, 2010).

11. UPA

A universal DNA barcode for the identification of photosynthetic organisms was recently proposed. Sherwood and Presting (2007) used the plastid 23S rDNA gene (UPA – universal amplicon) as a universal algal marker, claiming that a single pair of primers could discover sequences from a wide taxonomic range of photosynthetic organisms, including higher plants, diatoms, green algae, brown algae, red algae and even cyanobacteria. This marker has been used in biodiversity research of natural algal communities, both benthic and planktonic, and the distribution of species. Indeed, in 2014, Sherwood and colleagues based on the 23S rDNA markers conducted research on the identification of algae in the Koolau mountain range in Hawaii. The marker has also been successfully used to evaluate the distribution of algae and macrophytes (Wallace et al., 2015). In addition, the tested marker was also used to identify species present in fossil sludge, also with good results (Hou et al., 2014). The UPA marker showed great success of amplification and sequencing in the case of red algae also, but due to the more conservative nature it was not sufficient to distinguish closely related species, which is crucial for barcoding DNA (Sherwood et al., 2010). The UPA marker was also used in studies on the spread of species in the coast region

of Brazil. The review of species, based on the analysis of molecular marker sequences and supported with morphology, revealed the occurrence of five different taxa in the *Pyropia* genus (Milstein et al., 2015).

The specificity of UPA universal primers is undoubtedly influenced by the origin of plastids (Sherwood and Presting, 2007). This marker has a very high amplification success for different groups of micro and macroalgae. Due to its universality, it is more often used in research on plant communities, algae used in industry or bioindication (Hou et al., 2014; Milstein et al., 2015; Wallace et al., 2015). Although UPA had a relatively high success in PCR amplification, it turned out that some of the samples were contaminated or the resulting sequences were not of satisfactory quality. It is possible that the epiphytic or endophytic contaminations present in the samples were preferentially or simultaneously amplified with the target organism for one of the following reasons: the UPA primers matched the contaminations better than the target organism or the DNA extraction was more effective for contaminations than for the target organism. Markers with highly universal primers, such as UPA, are particularly susceptible to these problems and effective isolation of target DNA becomes extremely important (Saunders and Kucera, 2010). However, the possibility remains that the UPA may not have a differentiating power among closely related species, which may lead to underestimation of diversity, and although this has not been observed yet, this is an attribute that requires further research for UPA as a potential marker for DNA barcoding (Clarkston and Saunders, 2010).

The usability of the 23S rDNA gene is potentially limited due to the relatively small amount of information in the databases. Their constraints are likely to fall sharply with increasing genomic sequencing, particularly given the interest in using algae for biofuel production (Grossman, 2007). Therefore, the ability to use 23S rDNA gene sequencing to describe complex algal groups should grow rapidly. This increase will occur with the development of descriptions of genetic diversity of algae.

12. ITS

Most phytogeographic studies used internal transcribed spacer (ITS), a non-coding domain located between the small and large ribosome subunit. The popularity of the ITS region can be attributed to its relatively high nucleotide substitution rate, which allows comparisons of relatively recent divergent taxa. In addition, the ITS region can be readily amplified by PCR and sequenced with conservative primers. However, many earlier studies used a set of primers published by White et al. (1990) for fungi that frequently amplified false algae sequences (Provan et al., 2004). Plant Working Group excluded this sequence as a standard DNA barcode in view of possibility of multiple copies of this domain in one organism and the uncertainty in its evolution (Mirek et al., 2007).

ITS was also proposed as a barcode for algae and terrestrial plants and was widely used in phylogenetics of green algae species. The ITS1 and ITS2 markers are sufficiently variable to differentiate the strains of algae

within species (Evans et al., 2007; Hall et al., 2010; Mann et al., 2010; Moniz and Kaczmarska, 2010). The success of sequencing and amplification is low for the ITS marker in green algae, while for other taxonomic groups the universality of primers has been more successful.

In contrast to the 18S, ITS is highly variable and is often chosen for intra-population diatom studies. However, direct sequencing of ITS was not possible in several species, including *Nitzschia*, *Achnanthes*, *Eunotia*, *Navicula*, *Pseudo-nitzschia* and *Sellaphora* (Mann et al., 2010). Furthermore, matching the rDNA sequence is not straightforward and becomes increasingly difficult, because there are many copies of ITS sequences in plant genomes, it is questionable whether the sequence obtained by PCR would be stable and representative (Gao et al., 2010). The main problem is that ITS alignment can be reached with insertion/deletions therefore the alignment of variably sized fragments can be highly biased. Multiple or banded PCR products are often obtained, which makes sequencing impossible. It is likely that these primers strengthen contaminating organisms, leading to a multiband or banded PCR product (Saunders and Kucera, 2010). However, when the DNA barcodes are used to discover the species, and for identification, which is desirable given the current state of the taxonomy of diatoms, the question of sequences alignment is an important practical matter (Mann et al., 2010).

13. rbcL

Region of genes between the large and small subunits of RuBisCO (rbcL–rbcS spacer) is widely used in molecular studies of algae (red, green and brown algae) (Brodie et al., 1998). The protein coding for the plastid gene rbcL was proposed as a potential barcode of DNA for plant organisms by many researchers usually combined with one or more other markers (Chase et al., 2007). One of the advantages of this region is the rich library of information (Chase et al., 2007). This sequence allows for a high probability of placing the examined organism into the genus or family (Kress and Erickson, 2007).

The rbcL marker is sufficiently variable to differentiate the strains of green algae within species. However, according to Hall et al. (2010) it is not enough to identify all tested strains. Furthermore, if rbcL is chosen as the preferred marker used in diatom DNA barcoding (or any other plastid marker whose function is to code the proteins involved in photosynthesis), it will not work for species that lack a functional plastid and are obligatory heterotrophs such as: *Nitzschia alba* (Mann et al., 2010).

The usefulness of rbcL as a barcode DNA was the basis for several taxonomic and phylogenetic studies in marine macroalgae (Hommersand et al., 1994; Wang et al., 2010; Saunders and Kucera, 2010; Du et al., 2014). Unfortunately, the presence of introns in the rbcL of some marine species may negatively affect the universality of rbcL as a barcode marker, because obtaining the whole fragment of 1300 bp using only both way sequencing may be difficult and use of internal primer may be necessary (Saunders and Kucera, 2010; Du et al., 2014). Considering the fact that green algae

are particularly susceptible to the acquisition of intron sequences (Van Oppen et al., 1993; Bhattacharya et al., 1996) and that the extent to which introns are present in their genes is unknown. In spite of that, according to Saunders and Kucera (2010), *rbcl* is still a good candidate for barcode DNA of green algae. Due to the high degree of universality and low levels of contamination and high sequence quality in comparison with other tested markers.

In the case of red algae, *rbcl* shows a low level of sequence discrepancy between closely related taxa. Furthermore, the lack of insertion or deletion mutations eliminates the problems of alignment in the case of red algae (Freshwater and Rueness, 1994). According to Poong et al. (2014), the *rbcl* loci are more useful in explaining phylogeny than in DNA barcoding.

There is no one specific sequence that would be suitable as a standard DNA barcode for algae. Nevertheless, there are some promising barcode sequences candidates, as those mentioned above which have proven to be useful in many cases.

14. Conclusion

The development of DNA barcoding has helped taxonomic research and is of great value because the molecular assessment of the diversity of organisms plays an increasingly important role in controlling and detecting invasive species, identifying organisms for environment monitoring, testing commercial algae products and characterizing organisms (Hebert and Gregory, 2005; Sherwood and Presting, 2007).

There is no doubt that algae DNA barcodes will improve as PCR amplification and DNA sequencing progresses. Sequencing technology has improved dramatically over the past 25 years, and more recently sequencing systems have been available to produce large quantities of DNA sequences in a very short period of time. These techniques are very suitable for DNA barcoding of environmental samples consisting of a mixtures of multiple species, such as sediment, soil or seawater samples. Krishnamurthy and Francis (2012) point out that establishing robust species identification thresholds is a key element of the DNA barcoding process because it is only after determining the threshold in the target group that cryptic species can be identified. Today it is known that no single classification technique can be widely used to identify algal species. However, in cases where a single DNA area is not sufficient for species identification, a combination of two or more DNA regions should be used. New sequencing techniques enable faster and cheaper analysis of DNA barcodes, which in turn will become available in other branches of science, such as medicine or pharmacy. The development of new sequencing techniques made it possible to obtain whole genomes (nuclear, mitochondrial, plastid, whole transcripts) for wide range organisms and these techniques are nowadays used on daily bases generating a lot of data. This data can provide valuable information about more suitable genes/markers for DNA barcoding. The usefulness of these techniques requires the development of the proper barcode libraries.

Conflict of interest

None declared.

Ethical statement

The research was done according to ethical standards.

Funding Body: None.

References

- Armeli Minicante, S., Genovese, G., Morabito, M., 2014. Two new alien macroalgae identified by the DNA barcoding. *Biologia Marina Mediterranea* 21 (1), 70–72.
- Baldauf, S.L., Manhart, J.R., Palmer, J.D., 1990. Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proceedings of the National Academy of Sciences of the United States of America* 87 (14), 5317–5321.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizarachi, I., Ostell, J., Pruitt, K.D., Sayers, E.W., 2018. GenBank. *Nucleic Acids Research* 46 (D 1), D41–D47.
- Bhattacharya, D., Friedl, T., Damberger, S., 1996. Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. *Molecular Biology and Evolution* 13 (7), 978–989.
- Brodie, J., Hayes, P.K., Barker, G.L., Irvine, L.M., Bartsch, I., 1998. A reappraisal of *Porphyra* and *Bangia* (Bangiophyceidae, Rhodophyta) in the Northeast Atlantic based on the *rbcl*–*rbcs* intergenic spacer. *Journal of Phycology* 34 (6), 1069–1074.
- Brown, B., Emberson, R.M., Paterson, A.M., 1999. Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification. *Bulletin of Entomological Research* 89, 287–294.
- Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madriñán, S., et al., 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56 (2), 295–299.
- Clarkston, B.E., Saunders, G.W., 2010. A comparison of two DNA barcode markers for species discrimination in the red algal family Kallymeniaceae (Gigartinales, Florideophyceae), with a description of *Euthora timburtonii* sp. nov. *Botany* 88 (2), 119–131.
- Collins, R.A., Cruickshank, R.H., 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* 13, 969–975.
- Costa, F.O., Carvalho, G.R., 2007. The Barcode of Life Initiative: synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics Society and Policy* 3 (2), 29.
- Del Campo, E.M., del Hoyo, A., Royo, C., Casano, L.M., Álvarez, R., et al., 2010. A single primer pair gives a specific ortholog amplicon in a wide range of Cyanobacteria and plastid-bearing organisms: applicability in inventory of reference material from collections and phylogenetic analysis. *Molecular Phylogenetics and Evolution* 57 (3), 1323–1328.
- Du, G., Wu, F., Mao, Y., Guo, S., Xue, H., et al., 2014. DNA barcoding assessment of green macroalgae in coastal zone around Qingdao, China. *Journal of Ocean University of China* 13 (1), 97–103.
- Evans, K.M., Wortley, A.H., Mann, D.G., 2007. An assessment of potential diatom “barcode” genes (*cox1*, *rbcl*, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist* 158 (3), 349–364.
- Freshwater, D.W., Fredericq, S., Butler, B.S., Hommersand, M.H., Chase, M.W., 1994. A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbcl*. *Proceedings of the National Academy of Sciences of the United States of America* 91 (15), 7281–7285.
- Freshwater, D.W., Rueness, J., 1994. Phylogenetic relationships of some European Gelidium (Gelidiales, Rhodophyta) species, based on *rbcl* nucleotide sequence analysis. *Phycologia* 33 (3), 187–194.
- Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., Ma, X., Chen, S., 2010. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology* 130 (1), 116–121.
- Guiry, M.D., 2012. How many species of algae are there? *Journal of Phycology* 48 (5), 1057–1063.
- Grossman, A., 2007. In the grip of algal genomics. *Advances in Experimental Medicine and Biology* 616, 54–76.
- Gutell, R.R., Larsen, N., Woese, C.R., 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews* 58 (1), 10–26.

- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* 6 (4), e17497.
- Hall, J., Fucikova, K., Lo, C., Lewis, L., Karol, K., 2010. An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie Algologie* 31 (4), 529–555.
- Hammond, P., 1992. Species inventory. Living resources. In: Groombridge, B. (Ed.), *Global Biodiversity*. Chapman & Hall, London, pp. 17–39.
- Harper, J.T., Saunders, G.W., 2001. Molecular systematics of the Florideophyceae (Rhodophyta) using nuclear large and small subunit rDNA sequence data. *Journal of Phycology* 37, 1073–1082.
- Heather, J.M., Chain, B., 2016. The sequence of sequencers: the history of sequencing DNA. *Genomics* 107, 1–8.
- Hebert, P., Cywinska, A., Ball, S., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270 (1512), 313–321.
- Hebert, P.D., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54 (5), 852–859.
- Hollingsworth, P.M., Li, D.Z., van der Bank, M., Twyford, A.D., 2016. Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702), pii 20150338.
- Hommersand, M.H., Fredericq, S., Freshwater, D.W., 1994. Phylogenetic systematics and biogeography of the Gigartinales (Gigartinales, Rhodophyta) based on sequence analysis of rbcL. *Botanica Marina* 37 (3), 193–204.
- Hou, W., Dong, H., Li, G., Yang, J., Coolen, M.J., et al., 2014. Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Scientific Reports* 4, 6648.
- Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17 (1), 239.
- Knowlton, N., 1993. Sibling species in the sea. *Annual Review of Ecology and Systematics* 24, 189–216.
- Kooistra, W.H., Sarno, D., Balzano, S., Gu, H., Andersen, R.A., Zingone, A., 2008. Global diversity and biogeography of *Skeletonema* species (bacillariophyta). *Protist* 159 (2), 177–193.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102 (23), 8369–8374.
- Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS One* 6 (e508), 1–10.
- Kress, W.J., Erickson, D.L., 2012. *DNA Barcodes: Methods and Protocols*. Humana Press, New York.
- Krishnamurthy, P.K., Francis, R.A., 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation* 21 (8), 1901–1919.
- Kurtzman, C.P., 1994. Molecular taxonomy of the yeasts. *Yeast* 10, 1727–1740.
- Le Gall, L., Saunders, G.W., 2010. DNA barcoding is a powerful tool to uncover algal diversity: a case study of the Phylloporaceae (Gigartinales, Rhodophyta) in the Canadian flora. *Journal of Phycology* 46 (2), 374–389.
- Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., et al., 2016. DNAqua-Net: developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe.
- Leese, F., Bouchez, A., Abarenkov, K., Altermatt, F., Borja, A., Bruce, K., et al., 2018. Why we need sustainable networks bridging countries, disciplines, cultures and generations for aquatic biomonitoring 2.0: a perspective derived from the DNAqua-Net COST action. *Advances in Ecological Research*, vol. 58. Academic Press, pp. 63–99.
- Lefrançois, E., Apothéoz-Perret-Gentil, L., Blancher, P., Botreau, S., Charodon, C., Crepin, L., et al., 2018. Development and implementation of eco-genomic tools for aquatic ecosystem biomonitoring: the SYNQUA French-Swiss program. *Environmental Science and Pollution Research* 1–9.
- Mallet, J., Willmott, K., 2003. Taxonomy: renaissance or Tower of Babel? *Trends in Ecology & Evolution* 18 (2), 57–59.
- Manghisi, A., Morabito, M., Bertuccio, C., Le Gall, L., Couloux, A., et al., 2010. Is routine DNA barcoding an efficient tool to reveal introductions of alien macroalgae? A case study of *Agardhiella subulata* (Solieriaceae, Rhodophyta) in Cape Peloro lagoon (Sicily, Italy). *Cryptogamie Algologie* 31 (4), 423.
- Mann, D.G., Sato, S., Trobajo, R., Vanormelingen, P., Souffreau, C., 2010. DNA barcoding for species identification and discovery in diatoms. *Cryptogamie Algologie* 31 (4), 557–577.
- Mann, D.G., Vanormelingen, P., 2013. An inordinate fondness? The number, distributions, and origins of diatom species. *The Journal of Eukaryotic Microbiology* 60, 414–420.
- McDevit, D.C., Saunders, G.W., 2009. On the utility of DNA barcoding for species differentiation among brown macroalgae (Phaeophyceae) including a novel extraction protocol. *Phycological Research* 57 (2), 131–141.
- Mikheyev, A.S., Tin, M.M., 2014. A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources* 14 (6), 1097–1102.
- Milstein, D., Medeiros, A.S., Oliveira, E.C., Oliveira, M.C., 2015. Native or introduced? A re-evaluation of *Pyropia* species (Bangiales, Rhodophyta) from Brazil based on molecular analyses. *European Journal of Phycology* 50 (1), 37–45.
- Mirek, Z., Bierniek, W., Sztorc, A., 2007. Barkoding DNA—nowe narzędzie do opisu bioróżnorodności. *Wiadomości Botaniczne* 51, 41–50.
- Moniz, M.B., Kaczmarska, I., 2009. Barcoding diatoms: is there a good marker? *Molecular Ecology Resources* 9 (s1), 65–74.
- Moniz, M.B., Kaczmarska, I., 2010. Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* 161 (1), 7–34.
- Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany* 84, 335–341.
- Pečnikar, Ž.F., Buzan, E.V., 2014. 20 years since the introduction of DNA barcoding: from theory to application. *Journal of Applied Genetics* 55 (1), 43–52.
- Pniewski, F.F., Friedl, T., Latała, A., 2010. Identification of diatom isolates from the Gulf of Gdansk: testing of species identifications using morphology, 18S rDNA sequencing and DNA barcodes of strains from the Culture Collection of Baltic Algae (CCBA). *Oceanological and Hydrobiological Studies* 39, 3–20.
- Poong, S.W., Lim, P.E., Phang, S.M., Sunarpi, H., West, J.A., et al., 2014. A molecular-assisted floristic survey of crustose brown algae (Phaeophyceae) from Malaysia and Lombok Island, Indonesia based on rbcL and partial cox1 genes. *Journal of Applied Phycology* 26 (2), 1231–1242.
- Provan, J., Murphy, S., Maggs, C.A., 2004. Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology* 39 (1), 43–50.
- Pruitt, K.D., Tatusova, T., Klimke, W., Maglott, D.R., 2008. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Research* 37 (Suppl. 1), D32–D36.
- Ragan, M.A., Bird, C.J., Rice, E.L., Gutell, R.R., Murphy, C.A., et al., 1994. A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. *Proceedings of the National Academy of Sciences of the United States of America* 91 (15), 7276–7280.
- Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Resources* 7 (3), 355–364.
- Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., et al., 2016. R-Syst: diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. *Database*, 2016.
- Robba, L., Russell, S.J., Barker, G.L., Brodie, J., 2006. Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta). *American Journal of Botany* 93 (8), 1101–1108.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74 (12), 5463–5467.
- Saunders, G., 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 360 (1462), 1879–1888.
- Saunders, G.W., Kucera, H., 2010. An evaluation of rbcL, tufA, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Cryptogamie Algologie* 31 (4), 487.
- Sharma, S., Ciuffo, S., Starchenko, E., Darji, D., Chlumsky, L., Karsch-Mizarchi, I., Schoch, C.L., 2018. The NCBI BioCollections Database. *Database*, 2018.
- Sherwood, A.R., Presting, G.G., 2007. Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. *Journal of Phycology* 43 (3), 605–608.
- Sherwood, A.R., Sauvage, T., Kurihara, A., Conklin, K.Y., Presting, G.G., 2010. A comparative analysis of COI, LSU and UPA marker data for the Hawaiian florideophyte Rhodophyta: implications for DNA barcoding of red algae. *Cryptogamie Algologie* 31 (4), 451.
- Sherwood, A.R., Conklin, K.Y., Liddy, Z.J., 2014. What's in the air? Preliminary analyses of Hawaiian airborne algae and land plant spores reveal a diverse and abundant flora. *Phycologia* 53 (6), 579–582.
- Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21 (8), 1794–1805.

- Stoeckle, M.Y., Hebert, P.D., 2008. Barcode of life. *Scientific American* 299 (4), 82–89.
- Strugnell, J.M., Lindgren, A.R., 2007. A barcode of life database for the Cephalopoda? Considerations and concerns. *Reviews in Fish Biology and Fisheries* 17 (2–3), 337–344.
- Taberler, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. *Molecular Ecology* 21 (8), 1789–1793.
- Van Oppen, M.J., Olsen, J.L., Stam, W.T., 1993. Evidence for independent acquisition of group I introns in green algae. *Molecular Biology and Evolution* 10 (6), 1317–1326.
- Vasselon, V., Rimet, F., Tapolczai, K., Bouchez, A., 2017. Assessing ecological status with diatoms DNA metabarcoding: scaling-up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators* 82, 1–12.
- Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, K., Sterken, M., Hodgson, D.A., Flower, R., 2007. Historical processes constrain patterns in global diatom diversity. *Ecology* 88 (8), 1924–1931.
- Wallace, J., Champagne, P., Hall, G., Yin, Z., Liu, X., 2015. Determination of algae and macrophyte species distribution in three wastewater stabilization ponds using metagenomics analysis. *Water* 7 (7), 3225–3242.
- Wang, J., Jiang, P., Cui, Y., Li, N., Wang, M., et al., 2010. Molecular analysis of green-tide-forming macroalgae in the Yellow Sea. *Aquatic Botany* 93 (1), 25–31.
- White, T.J., Bruns, T., Lee, S.J.W.T., Taylor, J.L., 1990. Amplification and direct sequencing of fungal ribosomal RNA GENES for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* Academic Press, New York.
- Wilson, K.H., 1995. Molecular biology as a tool for taxonomy. *Clinical Infectious Diseases* 20, 192–208.
- Zimmermann, J., Jahn, R., Gemeinholzer, B., 2011. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution* 11 (3), 173.
- Zuccarello, G.C., Burger, G., West, J.A., King, R.J., 1999. A mitochondrial marker for red algal intraspecific relationships. *Molecular Ecology* 8, 1443–1447.

Further reading

- Benson D.A., Boguski M., Lipman D.J. and Ostell J., GenBank, *Nucleic Acids Research* 22 (17), 1994, 3441.