Environ Sci Pollut Res
DOI 10.1007/s11356-012-1051-5

GPR-Analyzer: a simple tool for quantitative analysis of hierarchical multispecies microarrays

Simon M. Dittami · Bente Edvardsen

Received: 14 March 2012 / Accepted: 21 June 2012
© Springer-Verlag 2012

Abstract Monitoring of marine microalgae is important to predict and manage harmful algae blooms. It currently relies mainly on light-microscopic identification and enumeration of algal cells, yet several molecular tools are currently being developed to complement traditional methods. MIcroarray Detection of Toxic ALgae (MIDTAL) is an FP7-funded EU project aiming to establish a hierarchical multispecies microarray as one of these tools. Prototype arrays are currently being tested with field samples, yet the analysis of the large quantities of data generated by these arrays presents a challenge as suitable analysis tools or protocols are scarce. This paper proposes a two-part protocol for the analysis of the MIDTAL and other hierarchical multispecies arrays: Signal-to-noise ratios can be used to determine the presence or absence of signals and to identify potential false-positives considering parallel and hierarchical probes. In addition, normalized total signal intensities are recommended for comparisons between microarrays and in order to relate signals for specific probes to cell concentrations using external calibration curves. Hybridization- and probe-specific detection limits can be calculated to help evaluate negative results. The suggested analyses were implemented in "GPR-Analyzer", a platform-independent and graphical user interface-based application, enabling non-specialist users to quickly and quantitatively analyze hierarchical multispecies microarrays. It is available online at http://folk.uio.no/edvardse/gpranalyzer.

Keywords Multispecies DNA microarrays · Hierarchical probes · Phylochips · Cell concentrations · Analysis software

Responsible editor: Philippe Garrigues

S. M. Dittami (✉) · B. Edvardsen
Marine Biology, Department of Biology, University of Oslo,
P.O. Box 1066, Blindern,
0316, Oslo, Norway
e-mail: simon.dittami@bio.uio.no

Introduction

Monitoring of marine microalgae is important to predict harmful algal blooms and mitigate their consequences. Manual identification and enumeration of algal cells using light microscopy is the standard method for this purpose (Humbert et al. 2010), although it is labor-intensive and frequently insufficient to reach definitive conclusions about species and thus toxicity. To complement these efforts molecular methods such as PCR (e.g., Penna et al. 2006; Costas et al. 2007), qPCR (e.g., Galluzzi et al. 2004; Penna et al. 2006), FISH (e.g., Simon et al. 2000; Not et al. 2002), dot blots (e.g., John 2004), sandwich hybridizations (e.g., Scholin et al. 1996; Ayers et al. 2005; Diercks et al. 2008), high throughput sequencing (e.g., Behnke et al. 2011; Massana et al. 2011), and microarrays (Metfies and Medlin 2005; Ahn et al. 2006; Gescher et al. 2008) have been adapted. The basic principle underlying these methods is the detection of species-specific nucleic acid sequences usually located in regions of the ribosome, the presence of which in a sample is then correlated to the presence of the corresponding species.

While most molecular approaches usually are designed for one or a few species, microarrays can target large numbers of species in parallel at reasonable cost. Moreover, because the number of probes on an array is not limiting, they can be constructed in a hierarchical manner, accommodating probes at different phylogenetic levels (e.g., species, genus, phylum), which can be used as an internal control to identify possible false-positives (Lange et al. 1996; Groben et al. 2004; Metfies et al. 2008). Multispecies microarrays have been successfully used for the analysis of microbial communities in a range of different habitats (Loy and Bodrossy 2006; Wagner et al. 2007) from the human body (e.g., Palmer et al. 2006; Huyghe et al. 2008) via soil extracts (e.g., Small et al. 2001) to activated sludge from sewage treatment plants (Loy et al. 2005). In marine environments, Peplies et al.

Published online: 06 July 2012

(2004) and Marcelino et al. (2006) have used DNA microarrays to identify and, in the latter case also quantify different bacterial taxa in field samples. With respect to harmful algae a few pilot studies have already tested microarrays for selected genera (Ahn et al. 2006; Gescher et al. 2008).

MIcroarray Detection of Toxic ALgae (MIDTAL, http://www.midtal.com) is an FP7-funded EU project including scientists from seven European countries and the USA aiming to establish an RNA-based hierarchical multispecies microarray for a large number of species, in order to strengthen the EU's capability of monitoring harmful algae. The project is currently in its final phase: a third-generation prototype array comprising 157 probes and controls targeting the 18S and 28S ribosomal subunits of 44 species has been produced by Scienion AG (Berlin, Germany) and is currently being tested with field samples in the different partner labs. Furthermore, a detailed manual on this chip and the protocols for its use is currently in press (Lewis et al. 2012), and an agreement about the commercialization of the chip is in the final phase of negotiation. The completion of the MIDTAL project will lead to an increasing number of MIDTAL microarray datasets, the analysis of which is complex and time-consuming, especially because the data differs strongly from microarray data generated in gene expression experiments.

"GPR-Analyzer" is a JAVA application designed to guide and simplify the analysis of these hierarchical multispecies data sets, starting with the output files generated by the scanner software (GPR stands for GenePix Result). It provides two main functions: First, like previous programs for the analysis of phylochips such as ChipChecker (Loy et al. 2002) and PhylochipAnalyzer (Metfies et al. 2008), it makes use of hierarchical or parallel probes to identify potential false-positives in the dataset. Secondly, it facilitates the quantitative comparison of microarray data with traditional light-microscopic cell counts by incorporating data from calibration experiments and providing an estimate of the cell concentration in a sample or the relevant detection limit. Both analyses were combined in an easy-to-use, platform-independent, and graphical user interface-based application.

Theory

Most scanner software packages have well-established algorithms and procedures to interpret raw scanner data, identify spots (i.e., physical representations of a probe on an array), and use the image data from the scanner to assign fluorescence values to spots (also called features) and the surrounding background. Axon-based software solutions such as the Genepix 6.0 software used in the MIDTAL project generally output these data in the GPR format. These GPR files constitute the entry point for the present software. In order to interpret the results contained in a GPR file, a number of calculations need to be performed. The necessary steps include the selection of a suitable measure as a basis for the analysis, averaging of replicate spots, normalization, use of calibration data to link signal intensities to cell numbers, and the use of internal controls to identify potential false-positives. The following sections will give a brief overview of the theoretical considerations underlying each of these steps.

Signal-to-noise ratios and total signal intensities

The choice of a suitable measure as a basis for further analyses depends on the type of analysis. Currently available tools for the analysis of phylochips such as ChipChecker and PhylochipAnalyzer examine microarray data mainly on a basis of presence or absence of signals (Loy et al. 2002; Metfies et al. 2008). Such presence/absence decisions rely on a threshold criterion, the most commonly used being the ratio between the signal of the spot in question to the local background or a negative control, i.e., the signal-to-noise ratio (SNR). This ratio provides a simple and intuitive measure to determine if a spot differs significantly from its background. Like ChipChecker and PhylochipAnalzer, GPR-Analyzer therefore uses the SNR with a custom threshold to determine if a signal is present.

The SNR, however, has an important drawback: it is strongly dependent on the background and thus the (local) quality of the hybridization. This has two implications on the analysis: first, the detection limit for a probe and by extension an algal species depends strongly on the quality of the hybridization. This is discussed in detail in the "Hybridization-specific detection limits" section below. The second implication is that the SNR is less suited for quantitative comparisons, as especially for hybridizations with low background noise, small local changes in the background can result in significant changes in the SNR. Although this effect is reduced by the fact that, on the MIDTAL array, each probe is spotted a total of eight times in four different areas of the chip, and although the variability between these replicates can be assessed (e.g., by displaying the standard deviation), for quantitative measurements, GPR-Analyzer uses a second, more robust, metric referred to as total signal intensity (TSI). The TSI is defined as the mean signal minus the local background multiplied by the total area of a spot. It is less dependent on the local background noise, especially for low backgrounds compared to the signal from the probe, and therefore more suited for quantitative and comparative analyses. On the downside, it bears less information about the quality of the signal. GPR-Analyzer uses the TSI in addition to the SNR, averaging both measures over all replicated spots present on the array for further processing.

Normalization

In spite of its reduced dependence on the local background, the TSI is still subject to a number of factors during the sample- and array preparation that may differ between hybridizations, users, and laboratories, and which would bias the analysis if not accounted for (Fig. 1a). These factors include extraction efficiency, labeling efficiency, hybridization efficiency, as well as losses or the possibility to use only subsamples from one step to the next (Fig. 1b), and possible degradation of the RNA during the preparation process. Several of these factors may be accounted for implicitly by using proper controls for normalization. Normalization is achieved by dividing the TSI obtained for a probe of interest by the average TSI for the probe used for normalization. This ratio is termed normalized signal here.

The standard MIDTAL protocol foresees several controls suitable for normalization (Fig. 1c), the earliest being the addition of 500,000 cells of the green alga *Dunaliella tertiolecta* prior to RNA extraction. If TSIs are normalized to a probe that binds specifically to *D. tertiolecta* RNA (a number of which are available on the MIDTAL array), all factors from extraction efficiency until scanner settings (Fig. 1b) would be implicitly accounted for by the normalization. An alternative control frequently used in MIDTAL is the hybridization control, a pre-labeled 247-bp amplicon of the *Saccharomyces cerevisiae* TATA-binding protein-coding sequence added to the hybridization just prior to loading the slide. Using the probe corresponding to this sequence ("Positive_25_dT") for normalization would also account for differences in hybridization and scanning, but any earlier factors such as extraction- and labeling efficiency, or use of subsamples in different steps would need to be accounted for separately.

GPR-Analyzer offers the possibility to choose any probe for normalization, to customize subsequent calculations for different types of standards, and to incorporate a number of correction factors.

Inferring target RNA concentrations and cell numbers

For the microarray data to be comparable with data obtained using other methods, normalized signals need to be associated to specific RNA concentrations or cell numbers of the target species in the environmental sample. This can be achieved via standard curves recording the normalized signal obtained for known samples with different RNA concentrations or cell numbers of a target species. Individual standard curves are required for each target species and for each normalization probe. In MIDTAL, these curves are based on cell numbers rather than RNA quantity (see Fig. 2a for an example), and although GPR-Analyzer supports both the calculation of RNA quantity and cell number (with suitable calibration data), only cell number-based calculations are described here. The use of cell numbers facilitates the comparison of microarray data with light microscopy data, but has the disadvantage that it does not take into consideration possible variations in the per-cell RNA content (see, e.g., Dittami and Edvardsen 2012), thus adding to the overall

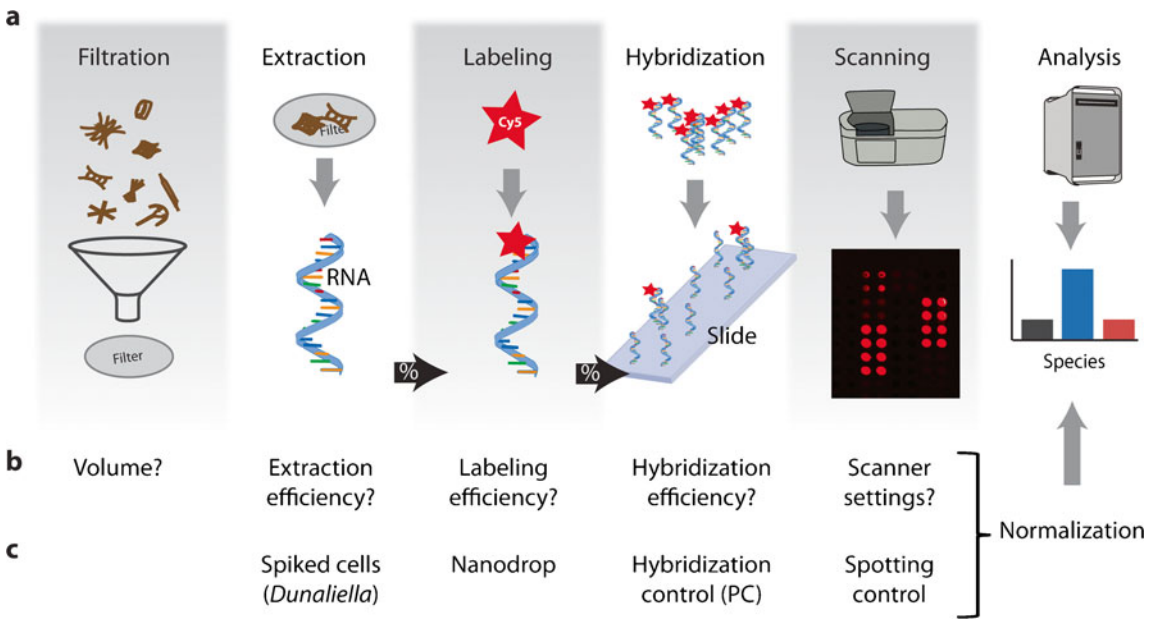

Fig. 1 Important steps in the MIDTAL protocol (**a**), associated factors to consider in the analysis (**b**), and available controls (**c**). *Black arrows* with a "%" sign indicate that frequently only a fraction of the sample is used for labeling and hybridization

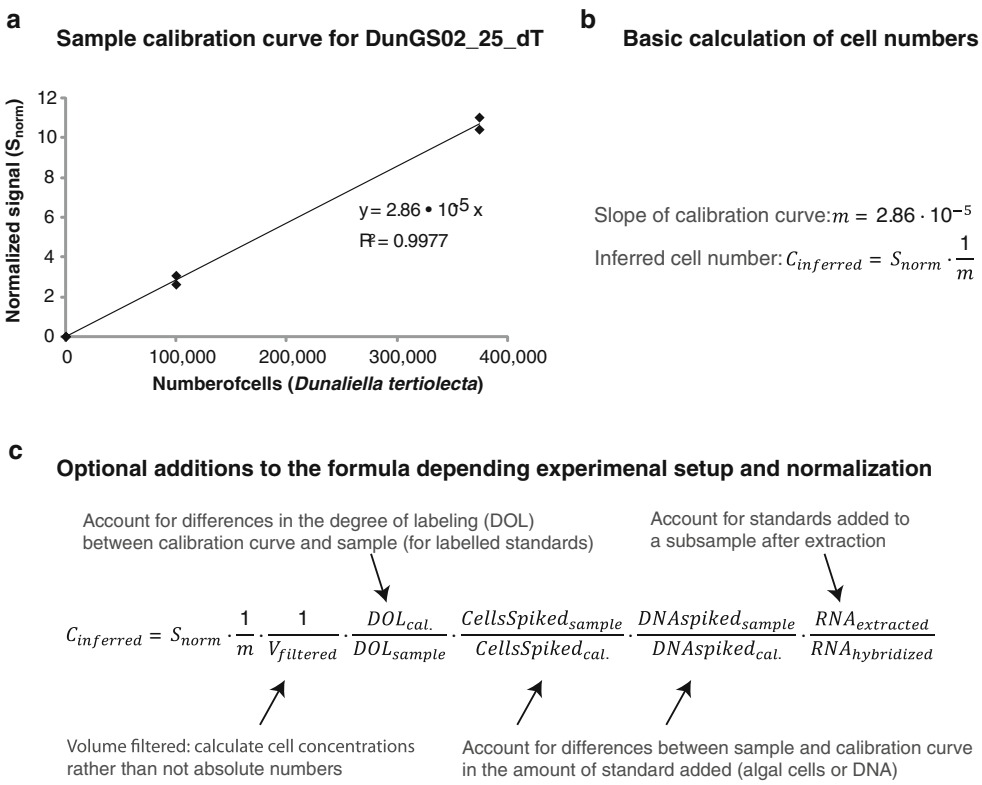

Fig. 2 Inferring cell numbers based on microarray signals. Panel (**a**) shows a standard calibration curve for the probe DunGS02_25_dT plotting the normalized microarray signal (S_{norm}) against the number of cells. In this case Positive_25_dT was used as probe for normalization. Such curves are required for all species represented on the array. If calibration curves and the sample of interest were prepared under identical conditions and with identical settings, cell numbers can be inferred directly from the microarray signal (**b**). However, depending on the experimental setup, additional factors may need to be accommodated (**c**). Only one of the two factors "DNAspiked" and "CellsSpiked" is used at a time depending on the normalization probe, but both factors are kept separate in the program to facilitate quick switching between normalization probes

variance in the data. Here, it is assumed that for all calibration curves, the normalized signal is proportional to the added cell number, and that it is zero when no target cells are present. Both assumptions have been confirmed for a number of species and over a wide range of concentrations within the MIDTAL project and probes that did not meet these criteria were excluded from the array. All calibration curves can thus be described by a simple equation in the form $y=mx$ (Fig. 2a).

Under the assumption that microarray experiments to generate the calibration curves and microarray experiments to analyze the samples of interest have been carried out under identical conditions, only the knowledge of "m" is sufficient to infer cell numbers in a sample based on a normalized signal (Fig. 2b). In some cases, certain experimental factors may, however, differ between the calibration curves and the experiment being analyzed. For example, the amount of standard added may vary. In addition, some factors such as labeling efficiency or the use of only a subsample for the hybridization, may not be accounted for by the probe used for normalization (e.g., "Positive_25_dT", see above), and thus need to be considered for the calculation of cell numbers. GPR-Analyzer therefore allows the formula used to infer cell numbers to be extended by a number of factors (Fig. 2c), depending on the probe used for normalization. In addition, it considers the volume sampled to calculate cell concentrations rather than cell numbers.

Hybridization-specific detection limits

Cell numbers can be inferred for all probes with available calibration data and signals above the detection limit. However, for probes below the detection limit, it is useful to know what number of cells would have been necessary to obtain a positive signal. Important factors limiting the detection limit are the volume of water sampled, the proportion of RNA used for the hybridization, labeling efficiency, and also the purity and integrity of the RNA. Since detection limits are calculated based on the SNR, they are furthermore strongly dependent on the local background noise. An estimation of the detection limit can be obtained by examining the local background of a probe, and extrapolating what TSI would have been necessary to obtain an SNR corresponding to the detection threshold. This hypothetical TSI can then be normalized and used to infer a cell number corresponding to the hybridization-specific detection limit (see calculations described in "Normalization" and "Inferring target RNA concentrations and cell numbers" sections). Detection limits calculated in this way combine information about (1) the (local) quality of the hybridization, (2) the procedure used for the hybridization and possible losses of RNA, and (3) the properties of the probe (obtained from the calibration curve). They will help users to assess the possibility of false negatives caused by poor hybridizations or the presence of large amounts of non-target RNA in a sample. An estimation of the hybridization-specific detection limit for each calibrated probe is implemented in GPR-Analyzer.

Hierarchical and parallel probes

Another factor that needs to be considered on multispecies microarrays is that of false-positive signals. An efficient answer to this problem is to include multiple parallel probes targeting the same species as well as higher-level (hierarchical) probes targeting, e.g., the genus or phylum as controls (Groben et al. 2004; Eller et al. 2007; Metfies et al. 2008). This approach has been taken for the MIDTAL microarray. If a signal is detected for a species-specific probe but one or several of the parallel or corresponding hierarchical probes are negative, the species-specific signal may constitute a false positive and is highlighted for detailed examination by the user. Two stand-alone tools are currently available to help users identify potential false-positives in hierarchical microarrays: PhylochipAnalyzer (Metfies et al. 2008) and ChipChecker (Loy et al. 2002). PhylochipAnalyzer is a graphical user interface-based windows application implementing a strictly hierarchical approach that does not consider parallel probes (i.e., probes for the same species on the same level), although these may be placed into an artificial hierarchical structure. ChipChecker is a command-line based tool implementing a loose hierarchical approach allowing parallel probes.

GPR-Analyzer uses a loose hierarchical representation of the relationship between probes similar to ChipChecker, but also incorporates a parser for XML-based hierarchy files generated using PhylochipAnalyzer.

Results and discussion

Scope and main features

GPR-Analyzer is a platform-independent application written in JAVA to simplify the complex task of analyzing hierarchical multispecies microarrays taking into account the theoretical considerations described in the "Theory" section. Its main functionalities comprise parsing scanner software output files in the GPR format, spot averaging, normalization, inferring cell concentrations (based on calibration curves), calculating hybridization-specific detection limits, and highlighting potential false-positives using hierarchical and parallel probes (Fig. 3). While the application was designed for use with the MIDTAL microarray, it is largely customizable and open-source, making it easy to adapt to other projects.

The primary objective in designing GPR-Analyzer was to enable the rapid and comprehensive analysis of the MIDTAL array for users without special training. Therefore, special emphasis was placed on user-friendliness. The main window (Fig. 4) provides access to all functionalities required for the analysis of MIDTAL arrays. Upon loading a GPR file, the user is prompted to enter basic metadata about the experiment, and the results of the analysis including cell numbers and highlighting of false-positives are immediately shown. Additional GPR files of replicate scans can be added to the analysis at any time. Species names corresponding to different probes can be displayed if they are available in the calibration data file. All

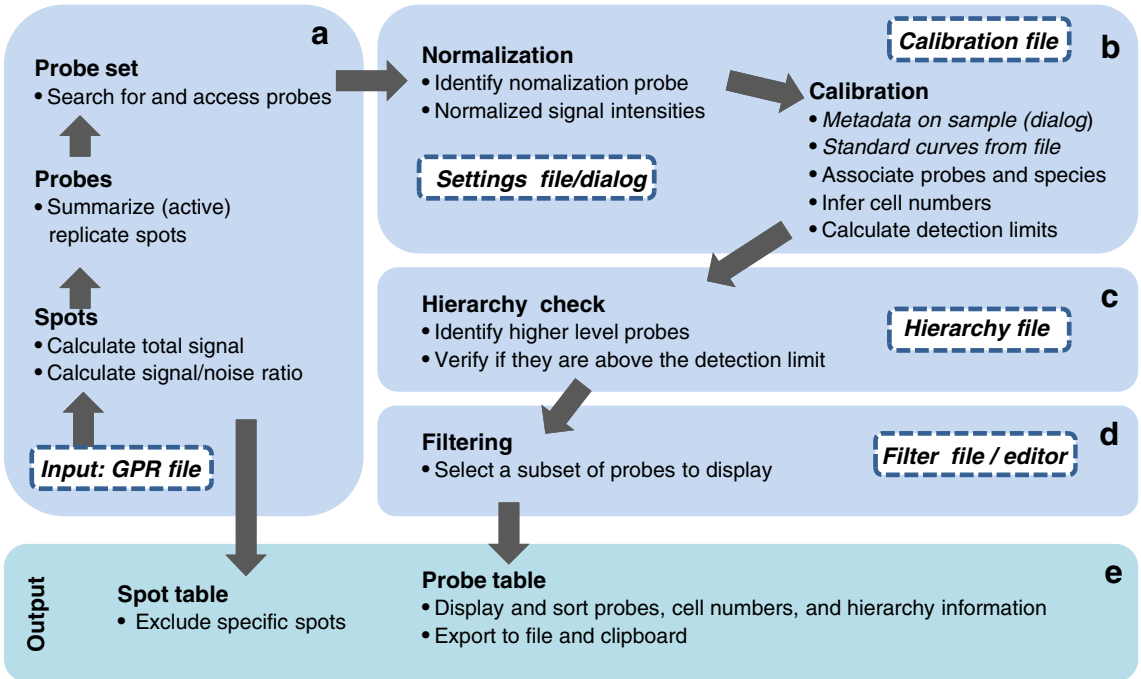

Fig. 3 Structure of GPR-Analyzer with major components: internal representations of the GPR file (**a**), normalization and calibration (**b**), hierarchy checking (**c**), filtering (**d**), and output (**e**). *Italics* indicate user/file input

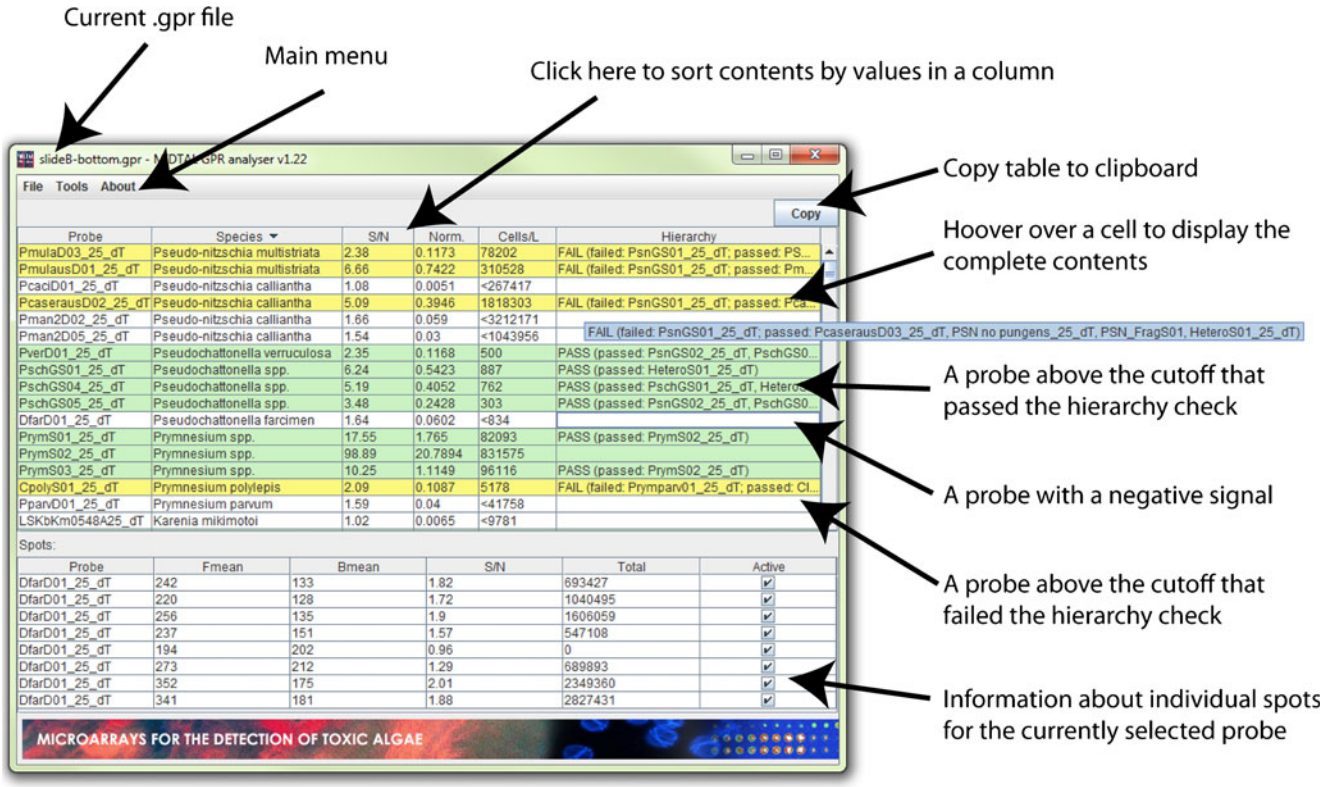

Fig. 4 Screenshot of the main window of GPR-Analyzer displaying a sample GPR file. *Green* color-coding indicates probes for species likely to be present, *yellow* potential false-positives based on hierarchical or parallel probes, and *white* negative probes

results may be exported for use in other applications via the system's clipboard or a tab-separated text file, or they may be explored within the application itself.

Most settings are set to reasonable values by default, although a number of parameters can be adjusted, allowing users, e.g., to select the data displayed, change the normalization probe, or set the detection threshold. These changes are immediately activated and automatically saved, aiding the routine analysis of similar samples following the same protocol. For advanced users, there are options to view and load alternative hierarchy files (used to define hierarchical and parallel probes) as well as alternative sets of calibration data.

A "filter" has been implemented in GPR-Analyzer enabling users to limit the output to a selection of probes of interest. This makes it easy to focus, e.g., on species-specific probes, or probes of a genus of interest, and will thus help to quickly identify relevant data and generate standardized reports for different applications. Filters can either be generated within GPR-Analyzer using the integrated filter editor, or obtained from an external source as a simple text file.

Finally, a semi-automated update system has been put into place. The update function also comprises calibration data, which is still in the process of being assembled by the MIDTAL consortium. All of these and some additional features are described in detail in the GPR-Analyzer manual provided with the application.

Customizability

GPR-Analyzer was designed primarily for use with the MIDTAL microarray, and a number of features have been tailored to suit this purpose. However, the MIDTAL project is now approaching its end and the commercialization of the MIDTAL microarray is well on its way. Thus, the future development of the array will gradually move from members of the consortium to the private industry. To account for these developments, but also to make GPR-Analyzer useful for other projects, an attempt was made to keep GPR-Analyzer as customizable as possible. To this purpose, calibration data is stored in an easily editable text format, comprising meta-information about the conditions used to generate these data and the formulas for the calibration of microarray signals. Likewise, information about parallel and hierarchical probes is stored in a simple text-based file editable in any text editor or spreadsheet program. Moreover, the XML format developed for use in PhylochipAnalyzer (Metfies et al. 2008), which comprises a dedicated editor for this type of file, can be imported. These two files are essentially all that needs to be modified in order to adapt GPR-Analyzer for use with future versions of the MIDTAL chip or in other projects using different platforms, provided that the scanner software used supports the GPR format. In addition, alternative update servers can easily be set in the settings file, making it unproblematic for chip providers to offer

customized versions of the software with their own updates. Finally, the entire source code is freely available and can be extended, modified, and redistributed. This could allow, e.g., adapting GPR-Analyzer for use with other chip platforms and scanner software that does not support the GPR format (e.g., the Agilent platform).

Conclusion

In conclusion, GPR-Analyzer will simplify the analysis of MIDTAL microarrays and hopefully also benefit other related projects. In comparison to PhylochipAnalyzer (Metfies et al. 2008), the pioneer for GUI-based analysis of hierarchical microarrays, it provides a number of additional functionalities that proved valuable in day-to-day use with the MIDTAL array. For example, the parallel use of SNRs and TSIs allows both screening for potential false-positives and normalization for quantitative purposes in a single process, and the automatic integration of calibration curves into the software saves a tremendous amount of time, because it spares the user the effort of looking up calibration data for each probe individually. Being able to infer cell counts within the application also forms the basis for the calculation of hybridization-specific detection limits described above. These are particularly useful for the monitoring of harmful algae, where decisions need to be made about whether a concentration of a toxic alga is below a set limit, and where it is crucial to have an indication as to whether a negative result may be due to poor hybridization quality or the presence of large quantities of non-target RNA. Finally, the platform-independent architecture, together with the use of a free programming language, makes it easy also for non-Windows users to run and modify the software. GPR-Analyzer, its source code, and sample data files are available from http://folk.uio.no/edvardse/gpranalyzer.

Acknowledgments The members of the MIDTAL consortium (http://www.midtal.com/partners.php) are acknowledged for helpful discussions and support, for providing calibration data, for testing the software, and for critical comments on the manuscript especially Jixin Chen, Linda K Medlin, Johannes Hagstrøm, Jessica Kegel, Joe Taylor, Gary McCoy, and Lucia Barra. Linda K Medlin and Jessica Kegel provided the hierarchy file used as a basis for the file included in the GPR-Analyzer package. Some of the artwork used in Fig. 1 was provided courtesy of the Integration and Application Network, University of Maryland Center for Environmental Science (http://ian.umces.edu/symbols). This work was funded by the EU's 7th Framework Program MIDTAL (*FP7*-ENV-2007-1-*MIDTAL*-201724).

References

Ahn S, Kulis DM, Erdner DL, Anderson DM, Walt DR (2006) Fiber-optic microarray for simultaneous detection of multiple harmful algal bloom species. App Env Microbiol 72:5742–5749. doi:10.1111/j.1550-7408.2000.tb00014.x

Ayers K, Rhodes LL, Tyrrell J, Gladstone M, Scholin C (2005) International accreditation of sandwich hybridisation assay format DNA probes for microalgae. N Z J Mar Freshwat Res 39:1225–1231. doi:10.1080/00288330.2005.9517388

Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. Env Microbiol 13:340–349. doi:10.1111/j.1462-2920.2010.02332.x

Costas BA, McManus G, Doherty M, Katz LA (2007) Use of species-specific primers and PCR to measure the distributions of planktonic ciliates in coastal waters. Limnol Oceanogr Meth 5:163–173. doi:10.4319/lom.2007.5.163

Diercks S, Metfies K, Medlin LK (2008) Molecular probe sets for the detection of toxic algae for use in sandwich hybridization formats. J Plankton Res 30:439–448. doi:10.1093/plankt/fbn009

Dittami SM, Edvardsen B (2012) Culture conditions influence cellular RNA content in ichthyotoxic flagellates of the genus *Pseudochattonella* (Dictyochophyceae). J Phycol in press. doi:10.1111/j.1529-8817.2012.01183.x

Eller G, Tobe K, Medlin LK (2007) Hierarchical probes at various taxonomic levels in the Haptophyta and a new division level probe for the Heterokonta. J Plankton Res 29:629–640. doi:10.1093/plankt/fbm045

Galluzzi L, Penna A, Bertozzini E, Vila M, Garces E, Magnani M (2004) Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate). App Env Microbiol 70:1199–1206. doi:10.1128/AEM.70.2.1199-1206.2004

Gescher C, Metfies K, Medlin LK (2008) The ALEX CHIP—development of a DNA chip for identification and monitoring of *Alexandrium*. Harmful Algae 7:485–494. doi:10.1016/j.hal.2007.11.001

Groben R, John U, Eller G, Lange M, Medlin LK (2004) Using fluorescently-labelled rRNA probes for hierarchical estimation of phytoplankton diversity—a mini-review. Nova Hedwigia 79:313–320. doi:10.1127/0029-5035/2004/0079-0313

Humbert JF, Quiblier C, Gugger M (2010) Molecular approaches for monitoring potentially toxic marine and freshwater phytoplankton species. Analyt Bioanalyt Chem 397:1723–1732. doi:10.1007/s00216-010-3642-7

Huyghe A, Francois P, Charbonnier Y, Tangomo-Bento M, Bonetti E-J, Paster BJ, Bolivar I, Baratti-Mayer D, Pittet D, Schrenzel J (2008) Novel microarray design strategy to study complex bacterial communities. Appl Environ Microbiol 74:1876–1885. doi:10.1128/AEM.01722-07

John U (2004) Development of specific rRNA probes to distinguish between geographic clades of the *Alexandrium tamarense* species complex. J Plankton Res 27:199–204. doi:10.1093/plankt/fbh160

Lange M, Guillou L, Vaulot D, Simon N, Amann RI, Ludwig W, Medlin LK (1996) Identification of the class *Prymnesiophyceae* and the genus *Phaeocystis* with ribosomal RNA-targeted nucleic acid probes detected by flow cytometry. J Phycol 32:858–868. doi:10.1111/j.0022-3646.1996.00858.x

Lewis J, Medlin LK, Raine R (2012) MIDTAL (Microarrays for the Detection of Toxic Algae): A protocol for a successful microarray hybridisation and analysis. Koeltz, Königstein, Germany

Loy A, Bodrossy L (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. Clin Chim Act 363:106–219. doi:10.1016/j.cccn.2005.05.041

Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J, Schleifer K-H, Wagner M (2002) Oligonucleotide microarray for 16 S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. App Env Microbiol 68:5064–5081. doi:10.1128/AEM.68.10.5064-5081.2002

Loy A, Schulz C, Lücker S, Schöpfer-Wendels A, Stoecker K, Baranyi C, Lehner A, Wagner M (2005) 16S rRNA gene-based

oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "*Rhodocyclales*". Appl Environ Microbiol 71:1373–1386. doi:10.1128/AEM.71.3.1373-1386.2005

Marcelino LA, Backman V, Donaldson A, Steadman C, Thompson JR, Preheim SP, Lien C, Lim E, Veneziano D, Polz MF (2006) Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. PNAS 103:13629–13634. doi:10.1073/pnas.0601476103

Massana R, Pernice M, Bunge JA, del Campo J (2011) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. ISME J 5:184–195. doi:10.1038/ismej.2010.104

Metfies K, Medlin LK (2005) Ribosomal RNA probes and microarrays: their potential use in assessing microbial biodiversity. Meth Enzymol 395:258–278. doi:10.1111/j.0022-3646.1996.00858

Metfies K, Borsutzki P, Gescher C, Medlin LK, Frickenhaus S (2008) PhylochipAnalyser—a program for analysing hierarchical probe sets. Mol Ecol Resour 8:99–102. doi:10.1111/j.1471-8286.2007.01927.x

Not F, Simon N, Biegala IC, Vaulot D (2002) Application of fluorescent in situ hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. Aquat Microb Ecol 28:157–166. doi:10.3354/ame028157

Palmer C, Bik EM, Eisen MB, Eckburg PB, Sana TR, Wolber PK, Relman DA, Brown PO (2006) Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res 34:e5. doi:10.1093/nar/gnj007

Penna A, Bertozzini E, Battocchi C, Galluzzi L, Giacobbe MG, Vila M, Garces E, Luglie A, Magnani M (2006) Monitoring of HAB species in the Mediterranean Sea through molecular methods. J Plankton Res 29:19–38. doi:10.1093/plankt/fbl053

Peplies J, Lau SCK, Pernthaler J, Amann R, Glöckner FO (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. Environ Microbiol 6:638–645. doi:10.1111/j.1462-2920.2004.00588.x

Scholin CA, Buck KR, Britschgi T, Cangelosi G, Chavez FP (1996) Identification *of Pseudo-nitzschia australis* (Bacillariophyceae) using rRNA-targeted probes in whole cell and sandwich hybridization formats. Phycologia 35:190–197. doi:10.2216/i0031-8884-35-3-190.1

Simon N, Campbell L, Ornolfsdottir E, Groben R, Guillou L, Lange M, Medlin LK (2000) Oligonucleotide probes for the identification of three algal groups by dot blot and fluorescent whole-cell hybridization. J Euk Microbiol 47:76–84. doi:10.1111/j.1550-7408.2000.tb00014.x

Small J, Call DR, Brockman FJ, Straub TM, Chandler DP (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Appl Environ Microbiol 67:4708–4716. doi:10.1128/AEM.67.10.4708-4716.2001

Wagner M, Smidt H, Loy A, Zhou J (2007) Unravelling microbial communities with DNA-microarrays: challenges and future directions. Microb Ecol 53:498–506. doi:10.1007/s00248-006-9197-7