Meeting Report: Molecular Ecology Workshop. Detection of Microbial Biodiversity in Environmental Samples, Camerino, Italy, September 19—21, 2005

A molecular ecology workshop for the detection of microbial diversity using microarray technology was held in Camerino, Italy from September 19—21, 2005 to present the achievements of the 5th FP EU MICROPAD. This EU project focused on the development of DNA microarrays for the detection of pathogenic protozoa, diatoms and flagellated algae. The identification of diatoms and flagellated algae with conventional methods, e.g. electron microscopy, requires broad taxonomic expertise, and monitoring field samples is both labor and time-consuming. Pathogenic protozoa are also difficult to separate from non-pathogenic relatives. Species-specific probes can be used to monitor the biodiversity of these organisms and the lessons learned in this project can be applied in a more general sense to other genes used in a microarray format.

In our workshop, we provided an introduction into the application and design of ribosomal RNA (rRNA) probes and DNA microarrays for the assessment of biodiversity. The application of molecular methods to answer ecological questions permits issues of biodiversity to be addressed at all levels. rRNA probes contribute significantly to the assessment of biodiversity at the molecular level because a species-specific probe can be made to recognize any species or higher taxon. It represents a powerful augmentation to traditional taxonomy, which is based on identifying species primarily by morphology; however, it is not the only gene that can be used to provide identification of species. Barcoding for Life uses the COX1 mitochondrial gene as its molecular marker. When these species probes, regardless of the gene from which they were designed, are applied to DNA microarray technology, then a powerful tool is created to assess biodiversity. DNA microarrays offer a great potential to facilitate the application of molecular probes to answer ecological and biodiversity questions through fast through-put of samples.

The workshop in Camerino was designed to bring together leading world experts in the development of these probes to generate a species-specific sequence or barcode and the application of these probes to a microarray for fast through-put analysis. Thus, a workshop summarizing our state of knowledge in the development of microarrays for phylogenetic analysis, the problems inherent to the method, the potential solutions to these problems, and prospects for the future was timely and should be of immense value to the wider scientific community, especially those institutions who belong to the Global Biodiversity Information Facility (GBIF) because phylochips can help achieve the goals of cataloging biodiversity.

Below is a list of the invited speakers as well as the title of their presentation:

- Linda Medlin, AWI, Bremerhaven, Germany
  The ARB Program and Probe Development
- Katja Metfies, AWI, Bremerhaven, Germany
  Microarrays for the Identification of Flagellated Algae
- Antonella Penna, Università di Urbino, Urbino, Italy
  Mediterranean Dinoflagellate Biodiversity: A Molecular Phylogeographic Approach
- Nina Silkenbeumer, University of Bremen, Bremen, Germany
  Microarrays for the Identification of Fish Larvae to aid in the Assessment of Fishery Stocks
- Georg Nies, University of Cologne, Cologne, Germany
  Implementation of a DNA-taxonomy Concept on Microarrays
The application of species identification arrays is the agreed that the biggest challenge with the identification of organisms. In contrast, the speakers reported positive results when microarrays were used to identify organisms. The majority of the talks reported about the state-of-the-art in respect to microarrays. At the end of the meeting, each talk highlighted different problems associated with microarrays. The workshop began with an introduction into the ARB program, which is the only program available for handling extensively large databases and tools for searching that database for species-specific regions. This was followed by a series of talks showing the current state of research for the application of probes using fast through-put methods to identify taxa. Each talk highlighted different problems associated with microarrays. At the end of the meeting, not only did the audience have a good overview about the state-of-the-art in respect to microarrays for species identification, but also about challenges and problems that co-occur with its application. The majority of the talks reported positive results when microarrays were used to identify organisms. In contrast, the speakers agreed that the biggest challenge with the application of species identification arrays is the development of suitable sets of probes that work under the same conditions on a single DNA-chip.

Following the presentations, there was a general discussion in which the problems with probe development were discussed. Potential solutions to each problem were identified. Below we present a summary of the most relevant problems discussed by members of the workshop and their suggestions as to how to solve them.

**Hybridization Efficiency and Specificity of the Microarray**

The hybridization efficiency of the probes on the chip is of crucial importance to its success for species identification. The main goal of a successful hybridization is to minimize cross-hybridization between arrayed elements on the chip and non-target nucleic acids. Besides that, it is of significant importance to achieve high signal-to-noise ratios and to ensure that the signal intensity on the chip is proportional to the amount of nucleic acid bound to it. These requirements have to be borne in mind during the design of the molecular probes. In her talk, Linda Medlin from the Alfred Wegener Institute in Bremerhaven introduced the ARB-software and a number of parameters that have to be considered during probe design.

All speakers shared the experience that very often the theoretical properties of a probe cannot be observed in practice. Therefore, each probe has to undergo intensive testing before it can be applied to real samples. This includes careful testing of the probe specificity and the evolution of the hybridization signal. If probes show unspecific binding to non-target nucleic acids, even though the specificity was confirmed in theory previously, it was suggested to resynthesize the probe from a different company, because it sometimes occurs that incorrectly synthesized probes may be the reason for unspecific binding. Also, all probes should be HPLC purified prior to use on DNA-chips.

In fact, unspecific binding is the biggest challenge of species identification microarrays. This is particularly important in microorganisms where the number of sequenced species that can be used to test the specificity of the probes is minimal in comparison to the estimated number of unknown species. Nevertheless, specificity can be assured. This could be done using a multi-probe approach, including hierarchical sets of probes.
that target species at different taxonomic levels or multiple probes for one target species. In this context, a species is only considered present if all probes that target the species give a positive signal. Also, if for any one group not all of its biodiversity is known, then hierarchical probes can help monitor the biodiversity of that group at a higher taxonomic level.

To draw reliable conclusions on species composition in field samples, reproducibility of the data has to be assured. The quality of all microarray data is heavily dependent on the quality of the spotted probes on the array. Moreover, the concentration and the amount of the probes on different chips should be the same. It was suggested to check this by either hybridization with random oligonucleotides or staining the DNA on the chip with specific dyes. In most talks, the microarray analysis was PCR-based. Therefore, it has to be kept in mind that the amplification of target nucleic acids from field samples introduces a bias to the analysis, because some DNAs are preferentially amplified in comparison to others. To avoid such PCR biases, it was suggested to use RNA for the analysis of field samples when possible. The speakers agreed that in order to assure the quality of the results of a microarray analysis, experiments should be replicated. This includes a replication of the experiment starting with the PCR, as well as hybridizations on different slides rather than duplications of hybridizations on the same slide.

Alexander Loy from the University of Vienna pointed out in his talk that the $\Delta G$ of a probe is of interest to the success of a probe. His experiments indicated a correlation of the hybridization efficiency of a probe and $\Delta G$. However, his experiments are currently only an indication. Therefore, new algorithms for calculating the $\Delta G$ on glass slides are needed.

Katja Metfies from the Alfred Wegener Institute in Bremerhaven reported a correlation of the binding loci of probes to their hybridization efficiency. The results indicate that probes that target the 18S rDNA of phytoplankton result in insufficient signal intensities if they bind in an area $\sim$900 bp downstream. This observation could be linked to the secondary structure, i.e. decreasing the size of the target molecule can minimize the influence of the secondary structure on the hybridization efficiency. Smaller nucleic acids, preferably with a size between 200 and 400 bp have a lower tendency to form secondary structures that block the probe-binding site. Different strategies to obtain small target molecules were proposed. First, it was suggested to use nick-translation labeling of target to obtain an average size of 500 bp. Second, DNases can be used to break the target DNA into smaller pieces. Finally, it was proposed to use sonication in order to generate smaller DNA pieces. However, the reproducible generation of DNA fragments of the right size by sonication requires extensive experience.

Standardization of Microarrays for Species Identification

At the end of the discussion, it was raised if and how the MIAME guidelines should be incorporated to the application of microarrays for species identification. The MIAME guidelines try to provide a conceptual structure for microarray experiment descriptions and aim to guide the development of microarray databases and data management software. However, currently, the guidelines only apply to the application of microarrays for expression analysis. There is no standardized protocol for the application of microarrays for species identification. Therefore, it was proposed that the scientific community should develop a standardized protocol, including a means of data storage that all laboratories agree on. Such a protocol would be required for the evaluation and comparison of species identification data that were generated with microarrays developed and used in different laboratories.

Conclusion

The molecular ecology workshop for the detection of microbial diversity using microarray technology was a small workshop with $\sim$20 participants. However, among the participants, there was broad experience with the application of microarrays for species identification. Therefore, a fruitful discussion took place and we found it particularly useful that both the strengths and the weaknesses of the approach were discussed in-depth. The overall conclusion from the workshop is that it was generally agreed that despite the problems associated with this technique, the future was optimistic for its general application in monitoring and biodiversity studies. The information applicable to phylochips will also be applicable to investigators using chips for bar coding, regardless of the gene used on the chip for taxon identification as
problems are common to the method and not just restricted to the gene of choice.

Katja Metfies\textsuperscript{a}, Marco Berzano\textsuperscript{b},
Claudio Gualerzi\textsuperscript{c}, Gerard Muyzer\textsuperscript{c}, and
Linda Medlin\textsuperscript{a,1}
\textsuperscript{a}Alfred Wegener Institut, Am Handelshafen 12,
27570 Bremerhaven, Germany
\textsuperscript{b}Laboratorio di Genetica, Dipartimento di Biologia
M.C.A., Università di Camerino, Via Camerini 2, I-62032 Camerino, Italy
\textsuperscript{c}Afdeling Biotechnologie, Technische Universiteit
Delft, Julianalaan 67, 2628 BC Delft, The Netherlands
\textsuperscript{1}Corresponding author;
fax: +49 471 4831 1425
e-mail lkmedlin@awi-bremerhaven.de

Available online at www.sciencedirect.com