Faculty of Science of Eötvös Loránd University
Biology Doctoral School
Program of Theoretical and Evolutionary Biology

INVESTIGATION OF THE SPECIES COMPOSITION OF MICROBIAL COMMUNITIES: THE ANALYSIS OF THE MULTITEMPLATE PCR AND DGGE

DOCTORAL THESIS

SIPOS RITA

Consultant:	Head of Doctoral School:	Head of Doctoral Program:
DR. MÁRIALIGETI KÁROLY	**DR. ERDEI ANNA**	**DR. SZATHMÁRY EÖRS**

Department of Microbiology

Budapest
2009

INTRODUCTION

In addition to uncovering the species composition of communities characteristic of different habitats, the microbial ecologist is driven by the urge to realistically map the metabolic relationships within the community, and its spatial and temporal dynamics. The small number of cultivable species (ca. 0.1-1%) forced researchers to apply cultivation-independent, molecular biological methods, with which a lot more samples can be processed at the same time. The diversity analysis of microbial communities is based on the analysis of various functional and/or phylogenetic marker genes. A precondition of many of the nucleic-acid-based methods is the enzymatic amplification of DNA or RNA, extracted from environmental samples, with polymerase chain reaction (PCR or RT-PCR). Even today, the most examined area of microbial diversity examinations is the analysis of the 16S ribosomal RNA gene.

The PCR-based molecular fingerprinting techniques provide the possibility to compare communities, even taxonomically identify members of the community in some cases. The individual methods are suitable to distinguish the quantitative relationships of some phylospecies to various degrees. Different methods are applied to separate the resulting mixed PCR product composed of amplicons of several species. The most often used method is molecular cloning and the base order analysis of each unique clone. Earlier it was assumed that the dominant sequences of a given clone library represented the dominant members of the analyzed community. However, one must keep in mind that, when drawing conclusions about the composition of a community during PCR-based analyses, the results must be handled with care as every step of the molecular sample processing holds possibilities for error. The researcher is faced with problems of PCR-based molecular methods during every community analysis, when statements whether a taxon is a dominant or marginal member in a given community from the points of view of abundance, or activity demonstrated in a metabolic process.

Multitemplate PCR is in many aspects the most critical element of the processing of environmental samples with molecular methods. The exponential kinetics of the PCR technique is very well studied in the case of the genomic DNA of a single species. However, in microbial ecology, we still have little knowledge of the quantitative changes during PCR in the case of the DNA template of sometimes the mixture of several thousand species in a single community. Due to the exponential nature of the PCR reaction, the distorting effects emerging in the beginning cycles may multiply in later cycles, just like any changes that take

place in previous steps, such as sampling and DNA isolation. These may all alter the results of the composition of the original microbial community. The exact study of the processes taking place in multitemplate PCR should be set as an objective so that PCR conditions could be determined not only on the basis of the expected yield, but also to avoid, or at least mitigate, incidental effects distorting the original composition.

OBJECTIVES

Our objective was the formulation of such recommendations to be used in everyday laboratory practice that would minimize the distortions emerging during the PCR step of sample processing. To this end, the distorting effect on species composition of multitemplate PCR was investigated in an artificial model system and environmental samples; in addition, one of the most widespread molecular fingerprinting method, PCR-DGGE was evaluated from the aspect of its potential of diversity mapping, with special emphasis on the application of various "universal" *Bacteria* domain-specific 16S rRNA gene primers (Fig. 1). Our objectives are detailed in the following points:

1. The description of a specific case study to demonstrate a problem when the selection of suboptimal parameters may lead to false conclusions. The renewed interpretation of earlier mushroom compost diversity analyses. Database preparation with the analysis of alignment of various *Bacteria* domain-specific primers on the basis of 920 643 aligned 16S rRNA gene sequence of *Ribosomal Database Project* version 10.12 (RDP-II, *Release* 10, *Update* 12) (Fig. 1/D).

2. The creation of a simple model system with the use of bacterial strains, where the changes in the template-to-product ratio taking place during multitemplate PCR may be detected reliably and quantitatively. The fast and cost-effective detection system of T-RFLP was applied for the evaluation of the composition of the PCR product, and reliability measurements were made in order to verify the usability of the model system (Fig. 1/A).

3. The determination and detailed analysis of the parameters of multitemplate PCR that may greatly influence the modifications of template-to-product ratios in multitemplate PCR. In the case of factors demonstrating the most significant deviations, an attempt is made to characterize the distortion and describe its extent (Fig. 1/A).

4. The testing of PCR factors investigated in the model system on environmental samples (rhizosphere and rhizoplane). The formulation of specific recommendations for the processing of diverse environmental samples with molecular methods (Fig. 1/B).

5. The testing of the application of the PCR-DGGE community fingerprinting method, with respect to the above recommendations and from the point of view of the usability of "universal" primers. Is it worthwhile to use several "universal" primer pairs for the PCR-DGGE-based mapping of the diversity of a bacterial community? (Fig. 1/C).

6. An experiment for the more complete investigation of bacterial diversity found in rhizosphere samples: is the detection of rare phylospecies present in small quantities possible from DNA recovered from several parts (background) of the DGGE gel in the case of a community pattern (Fig. 1/E).

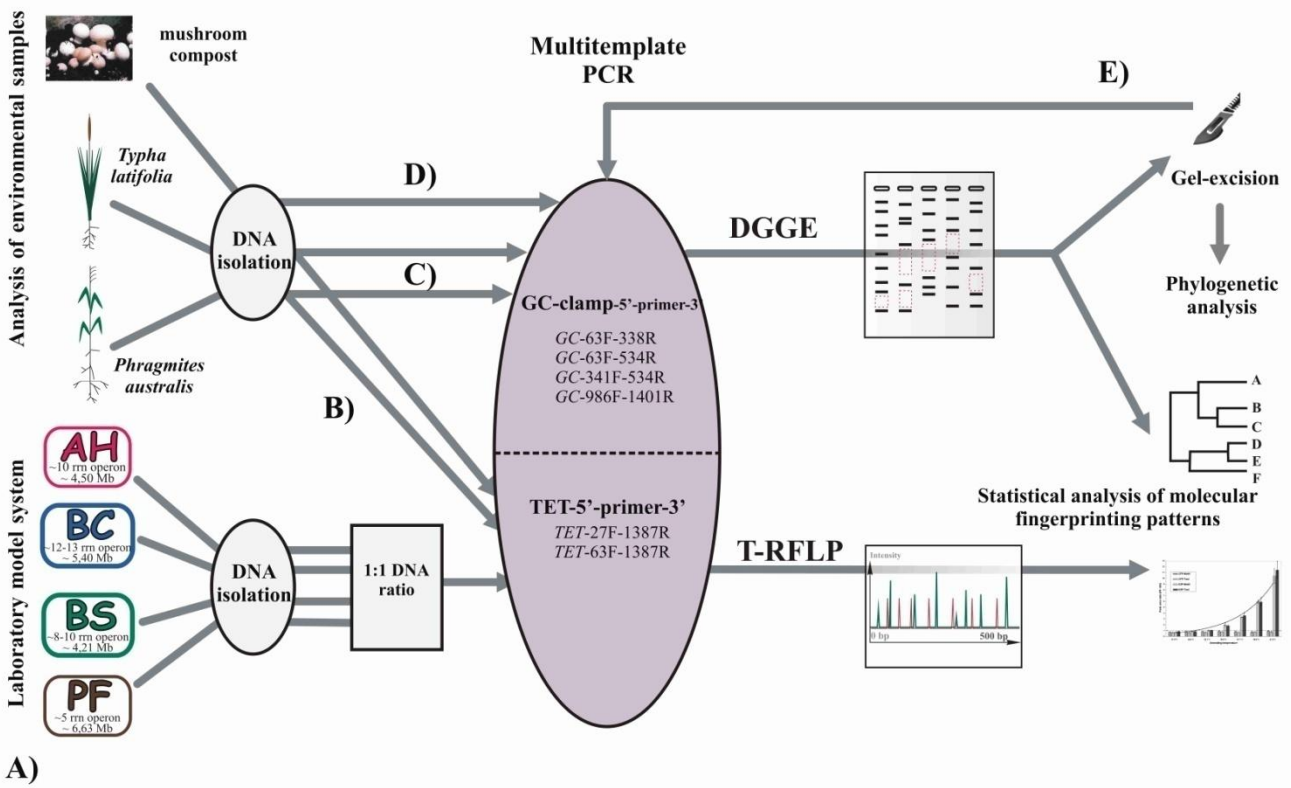

Figure 1. Outline of the molecular biological investigations carried out on bacterial strains of a model community and environmental samples.

A) Analysis of the template-to-product ratios of the multitemplate PCR using defined DNA template ratios, by the T-RFLP method. The strains are: AH (*Aeromonas hydrophila*), BC (*Bacillus cereus*), BS (*Bacillus subtilis*) and PF (*Pseudomonas fluorescens*), respectively. **B)** Testing the results of the model community on bulrush and reed rhizosphere and rhizoplane environmental samples, also with T-RFLP. **C)** Application of four "universal" 16S rRNA specific primer pairs in parallel for the DGGE fingerprinting analysis of 16 rhizosphere and rhizoplane samples. **D)** Analysis of mushroom compost microbiota diversity with DGGE using several primer pairs, followed by phylogenetic analysis of discrete bands. **E)** During the survey of the "background smear" of DGGE gel areas in order to find bacterial taxa of low abundance, gel areas were excised and reanalyzed by DGGE.

APPLIED METHODS

Model system experiments and testing on environmental samples

We have designed a model system in which the terminal fragments of the four bacterial strains (*Aeromonas hydrophila* [AH] *Bacillus cereus* [BC]; *Bacillus subtilis* [BS] és *Pseudomonas fluorescens* [PF]) could be well distinguished. The complete 16S rRNA sequence of the strains and the length of the theoretical (*in silico*) T-RFs were determined with the use of computer simulation, and the empirical terminal fragment lengths were assessed by T-RFLP for the combinations of the applied universal primer pairs (TET-27F-1387R and TET-63F-1387R) and five restriction enzymes (*Hin*6I, *Alu*I, *Csp*6I, *Bsh*1236I és *Tas*I). The fluorescently labeled T-RFs were electrophoretically separated on a model ABI PRISM 310 Genetic Analyzer automatic capillary instrument.

The isolated genomic DNA of the model strains was mixed in all possible combinations and at several predefined concentration ratios (1:1; 1:10; 1:50), then used as templates for multitemplate PCR. The alteration of the amplicon ratios of the obtained PCR product was determined with the T-RFLP electropherograms. The mixing of the template ratios was repeated in four parallels, a hot start *Taq* polymerase enzyme was used for PCR and T-RFLP electrophoresis was also conducted in three parallels. The peak area data were used for the determination of the relative abundance of the detected T-RFs in the PCR product. The effect of annealing temperature was investigated at 10 temperatures (47°C–61°C temperature range) with the use of a gradient PCR instrument, and the effect of PCR cycle numbers was studied with 5 cycles (12, 18, 24, 32 and 48 cycles, respectively). In the case of testing the model system results on environmental samples, directly isolated genomic DNA of bulrush rhizosphere and rhizoplane samples was used and the effect of several PCR parameters was investigated through the statistical analysis of T-RFLP patterns.

DGGE analysis of mushroom compost microbiota

A complete composting cycle with various sampling points was monitored during November and September, 2003. Sampling of the heap, the bunker and the mature phase II mushroom composts (at the end of the indoor composting tunnel) was carried out. Genomic DNA was isolated from the compost samples and the first PCR round was conducted with the 27F-1492R primer pair. In a second, nested step, the primer pairs *GC*-63F-338R and *GC*-968F-1401R were used with touch down thermal protocols. The electrophoretic separation of the multitemplate PCR amplicons was carried out in a 7-8% PAA gel with a 40-70%

denaturing gradient at 60°C, 80V, for 15 hours (INGENYphorU instrument). Following electrophoresis, the PAA gels were stained with ethidium-bromide and the obtained banding patterns were recorded under UV illumination and documented with digital imaging.

Discrete DGGE bands were excised; their DNA sequence was recovered and re-amplified with PCR and the 16S rRNA gene nucleotide base order was determined. The phylogenetic analyses of obtained sequences were conducted in MEGA 4.0, the phylogenetic relationships were depicted with a *Neighbor-Joining* tree, in which evolutionary distances were computed using the Maximum Composite Likelihood method using 1000 bootstrap replicates.

DGGE analysis of rhizosphere fractions originating from a wastewater treatment constructed wetland

Rhizome sampling from the first, bulrush unit and the second, reed bed unit was obtained on 28 May and 7 September, 2003. Sample processing yielded 16 samples with respect to sample location and time, which is summarized in Table 1.

Table 1. Characteristics of bulrush and reed rhizosphere fractions**.** *Bulrush samples originate from both wastewater inflow (I) and wastewater outflow areas (O).

Sample		Plant		Sampling time		Rhizosphere fraction*		Place of origin	
#	*Marking*	*Typha latifolia*	*Phragmites australis*	*May*	*September*	*rhizosphere/rhizoplane*		*constructed wetland*	*natural wetland*
1	Typ-IRS-m	Typ		m		IRS		+	
2	Typ-ORS-m	Typ		m		ORS		+	
3	Typ-IRP-m	Typ		m			IRP	+	
4	Typ-ORP-m	Typ		m			ORP	+	
5	Typ-IRS-sz	Typ			sz	IRS		+	
6	Typ-ORS-sz	Typ			sz	ORS		+	
7	Typ-IRP-sz	Typ			sz		IRP	+	
8	Typ-ORP-sz	Typ			sz		ORP	+	
9	Phr-RS-m		Phr	m		RS		+	
10	Phr-RP-m		Phr	m			RP	+	
11	Phr-RS-sz		Phr		sz	RS		+	
12	Phr-RP-sz		Phr		sz		RP	+	
13	Phr-RS-mo		Phr	m		RS			o
14	Phr-RP-mo		Phr	m			RP		o
15	Typ-RS-mo	Typ		m		RS			o
16	Typ-RP-mo	Typ		m			RP		o

The series of bulrush and reed rhizosphere and rhizoplane samples, different variable regions of the 16S rRNA genes (V1+V2; V1–V3; V3; V6–V8) were amplified by PCR applying four universal primer pairs. We took into consideration our results obtained from the model system experiments, when the PCR-DGGE processing of the rhizosphere samples was conducted: we applied low annealing temperature, direct/one step PCR protocols omitting the previously routinely used nested and touch down protocols. DGGE electrophoresis was carried out similarly to the compost samples, and the DGGE banding patterns were converted

to binary matrix data by *TotalLab* gel-analysis software. A Jaccard index based 16 x 16 similarity matrix was created with the PAST program, and hierarchical cluster analysis (UPGMA) and 3D NMDS ordination using the STATISTICA software package were conducted. These evaluations were carried out separately for the four primer pairs, as well as in combination (concatenated data) of the four, in an artificially generated combined binary data matrix.

Alignment of universal primers specific for the 16S rRNA gene

We created an own database of the 10 most widely used universal primers using the RDP 10.12 sequence database and the *Probe Match* function. The *in silico* sequence alignment of the 10 primers with that of the target sequences of the database was conducted for *Bacteria* and *Archaea* domains alike and with the application of different number of mismatches.

RESULTS AND CONCLUSION OF THE THESIS

Summary of the most important results

1. We have successfully set up a model system composed of four bacterial strains, in which we detected – with the use of T-RFLP method – the changes in the artificially set genomic DNA ratios of the strains with respect to the PCR product (with the use of more primers and restriction enzymes). We have determined the margin of error of T-RF detections for specific strain-primer-restriction enzyme combinations.

2. We have demonstrated that the most accurate evaluation of template-to-product ratios with the ABI 310 capillary gel-electrophoresis system used at our facility is definitely provided by the peak area data obtained from the electroferograms (especially in the case of the separation of terminal fragments with a large length difference [> 50 nucleotides])

3. We were able to demonstrate with respect to "universal" primer pairs – 63F-1387R and 27F-1387R – and several strain combinations that templates showing 100% nucleotide match were amplified to a significantly larger extent during PCR than templates showing 3 mismatches (63F); thus resulting in serious preferential amplification. The changes in the original template ratios gave an almost exponential relationship with annealing temperature, although the rate of preferential amplification could be decreased, and approached the results of an exact match at some low annealing temperatures.

4. The changes in PCR cycle number had only a small effect on the modification of the template-to-product ratio in the model system, yet the effect of the three mismatches appeared significant at almost any cycle number.

5. The experience gained from predefined genomic template DNA ratios was examined on environmental bulrush rhizosphere samples. In the case of both primer pairs, the decrease of the annealing temperature yielded a more complex T-RFLP pattern with several peaks. The ratio of individual T-RF peaks showed significant deviations at different temperatures; in extreme cases, some peaks completely disappeared from the community patterns at higher annealing temperatures. Our assumption was that preferential amplification originating from mismatch may have played a part in the changes of the ratio of some terminal fragments. The application of different cycle numbers did not result in significant deviations in the T-RFLP fingerprint pattern of the bulrush rhizoplane microbial community, supporting former hypotheses that the effect of cycle number is small in the case of diverse environmental samples.

6. The sequence complementarity of 10 universal 16S rRNA gene-specific primers was examined against the 920 643 aligned 16S rRNA gene sequences contained in *Ribosomal Database Project* version 10.12 (June 2009). An own database, distributed into taxonomical groups (*phylum*) was created on the basis of alignment analyses.

7. Different universal primers have revealed completely different bacterial diversity during the PCR-DGGE analysis of mushroom compost microbiota. We demonstrated in the case of two *Bacteria* divisions that the different degree of primer matching in the case of different taxons stood in the unsuccessful detection of groups.

8. The four 16S rRNA gene-specific primer pair yielded different DGGE banding patterns with and resolution yet individually reproducible and high diversity. The comparison of the patterns evaluated individually with the results of the evaluation of the four patterns together led to the conclusion that concatenated data gave the most realistic picture, and this method was also useful to exclude outlier data. In addition, we demonstrated that the different statistical evaluation (hierarchical analysis and ordination) of binary data matrices generated from DGGE patterns yielded extra information and gave a more detailed picture of real similarity relationships.

9. We could not prove our hypothesis that the parts of the DGGE gel that contained no bands contained taxa in small abundance; because we could not detect unique bands within these gel areas that were different from the original pattern.

10. The separation of amplicons in the PAA gel with DGGE method is far from perfect, since with repeated DGGE separation of DNA obtained from different gel pieces, we were able to reproduce a banding pattern characteristic of the complete sample, independent of denaturing concentrations. Our results indicate the possibility that some dominant PCR amplicons appear not only at defined, discrete bands but anywhere in the gel.

Recommendations for the 16S rRNA gene-based molecular processing of environmental samples

The fact that the ecological conclusions, and thus the picture of a microbial community, that may be derived from the results of molecular diversity-investigation methods based on the examination of the 16S rRNA gene is significantly garbled by the application of incorrectly controlled PCR primers. Although newly designed primers are tested for PCR effectiveness against strains and the genomic DNA of environmental samples, but any primer mismatch problems that has been revealed and judged significant should also be taken into account. A possible solution would be the use of a mix of variants of the slight modifications of the same primer.

PCR optimization in everyday laboratory practice often only extends to the production of sufficient amount of specific PCR product. Due to the specificity of PCR, usually high-temperature, often *"touch-down"* protocol is used. On the basis of our experiments conducted on the model system and on the environmental rhizosphere samples on the other hand, we recommend the opposite: in order to decrease preferential amplification, only use *"touch-down"* thermal profiles if the production of aspecific products cannot be avoided otherwise, or when the aim is the specific detection of a lower taxonomical group.

DNA isolation

- We recommend the combination of physical, chemical and enzymatic steps in order to get adequate access to the genomic DNA material of bacteria with different cell wall structures.

Multitemplate PCR conditions

- In case our objective is the mapping of the entire bacterial diversity, PCR reaction should be conducted with **several universal primer pairs** in parallel. The exact taxonomical specificity of primers should be verified with the most recent sequence database data. Primer pairs should be selected so that they realize the widest possible coverage of bacterial groups. The repeated testing of oligonucleotides is indispensable even in the case of primers with a narrower spectrum.
- PCR reaction should be optimized with respect to **annealing temperature** with each primer pair. The **lowest possible** temperature, with which the production of specific products without the appearance of aspecific products is feasible, should be determined. Then higher PCR cycle numbers (30-32 cycles) can still be used.

- The PCR should be repeated at the **optimal annealing temperature** in parallel reactions, and a **smaller cycle number** (ca. 25 cycles) should be used in order to minimize the production of PCR by-products. In the case of small yields, the **products of several parallel PCR reactions may be combined** to have an adequate amount of DNA for further analyses.
- The **final extension** at the optimum temperature determined by the polymerase enzyme should last **for 30 minutes**, which serves the purpose of completing incomplete strands and excluding double bands at the DGGE separation of different amplicons.
- The **concentration** of the PCR product is useful, which may be achieved with an adequate **purification step**. The removal of single-stranded DNA from the PCR product by an enzymatic method is possible (*Mung Bean nuclease* treatment).

DGGE fingerprinting pattern analysis

- The use of **non-symmetrical similarity indices** or the Raup and Crick probability variable is recommended for the binary evaluation of DGGE banding patterns.
- The results obtained with different primer pairs may also be evaluated in **a combined (concatenated) data matrix**, which may also mitigate the preferential effect of primers with respect to different taxonomical groups.
- A more exact picture may be produced of microbial communities if the statistical analysis of similarity data matrices is carried out with different methods (hierarchical clustering and ordination) and the results are compared.

The 16S rRNA-based investigation of microbial communities has been the most thoroughly researched area of microbial ecology, and the ribosomal gene-based taxonomy is still justified. The 16S rRNA gene-based molecular community analyses provide the possibility for the combined processing of several samples and extremely fast mapping. Many times, preliminary investigations with universal 16S rRNA gene-specific primers is necessary also in the case of investigations with narrower specificity (specific investigations for a given taxonomical group or biological function) in order to be able to select specific samples from a given sample series for the purpose of more detailed processing (clone library preparation, metagenome analysis). It is impossible to draw far-reaching ecological conclusions with the use of a single primer. A condition of the proper selection and control of primers used during PCR is that the sequences found in public databases are continuously monitored for quality, because newly designed primers can exclusively be based on the data of known sequences, and a nomenclature applied by everybody in unison is also needed.

PUBLICATIONS LAYING THE FOUNDATION OF THE THESIS

Referred scientific research papers with impact factors:

Székely, A. J., **Sipos, R.,** Berta, B., Vajna, B., Hajdu, Cs., Marialigeti, K. (2009): DGGE and T-RFLP Analysis of Bacterial Succession during Mushroom Compost Production and Sequence-aided T-RFLP Profile of Mature Compost. Microb Ecol 57:522-33.

Sipos, R., Székely, A. J., Palatinszky, M., Revesz, S., Marialigeti, K., Nikolausz., M. (2007): Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol Ecol 60:341-50.

Nikolausz, M., **Sipos, R.,** Revesz, S., Szekely, A., Marialigeti, K. (2005): Observation of bias associated with re-amplification of DNA isolated from denaturing gradient gels. FEMS Microbiol Lett 244:385-90.

Book Section:

Sipos, R., Székely A., Révész, S., Márialigeti, K. (2010): Addressing PCR Biases in Environmental Microbiology Studies. *In* S. P. Cummings (ed.), *Bioremediation: Methods and Protocols* vol. 599. Springer-Verlag New York, LLC. ISBN: 978-1-60761-438-8

OTHER PUBLICATIONS

Referred scientific research papers with impact factors:

Borsodi, A. K., Rusznyak, A., Molnar, P., Vladar, P., Reskone, M. N., Toth, E. M., **Sipos, R.,** Gedeon, G., Marialigeti, K. (2007): Metabolic Activity and Phylogenetic Diversity of Reed (*Phragmites australis*) Periphyton Bacterial Communities in a Hungarian Shallow Soda Lake. Microb Ecol 53:612-20.

Révész, S., **Sipos, R.,** Kende, A., Rikker, T., Romsics, C., Mészáros, E., Mohr, A., Táncsics, A., Márialigeti, K. (2006): Bacterial community changes in TCE biodegradation detected in microcosm experiments. Int. Biodeterioration and Biodegradation 58:239-247.

Relevant Conference Lectures:

Sipos, R., Nikolausz, M., Micsinai, A., Nyírő, G., Vladár, P., Márialigeti, K.: Molekuláris ujjlenyomat módszerek összehasonlító elemzése széleslevelű gyékény (*Typha latifolia*) rizoszféra baktériumközössége faji diverzitásának megismerésében. Magyar Mikrobiológiai Társaság 2002. évi Nagygyűlése, Balatonfüred, 2002. október 8-10.

Sipos, R., Nikolausz, M., Palatinszky, M., Székely, A., Márialigeti, K.: Biases in template-to-product ratios of multitemplate PCR in a model community. 14th International Congress of the Hungarian Society for Microbiology, Balatonfüred, Hungary, 9-11. October, 2003.

Sipos, R., Székely, A.J**.,** Berta, B., Bujdosó, L., Hajdú Cs., Márialigeti, K.: Tracking microbial communities during mushroom compost maturing with the prospect of selecting potential inoculants. 10th International Symposium on Microbial Ecology (ISME-10), Cancun, Mexico, August 22-27, 2004.

Sipos, R., Nikolausz, M., Palatinszky, M., Székely, A.J., Révész, S., Márialigeti, K.: Molekuláris ujjlenyomat módszerek (DGGE, TRFLP) tesztelése különböző univerzális primer párokkal. Magyar Mikrobiológiai Társaság 2004. évi Nagygyűlése, Keszthely, 2004. október 7-9.

Sipos, R., Mohr A., Mészáros, É., Cebe, G., Révész, S., M., Márialigeti, K.: The application of the MDA method in microbial community analysis of TCE contaminated groundwater. 2nd Central European Forum for Microbiology (CEFORM), Keszthely, Hungary, 7-9. October, 2009. Acta Microbiologica et Immunologica Hungarica 2009, 56 (suppl.): 240.