Faculty of Science of Eötvös Loránd University
Biology Doctoral School

Program of Experimental Plant Biology

Vajna Balázs

Characterization of bacterial community changes during oyster mushroom substrate preparation using molecular methods

The optimization and applicability of the T-RFLP data processing

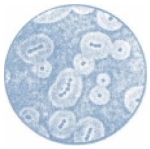
– a b s t r a c t o f P h D t h e s i s –

Supervisor:
Márialigeti Károly
(Department of Microbiology, ELTE)

Head of Doctoral School:
Dr. Erdei Anna
professor, head of department
(Department of Immunology, ELTE)

Head of Doctorate Program:
Dr. Szigeti Zoltán
professor, head of department
(Department of Plant Physiology and Molecular Plant Biology, ELTE)

Department of Microbiology, ELTE, 2010.



Introduction

The basis for the better understanding of the role and function of bacterial communities is the description of community composition. Traditionally this is done by culturing, where pure cultures are prepared and characterized. Culture-based methods are still needed today since they provide valuable information about the description of bacterial species. The main problem of present day microbial ecology is the question of the „uncultivable" bacteria, mainly the problem that traditional techniques might only show us „the tip of the microbial iceberg". While molecular methods provide new insights into revealing the „silent majority" of microbial diversity.

In investigation of bacterial diversity, molecular biological methods use partly the entire bacterial cell (e.g. *fluorescent in situ hybridization*), partly compounds extracted from the samples that are characteristic for the microbial community, mainly nucleic acids (DNA and RNA), fatty acids and quinones.

The most investigated region of the bacterial genome is the 16S rRNA gene for community analysis. Detailed investigation of this gene is usually carried out by clone library analysis. However, when handling large number of samples this procedure is costly and time consuming, thus in the last decade the use of several community fingerprinting techniques became widespread, such as DGGE (*denaturing gradient gel electrophoresis*), RISA (*ribosomal RNA intergenic spacer analysis*), SSCP (*single strand conformation polymorphism*) and T-RFLP (*terminal restriction fragment length polymorphism*). Among these techniques T-RFLP stands out for its precision, reproducibility and high resolution.

T-RFLP analyses can be divided into two principal stages. The first stage is the construction of the T-RFLP fingerprint consisting of three parts. (1) The appropriate region of the isolated DNA is amplified by 5' end labeled fluorescent primers. The obtained PCR products carry a label at one end. (2) These products are then digested with restriction endonuclease enzymes, resulting labeled — the terminal part of the cleaved DNA amplicon (T-RF, *terminal restriction fragment*) — and unlabeled DNA fragments of different sizes. The reaction mixture is then purified (e.g. by ethanol precipitation). (3) In the end, the labeled and unlabeled DNA fragments are separated by high resolution capillary gel electrophoresis, from which only the labeled DNA segments (T-RFs) are detected by laser induced fluorescence. The detection of T-RFs generates the T-RFLP chromatogram. The second part of the T-RFLP analysis contains the processing and interpretation of the results (the raw T-RFLP chromatograms) (Fig. 1.).

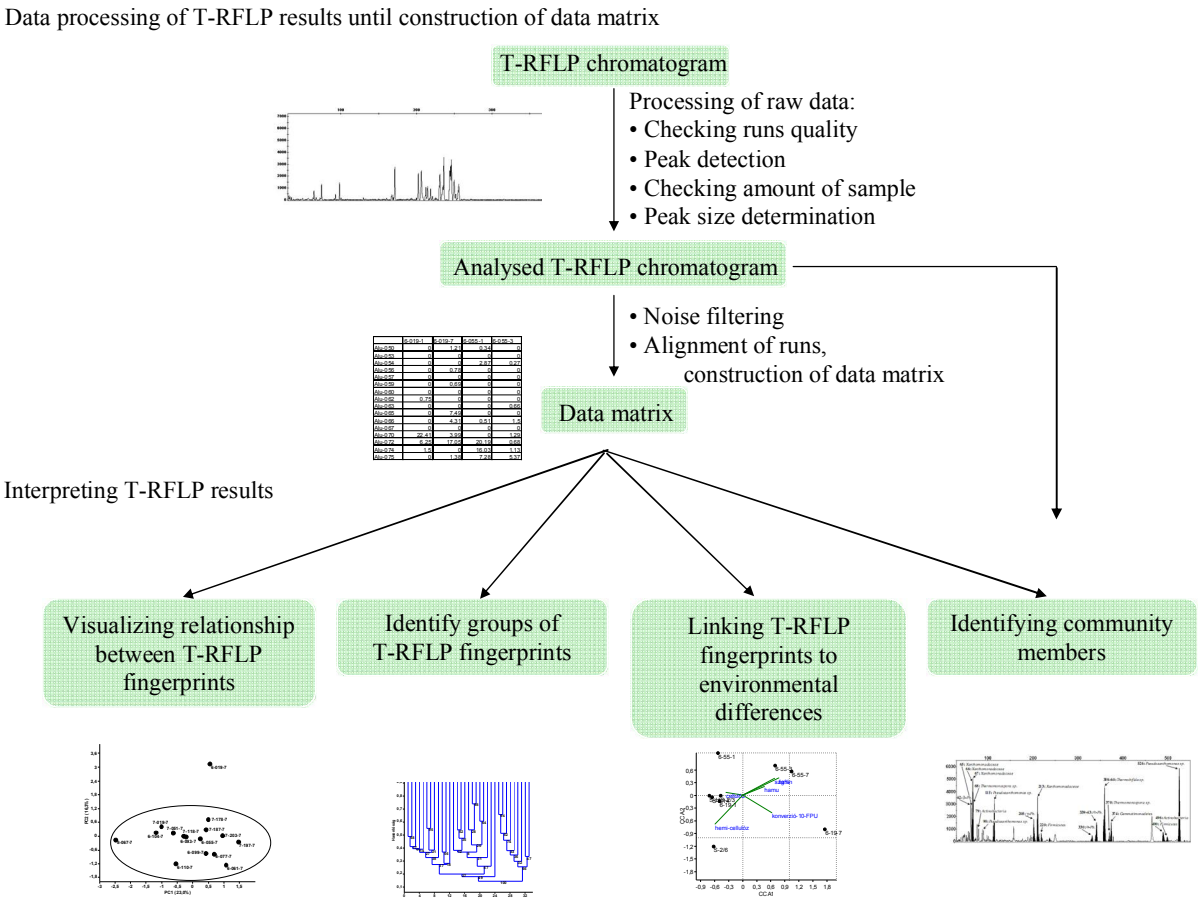


Figure 1.: The main steps of processing and interpretation of T-RFLP results.

Researchers applying the T-RFLP method often lack sufficient technical background information mostly for proper data processing. Therefore, we investigated the long established T-RFLP method in more detail, and used it for the monitoring of bacterial communities during oyster mushroom substrate production.

Oyster mushroom (*Pleurotus* spp.) production is second to that of button mushroom in Hungary. Although, oyster mushroom has long been produced, we acquired limited knowledge of the microbiota of the substrate, which is the main parameter of the quality of the substrate. Although, the microbial succession taking place during composting has long been investigated, this differs significantly from that of the quicker, more regulated mushroom substrate production, that is based on partial composting, pasteurization and conditioning.

Objectives

In the first part of the thesis, we investigated the following points:

1. Optimization of the T-RFLP data processing, which constitutes the examination of the following stages:
    a. Determination of the criteria, which have to be realized in order for a T-RFLP run to be considered for further analysis.
    b. Analysis of whether the processing of several parallel runs is necessary, or whether data of a single run per sample is sufficient.
    c. Assigning the principles of choosing the appropriate noise filtering method, that determines which peaks represent true peaks or whether they are considered as noise.
    d. Optimization of the alignment of data of individual T-RFLP runs, and data matrix construction; the process by which corresponding T-RF peaks could be identified between two different runs.
    e. Interpretation of the coded differences within the data matrix, corresponding to differences of T-RFLP runs. Presenting the technique of identifying the dominant T-RFs, causing the differences.
2. Generating a standardized protocol for T-RFLP data processing as a result of optimization. This protocol should contain all the methods available at each step of the analysis, highlighting the cases where other methods should be applied. The standardized protocol could serve as a strong scientific basis for future T-RFLP analysis in microbial ecological investigations.

In the second part of the thesis we have examined the microbiological aspects of oyster mushroom substrate production:

3. The investigation of the bacterial community of the oyster mushroom substrate and its changes during composting using the standardized T-RFLP data processing protocol.
4. Exploring the relationship between the amount of harvested mushroom from a given substrate and the microbiological and physico-chemical properties of the mature substrate.
5. The identification of the dominant bacteria characteristic of each substrate production phase with the help of T-RFLP and clone libraries.

Applied methods

The investigated oyster mushroom substrate was produced by the partial composting of wheat straw by Pilze-Nagy Ltd. During the period of 2006-2008 we have examined 16 substrate production series, taking samples at 3 stages of each series: (phase 1.) chopped and wetted wheat straw at the beginning of production; (phase 3.) substrate from the end of heap composting, before loading into the tunnel; (phase 7.) ready substrate from the end of the tunnel, just before spawning. The pH, the nitrogen, moister and ash contents of the ready substrate (phase 7.), were determined in an accredited laboratory (Bács-ÁG Ltd.), while the amount of harvested oyster mushroom from a given substrate was determined by the workers of the mushroom producing company.

The bacterial communities were analyzed by 16S rRNA gene based molecular biological techniques. The samples were physically disrupted by liquid nitrogen to access community genomic DNA, and the crude lysate was further purified using a silicate matrix-based method. The first part of the bacterial 16S rDNA region was amplified using HEX-27F (fluorescently labeled) – 534R primers during PCR. The products were digested with the restriction endonuclease enzymes *Alu*I and *Hin*6I respectively. The digests were purified by ethanol precipitation and finally separated by capillary electrophoresis, and only the fluorescently labeled terminal fragments were detected by laser excitation.

The processing of the T-RFLP chromatograms was carried out by the Microsatellite mode of the GeneMapper v3.7 (Applied Biosystems) program, and noise reduction was achieved by the T-REX online T-RFLP data processing software. The alignment of the individual T-RFLP runs was accomplished by the built-in nearest integer rounding of the T-REX program and the T-align module, as well as scripts in the R program. The determination of the optimal bin size (thus the determination of whether two successive T-RF peaks should correspond to the same category) was carried out by observing the extent of the T-RF drift (peak size difference of parallel runs of the same sample) calculated in the Excel program. During alignment of the T-RFLP profiles, the size of the individual peaks was standardized by the peak area data (total fluorescence). For further analysis, we have coupled the data matrices of the two restriction enzymes (*Alu*I and *Hin*6I) in order to get a more robust result.

The interpretation of the obtained data matrices of the T-RFLP data processing (ordination and dendrograms) was statistically evaluated using the Past and R programs. We calculated the diversity of each sample (based on Shannon and Simpson-indices) and a box-plot was generated to identify outliers according to these diversity indices. We tested the significance of sample clustering by the ANOSIM method, while the T-RF peaks responsible for the separation of samples were identified by the SIMPER method. The result of the different

ordination methods were compared by Procrustes analysis using the GenStat for Windows program.

During canonical analysis correlations were computed with Past program between sample coordinates at the first principal components and physico-chemical properties, as well amount of harvested mushroom, and significance of these correlations was determined. The analysis was supplemented with "envfit" script of R program, which fits environmental vectors onto PCA ordination, and determines the correlation coefficient and significances of this fitting by means of random permutations of the data.

As T-RFLP is not directly applicable for species identification, clone libraries were constructed from 3 samples of one production series and further 3 samples from mature substrate. For that purpose the first part of 16S rRNA gene was used, the same as for community T-RFLP. The T-RFLP analyses of the clones were conducted, and clones were clustered according to restriction sites. Those clones were selected, which had the same T-RFs as dominant community T-RFs of each phase or group, the latter in case of phase 7 samples. Sequence of group representatives were analysed. Resulting chromatograms were manually corrected with Chromas program, potentially chimeric sequences found with Mallard program were excluded. Sequences were compared using Blast program with GenBank and EzTaxon databases, the latter containing sequences of prokaryotic type strains.

Results and conclusions of the thesis

In the first part of the thesis T-RFLP data processing was optimized, then applied for the investigation of the bacterial community of the oyster mushroom substrate and its changes during composting. Later the relationship between the amount of harvested mushroom from a given substrate and the microbiological (T-RFLP data) and physico-chemical properties of the mature substrate was investigated. Finally the bacterial composition of each phase, especially phase 7, was characterized using clone libraries.

The description of the optimized T-RFLP data processing

The developed protocol is summarized as the followings:

1) First task is the **selection of adequate runs**. GeneMapper program checked automatically the quality of each run, and we accept only runs with good quality. Then runs were checked manually, but detected peaks were changed only in case of major bias. Only fragments longer than 35 bp were included, but fragments longer than 500 bp were retained. Finally only runs without "offscale" peak and having total fluorescence higher than 100000 RFU were accepted. These runs were uploaded to T-

REX online T-RFLP data processing software, where noise filtering and alignment of runs were concluded.

2) Mainly due to the cost and time limitations, only one run from each sample was used for analysis. Therefore, the extent of error due to neglecting the **variance between parallel runs of the same sample** was determined. To reduce this kind of error, first the data of the *Alu*I and *Hin*6I T-RFLP runs were combined; second, on the basis of the variance among several parallel runs, the maximum difference in base pairs between two peaks was determined, when two peaks can still be aggregated into the same category (bin). These bins can be applied directly for other samples, or new bins can be determined as described here.

3) For **noise filtering** methods of proportional threshold and „statistical determination of the threshold" were compared. For determining the actual level of noise filtering, it was taking into account, that at parallel runs only peaks larger than 0.2% relative area were present in every runs, so smaller peaks can be regarded as noise. Finally the method of „statistical determination of the threshold" built in T-REX program with minor changes was selected, with 4 SD thresholds.

4) The **alignment** of the T-RFLP profiles from different runs was carried out with the built in T-Align module using the variable bins option determined through the analysis of parallel runs. As T-REX is not directly suitable for using variable bins, first alignments with all bins were carried out separately, than respective parts from the alignment were combined. Finally the two matrixes from *Alu*I and *Hin*6I T-RFLP runs were united.

5) Based on the resulting data matrix, the **similarities/dissimilarities among the samples were visualized** with principal component analysis (PCA). Samples having very low Shannon- and Simpson diversity can bias the ordination or tree, so they were excluded from visualization. For selecting the right ordination method the beta-diversity of the data matrix was determined using dCA and calculating heterogeneity. Finally the analysis was supplemented with hierarchical clustering based on Bray-Curtis similarities and Simper analysis, together with the identification of $Simper_{50}$ peaks with clone libraries.

The resulting T-RFLP data processing protocol is directly applicable to the analysis of other samples, or with the decision-making mechanisms provided more appropriate protocols can be determined. Using this protocol T-RFLP can be applied for microbial ecology researches with a more stable scientific background.

Summarizing the results of bacterial community analysis of oyster mushroom substrate production and further perspectives

- The developed T-RFLP data processing protocol was applied successfully for the data of oyster mushroom bacterial community. Consequently the protocol is adequate for selected aim. In the near future it would be worth applying it to other type of samples, e.g. samples with lower diversity, having less dominant T-RF, without apparent succession.
- The succession of bacterial communities during oyster mushroom substrate production was demonstrated. A microbial succession during composting is already known form other composts, and it was detected in button mushroom substrate production too, but in the case of oyster mushroom substrate it has not been described in such details yet. The succession could be well followed on the principal component analysis plots, confirmed by significant differences based on ANOSIM analysis among each phase (Fig. 2.).
- It was found, that the microbial community structure of consecutive production series are quite similar to each other. It was hypothesized, that in tunnels spore-forming and thermophilic microbes could survive – despite cleaning and disinfection between consecutive production series – whose amount and composition were changing continuously. These bacteria could inoculate substrate arriving in to the tunnels again and again. For confirming this hypothesis the microbiota of the tunnel after cleaning and disinfection should be investigated.
- Using canonical analysis no significant correlations could be detected among bacterial fingerprints of the mature substrate and measured physico-chemical properties as well as respective mushroom yield. This negative result can be explained by several factors but the most important is the inadequate number of samples from a statistical point of view. But using canonical analysis it was demonstrated, how T-RFLP fingerprints and other environmental parameters can be compared using PCA and correlation analysis.

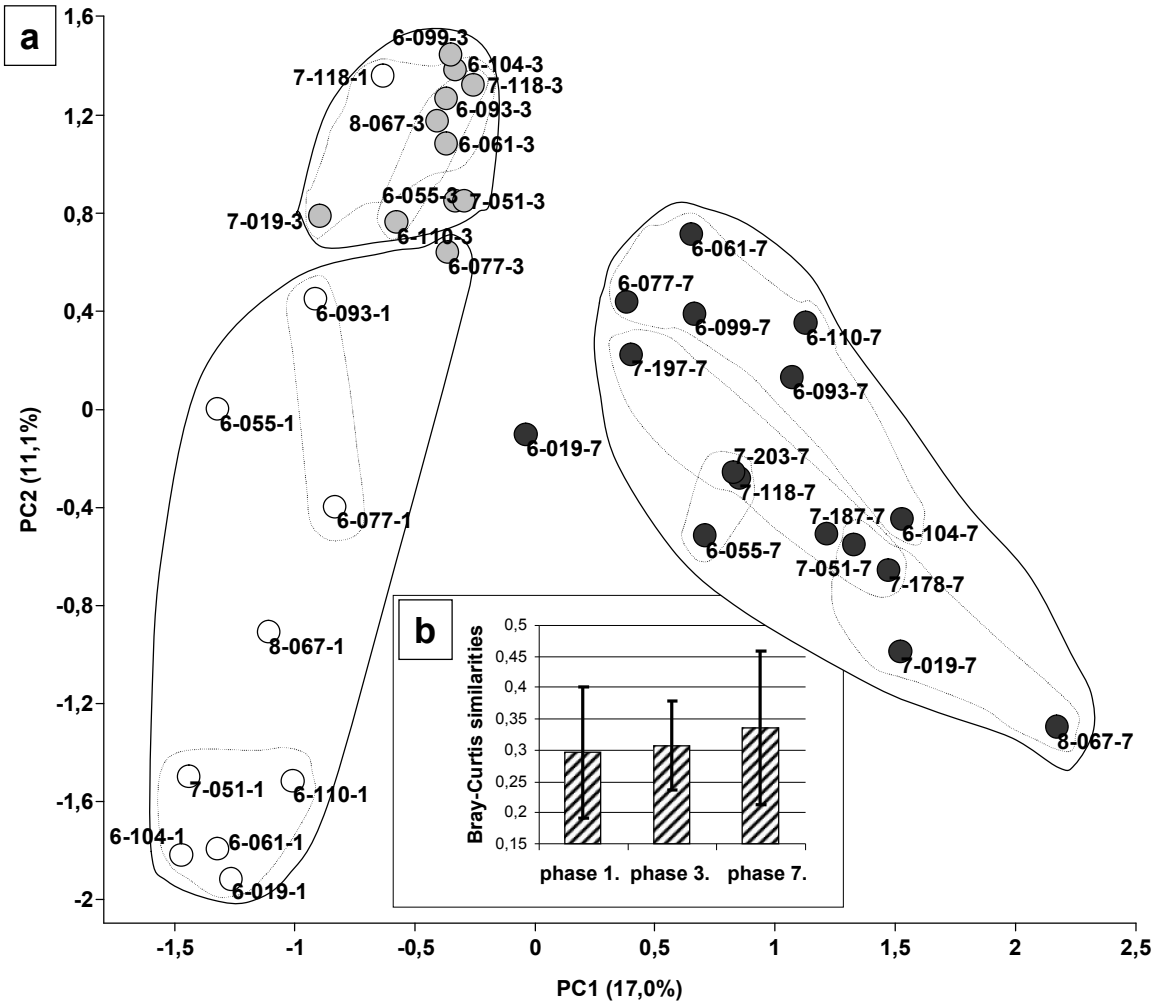


Figure 2.: (a) Comparing bacterial T-RFLP fingerprints of oyster mushroom substrate production with principal component analysis. Continuous and dotted lines refer to groups of UPGMA tree based on Bray-Curtis similarities. Groups surrounded by continuous lines have Bray-Curtis similarities more than 20%, groups surrounded by dotted lines have Bray-Curtis similarities more than 35%. Symbols: white – phase 1.; grey – phase 3.; black – phase 7. (b) Bray-Curtis similarities within each phase. Error bars indicate standard deviation.

- Most of the dominant T-RF-s were identified and their closest relatives were found using clone library analyses. At the beginning of substrate production ubiquitous mesophilic bacteria were typical. While at the end of the partial composting, *Bacillus* spp. and related species, as well as *Pseudoxanthomonas* spp. were the dominant microbes, that are well known from the thermophilic phase of other composts. Most of the dominant bacteria of the mature substrate could be assigned to Actinobacteria, Firmicutes and the genus *Thermus*.
- Dominant bacterial groups were present in every clone library constructed from mature substrate, although sometimes not the relatives of the same species were identified. This

support partially the theory of functional redundancy, which means that a given function of the community can be carried out by several different, but - in the case of the current study - related, bacteria.

- During substrate production the ratio of bacterial clones with no closer relatives was increasing constantly. Consequently it would be necessary to construct bacterial strain collection from the mature substrate for detailed analysis of unknown species.
- Furthermore the changing of the described bacterial community should be investigated during later part of mushroom production (substrate colonization by oyster mushroom mycelia and fruiting body production).

Publications laying the foundation of the thesis

Referred scientific research papers with impact factors:

Székely, A., J., Sipos, R., Berta, B., **Vajna, B.**, Hajdú, C., Márialigeti, K. 2009. DGGE and T-RFLP Analysis of Bacterial Succession during Mushroom Compost Production and Sequence-aided T-RFLP Profile of Mature Compost. Microbial Ecology 57, 522-33.

Vajna, B., Nagy, A., Sajben, E., Manczinger, L., Szijártó, N., Kádár, Zs., Bordás, D., Márialigeti, K. 2010. Microbial community structure changes during oyster mushroom substrate preparation. Applied Microbiology Biotechnology 86, 367-375.

Relevant conference lectures:

Vajna, B., Szili, D., Nagy, A., Márialigeti, K. 2008. Characterization of bacterial community changes during oyster mushroom substrate production. The 6th International Conference on Mushroom Biology and Mushroom Products, September 29. - October 3. 2008, Bonn, Germany, Book of Abstracts 52.

Other publications

Referred scientific research papers with impact factors:

Kovács, G., G., László, L., Bakos, A., Minarovits, J., Bishop, M., Ströbel, T., Mitrova, E., **Vajna, B.**, Majtényi, K. 2005. Increased incidence of genetic human prion disease in Hungary. Neurology 65, 1666-1669.

Mezei, M., Balog, K., Babic, D., Z., Toth, G., Cech, G., **Vajna, B.**, Tauber, T., Seme, K., Tomazic, J., Vidmar, L., Poljak, M., Minarovits, J. 2006. Genetic variability of *gag* and *env* regions of HIV-1 strains circulating in Slovenia. AIDS Research and Human Retroviruses 22, 109-113.

Borsodi, A., K., Makk, J., Rusznyák, A., **Vajna, B.**, Taba, Gy., Márialigeti, K. 2007. Phenotypic characterization and molecular taxonomic studies on *Bacillus* and related isolates from reed (*Phragmites australis*) periphyton. Aquatic Botany 86, 243-252.

Relevant conference lectures:

Vajna, B., Marialigeti, K. 2005. Determining dry matter- and nitrogen-content of phase II *Agaricus bisporus* compost by NIR-technique, as an example of characterizing the quality of mushroom compost. Acta Microbiologica et Immunologica. 52, S169.

Kanai, D., **Vajna, B.**, Márialigeti, K. 2009. Optimization of measuring cellulase and xylanase activity in oyster mushroom substrate. Acta Microbiologica et Immunologica Hungarica 56, 42-43.