Microarray-based Classification of Bacterial Strains

Goal. Goal of this proof-of-principle study is the classification of *Salmonella* strains based on genomic profiling. The method is intended to replace the established phage-based lysotypic method.

Introduction. A set of more than 60 arrays representing different bacterial strains of clinical and epidemiological relevance is investigated (Tschäpe, Robert Koch Institute, Wernigerode, Germany). A set of probes is established by differential amplification techniques (Hardt, ETH Zurich, Switzerland), Oligo-nucleotide microarrays are manufactured by Scienion AG, Berlin, Germany. Data analysis and classification methods are developed and performed by MicroDiscovery GmbH, Berlin, Germany.

Data Modelling. By suitable mathematical modelling of primary array data a discriminant function was derived. Based on this function, individual probes on the array could be assigned p-values resulting in a precise measure of reliability for each probe on the set of arrays. Starting from this description, a probability-based classification of the different strains is performed employing hierarchical clustering methods and a suitable distance measure.

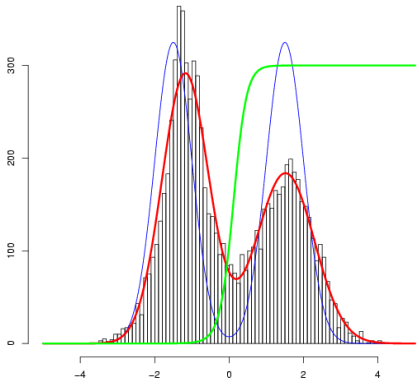

Fig. 1. Histogram and derived discriminant function obtained by fitting a two-component mixture distribution to the data

Results. Five reference strains appear as five distinct blocks in the upper left corner of the correlation diagram (orange arrows). Tested strains are forming two pronounced blocks, one of them being similar to the second reference strain (blue arrow), the other showing no significant similarity to the reference strains (pink arrow) (Fig. 2a). Hierarchical classification of the same data is shown in Fig 2b.

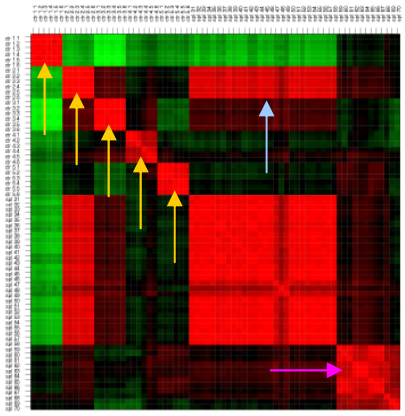

Fig. 2a. Correlation diagram of reference strains and a set of test data

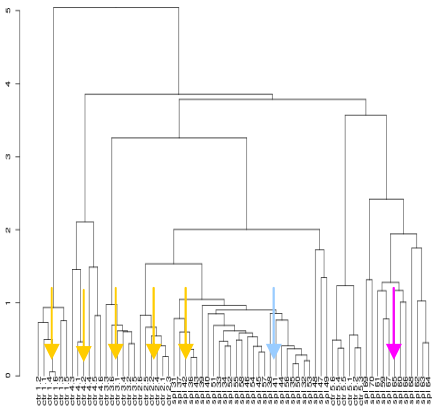

Fig. 2b. Hierarchical classification

Conclusion. This study shows the successful, efficient and robust classification of *Salmonella* strains. We were able to discover relevant markers, classify test samples according to similarity to the references and correctly flag classes that are not represented in the reference set. A novel customized data classification method developed by MicroDiscovery has proved crucial for this success. The approach will be developed further and extended to include more references and markers.

MicroDiscovery GmbH • Marienburger Straße 1 • 10405 Berlin
Tel: +49-(0)30-4435090-0 • Fax: +49-(0)30-4435090-10
johannes.schuchhardt@microdiscovery.de • www.microdiscovery.de