

Primary Biological DNA Database

D. Pickersgill^{1,2}, J. Wehking², I. Müller-Germann^{1,2}, U.Pöschl¹, V.R.Després²

¹Multiphase-Chemistry, Max Planck Institute for Chemistry, P.O. Box 3060, D-55020 Mainz, Germany

²General Botany, Johannes Gutenberg University, Saarstraße 1, D-55099 Mainz, Germany

Keywords: Biological Aerosol, database, pollen

This work focuses on the design and implementation of a Database (DB) for DNA fragments isolated from primary biological aerosol particle (PBAP) samples (Després et al. 2012 and references therein), along with all relevant information such as taxonomic, sampling, locational and meteorological information. A data structure was designed, according to the relational model (Codd 1970), that maximises flexibility, eliminates data redundancy and logically links all relevant data needed for analyses. The data structure was implemented using the open-source relational DB management system (RDBMS) MySQL.

For testing and subsequent analysis, data from 1179 plant sequences were imported, that were sampled in Mainz from March 2006 to April 2007. The sequences were attributed to 158 plant OTUs (Blaxter et al., 2005) which were used for taxonomic identification. Taxonomic tree information for the OTUs was stored according to the Nested-Set-Model (Celko, 2004).

Using SQL-queries, the plant data-set was analysed. Generalised queries were written that calculate Chao1, Shannon index, Shannon evenness, Simpson index, Simpson Diversity and the normalised species richness for the meteorological seasons including the normalised standard error. Using the generalised queries, values were extracted for various subsets of the data-set. Three methods were developed to normalise OTUs found on a sample, which were used for statistical analysis (Pickersgill 2013).

In addition data from a literature search concerning the discovered plant species was imported e.g. the pollen size, pollen count, or flowering period. This allows in-depth analysis e.g. the comparison of pollen size and species detection on coarse vs fine particle filter samples with a cut-off at 3 µm. Even though the literature shows that the pollen size starts at ~20 µm for the smallest pollen (Fig. 1), 42% of the total OTUs from Mainz were sampled on the fine particle filters. We assume that this plant-DNA originates either from other fragmented plant tissues or from pollen that have burst. This is interesting, as by this bursting process respirable allergens become free and cause pollinosis in humans.

There is a need for using DB management systems to research and analyze the growing data set under the background of pollinosis and other

interactions of bioaerosols with the biosphere and the atmosphere.

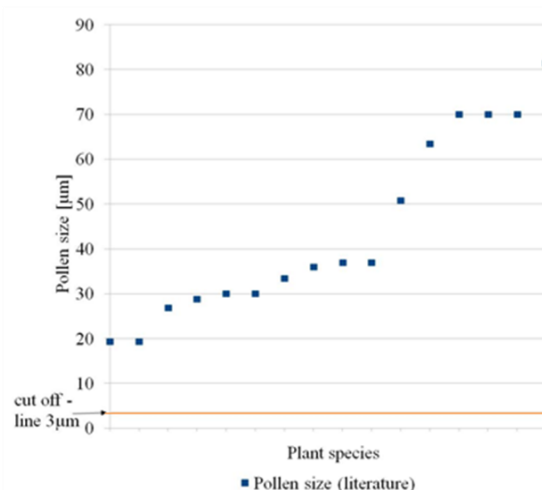


Figure 1. Pollen size (blue) of analyzed plant species of a fine particle filter with a cut-off of 3 µm (orange) in Mainz, Germany.

Acknowledgements

Max Planck Society are gratefully acknowledged for funding, J. Fröhlich-Nowoisky for sampling and DNA extraction, H. Schröder, M. Krupp and M. Geilke for informatics support.

References

- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). *Series B, Biological sciences*, 360(1462), 1935–43
- Celko, J. (2004). *Trees and Hierarchies in SQL for Smarties*. Elsevier Inc.
- Després, Viviane R., Alex Huffman, J., Burrows, S. M., Hoose, C., Safatov, A. S., Buryak, G., Fröhlich-Nowoisky, J., et al. (2012). *Tellus B*, 64.
- Pickersgill, Daniel. Design and Implementation of a Primary Bioaerosol DNA Database. Diploma thesis at the University Mainz, 2013.
- Solomon, P. A., Moyers, J. L., & Fletcher, R. A. (1983). *Aerosol Science and Technology*, 2(4), 455–464.