MS^n spectral tree databases: annotation and identification of plant metabolites

Justin J.J. van der Hooft[1,2,3], Miguel Rojas-Cherto[3], Thomas Hankemeier[3], Michael van Vliet[3], Theo Reijmers[3], Robert Hall[1,2,3] & Ric de Vos[1,2,3]

[1] Plant Research International, Wageningen-UR, Wageningen, The Netherlands
[2] Centre for BioSystems Genomics, Wageningen, The Netherlands
[3] Netherlands Metabolomics Centre, Leiden, The Netherlands
ric.devos@wur.nl

CBSG 2012 Centre for BioSystems Genomics | Netherlands Metabolomics Centre | TERPMED

Introduction and Aim

MS^n tools for annotation and identification of metabolites detected in LC-MS profiling are under development. Since MS^n spectral trees are robust,[1,2] MS^n databases and dedicated software able to process and handle MS^n data are key to efficiently implement MS^n in the metabolomics pipeline.

Here we examine the use of high mass resolution MS^n data as input for a plant spectral tree database that can be populated, visualized, and queried, using recently developed MetiTree software[3], encompassing MS^n processing tools like MEF[4] and a spectral tree comparison algorithm.[5]

MS^n data generation

High mass resolution MS^n fragmentation data was obtained by using a LC-PDA-LTQ MS-Orbitrap FTMS (Thermo) combined with a Nanomate fraction collector / injector (Advion) with a chip-based nanospray source (Figure 1).

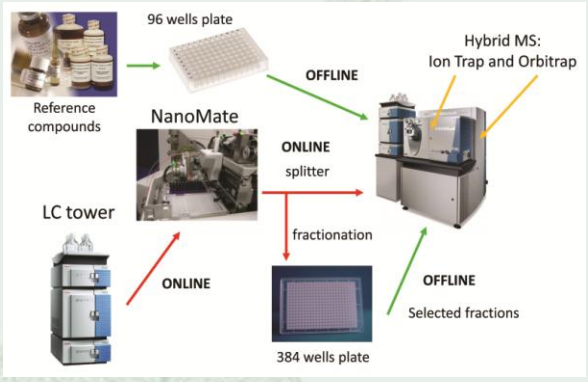

Figure 1: Schematic overview of MS^n spectral tree data generation.

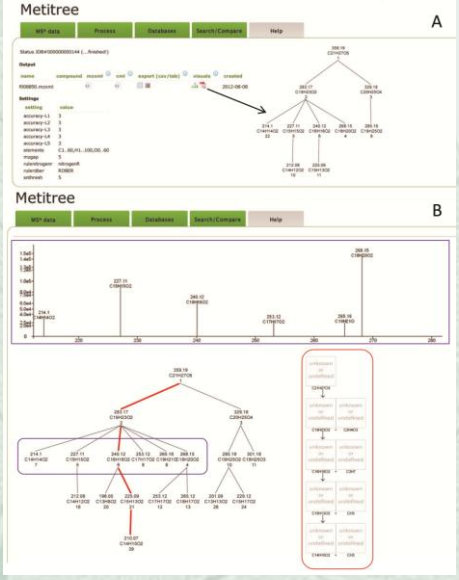

Figure 2: Metitree software tool. A) Metitree output and visualization of a processed spectral tree. B) Example of a tree of fragments (left) assigned with the corresponding elemental formulas of fragments (left) and losses (right) of a fragmentation path (marked in red) as obtained from a $C_{21}H_{28}O_5$ diterpene related compound. The MS^3 spectrum (marked in purple) is shown on top.

MS^n spectral tree processing

The MetiTree software assigns elemental formulas to both the parent ion of a fragmented metabolite and its daughter ions.[4] Using the hierarchical relation of the fragment ions, the elemental formulas are now the nodes of the spectral trees (Figure 2). Upon processing of the raw MS^n files, also chemical and electronic noise peaks are removed, which facilitates comparison between spectral trees. As proof of principle, a spectral tree database was created and populated with 15 (phenolic) diterpenes and related structures present in *Rosmarinus officinalis* trichomes for querying unknown structures (Figure 3).

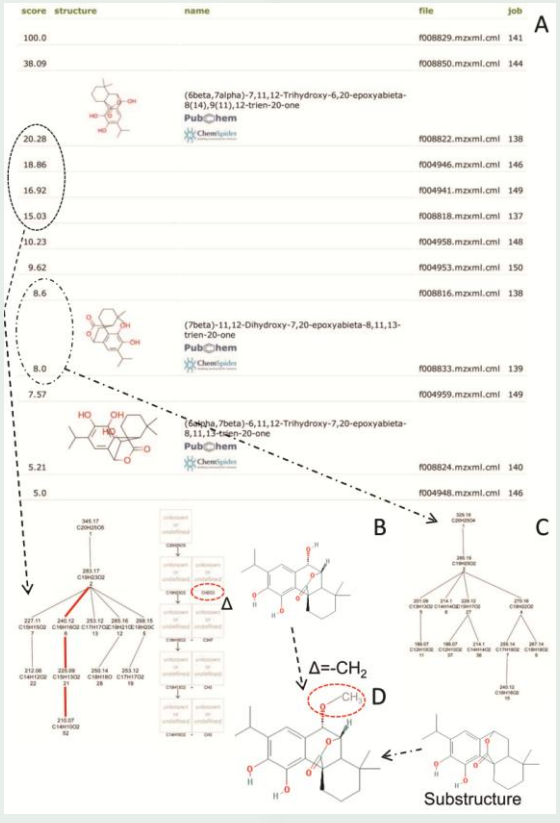

Figure 3: Automatic comparison of spectral trees by the Metitree software. A) A list of retrieved hits upon querying the database with an unknown $C_{21}H_{28}O_5$ terpene-like compound (Fig. 1B). Two groups of scores around 17.5 and 8.5 are highlighted, of which B) and C) are example spectral trees that belong to rosmanol ($C_{20}H_{26}O_5$) (B) and carnosol ($C_{20}H_{26}O_4$) (C). D) Based on structural and spectral comparisons, the structure in D is proposed.

Conclusions

- Robust MS^n data of plant metabolites can be used to populate spectral tree databases.
- Comparing spectral trees from yet unknown compounds to already fragmented metabolites using MS^n databases facilitates structural elucidation.

References

1. Van der Hooft, J. J. J. et al. (2011). "Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation." Anal. Chem. 83(1): 409-416.
2. Van der Hooft, J. J. J. et al. (2012). "Spectral trees as a robust annotation tool in LC–MS based metabolomics." Metabolomics 8(4): 691-703.
3. Rojas-Chertó, M. et al. (2012). "MetiTree: a web application to organize and process high resolution multi-stage mass spectrometry metabolomics data." Bioinformatics.
4. Rojas-Chertó, M. et al. (2011). "Elemental composition determination based on MSn." Bioinformatics 27(17): 2376-2383.
5. Rojas-Cherto, M. et al. (2012). "Metabolite identification using automated comparison of high resolution MSn spectral trees." Anal. Chem. 84(13): 5524-5534.

Centre for BioSystems Genomics P.O. Box 98 6700 AB Wageningen The Netherlands Tel: +31 317 481037 Email: secretariaat@cbsg.nl www.cbsg.nl