

Emma Veon & Bryan A. Hanson

Dept. of Chemistry & Biochemistry, DePauw University, Greencastle IN USA

Introduction

Plant metabolomics is a method of characterizing the state of plants and their associated chemical processes under a defined set of growth conditions. In such experiments, plants are grown in experimental conditions such as several levels of salinity or water availability. The tissue is then non-selectively extracted, and NMR spectra are obtained. The spectra reveal how the metabolome of the plant has changed in response to its environment. Developing a greater understanding of how plants respond to stressful conditions can give insight into how plants will cope with large-scale environmental changes, such as climate change.

Motivation

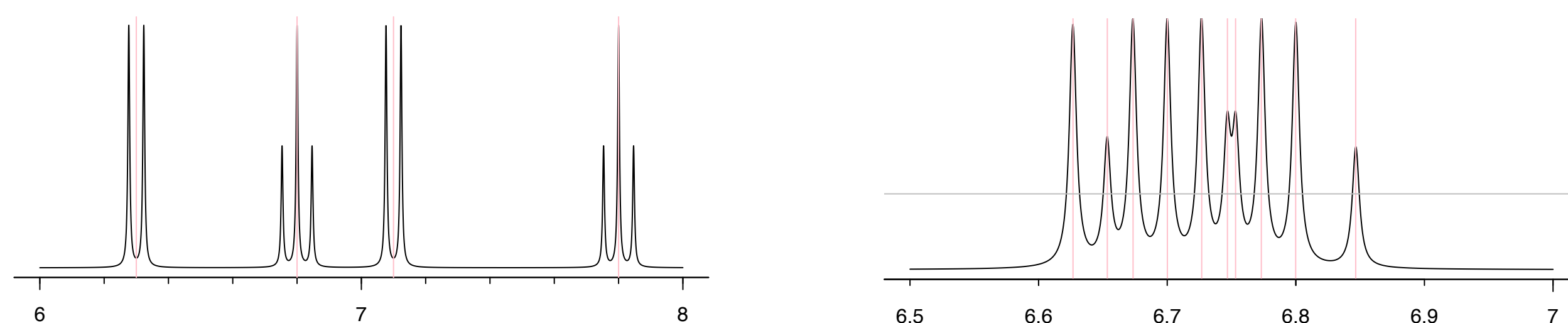
The NMR data collected from the plant tissue samples provides a statistical signature of all metabolites present. Through a chemometric analysis, one can determine which individual peaks – and later their corresponding metabolites – that have changed due to the growth conditions. In order to identify these key metabolites, one needs a database containing NMR spectra of known compounds to serve as a point of comparison. Currently, a number of public NMR databases exist, however they don't offer much in the way of search capabilities. Details of the search algorithms are generally unknown to the users, and basic search concepts such as "and" or "or" are not accessible.

Goal

The purpose of our research is to develop an NMR database containing ¹H NMR spectra of known metabolites. In order to overcome the limitations of currently available databases, we have built a package in the open source R language. The package `NMRdb` is equipped with appropriate search tools that give the user power and flexibility. The functions `searchNMRdb` and `guessFG` contained in the package allow for the greater control on the part of the user.

Strategy

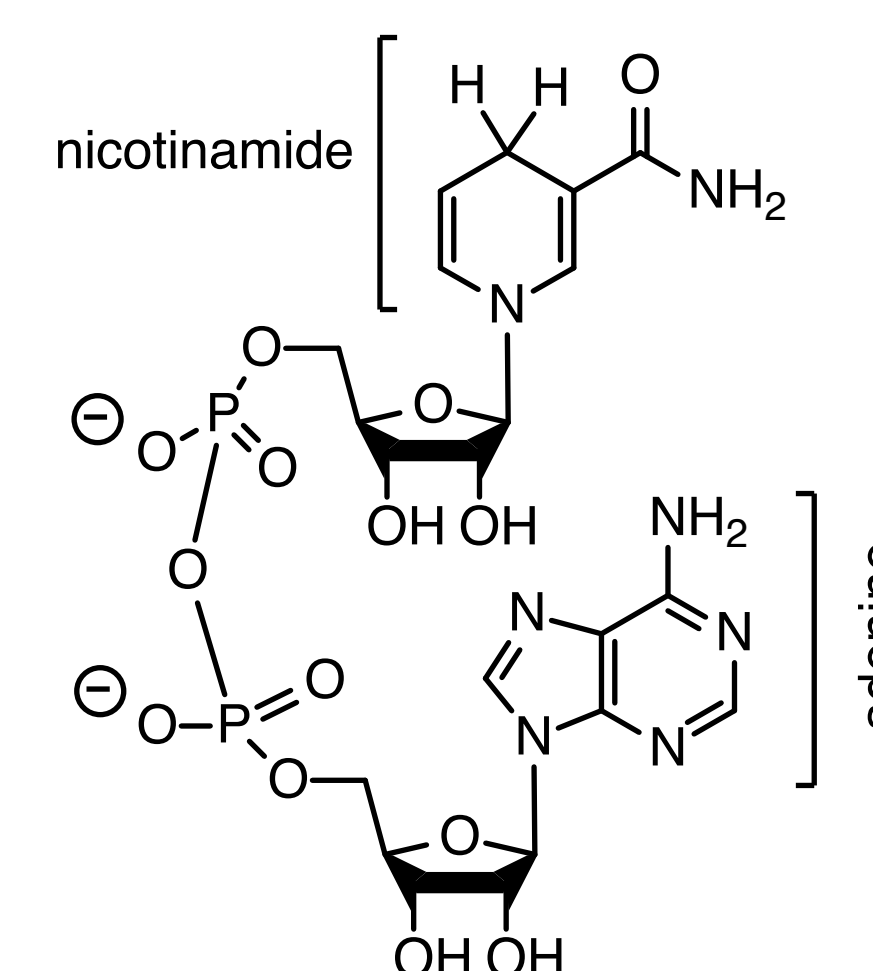
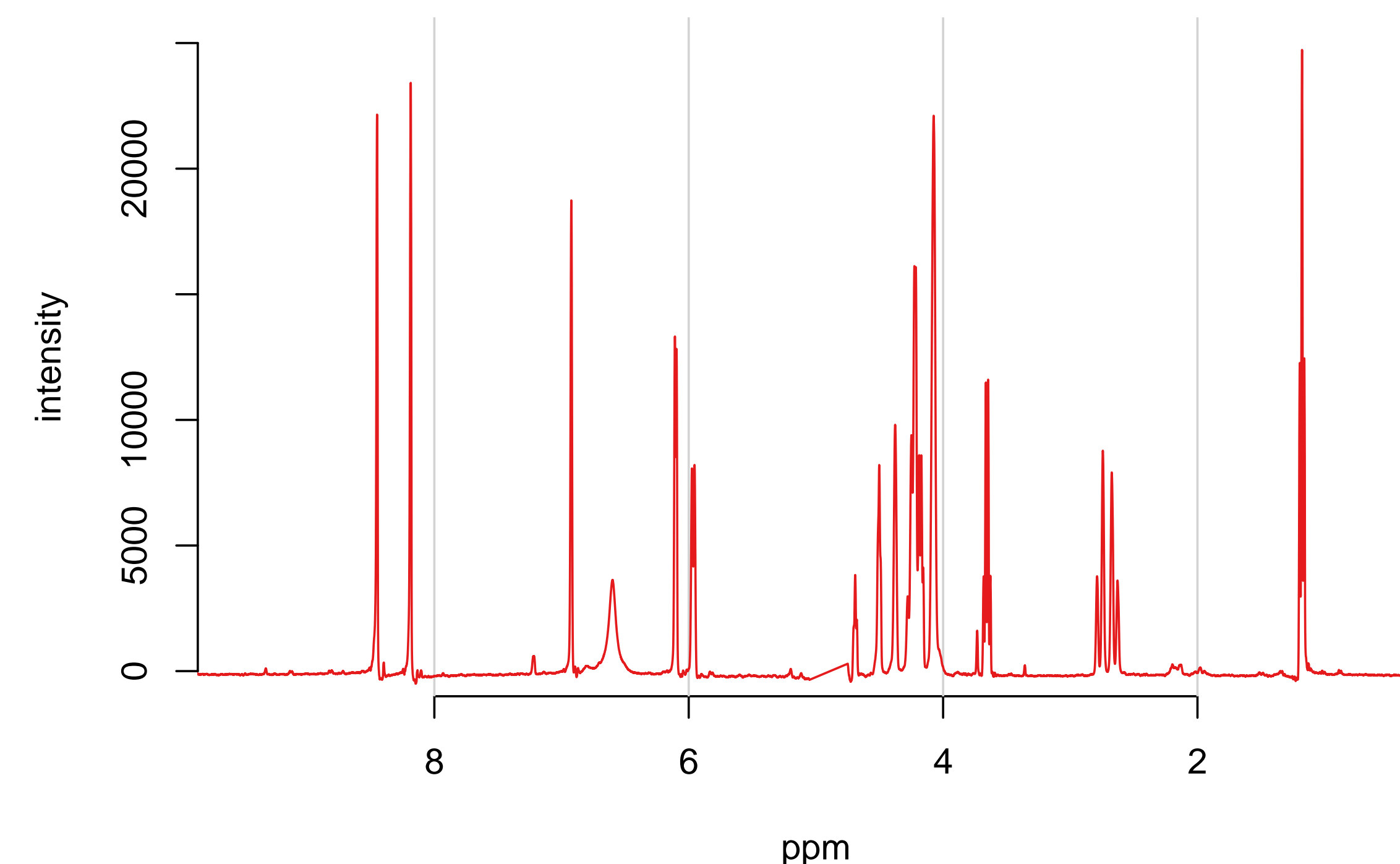
NMR spectra are typically reported as peak center, area, splitting, and sometimes coupling constant (J). This is fine for well-resolved and separated peaks, but useless for heavily overlapped peaks and mixtures of compounds seen in metabolomic studies. Since area and splitting information is lacking, we constructed our database by automatically detecting peaks above a threshold, without regard to whether the peak is part of a multiplet or not. While some information is discarded with this approach, it facilitates automatic matching against the database. The figure on the left below illustrates the traditional way of reporting peaks for an ortho aromatic ABCD pattern. However, if those peaks are not well separated, as on the right, it makes more sense to simply use the peak positions.



Sample Preparation, Acquisition & Processing

Briefly, samples were prepared with 5-10 mg or 10 μ L of each compound in 650 μ L of a 0.2 M pH 7.4 phosphate buffer. 50 μ L of 1% TSP was added. The ¹H NMR spectrum was collected using a presaturation sequence to suppress the water peak. The data were phased and referenced, then moved into the R package `ChemoSpec`. The partially suppressed water peak was removed, and the significant peaks identified using the `speak` package. Each spectrum was inspected for impurities and the peaks assigned.

Example: Structure and ¹H Spectrum of NADH



NMRdb Tools: Database Search

The `NMRdb` package is designed to allow users to search for multiple peaks in the curated NMR database. The function works by taking a vector of the NMR shifts, or peaks, and searching for matches in the database (peaks are actual peak maxima, not the center of idealized patterns). This is accomplished through the function `searchNMRdb`, which includes arguments for the shifts, the database to match against, and a window value which controls fuzzy vs. exact matching. The output is a data frame of possible matches. The table below shows the result of searching for peaks at 8.22, 6.01, 4.53, 3.72, 2.78, and 1.22 ppm. These are some of the peaks in the spectrum of NADH.

Compound	No. Hits	% Peaks Matched
citric acid	1	25.00
NADH	6	18.80
phytic acid	1	8.30
glycerol	1	7.70
NAD	2	6.50
spermidine	1	5.90
FAD	1	3.40
D-galactose	1	3.30
maltose	1	3.10
L-ribulose	1	3.00
D-ribose	1	2.80
D-erythrose	1	2.30

NMRdb Tools: Guessing the Functional Groups

If an solid identification cannot be made, one step to identifying a compound present in the sample is to identify the functional groups that are present. The function `guessFG` operates in a similar fashion to `searchNMRdb`, but returns a data frame of matching shifts and associated functional groups. The table below is based on searching the same peaks as in the previous example.

shift	functional group
4.53	CH-CH-PO4
4.53	CH-OH
3.72	CH-OH
6.01	CH-OR
2.78	CH2-COOH
2.78	CH2-NH2
3.72	CH2-OH
3.72	CH2-PO4
2.78	CH2-PO4
1.22	CH2-PO4
8.22	NR=CH=CR

Discussion

Thus far, ¹H NMR Spectra of 53 different compounds involved in the metabolic processes of plants have been collected and processed. Each spectrum was carefully inspected for impurities and assignments made. Spectra were processed as described above and incorporated into the database. The `NMRdb` package will allow users to search complex metabolomic data sets and identify compounds that may be up or down regulated in the experiment. Databases are always limited in that some metabolites are not available, not stable or not soluble. The `guessFG` function will allow users to identify potential functional groups in metabolites that are not present in the database.

Future Work

Future work will include:

- Collecting spectra of additional compounds involved in metabolic processes. This will increase the number and diversity of the compounds.
- Continue to expand the number of functions in the `NMRdb` package to make the database even more useful.
- Expand the types of spectra included in the database. Eventually we would like to develop methods which employ 2D NMR such as COSY and HSQC.

References

- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (ver. 3.4). www.R-project.org
- Bryan A. Hanson (2017) ChemoSpec: Exploratory Chemometrics for Spectroscopy (ver. 4.4.17). github.com/bryanhanson/ChemoSpec
- Bryan A. Hanson & Emma Veon (2017). NMRdb: NMR Spectra, Peak Database and Related Functions (ver. 0.1.95). github.com/bryanhanson/NMRdb
- Trung Nghia Vu, Kris Laukens and Dirk Valkenburg (2017). `speak`: Tools for Nuclear Magnetic Resonance Spectrum Alignment and Quantitative Analysis (ver. 1.2.3). CRAN.R-project.org/package=speak