# Getting Started

*Chase Clark*

*October 24, 2018*

Version 2 of IDBac is vastly different than IDBac version 1 (in a good way!). Here are some of the most important things to know.

### Converting Data

IDBac now accepts a variety of input formats including raw instrument files, mzML, mzXML, and txt.

### Experiments

Now IDBac organizes data into "Experiments".

**You can think of an "Experiment" as a collection of samples** For example: - *Experiment 1:* Bacterium_A, Bacterium_B, Bacterium_C - *Experiment 2:* Bacterium_X, Bacterium_Y, Bacterium_Z

So, instead of keeping track of "IDBac" folder, the software will store and keep track of all samples belonging to an "experiment" within a single SQLite file (these "experiment files" are transferable!). IDBac will show all available "experiments" in the "Select Experiment" tab.

This has the side-effect that it will be easy to extend this for use with our future, freely available, reference database.

### Protein Data Analyses

After in-house validation the "binning" algorithm has changed (for the better!) and so results from version 2 of IDBac will **NOT** be the same as those of previous versions of IDBac.

While Version 1 had Principal Components analysis (PCA), I have added Principal Coordinates Analysis (PCoA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) as options as well. Note: These are advanced statistical tools and users should be comfortable with their pros and cons before basing any conclusions on their representations.

### Small Molecule Analysis (Specialized Metabolite)

The binning algorithm had some issues with larger sample sizes. Therefore after many discussions we have taken the liberty of rounding $m/z$ values to help mitigate error, users should be aware of this.

IDBac now includes a 3-dimensional Principal Components plot (1st three dimensions) tha appears above the MAN.