TÀI LIỆU THAM KHẢO XÂY DỰNG TIÊU CHUẨN

Gỗ - Xác định loài bằng công nghệ quang phổ khối lượng

Phần 1: Thuật ngữ và định nghĩa [1], [2], [3] [4]

Phần 2: Phương pháp lấy mẫu [1]

Phần 3: Phương pháp xây dựng cơ sở dữ liệu [2], [5] [6]

Phần 4: Phương pháp xác định loại gỗ [2]

DANH MỤC TÀI LIỆU THAM KHẢO

[1] Schmitz, N., Blanc-Jolivet, C., Cervera, M. T., Chavesta, M., Cronn, R. C., Deklerck, V., ... & Wiemann, M. C. (2019). Mạng lưới theo dõi gỗ toàn cầu (GTTN) - *Hướng dẫn lấy mẫu*. Xây dựng tiêu chuẩn quốc tế và cơ sở dữ liệu của GTTN.

[2] Schmitz, N., Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J. W., Cabezas, J. A., ... & Zuidema, P. A. (2020). *Tổng quan về các phương pháp sử dụng trong giám định gỗ*. Hướng dẫn về các phương pháp truy xuất gỗ.

[3] US-WISC. *Giám định thực vật bằng công nghệ quang phổ khối lượng (DART-TOFMS)*.

[4] US-WISC. *Hướng dẫn phân tích giám định bằng DART TOFMS và Thu thập dữ liệu*.

[5] US-WISC. *Hướng dẫn đặt tên tệp dữ liệu*

[6] US-WISC. *Hướng dẫn tạo thư viện NIST*

GTTN
Global Timber
Tracking Network

General sampling guide

Task 2: Development of international standards and GTTN database
Activity 2.1: International standards
Deliverable 2.1.1: Reviewed GTTN guidelines on sampling of reference material
Date: 7 February 2019

GTTN
Global Timber
Tracking Network






Recommended citation:
Schmitz, N. (ed.), Blanc-Jolivet, C., Boner, M., Cervera, M.T., Chavesta, M., Cronn, R., Degen, B., Deklerck, V., Diaz-Sala, C., Dormontt, E., Ekué, M., Espinoza, E.O., Gasson, P., Gehl, D., Gehre, M., Haag, V., Hermanson, J.C., Honorio-Coronado, E., Koch, G., Lancaster, C., Lens, F., Liendo-Hoyos, E.P., Martínez-Jarquín, S., Montenegro, R., Paredes-Villanueva, K., Pastore, T., Ramananantoandro, T., Rauber-Coradin, V.T., Ravaomanalina, H., Rees, G., Sebbenn, A.M., Tysklind, N., Vlam, M., Watkinson, C., Wiemann, M. 2019. General sampling guide for timber tracking. Global Timber Tracking Network, GTTN Secretariat, European Forest Institute and Thuenen Institute.

DOI:10.13140/RG.2.2.26883.96806

The following document is an output from GTTN, which is coordinated by the European Forest Institute (EFI) and funded by the German Federal Ministry for Food and Agriculture (BMEL). The information expressed in this document are the views of the authors and do not necessarily represent those of the donor or of the European Forest Institute.

GTTN
Global Timber
Tracking Network

General sampling guide for timber tracking

How to collect reference samples for timber identification

Editor: Nele Schmitz

Authors*: Céline Blanc-Jolivet[1], María Teresa Cervera[2], Manuel Chavesta[3], Richard Cronn[4], Victor Deklerck[5], Carmen Diaz-Sala[6], Eleanor Dormontt[7], Peter Gasson[8], David Gehl[9], Volker Haag[10], John C. Hermanson[11], Eurídice Honorio Coronado[12], Cady Lancaster[13], Frederic Lens[14], Estephanie Patricia Liendo Hoyos[15], Sandra Martínez-Jarquín[16], Rolando Montenegro[3], Kathelyn Paredes Villanueva[17,18], Tereza Pastore[19], Tahiana Ramananantoandro[20], Harisoa Ravaomanalina[21], Alexandre Magno Sebbenn[22], Niklas Tysklind[23], Mart Vlam[18], Charlie Watkinson[24], Michael Wiemann[11]

With contributions from*: Markus Boner, Bernd Degen[1], Marius Ekué, Edgard O. Espinoza[25], Matthias Gehre[26], Gerald Koch[10], Vera T. Rauber Coradin[19], Gareth Rees[27]

We also thank Tom De Mil and Priwan Srisom for their photo contributions.

*Names are listed in alphabetical order.

Edition 7 Feb. 2019

[1] Thünen Institute of Forest Genetics, Grosshansdorf, Germany
[2] Centro de Investigación Forestal, El Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CIFOR), Madrid, Spain
[3] Laboratorio de anatomía e identificación de maderas, Universidad Nacional Agraria la Molina, Lima, Peru
[4] US Forest Service Research & Development, Washington, DC, USA
[5] Laboratory of Wood Technology, Ghent University, Belgium
[6] Dept. Ciencias de la Vida, Universidad de Alcalá, Madrid, Spain
[7] University of Adelaide, Adelaide, Australia
[8] Royal Botanic Gardens, Kew, UK
[9] Environmental Investigation Agency, Washington, DC, USA
[10] Thünen Institute of Wood Research, Bergedorf, Germany
[11] US Forest Service Forest Products Laboratory, Madison, WI, USA
[12] Instituto de Investigaciones de la Amazonía Peruana (IIAP), Iquitos, Peru
[13] Wood Identification & Screening Center, USFS International Programs, National Forensics Laboratory, U.S. Fish & Wildlife Service, Ashland, OR, USA
[14] Naturalis Biodiversity Center, Leiden, The Netherlands
[15] Autoridad de Fiscalización y Control Social de Bosques y Tierra, Santa Cruz de la Sierra, Bolivia
[16] Department of Biochemistry and Biotechnology, CINVESTAV Unidad Irapuato, Irapuato, Mexico
[17] Universidad Autónoma Gabriel René Moreno, Santa Cruz, Bolivia
[18] Wageningen University & Research, Wageningen, The Netherlands
[19] Laboratório de Produtos Florestais, Serviço Florestal Brasileiro, Brasília, Brazil
[20] Mention Foresterie et Environnement, Ecole Supérieure des Sciences Agronomiques, Université d'Antananarivo, Antananarivo, Madagascar
[21] Mention Biologie et Ecologie Végétales, Faculté des Sciences, Université d'Antananarivo, Antananarivo, Madagascar
[22] Instituto Floresta de São Paulo, São Paulo, Brazil
[23] Institut National de Recherche Agricole, Kourou, Guyane Française
[24] Agroisolab UK Ltd, York, UK
[25] National Forensics Laboratory, US Fish & Wildlife Service, Ashland, OR, USA
[26] Laboratory for Stable Isotopes, Department for Isotope Biogeochemistry, Helmholtz-Centre for Environmental Research, Leipzig, Germany
[27] Elementar UK Ltd, Stockport, UK

GTTN
Global Timber
Tracking Network

Table of Contents

Rationale

This is a guide for the collection of *reference samples* of trees to enable the **identification of species and/or geographical origin of woody material**. It is an update of the sampling section of the GTTN standards and guidelines (Ekué 2014) and builds further on a discussion initiated during a workshop held in Hamburg at the Thünen Institute for Wood Research in 2014. If you are looking for support on how to collect *test samples*, see the UNODC guide (UNODC 2016).

To enable the implementation of the different laws regulating the trade in illegal wood, **reference databases** for various timber tracking tools are urgently needed for at least the most traded and endangered tree species. The Global Timber Tracking Network (GTTN) is building a central database where not only the reference data can be stored but which will also function as a sample locator. Having a common sampling guide will facilitate meaningful exchange of samples.

In addition, to optimise the use of wood/wood product identification (taxonomic identity or geographic origin) in support of law enforcement, the guide anticipates upcoming developments to combine (Paredes Villanueva 2018) different timber identification methods (Dormontt *et al.* 2015, Lowe *et al.* 2016) such as **wood anatomy** (Koch and Schmitt 2015, Helmling *et al.* 2018), **DNA-based methods** (Jolivet and Degen 2012, Blanc-Jolivet *et al.* 2018, Chaves *et al.* 2018), **stable isotopes** (Paredes-Villanueva *et al.* in preparation, Vlam *et al.* 2018), **DART TOFMS** (Lancaster and Espinoza 2012, Espinoza *et al.* 2015, Deklerck *et al.* 2017, Paredes-Villanueva *et al.* 2018) and **NIRS** (Pastore *et al.* 2011, Bergo *et al.* 2016, Snel *et al.* 2018). This sampling guide is written to make sharing of samples between researchers specialised in different timber tracking methods possible, as samples should ideally come from the same location in the tree, from the same individual and from well-identified trees when combining methods.

This guide is intended for scientists, to provide all the information needed to get the most out of sampling campaigns for timber identification purposes. This information should allow setting up a sampling protocol adapted to the specific goal of the research project, the conditions of the sampling area and the background of the people who will do the sampling. Note that this guide is to collect reference samples and hence relatively high amounts of samples from different individuals are needed to take the variability of a species into account. Once reference data have been developed for a tree species for one or more identification methods, however, only one sample of an unidentified wooden object is often sufficient to determine its identity.

Abbreviations

AAC	Assiettes Annuelles de Coupe (Annual Cutting Area)
°C	Degrees Celsius
Ca.	Circa
CITES	Convention on International Trade in Endangered Species of wild fauna and flora
Ø	Diameter
DART TOFMS	Direct Analysis in Real Time Time-of-Flight Mass Spectrometry
DBH	Diameter at Breast Height
DF10	Document specifying the timbers extracted from the forest
DNA	DeoxyriboNucleic Acid
e.g.	for example
EUTR	EUropean Timber Regulation
GPS	Global Positioning System
GTTN	Global Timber Tracking Network
ID	Identification
Min.	Minimum
NGO	Non-Governmental Organisation
NIRS	Near InfraRed Spectroscopy
Pvc	Polyvinyl chloride
RH	Relative Humidity
Sample ID	Sample IDentity
UNODC	United Nations Office on Drugs and Crime

Quick guide

The ideal reference sample collection for timber identification

Preparation ➤ p. 8

Local support & Code of conduct
- Identify local expertise and knowledge to build a strong sampling team
- Follow local up to international regulations

Budget
- Often underestimated
- Plan well!

Sampling design
- Scientific: review on site & species
- Practical: time, tools, transport

Field work ➤ p. 17

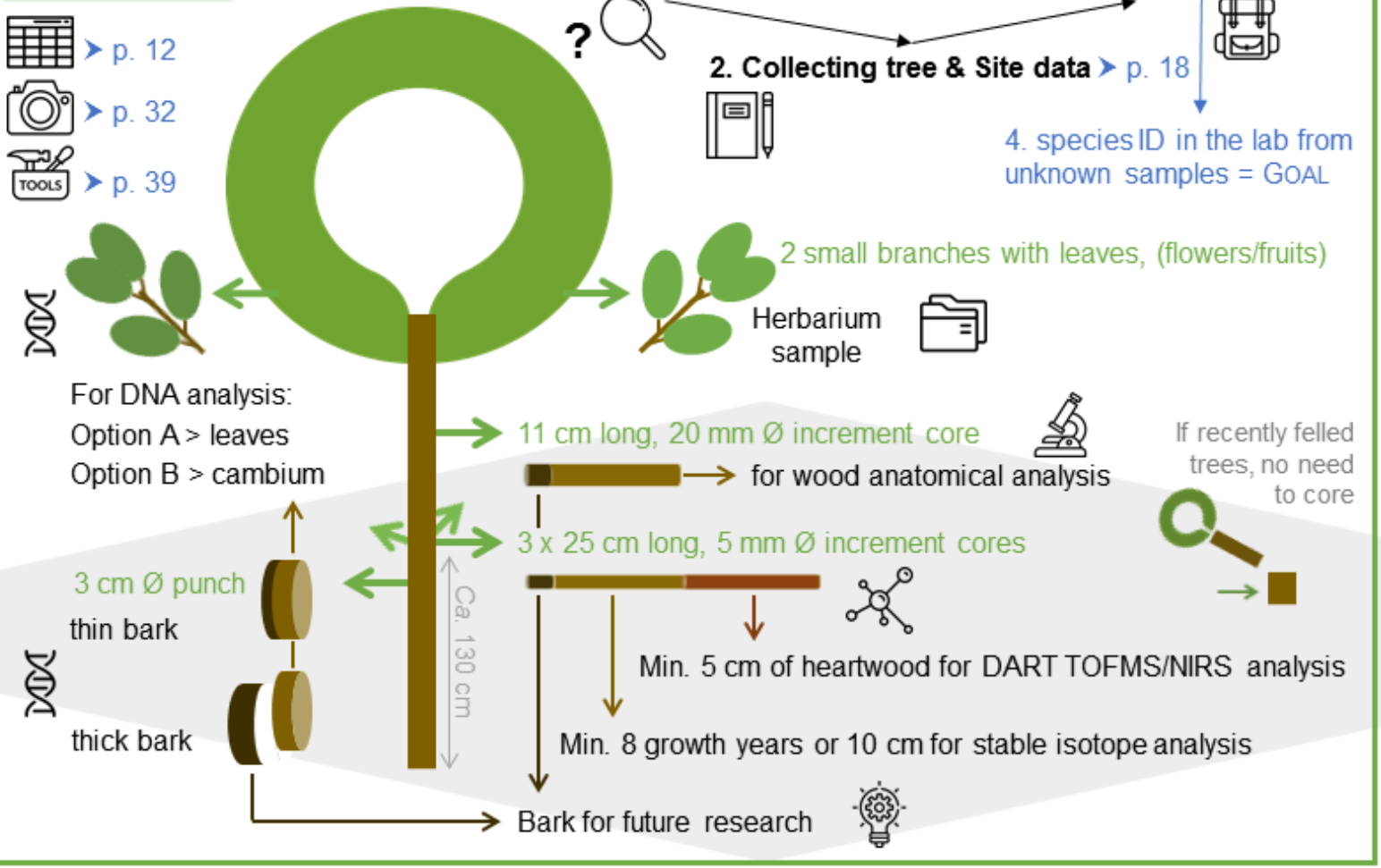
1. Species ID in the field ➤ p. 17 → 2. Collecting tree & Site data ➤ p. 18 → 3. Collecting samples ➤ p. 20

4. species ID in the lab from unknown samples = GOAL

➤ p. 12

➤ p. 32

TOOLS ➤ p. 39



Storage & Transport ➤ p. 25

Assuring sample quality
- Forest to lab sample chain
- Unique & consistent labels
- Correct storage in the field and long term

↓

Prevent:
- moulding
- bacterial or insect damage
- DNA degradation
- contamination
- tissue shrinkage (for wood anatomy)

- Verify and deposit herbarium vouchers and duplicate wood samples at public reference collections

Check lists

Checklist preparatory work

Before all else:

1. Did I consider costs for permits, transport of the sampling team, transport of samples back to the lab, payment of sampling team, accommodation and subsistence, sampling material and equipment? ➤ 1.1-1.2
2. Did I get permits to do research in the different sampling sites, to collect samples and to export and import them? ➤ 1.1-1.2
3. Did I explore the available local knowledge and expertise and find local partners to build a local sampling team? ➤ 1.3

Specifying the aim of the mission:

4. Did I clarify the research question of the sampling campaign? ➤ 1.4.1
5. Did I do a scientific literature review on the species and sites that will be sampled to collect all basic information required? ➤ 1.4.1

To decide beforehand:

6. Did I decide on how to select sites and trees within sites? ➤ Table 1, 1.4.1
7. Did I decide on the amount of material that will be sampled (based on budget and essential quantities)? ➤ Table 1, 2.3.1
8. Did I decide on the site and tree data that will be collected and how? ➤ 2.2, 3.1, appendix 3
9. Did I decide on how samples will be stored in the field, during transport and when back at the lab? ➤ 2.3.2-3, 3.2-3.4
10. Did I decide on the material and equipment to be used? ➤ Table 1, Appendix 2
11. Did I decide on a labelling code? ➤ 1.4.2
12. Did I decide on all other practicalities for the field work? ➤ 1.4.2

Checklist fieldwork

Packing:

1. Do I have all required material and equipment for the amount of samples that I want to sample? ➤ Appendix 2
2. Do I know how to label or is all material pre-labelled? ➤ 1.4.2
3. Do I have what is needed to identify the tree species of interest in the field? ➤ 2.1

At the field site:

4. Start recording the field trip in your notebook/on your template form ➤ 2.2, Appendix 3
5. Collect site information ➤ 2.2
6. Collect herbarium material and leaf samples ➤ 2.3.2
7. Collect wood samples ➤ 2.3.3
8. Collect and record all tree info ➤ 2.2

At the field station/camping area:

1. Dry wood cores/samples and change humid silica for fresh one ➤ 3.2
2. Assemble herbarium specimens if not done yet, change humid newspapers for dry ones or add alcohol if drying the herbarium material later ➤ 3.2
3. Check, complete and organise field notes where needed, digitise if already possible ➤ 3.1

1. Preparatory work

1.1 Code of Conduct

The first principle that has to be considered is the sovereign rights of states over their forest resources. Collection, transport, processing, management and storage of material from forest trees have to be performed in accordance with the **national and local regulations** (ask for information from *e.g.* your local partner(s), forester, concession/land owner, park authorities). In addition, the sampling campaign should be in line with the existing **regional regulations** such as the EUTR, the US Lacey Act and the Australia Illegal Logging Prohibition Act (see *e.g.* here for more information) and with **international regulations** such as CITES and the Nagoya protocol (an explanatory guide can be found here). For information about the requirements concerning CITES listed species you can contact national CITES authorities.

Accordingly, research **permits** for field collection, Material Transfer Agreements or other appropriate documentation must be requested well in advance to ensure the correct collection, transport and management of the forest tree material harvested and stored as reference samples. In addition, the **community/ies living in the area of sampling need to be informed** on the sampling campaign (as some might for example be worried the bore holes will damage the trees).

1.2 Budget

Sampling costs are often underestimated. Before planning your sampling campaign contact the GTTN network and the GTTN followers via the ResearchGate project page to find out if you can team up with others interested in sampling in the region to make the trip more cost-efficient. It is advisable to account for the following expenses when budgeting:

- **Any fees related to getting permission** and support from both national and local authorities for the planned sampling and for transportation of the samples from the field to the lab.
- **Transportation to the different sampling sites**: costs will be related to accessibility. Inform yourself on the means and duration of transportation required to reach the different sampling sites and the related costs (vehicle, driver, fuel costs).

- **Transportation and/or shipping of the samples** to the laboratory, including potentially required phytosanitary certificates.
- **Payment for assistance** by people knowing (i) the area and (ii) the tree species during the entire journey to and in the forest. Consider sampling efficiencies as low as 10 trees per day for tree species with low densities.
- **Accommodation and subsistence.**
- **Sampling material and equipment** (see Appendix 2).

TIP: If you will need a **car** and you have the choice, pick one with a functioning cigarette lighter (accessory power outlet). This will enable you to charge batteries (for GPS, electric increment borer, laser meter, camera, computer) in the car when needed.

TIP: To be able **to estimate the sampling work that can be done in one day** if samples are taken as described in *§2.3 Collecting samples*, it is advisable to do field tests with the sampling team. The duration of a sampling campaign will depend on variables such as: species density, available equipment (*e.g.* mechanical or hand borer), time needed to get to the canopy (to collect leaves), chosen intensity of herbarium specimen collection, number of timber identification methods material is collected for, experience of the field team.

1.3 Local support

1.3.1 Find a local partner institute

TIP: It is recommended to include local partners from the project design onwards to make sure that the project interests both sides and the local partner does not just serve as a collector.

Identification of **local partners** (universities, research institutes, NGOs, companies, …) which already have expertise and/or interest in timber identification techniques and/or have some infrastructure, material and trained personnel.

The local partner will be able to advise on a **local botanist/(para)taxonomist, an experienced driver and a field guide**, who know the area and its species as well as its dangers. They are an indispensable part of the field team as guides in the forest to find the targeted trees, facilitate interaction with local communities and to reduce the risk of attacks from animals or hostile people (*e.g.* illegal loggers, miners).

Get advice from your local partner on how to get the required **permit(s) to collect and export** samples and who should be contacted before arriving at the different sites you want to sample (*e.g.* community leaders, officials, company personnel). Check if some physical samples can be stored in a local herbarium (see Index

Herbariorum) and/or xylarium (see Index Xylariorum) and taxonomically identified by specialists (start with checking the GTTN network to find contacts).

Identify **local students** who are working or might work on the species of interest and might be interested in co-authoring the research papers and/or to participate in the expedition.

1.3.2 Set up a local sampling team

Create a base of trust both with the local community and within the sampling team before starting the sampling campaign and make sure everyone knows the role and responsibility of each other. In case the principal investigator cannot participate for the full length of the sampling campaign, his/her presence at the start of the sampling is necessary to train the people who will do the sampling and adapt the sampling protocol if necessary.

- Use the local knowledge on species identity, variability, density and sites of occurrence provided by botanists, ecologists, local guides and collaborators.
- At least one person should be scientifically trained and understand the reasoning behind the sampling design and be responsible for oversight of the sample collection accordingly, for note taking and for correct GPS reading.
- At least one person should be technically trained and responsible for sample collection according to protocol and maintenance of equipment.
- Depending on the conditions additional expertise might be necessary: a person that can use a gun, a driver used to the terrain that will be sampled, a tree climber, a person trained in using a sling shot.

1.4 Sampling design

1.4.1 Scientific set-up

To be able to set-up the sampling design a **scientific literature review** and general information search should be undertaken to collect as much information as possible on the species and geographic locations of interest. The thoroughness of the review on the geographic location(s) will depend on the **goal of the sampling**, species or origin identification and the required resolution of the origin identification. Table 1 gives an overview of the reference material that needs to be collected to allow species or origin identification using the different tracking methods.

Information that should be collected (where applicable for the specific wood identification goal of the sampling):

- **To decide on where to go sampling** (which countries and locations)
  - samples already available (check GTTN's reference database)
  - species distribution (focus on natural occurrence not on political borders)
  - intraspecific species diversity (genetic variation, which might also influence anatomical and chemical properties)
  - species abundance (a minimum of 20 individual trees per species of interest should be available for sampling in an area of 1 km^{2*})
  - spatial distribution of species in forest concession (forest inventory map)
  - environmental variation (include as much as possible)
  - chance of getting a permit to sample at the sites of interest
  - accessibility and feasibility (infrastructure)
  - safety (political situation, terrain)
  - relevance for the timber trade (areas where legal and/or illegal harvesting is currently happening, or where it is projected to happen)
  - risk of endangering the species population[†]
  - possibility to partner with a concession holder and to sample during or shortly after logging (within one week at most and with trees still lying at the felling site, to guarantee fresh wood and leaves and the leaves' origin)
- **To decide on when to go sampling**, balancing the ease to identify species (flowers or fruits available), the ease to do field work (dry season) and minimising tree injury by coring (faster compartmentalisation of the wound in the growing season[‡])
  - species phenology (months of leaf flushing, flowering, fruiting)
  - climatic conditions (see §1.4.2 practical set-up)
- **To decide on what to sample**
  - taxonomically closely-related species or cryptic species
  - trunk diameter found in trade and diameter at which the species starts forming heartwood in the location of interest
- **To anticipate potential identification issues**
  - potential association with rhizobia (can influence isotope profile)
  - seed/tree source of species in the forest concession

[*] For heavily harvested species where this might be impossible, select sites with the highest tree density available.
[†] *E.g.* Neo *et al.* (2017)
[‡] Grissino-Mayer (2003), Tsen *et al.* (2016)

Table 1. Overview of the essential and ideal amount of reference material that needs to be collected for species or geographic origin determination of wood via the currently available techniques.

Design questions	Wood anatomy	DNA	Multi-element stable isotopes	DART-TOFMS	NIRS
For all questions					
general requirements	Sample all material (leaves, wood) from mature trees (DBH larger than 20 cm), at breast height or 30 cm above buttresses[I], where no stains or damage from bacteria, fungi or insects are visible and from trees growing in as varied environments as possible (soil type, altitude, exposure, fresh water access, …). Assure an even distribution of the number of individuals among sampling sites, with a preference for more sampling sites with fewer trees per location.				
type of material	Sap- and/or heartwood	Leaves, needles, buds and/or cambium	Sap- or heartwood or both	Heartwood[II]	
amount of material per sample	Block of 1 cm^3 or a 20 mm diameter core or (ideal) 1 x 7 x 11 cm wood piece[III]	10 cm^2 of leaves/needles/buds or 3 cm diameter punch of cambium layer or (but less ideal) 1 cm^3 of sapwood	Min. 8 growth years or *ca.* 10 cm of a 5 mm diameter core (5 g of wood in shavings)	A small core (3-5 slivers, 10-20 optimal, with a sliver being of fingernail size is enough)	Blocks[IV] of min. 2 cm^2 in tangential or radial longitudinal direction
replicates[V]	1 per tree[VI]	3 per tree	3 per tree	1 per tree	3 per tree
preferred equipment	Increment borer, chisel and hammer, saw	Telescopic scissors or sharpened hook, sling shot, puncher and mallet	Increment borer[VII] (manual or mechanical)		
For species identification					
botanical material	1 herbarium specimen (branch with leaves, fruits and/or flowers and optional a piece of bark) per tree				
nr. of trees & sites (essential)	5 trees or 5 trees per site if environment changes	50 trees over the whole species range	*not possible with this method*	15 trees	20 trees
outgroup (ideal)	At least 5 trees should be collected from each species that could be confused with the species of interest (same genus).				
nr. of trees & sites (ideal)	20 trees over the whole species range (for machine vision)	10 trees per sampling site with a total min. of 50 if covering the whole species range. More sampling sites are better than more trees per site.	*not possible with this method*	20 trees	30 trees

Table 1. (continued)

Design questions	Wood anatomy	DNA	Multi-element stable isotopes	DART-TOFMS	NIRS
For origin tracking to a region or country					
botanical material	Pictures of trunk, leaves, and if possible fruits and/or flowers per tree and 1 herbarium specimen per site[VIII]. If one tree is difficult to identify, then a herbarium specimen should be taken.				
nr. of trees & sites (essential)	*not possible with this method*	20 trees per sampling site	5 trees per sampling site	50 trees[IX] in total for 1 region/country	50 trees[IX] in total for 1 region/country
nr. of trees & sites (ideal)		30 trees[X] (at least 200 m apart[XI]) per sampling site (at least 100 km apart) with a total of 1000 trees and sites covering the entire species range and all different environmental conditions	Each time 10 trees per sampling site and sampling sites covering entire species range	100 trees, sampling sites covering entire species range	100 trees, sampling sites covering entire species range
For origin tracking to a concession					
botanical material	Pictures of trunk, leaves, and if possible fruits and/or flowers per tree and 1 herbarium specimen per site[XII].				
nr. of trees & sites (essential)	*not possible with this method*	Per focus concession 200 trees at least 50 m apart[XI] (5 x 40 trees in the annual logging plot and 4 other well-distributed areas) and from each neighbouring concession 50 trees (can be along a transect)	50 trees per concession and from each neighbouring concession 25 trees	50 trees[IX]	50 trees[IX]
nr. of trees & sites (ideal)		Sample size depends on concession size and distance to neighbour concessions	Depending on the climatic or environmental variations in a sample site	100 trees	100 trees
For origin tracking to an individual tree					
botanical material	1 herbarium specimen (branch with leaves, fruits and/or flowers) per tree[XIII]				
nr. of trees & sites (essential)	*not possible with this method*	all trees which should be felled according to management plan	*not possible with this method*	*not possible with this method*	*not possible with this method*

[I] Wood characteristics change from roots to canopy and it is hence advisable to standardise the height of sample collection. Also near buttresses (and any other imperfections) wood characteristics are deviant.

[II] Heartwood in slivers, blocks, or sawdust is required for chemical analysis by DART TOFMS and NIRS. Heartwood has a higher content of extractives than sapwood which allows easier discrimination between species. In addition, sapwood contains sugars that confuse the spectra for identification. Different species of trees have varying degrees of depth at which heartwood forms so care should be taken to clearly identify and collect the heartwood.

[III] Only possible from already felled trees.

[IV] Also powder of 4 mm granulometry can be used to obtain a NIRS spectrum. Only on wood pieces, however, can the method be used in the field. Besides, during the milling process special care should be taken to not affect the chemical components in the wood.

[V] Replicates might be needed to collect enough material and to account for intra-tree variation.

[VI] For machine vision it is however useful to sample from different positions in the tree to include as much intra-tree and intra-specific variation as possible (but while sampling only mature wood).

[VII] Advice and tips for using an increment borer can be found in Grissino-Mayer (2003) and examples of mechanical borers are described at http://www.smartborer.com (Kagawa and Fujiwara, 2018) and in Krottenthaler *et al.* (2015).

[VIII] Ideally, each reference sample should be connected to a herbarium specimen (preferably branch with leaves, flowers and/or fruits) deposited in a public herbarium. However, this is not always possible (*e.g.* when sampling 1000 trees for provenance determination).

[IX] Origin tracking with DART and NIRS is currently under development. The required number of trees might thus lower in future.

[X] Double the number of individuals if congeneric species may confound species identification.

[XI] This condition is lifted when tree density is too low to otherwise reach the minimum sample size.

[XII] Also in concessions misidentifications can happen.

[XIII] Even here herbarium specimens are essential because (1) many journals won't accept wood identification papers that don't reference herbarium specimens and (2) when the material would ever be used in a court case, the absence of herbarium specimens would harm the case.

GTTN
Global Timber Tracking Network

Overview of current practices in data analysis for wood identification

A guide for the different timber tracking methods

June 2020

GTTN
Global Timber
Tracking Network

Overview of current practices in data analysis for wood identification

A guide for the different timber tracking methods

June 2020

Editor:
Nele Schmitz

Authors*:
Hans Beeckman[1], Céline Blanc-Jolivet[2], Laura Boeschoten[3], Jez W.B. Braga[4], José Antonio Cabezas[5], Gilles Chaix[6,7,8], Simon Crameri[9], Bernd Degen[2], Victor Deklerck[1,10], Eleanor Dormontt[11], Edgard Espinoza[12], Peter Gasson[13], Volker Haag[14], Stephanie Helmling[14], Micha Horacek[15,16], Gerald Koch[14], Cady Lancaster[12], Frederic Lens[17], Andrew Lowe[11], Sandra Martínez-Jarquín[18], Justyna Anna Nowakowska[19], Andrea Olbrich[14], Kathelyn Paredes-Villanueva[20], Tereza C.M. Pastore[21], Tahiana Ramananantoandro[22], Andriambelo R. Razafimahatratra[22], Prabu Ravindran[23,24], Gareth Rees[25], Liz F. Soares[4], Niklas Tysklind[26], Mart Vlam[3], Charlie Watkinson[25], Elisabeth Wheeler[27,28], Robert Winkler[18], Alex C. Wiedenhoeft[23,24,29,30], Valentina Th. Zemke[14], Pieter Zuidema[3].

*Names are listed in alphabetical order.

Recommended citation:
Schmitz, N. (ed.), H. Beeckman, C. Blanc-Jolivet, L. Boeschoten, J.W.B. Braga, J.A. Cabezas, G. Chaix, S. Crameri, B. Degen, V. Deklerck, E. Dormontt, E. Espinoza, P. Gasson, V. Haag, S. Helmling, M. Horacek, G. Koch, C. Lancaster, F. Lens, A. Lowe, S. Martínez-Jarquín, J.A. Nowakowska, A. Olbrich, K. Paredes-Villanueva, T.C.M. Pastore, T. Ramananantoandro, A.R. Razafimahatratra, P. Ravindran, G. Rees, L.F. Soares, N. Tysklind, M. Vlam, C. Watkinson, E. Wheeler, R. Winkler, A.C. Wiedenhoeft, V.Th. Zemke, P. Zuidema (2020). Overview of current practices in data analysis for wood identification. A guide for the different timber tracking methods. Global Timber Tracking Network, GTTN secretariat, European Forest Institute and Thünen Institute. DOI: 10.13140/RG.2.2.21518.79689

Cover: José Bolaños
Project leader: Jo Van Brusselen
Pictures: Unless mentioned otherwise, all pictures in this document were provided by the authors.

The following document is an output from GTTN, which is coordinated by the European Forest Institute (EFI) and funded by the German Federal Ministry for Food and Agriculture (BMEL). The information expressed in this document are the views of the authors and do not necessarily represent those of the donor or of the European Forest Institute.

Affiliations:

1-Royal Museum for Central Africa, Tervuren, Belgium

2-Thünen Institute of Forest Genetics, Grosshansdorf, Germany

3-Wageningen University & Research, Wageningen, The Netherlands

4-University of Brasília, Brasília, Brazil

5-Forest Research Centre, National Institute for Agricultural and Food Research and Technology (INIA-CIFOR), Madrid, Spain

6-CIRAD, UMR AGAP, Montpellier, France

7-AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

8-ESALQ-USP, Wood Anatomy & Tree-Ring Lab, Piracicaba, Brazil

9-Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland

10-Ghent University Woodlab, Ghent, Belgium

11-University of Adelaide, Adelaide, Australia

12-U.S. Fish & Wildlife Service, Ashland, USA

13-Royal Botanic Gardens, Kew, UK

14-Thünen Institute of Wood Research, Bergedorf, Germany

15-HBLFA Francisco-Josephinum, Wieselburg, Austria

16-Institute of Lithospheric Research, Vienna University, Vienna, Austria

17-Naturalis Biodiversity Center, Leiden, The Netherlands

18-Department of Biochemistry and Biotechnology, CINVESTAV Unidad Irapuato, Irapuato, Mexico

19-Cardinal Stefan Wyszynski University, Warsaw, Poland

20-Gabriel René Moreno University, Santa Cruz, Bolivia

21-Brazilian Forest Service, Brasília, Brazil

22-School of Agronomy, Department of Water and Forestry, University of Antananarivo, Antananarivo, Madagascar

23-USDA Forest Service, Madison, USA

24-University of Wisconsin, Madison, USA

25-Agroisolab UK Ltd, York, UK

26-National Institute of Agricultural Research (INRA), Kourou, Guyane Française

27-North Carolina State University, Raleigh, USA

28-North Carolina Museum of Natural Sciences, Raleigh, USA

29-Purdue University, West Lafayette, USA

30-São Paulo State University, São Paulo, Brazil

For more information about the authors see the GTTN project partner finder.

Table of Contents

4. DART TOF Mass Spectrometry

Definition DART TOFMS reference data: the chemical fingerprint of a wood sliver (and by extent species/provenance) based on the complete set of small chemical molecules found within the sample.

Authors: Victor Deklerck, Edgard Espinoza, Cady Lancaster, Sandra Martínez-Jarquín, Kathelyn Paredes-Villanueva, Robert Winkler

*Authors are in alphabetical order.

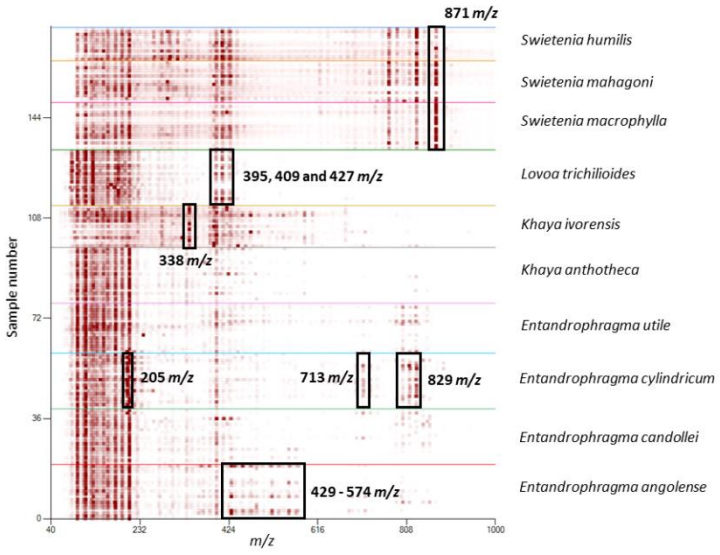


Fig. 16: Heatmap showing the presence of the ions for the different specimens per species.

4.1 RESOURCES REQUIRED

Access to reference material

The largest database currently available for DART TOFMS is the **Forensic Spectra of Trees Database©** (ForeST Database©). This database is held and curated by the National Fish and Wildlife Forensic Laboratory, NFWFL (Ashland, OR, USA). ForeST© contains thousands of species from hundreds of genera. The focus of the database is for species identification of CITES-listed species, lookalikes of the CITES-listed, and commercially significant timber species, especially from tropical regions. **A softwood specific database** is currently being refined and provides genus-level identification.

Reference applicability

For use of the ForeST Database©, NFWFL recommends that it be used in conjunction with **JEOL-line TOF MS**. Due to the chemical complexity of wood products, academic investigation of species separation and classification can use a variety of instrument parameters. **Use of the ForeST Database© requires tuning instruments to the database**. Contact NFWFL for more details on instrument validation.

NOTE: DART TOFMS is not the only mass spectrometer that can be used for timber identification, it is however, the most developed one with the most developed database and adjoined software.

Software

Table 5: Overview of the software used for wood identification via DART TOFMS.

Software	Performance	Accessibility
R	Can be used for any statistical computing task. Computer programming knowledge needed. Customization easily implementable.	Free software
Mass Mountaineer	Purpose-built software for DART TOFMS with all standard statistical packages included. More labour-intensive customization possibilities for other research explorations. Recommended for use in forensic science.	Commercial software
NIST Search Software	The National Institute of Standards and Technology (NIST) search software provides some of the most advanced search algorithms for qualitative analysis. Spectra can be quickly compared.	Commercial software

4.2 USING DART TOFMS FOR TAXON OR PROVENANCE IDENTIFICATION

4.2.1 DEVELOPMENT OF DART TOFMS REFERENCE DATA

NOTE: DART TOFMS is not regularly used for provenance identification at the time of this publication. However, although the variation in the results is subject of further investigations, the protocol below can be used as a start for provenance identification as well (Espinoza *et al.* 2014, Finch *et al.* 2017, Paredes-Villanueva *et al.* 2018).

4.2.1.1 SAMPLE SELECTION & PREPARATION

Collect samples of the species and/or geographic region (provenance) needed to be characterised and include **equivalent sample sizes per species/provenance**. Use only the **heartwood** (see GTTN sampling guide) and the non-contaminated parts of the wood (see Appendix 4). However, it is still good practice to also collect **reference material of known or suspected contaminants** (such as varnish, oil, coating). Molecular ion peaks of known contaminants can be subtracted from the mass spectra and contaminated wood samples can be salvaged in post-processing.

4.2.1.2 SPECTRA COLLECTION

Collect the spectra as per the instructions of the mass spectrometer instrument. **The mass spectrometers are calibrated** with a reference solution of known masses (for example polyethylene glycol solution). If the ForeST Database© is to be used, the DART TOFMS should be calibrated to the database (see §4.1 > *Reference applicability*).

For **archiving and sharing the mass spectrum**, it is important to declare the provenance of the reference material. One easy solution is to label the individual spectrum file names with explicit data that can describe the metadata easily. NFWFL uses the following strategy: GenusSpecies_AnalysisLabAccession_SourceAccession. For example:

File name: DalbergiaNigra_WD123456_MADw1234

Species: *Dalbergia nigra*

NFWFL accession number: WD123456

Original catalogue number: Forest Products Lab ID for Madison Wood Collection (MADw) 1234.

4.2.1.3 DATA PREPARATION

Before developing a statistical model there is a need to remove outliers from training sets. Hawkins (1980) describes 'an outlier' as *an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism*. **Cross-check the spectra with spectra already in the database and with spectra of known contaminants**.

- Heatmaps allow for a simultaneous comparison of all the spectra data and are ideal for identifying outliers (*e.g.* in *R* or *Mass Mountaineer*).
- Average spectra can be created (*e.g.* in *R*) for each species by combining all available spectra for this species. From this super spectrum outliers can be identified.
- Be cautious of contaminants. Use heatmaps of the known contaminants and cross-reference to the wood sample data set to avoid selecting contaminate molecular ion peaks in model training.

Before removing samples from the dataset, outliers should be further evaluated. One way to evaluate spurious spectra is to determine if suspected outlier specimens have a different and distinct chemotype from other spectra of the same species. Another way is to check if the intensity is similar for all the *m/z* ions. Samples from the same species will show similar intensities of the molecules detected.

Possible reasons for outliers are:

- The reference database does not yet represent the chemical intra-variability of the species or provenance[30].
- Mislabelled or misidentified samples.
- The measurement was not done properly.
- The tissue type was different from the other specimens (*e.g.* sapwood *vs.* heartwood)
- There is a contaminant present on the sample.

[30] When is an outlier not an outlier? Whichever approach you take to determine this, it is key to know your data and your research area well. For more info on data interpretation see Appendix 6.

It is recommended that outlier samples be re-analysed by DART TOFMS to be able to exclude the possibility of an erroneous measurement.

4.2.1.4 MODEL DEVELOPMENT

- ☑ Only use reference spectra which are in consensus.
- ☑ Hold out reference spectra to check later for model overfitting and model validation.
- ☑ Build statistical models for species/provenance classification using (i) the same number of spectra for each species/provenance included, and (ii) a suitable kind of statistical analysis. What is suitable will depend on the species/provenance (group) being investigated.
  - Multivariate statistics such as Principal Component Analysis (PCA), Kernel Discriminant Analysis (KDA) or Discriminant Analysis of Principal Components (DAPC) with *Mass Mountaineer* are **recommended for forensic analysis** because this software easily produces quantitative results and provides probability estimates, which are used in court to describe certainty of the analysis.
  - *Random forest* (Deklerck *et al.* 2017, Finch *et al.* 2017, Paredes-Villanueva *et al.* 2018, Deklerck *et al.* 2019) has the advantage that it **can be used when there are less than 4 classes**.
  - PAM clustering or Adaptive Boosting are two robust classifiers for discrimination between **two groups**.

The application of DART TOFMS to wood identification is a relatively new technique (since 2012), and therefore model development options are not limited to the above list since new mathematical tools are continuously being explored (such as dynamic time warping and deep learning algorithms).

4.2.1.5 MODEL OPTIMISATION & VALIDATION

Before constructing the final model, it is proposed to **screen the spectral data pre-processing parameter settings for optimal classification accuracy**. Depending on the classification algorithm this can be done as follows:

- Reduce the number of variables (combinations of ions) by carefully selecting components in a PCA.
- Select ions for model building using Fisher Ratio analysis or by careful visual examination of heatmaps.
- Mass tolerance for binning and relative abundance cut-off threshold settings (pre-processing parameters) and number of ions (classification parameter) can be screened in an automated way, when using the *random forest* algorithm or any algorithm that is data frame dependent and does not work with individual text files for model-building, as described in Deklerck *et al.* (2019).

Test the model for overfitting and classification accuracy by removing reference spectra (leave one out cross validation, LOOCV) and by analysing unused reference spectra that were not used in the model development (sometimes called 'hold-out set' or 'validation set'). Model parameters are optimised to obtain (i) the highest LOOCV possible, and (ii) accurate classification of the hold-out specimens.

4.2.2 ANALYSIS OF DART TOFMS DATA FOR TEST SAMPLES

NOTE: The **purpose of this protocol** is to provide a procedure to analyse and identify specimens of wood by comparing them with a curated database of known species or a specific set of reference samples. The intent of this method is not to identify all the compounds found in the wood samples, but to infer from specific ions present that a given wood sample did or did not originate from a known species. An experienced analyst may occasionally need to vary the procedure to accommodate a particular sample. The different data analysis steps:

- **Wood anatomical analysis** of the unknown sample to determine the genus.
- **Collection of spectra** as described in §4.2.1.2.
- **Library search for preliminary classification**. Determine if the unknown spectra match a species held in the ForeST reference database. The top library hits are then used to create the subsequent heatmaps and multivariate statistical models. This step is implemented in *Mass Mountaineer*.
- **Creation of a heatmap and evaluation of the unknown**. Frequently the library search will indicate that two or three species have spectra similar to the unknown. These taxa should be used to create the heatmap and to determine if the chemotypes for each species selected are similar to that of the unknown spectra.
- **Selection of the variables (ions) for the multivariate modelling**. Use the full set of curated reference spectra available for the species/provenance of interest as the training set. For *random forest*, variable selection is automatically included in the set-up and can be optimized (see §4.2.1.5). This is similar for other machine learning techniques or classification algorithms in *R*.
- **Taxonomic assignment of the unknown** once a model performs satisfactorily. When using *Mass Mountaineer*, the software has a utility to automatically classify the unknown spectra against the multivariate model and the program provides a probable estimate of accuracy. For *random forest*, first build the model according to §4.2.1.4, then use the full reference set as training and the unknowns as test set.
- **Final classification decision** is obtained when (i) the library results, (ii) the heatmap, and (iii) the multivariate analysis all agree with each other.
- **Classification validation** as described in §4.2.1.5.

4.2.3 STRENGTHS & LIMITATIONS

Strengths

- **Very small sample** is needed (wood slivers).
- **Versatile sampling**. As only a wood sliver is needed, samples can be easily collected from a range of products (guitars, watches, logs, …).
- Analyses can be done at **low cost**. Once you have the mass spectrometer, running costs are low and no other equipment, expensive chemicals or software is required for wood identification.
- **Tailor-made software** for data analyses (*Mass Mountaineer*, see Table 5) comes together with the purchase of a JEOL DART TOFMS.
- Analyses can be performed in few minutes, allowing for **high-throughput** screening. Once models are built for certain (groups of) species/provenances wood identification can be done within a short time.
- A **curated reference database** (ForeST Database©) of about 2000 species, representing 630 genera is currently available (Nov. 2019).

Limitations

- **Extreme physical or chemical processes** in the wood (caused by *e.g.* temperature, micro-organisms) could alter the DART MS profile.
- **Possible interference of chemical contaminants** such as glue and packaging.
- **Wood panels and plywood** might be difficult to identify due to the adhesives used for the manufacturing. The mix of species used for their fabrication might also be a limitation.
- **Some tree species** present few signals in DART MS and therefore provide insufficient data for evaluation.
- **Sap- and heart-wood can show different profiles**. As the current database is based on heartwood only (being the best identifier and the most used in wooden objects), caution is therefore needed when an unknown sample might contain sapwood.

4.3 KEY LITERATURE FOR DART TOFMS DATA ANALYSIS

Baxter, M.J. (1999). On the Multivariate Normality of Data Arising from Lead Isotope Fields. Journal of Archaeological Science 26(1): 117–124.

Cody, R.B., J.A. Laramee and H.D. Durst (2005). Versatile new ion source for the analysis of materials in open air under ambient conditions. Analytical Chemistry 77(8): 2297-2302.

Cody, R.B., A.J. Dane, B. Dawson-Andoh, E.O. Adedipe and K. Nkansah (2012)."Rapid classification of White Oak (*Quercus alba*) and Northern Red Oak (*Quercus rubra*) by using pyrolysis direct analysis in real time (DART (TM)) and time-of-flight mass spectrometry. Journal of Analytical and Applied Pyrolysis 95: 134-137.

Deklerck, V., K. Finch, P. Gasson, J. Van den Bulcke, J. Van Acker, H. Beeckman and E. Espinoza (2017). Comparison of species classification models of mass spectrometry data: Kernel Discriminant Analysis vs Random Forest; A case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). Rapid Communications in Mass Spectrometry 31(19): 1582-1588.

Deklerck, V., C. Lancaster, J. Van Acker, E. Espinoza, J. Van den Bulcke and H. Beeckman (under review). Chemical fingerprinting of wood sampled along a pith-to-bark gradient allows individual distinction and provenance identification. Forestry.

Deklerck, V., T. Mortier, N. Goeders, R.B. Cody, W. Waegeman, E. Espinoza, J. Van Acker, J. Van den Bulcke and H. Beeckman (2019). A protocol for automated timber species identification using metabolome profiling. Wood Science and Technology https://doi.org/10.1007/s00226-019-01111-1.

Espinoza, E.O., C.A. Lancaster, N.M. Kreitals, M. Hata, R.B. Cody and R.A. Blanchette (2014). Distinguishing wild from cultivated agarwood (*Aquilaria* spp.) using direct analysis in real time and time of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 28(3): 281-289.

Espinoza, E.O., M.C. Wiemann, J. Barajas-Morales, G.D. Chavarria and P.J. McClure (2015). Forensic Analysis of Cites-Protected *Dalbergia* Timber from the Americas. IAWA Journal 36(3): 311-325.

Evans, P.D., I.A. Mundo, M.C. Wiemann, G.D. Chavarria, P.J. McClure, D. Voin and E.O. Espinoza (2017). Identification of selected CITES-protected Araucariaceae using DART TOFMS. IAWA Journal 38(2): 266-S3.

Finch, K., E. Espinoza, F.A. Jones and R. Cronn (2017). Source Identification of Western Oregon Douglas-Fir Wood Cores Using Mass Spectrometry and Random Forest Classification. Applications in Plant Sciences 5(5): apps.1600158.

D., Hawkins (1980). Identification of Outliers. Chapman and Hall, London.

Lancaster, C. and E. Espinoza (2012). Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid Communications in Mass Spectrometry 26(9): 1147-1156.

Lancaster, C. and E. Espinoza (2012). Evaluating agarwood products for 2-(2-phenylethyl) chromones using direct analysis in real time time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 26(23): 2649-2656.

McClure, P.J., G.D. Chavarria and E. Espinoza (2015). Metabolic chemotypes of CITES protected *Dalbergia* timbers from Africa, Madagascar, and Asia. Rapid Communications in Mass Spectrometry 29(9): 783-788.

Paredes-Villanueva, K. (2018). Tropical timber forensics: a multi-methods approach to tracing Bolivian *Cedrela* (Doctoral dissertation, Wageningen: Wageningen University).

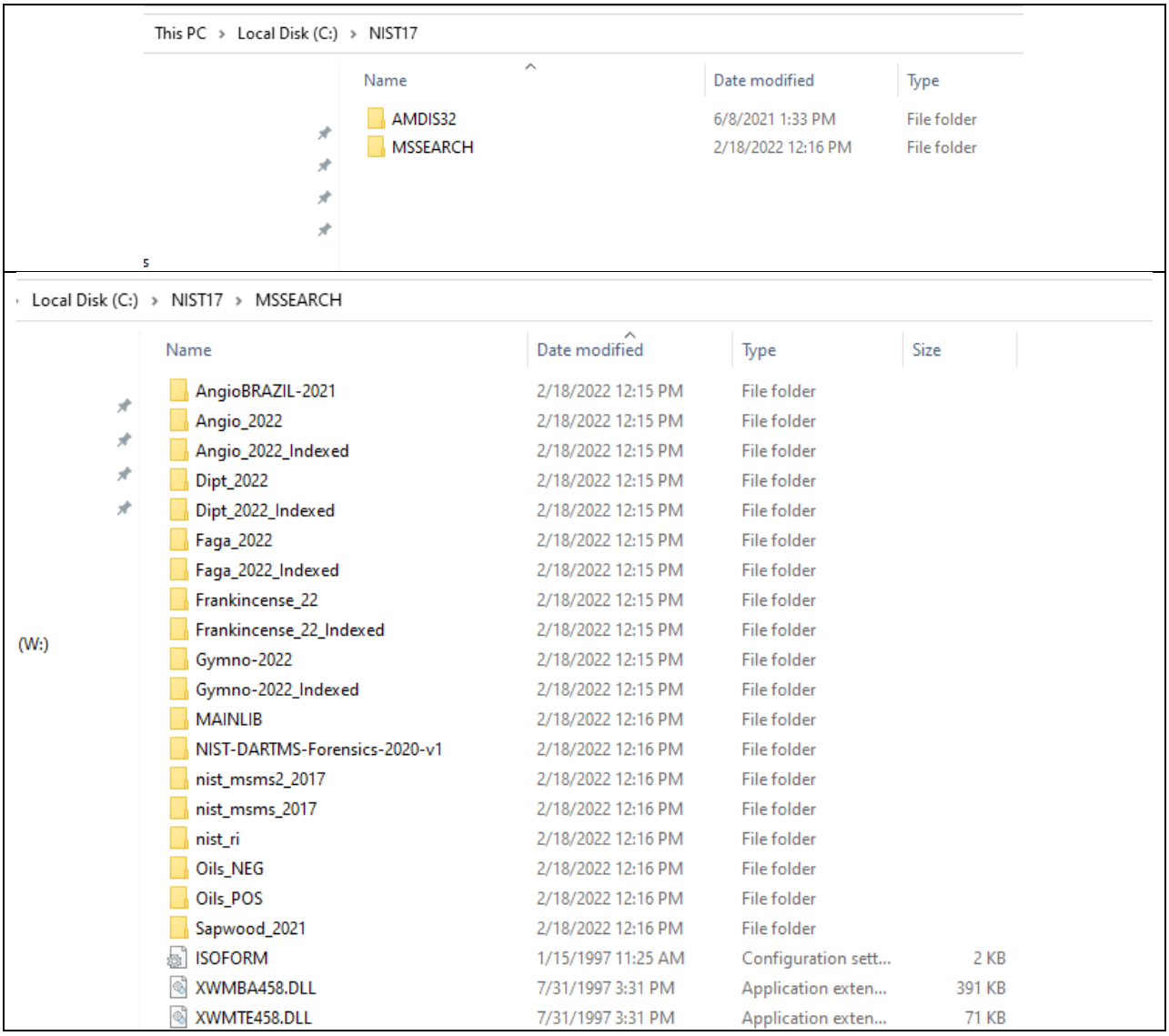
Paredes-Villanueva, K., E. Espinoza, J. Ottenburghs, M.G. Sterken, F. Bongers and P.A. Zuidema (2018). Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. Forestry: An International Journal of Forest Research 91(5): 603-613.

Library Search & General Classification Scheme Using Mass Mountaineer

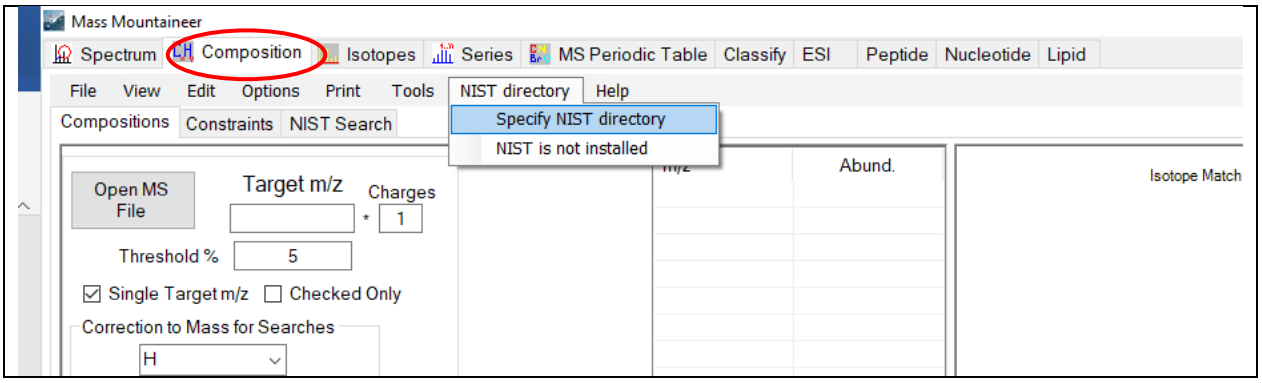
This document is intended as a general guide, refer to the Mass Mountaineer manual for complete steps

If using Mass Mountaineer for the first time:

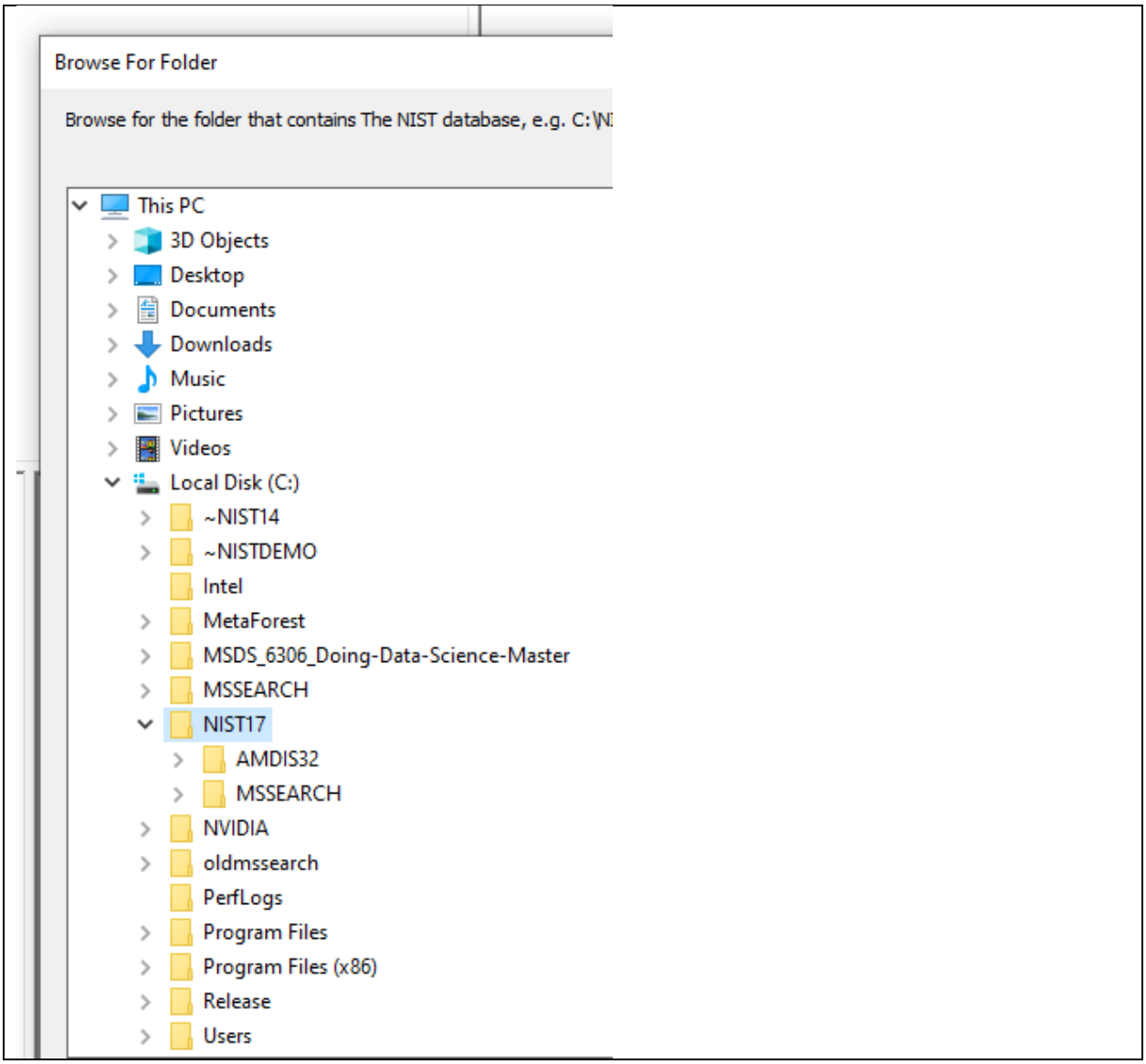
1. Download NIST17 and the libraries into your C: drive.
2. Cut and paste all of the libraries into the MSSEARCH folder within NIST17



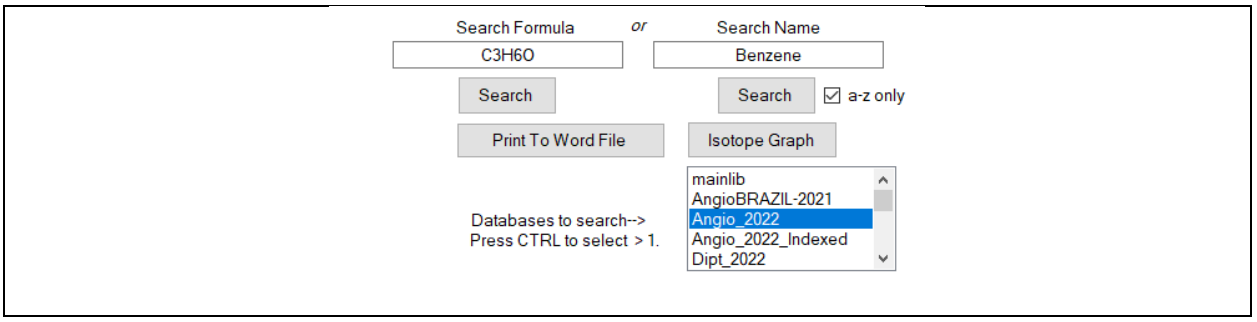
3. Open Mass Mountaineer and click the Composition tab (red circle), then click the NIST directory button and select Specify NIST directory



4. Locate the NIST17 folder and click OK



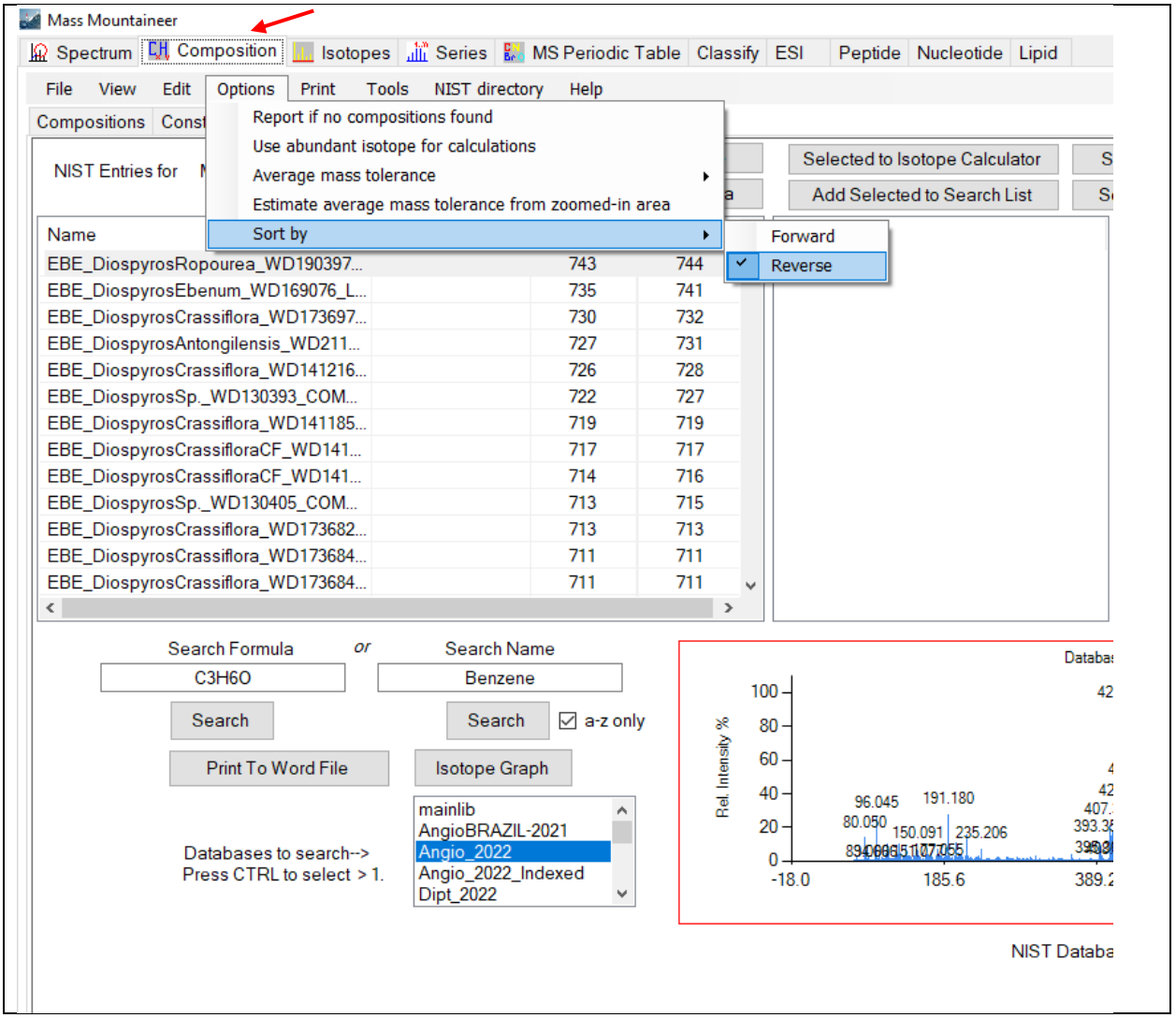
5. Click the NIST Search tab at the top of Mass Mountaineer and check to see if the libraries are now visible. If you do not see the libraries, restart Mass Mountaineer.
6. Select the Angio_2022 library by simply clicking on it.



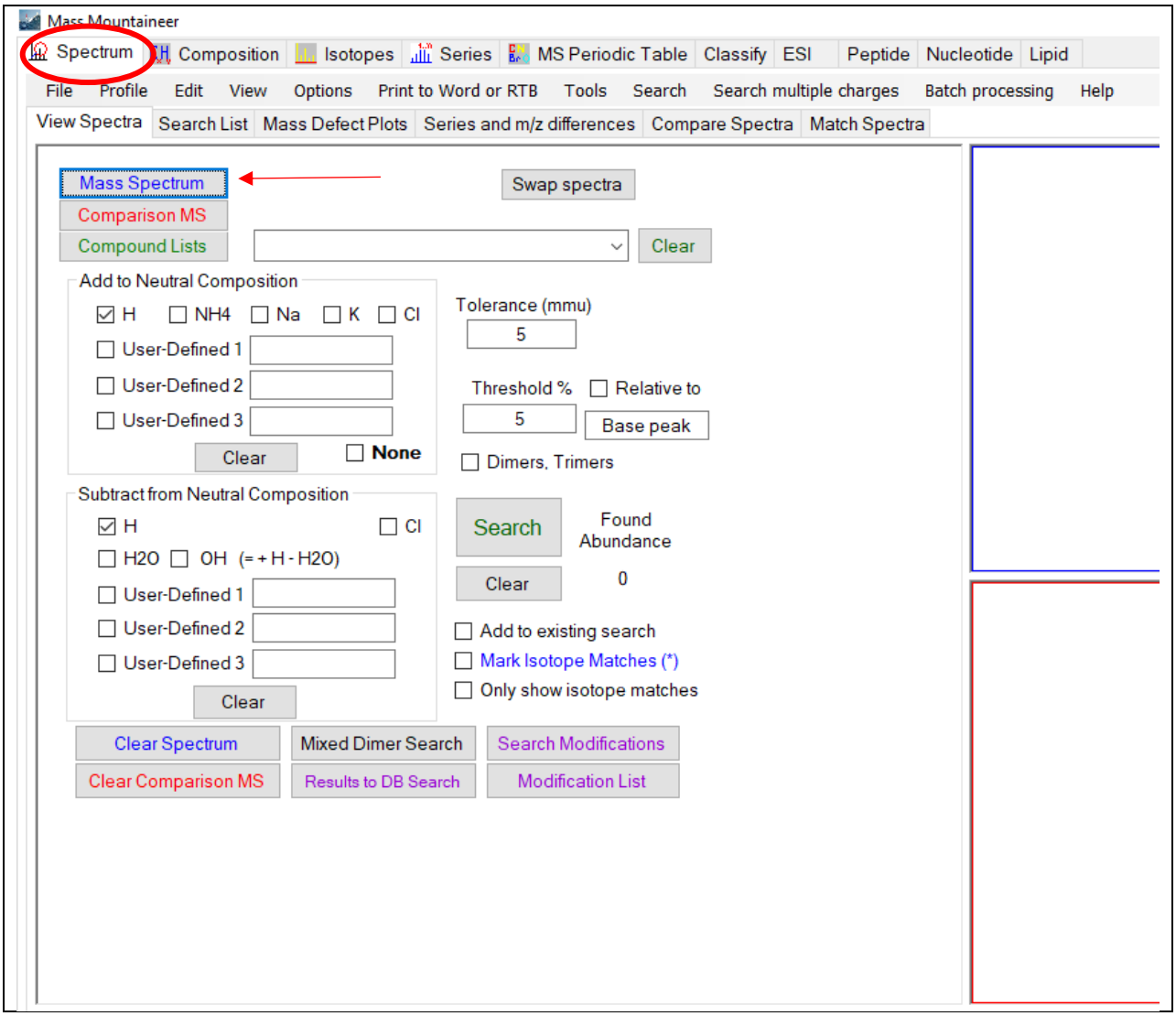
7. Now you are ready to perform a search against the library

Search a Single Spectrum File

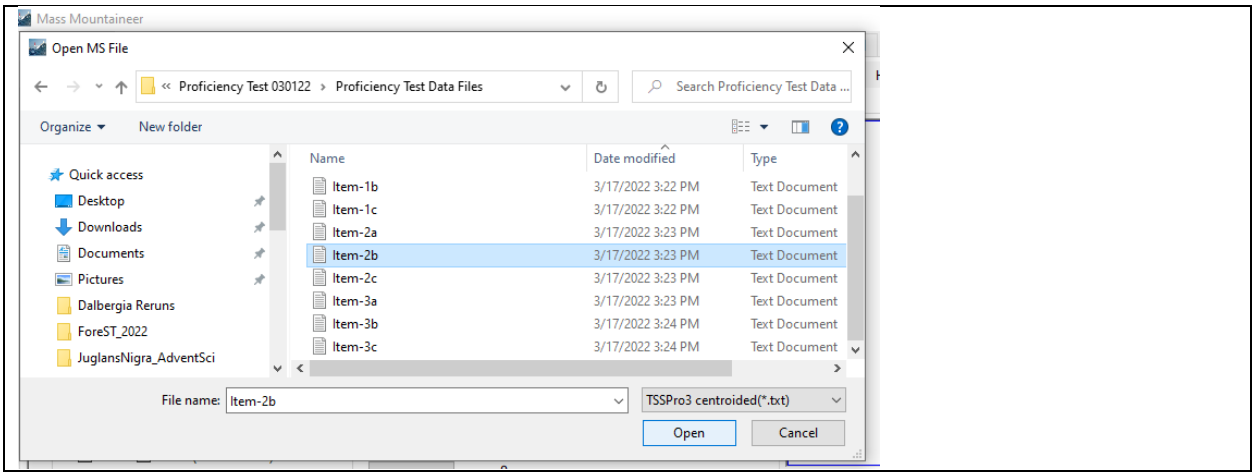
1. Open Mass Mountaineer and click the Composition tab (red arrow), then click Options, then hover over "Sort by" and select Reverse



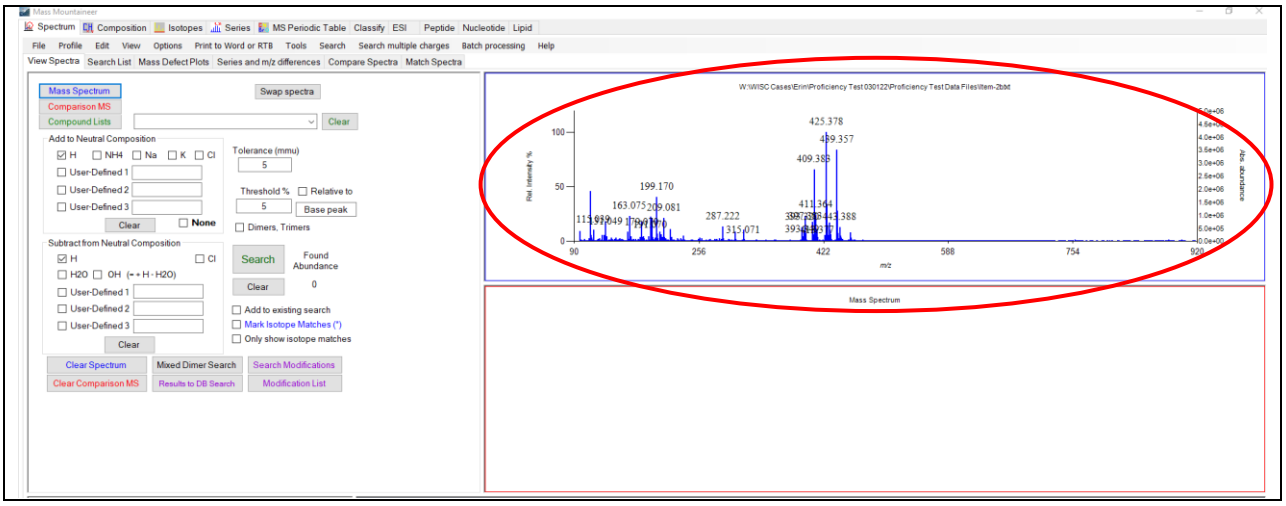
2. Click the Spectrum tab (red circle), then click the Mass Spectrum button (red arrow)



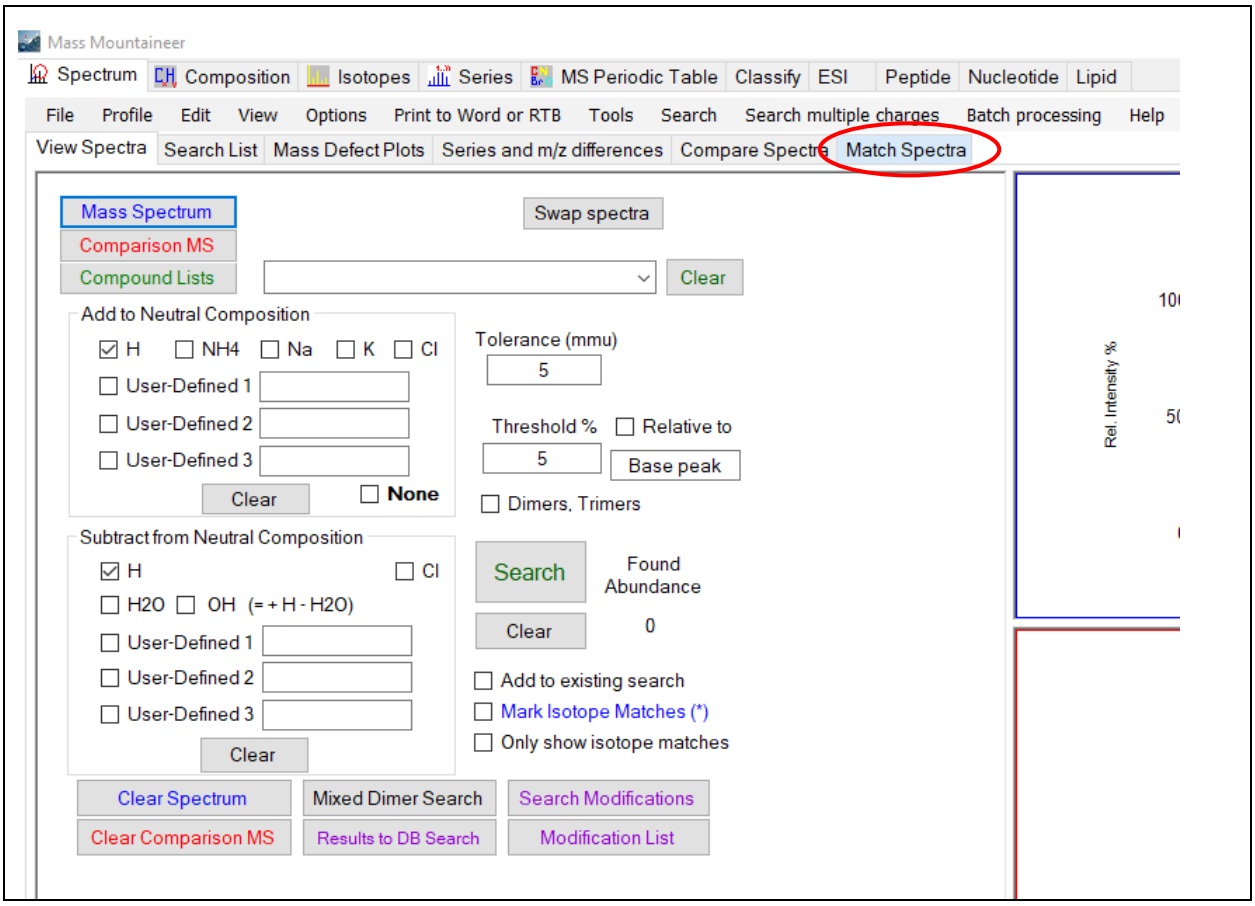
3. Locate the spectrum you want to analyze and click Open



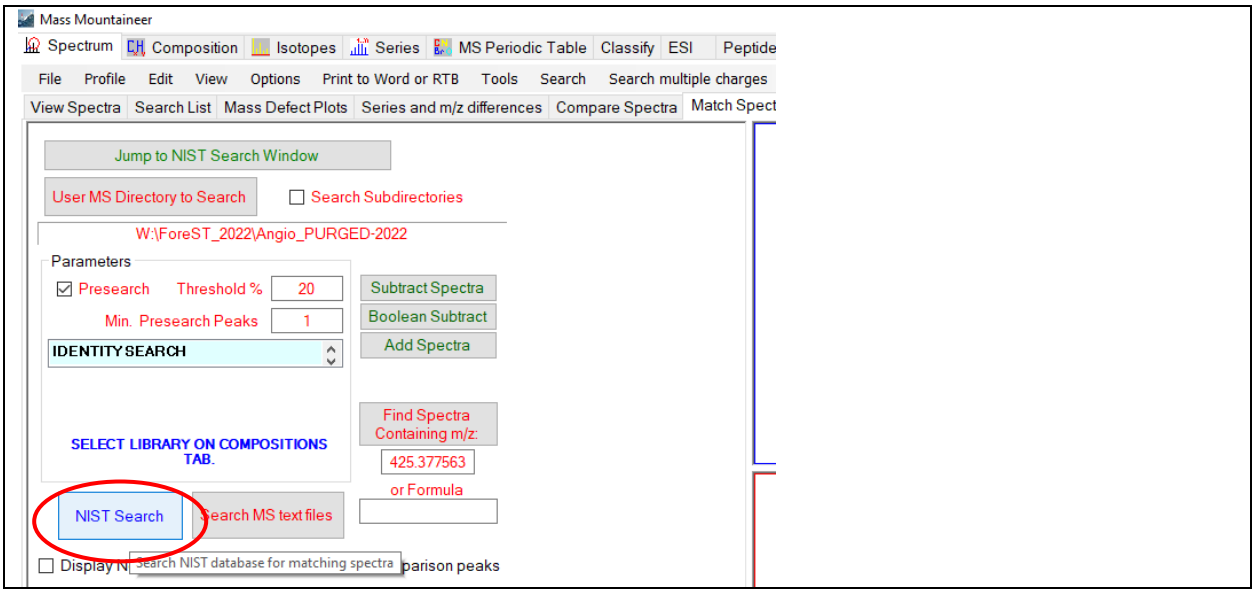
4. You will see the spectrum in the top window on the right side of the screen (red circle)



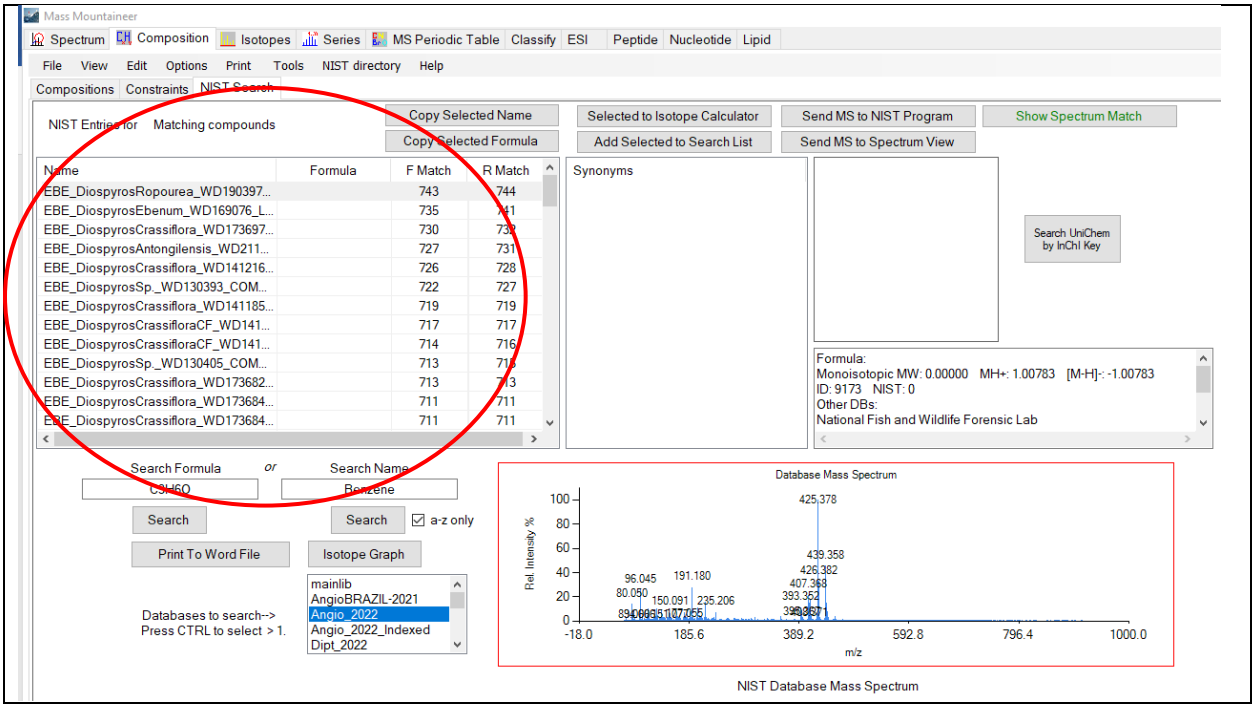
5. Click the Match Spectra tab



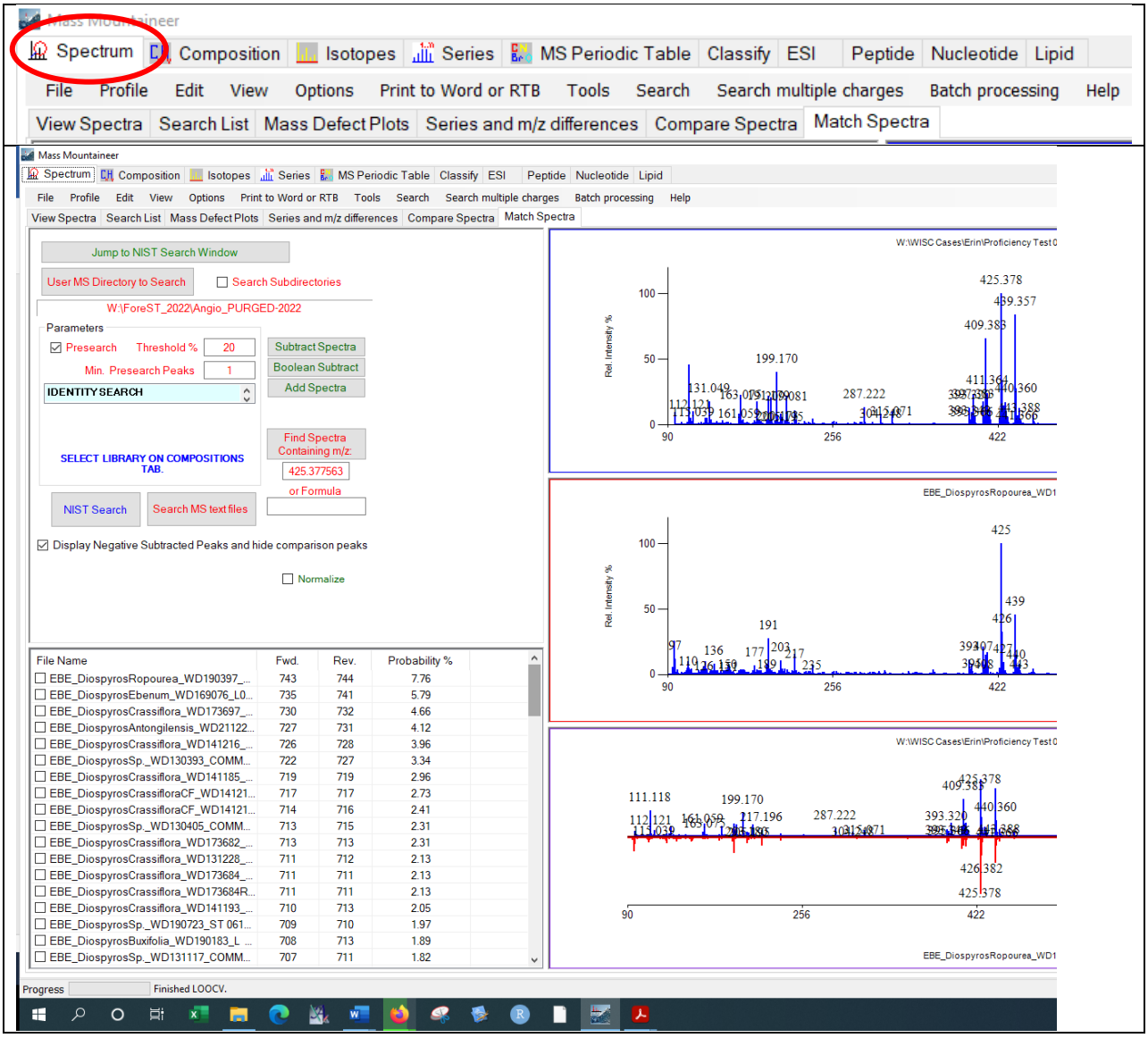
6. Click NIST Search



7. Mass Mountaineer will show the species in ForeST that matched best with the spectrum (red circle)

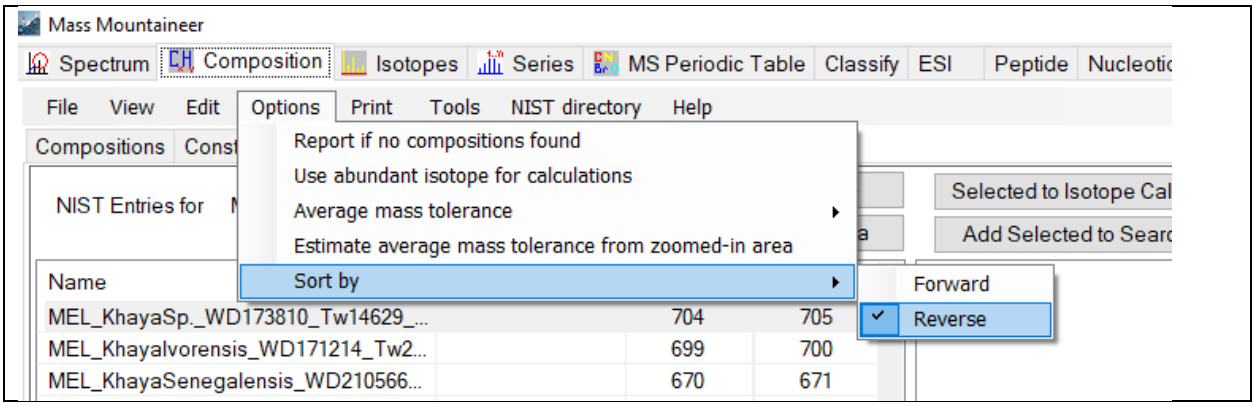


8. In order to compare the matches, click Spectrum (red circle), and you will see all species/sample matches.

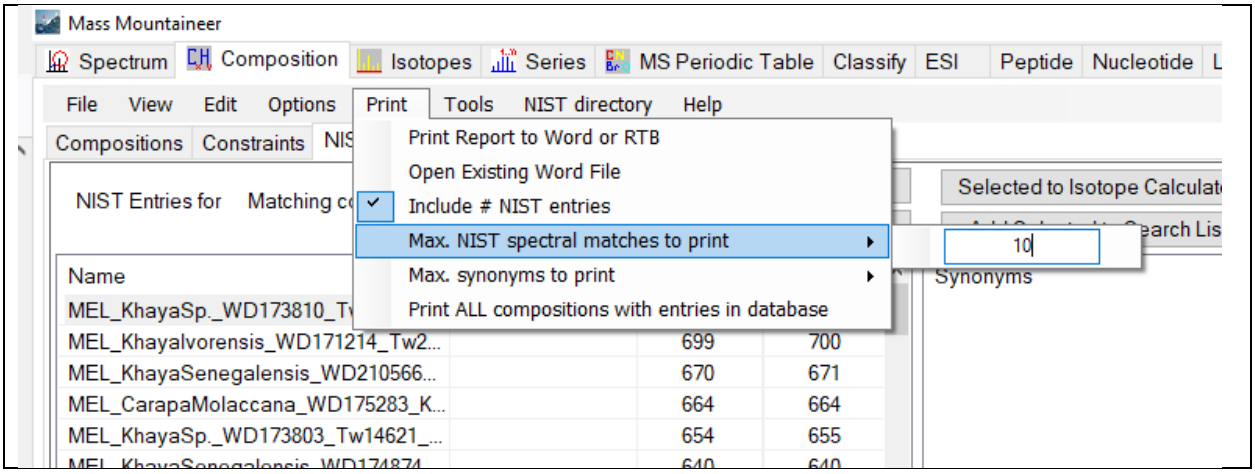


Batch searching ForeST Libraries

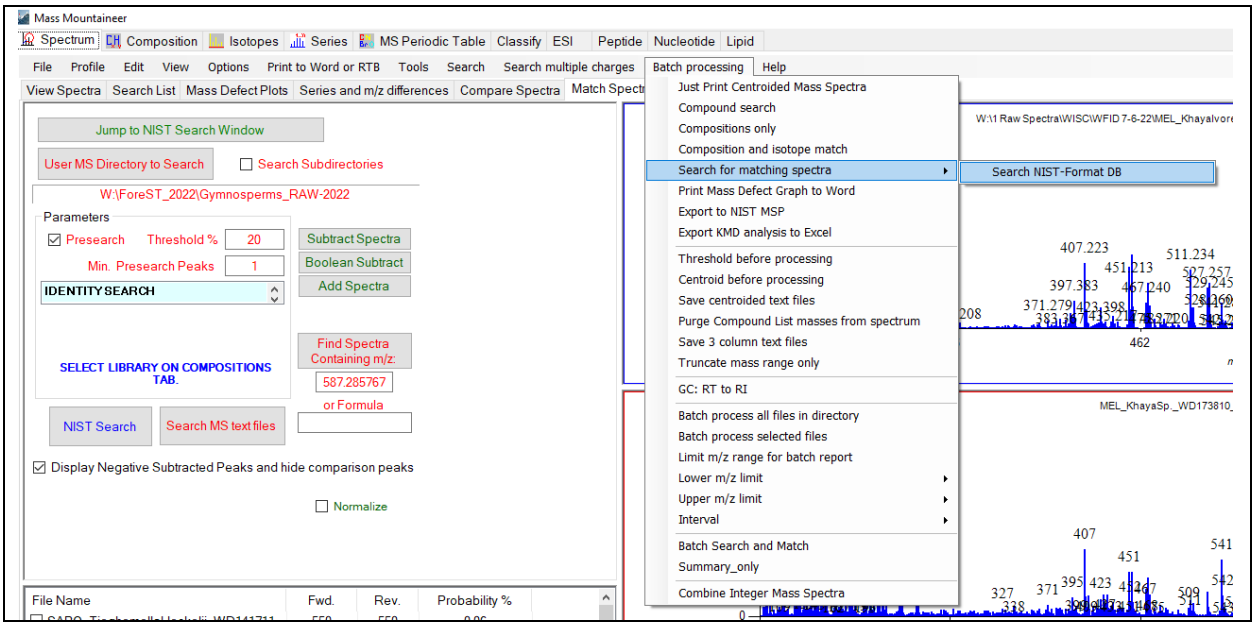
1. Set search results to Reverse by navigating to the Composition tab, select Options, hover over “Sort by” and click Reverse.



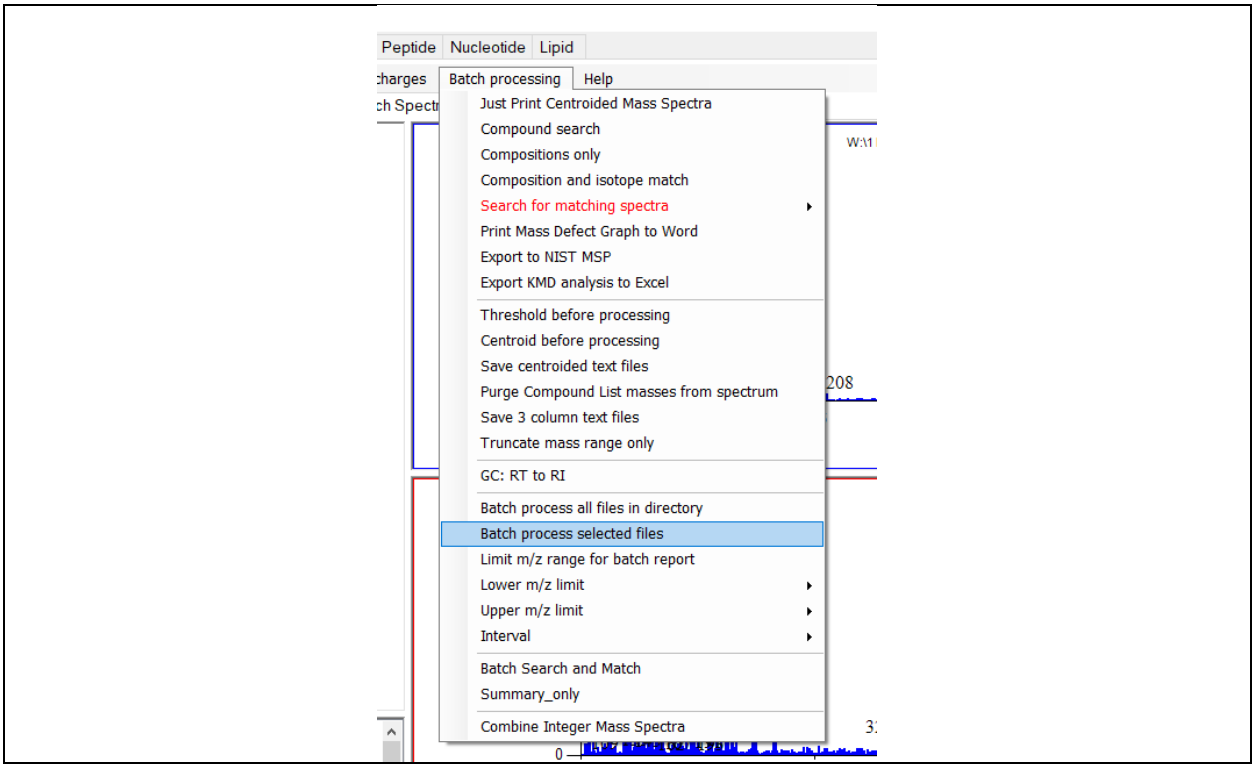
2. From the Composition tab, select Print, then hover over “Max. NIST spectral matches to print” and input the desired number of results to be shown.



3. From the Spectrum tab, click Batch processing, and hover over “Search for matching spectra” to click “Search NIST – Format DB”
    a. **NOTE: Be sure that Compound search is not selected, the drop-down list should look as shown below.**



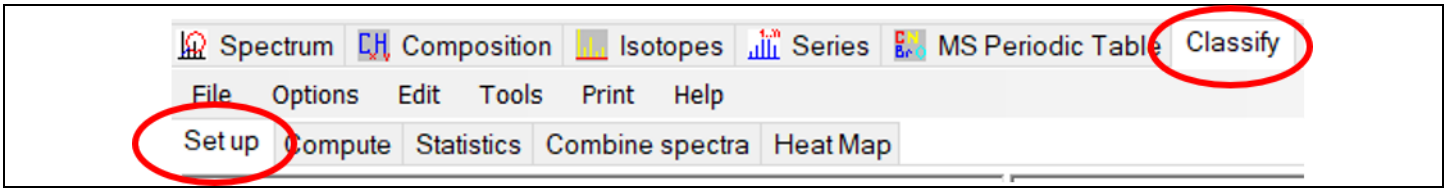
4. Click Batch processing again, and select “Batch process selected files”



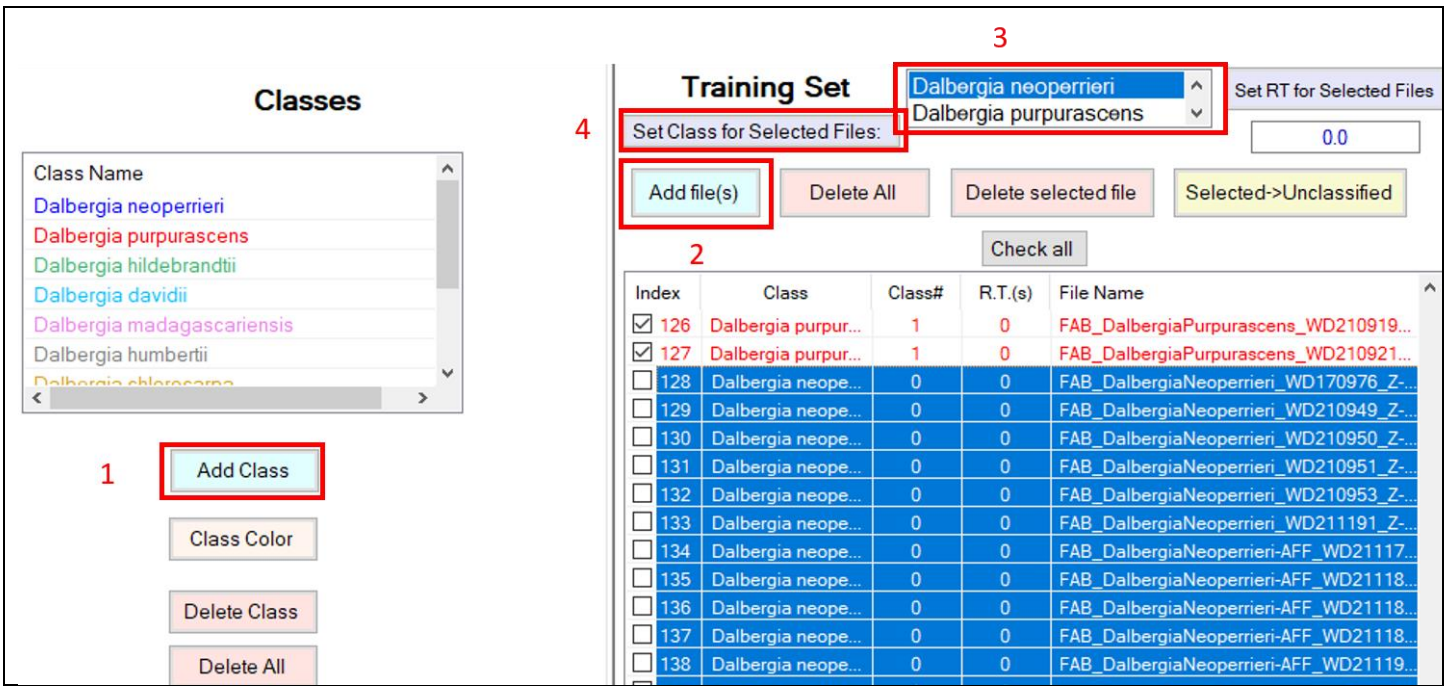
5. Select the spectra to be searched and click Open, after a few moments a Word document will appear and begin populating with the search results.

General scheme for classification

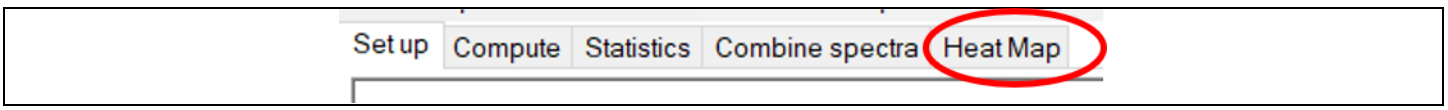
1. After determining which species should be included in a model, navigate to the Classify tab, and then click Set Up.



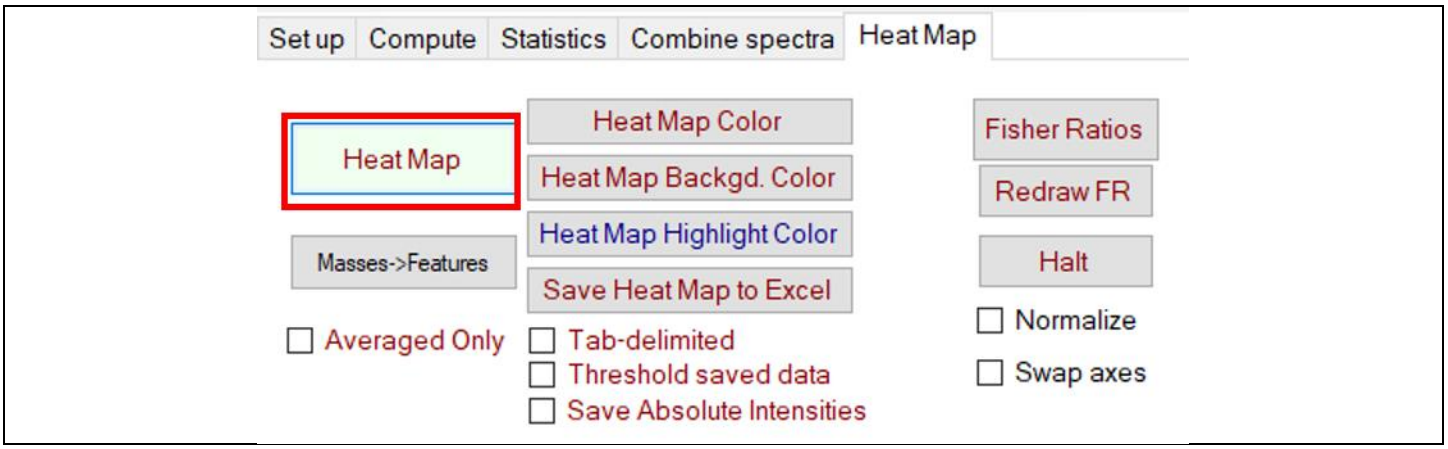
2. Now it is time to create a heat map using spectra from the species that matched the unknown/evidence item.
    1. Click Add Class for as many species as needed and type in each species name.
    2. Click Add file(s) and select approximately 20 spectra per species, click Ok
    3. Highlight all spectra for a single species (click the topmost spectrum, hold down Shift, and click the bottommost spectrum) and click the corresponding species in the box at the top.
    4. Click Set Class for Selected Files:



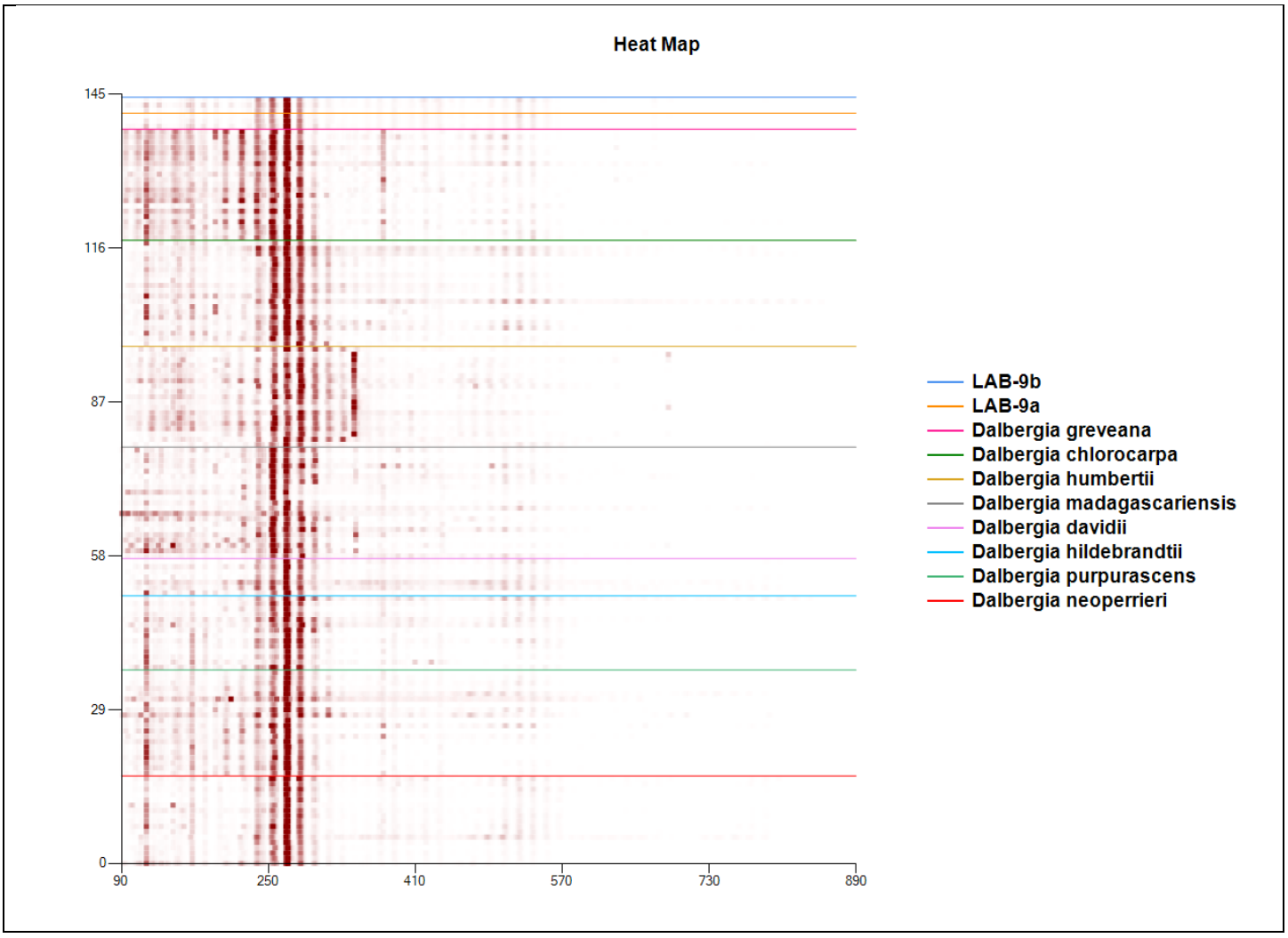
3. Construct a heatmap of the species matches from the reverse search and the unknown/evidence item by clicking the Heat Map tab.



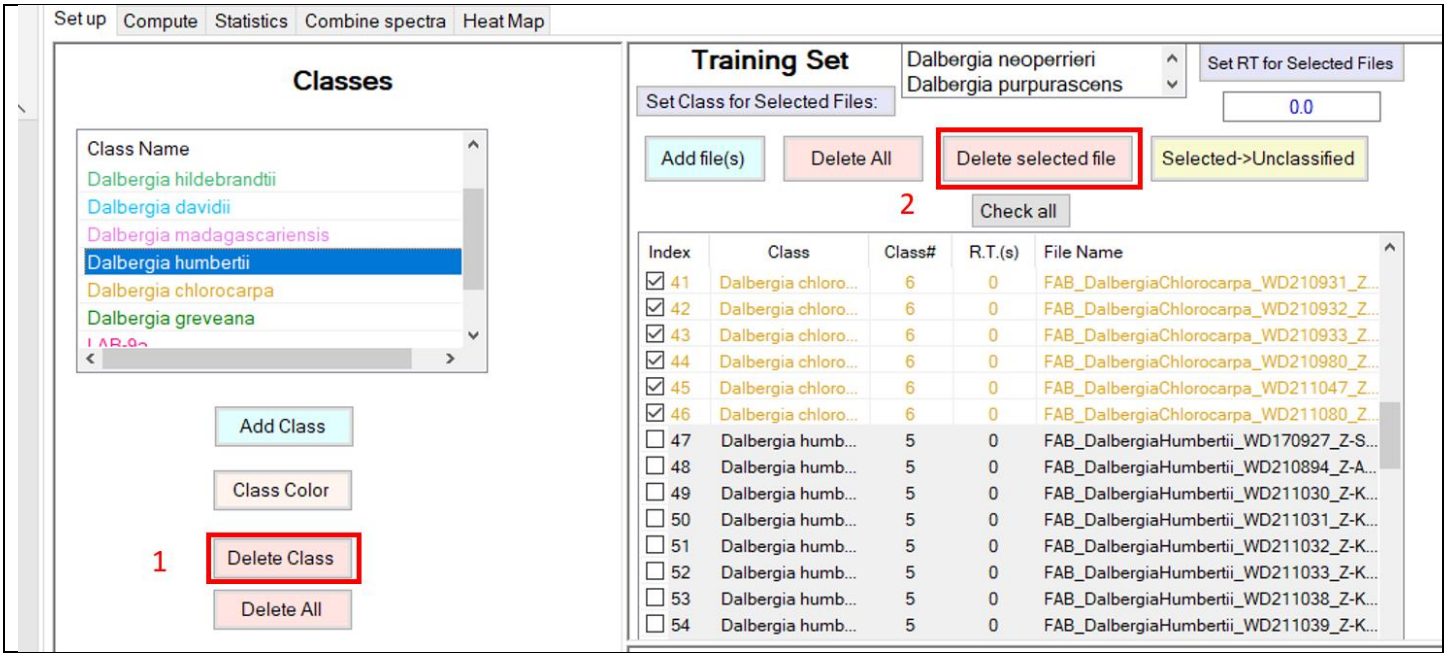
4. Then click the Heat Map button.



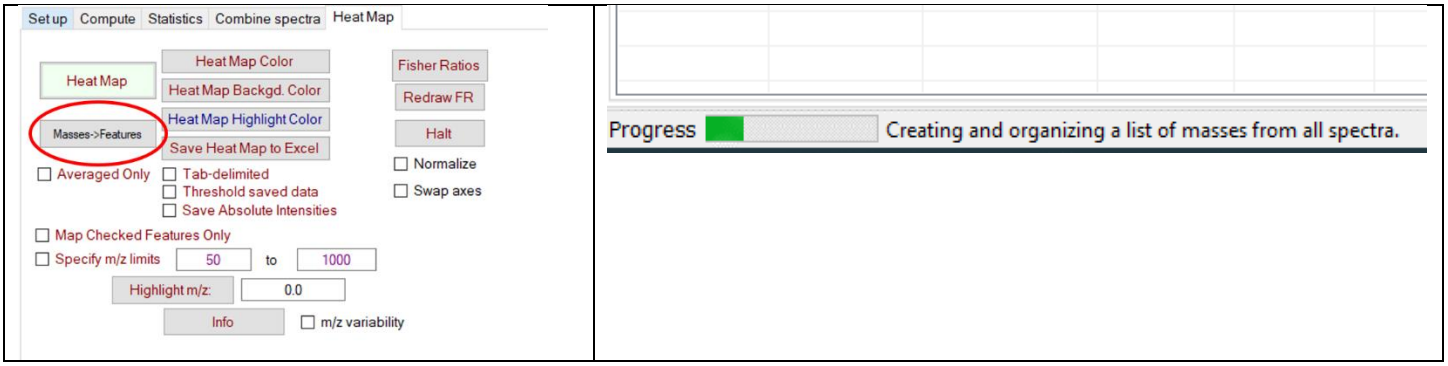
5. Visual analysis of the heat map can allow the user to discard those species that are not good matches. We see in the heat map above that *D. greveana*, *D. humbertii,* and *D. madagascariensis* can be removed from further analysis.
    1. **Note:** More than one heat map may be required, especially if there are more species matches than can easily be seen in a single heat map.



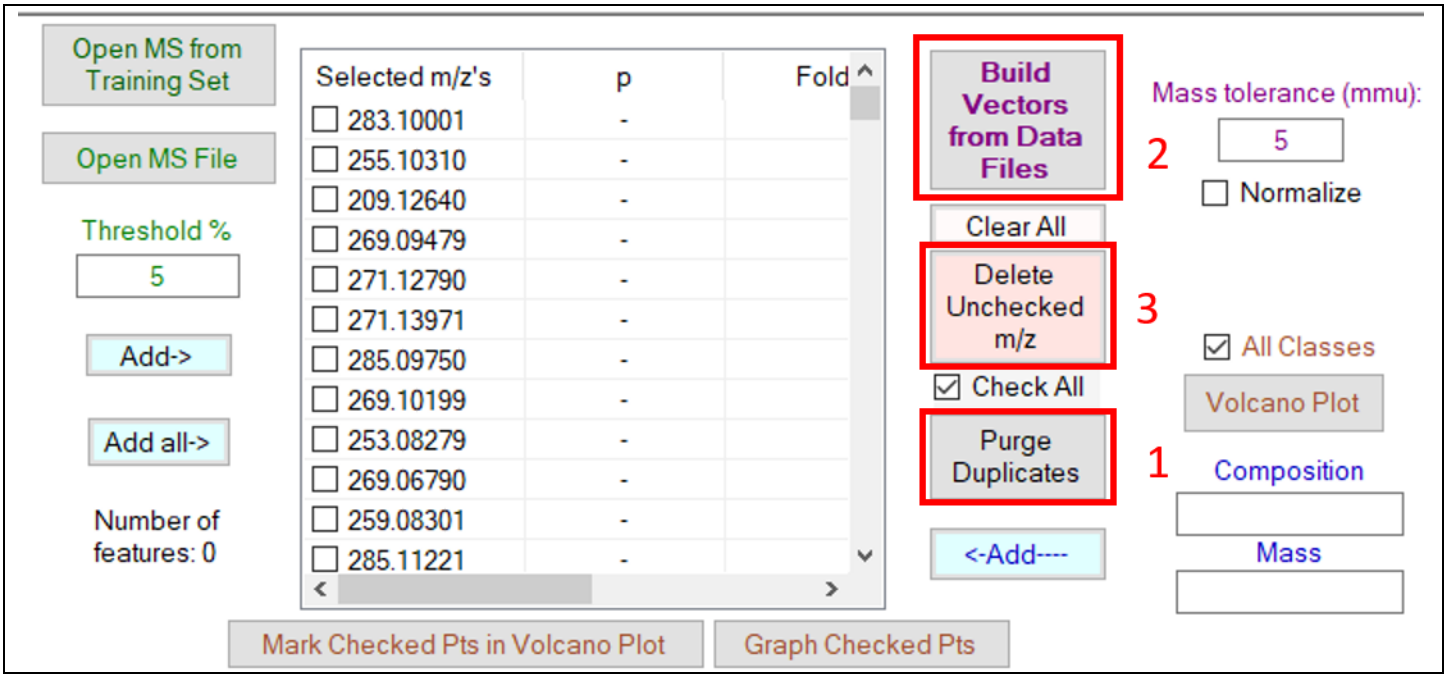
6. Remove species that do not match the unknown/evidence item by navigating to the Set Up tab.
    1. Highlight the species you want to remove under Classes and click Delete Class.
    2. Select the accompanying spectra files from the Training Set and click Delete Selected File.



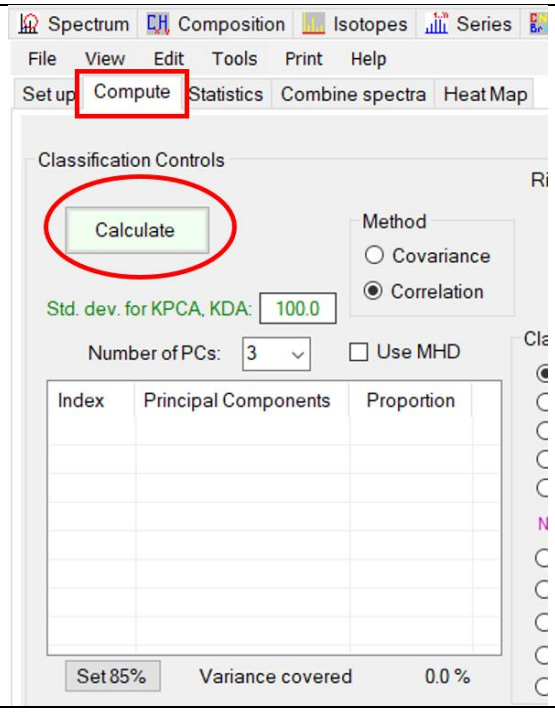
7. Once you are satisfied with your training set, remove the unknown/evidence item spectra and classes, then rebuild the heat map by navigating to the Heat Map tab and pressing the Heat Map button (Steps 3 & 4)
8. Press the Masses -> Features button, this may take a few minutes depending on the size of the training set, you can monitor the progress in the bottom left corner of the screen.



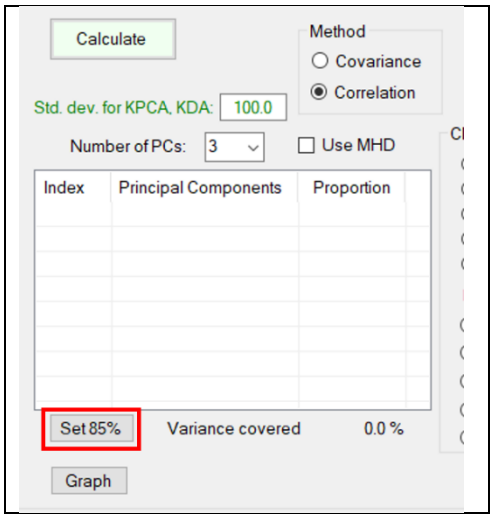
9. Navigate back to the Set Up tab and you will see that there is now a list of ions in the box at the bottom of the screen.
    1. Click Purge Duplicates ~5 times (this deletes any identical entries).
    2. Click Build Vectors from Data Files (this applies ANOVA to the ions to find statistically relevant values).
    3. Click Delete Unchecked m/z (this removes those ions that were not statistically relevant).
        i. **NOTE: If you wish to delete all ions click Clear All.**



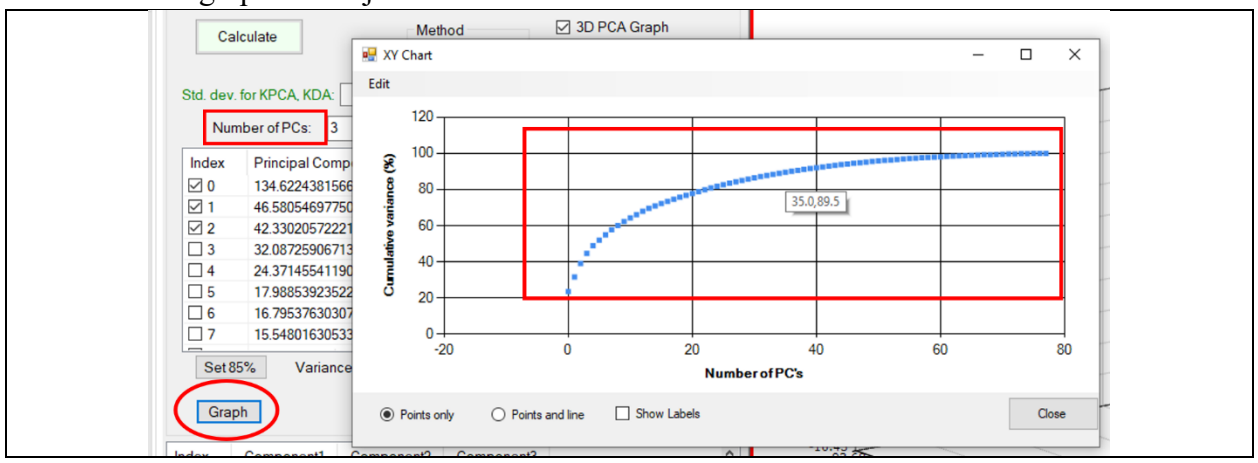
10. Navigate to the Compute tab and click Calculate.



11. A good way to first explore the model is to create a PCA (Principal Component Analysis) model. The default number of PCs (Principal Components) is 3. There are two ways to adjust this number to cover ~85-90% of the variance:
    1. Click the Set 85% button

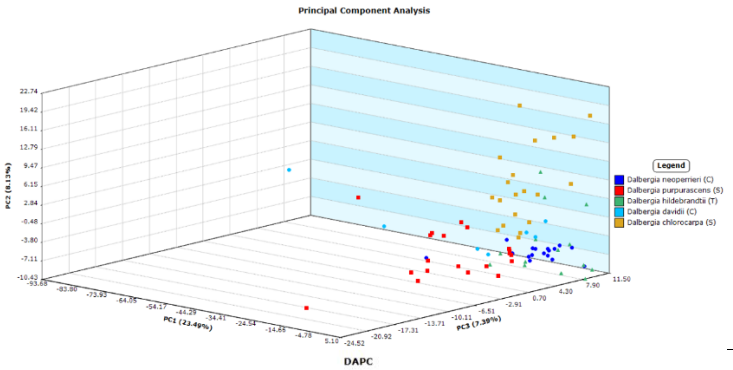
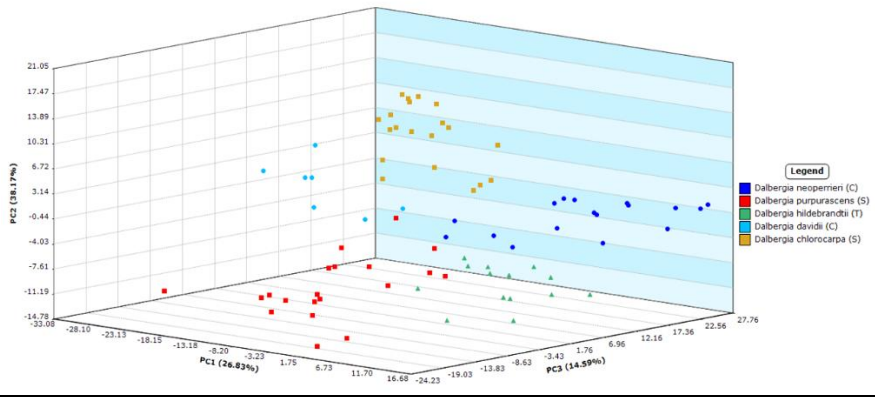


    2. Or, click the Graph button and hover the cursor over the blue diamonds. The values will show up automatically, in the case below we can see that we need 35 PCs to cover 89.5% of the variation. Then close the graph and adjust the Number of PCs value.

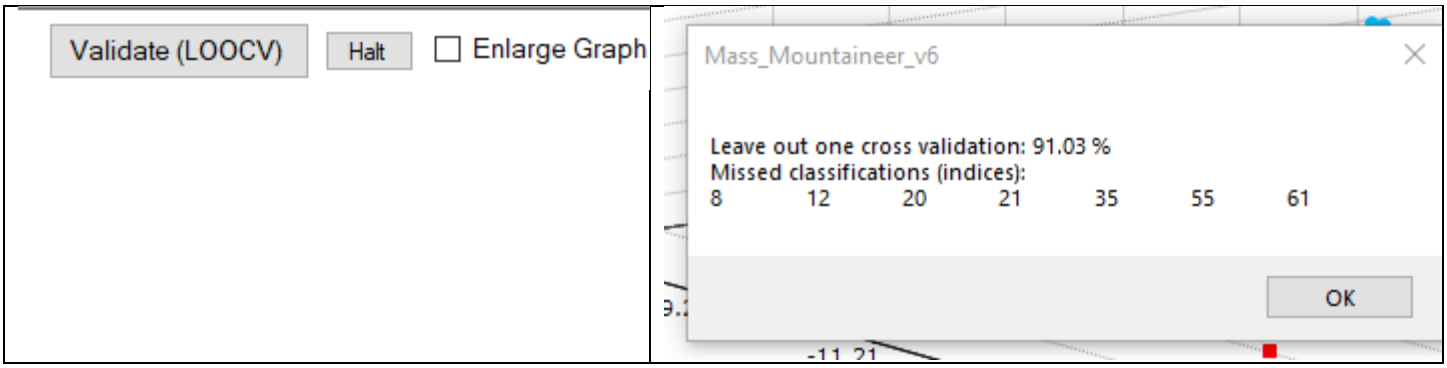


NOTE: PCA is <u>NOT</u> a classification tool, by itself useful for data exploration.

12. Click Calculate once more to see the trends in your data. In order to make a classification model, click the box next to DAPC (Discriminant Analysis of Principal Components) and click Calculate once more. If there are differences between the species classes, the result of DAPC should show some grouping trends.

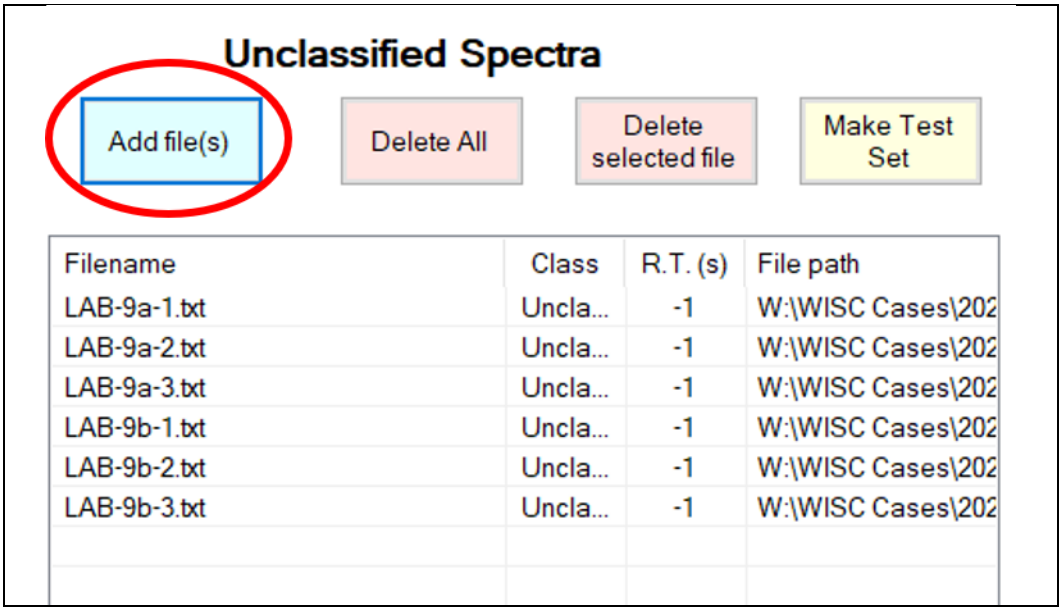
PCA	
DAPC	

13. To find the accuracy of your model, click the Validate (LOOCV) button. LOOCV (Leave One Out Cross Validation) takes each sample out of the model and calculates probability that it belongs to a specific class. An LOOCV value of 100% indicates that all samples were assigned to their correct class, while an LOOCV value below 85% indicates that the model may be unreliable. This generally happens with spectra from closely related species or if a misidentified sample/s is in the training set, sometimes this can be remedied by cleaning up the training set samples through heat map analysis.



NOTE: There are other classification algorithms that are available in Mass Mountaineer, read about them before using so that you know their purpose and usage.

14. Navigate back to the Set Up tab, find the Unclassified Spectra table on the right side of the screen and click Add File(s). Locate your unknown/evidence spectra and click Open.

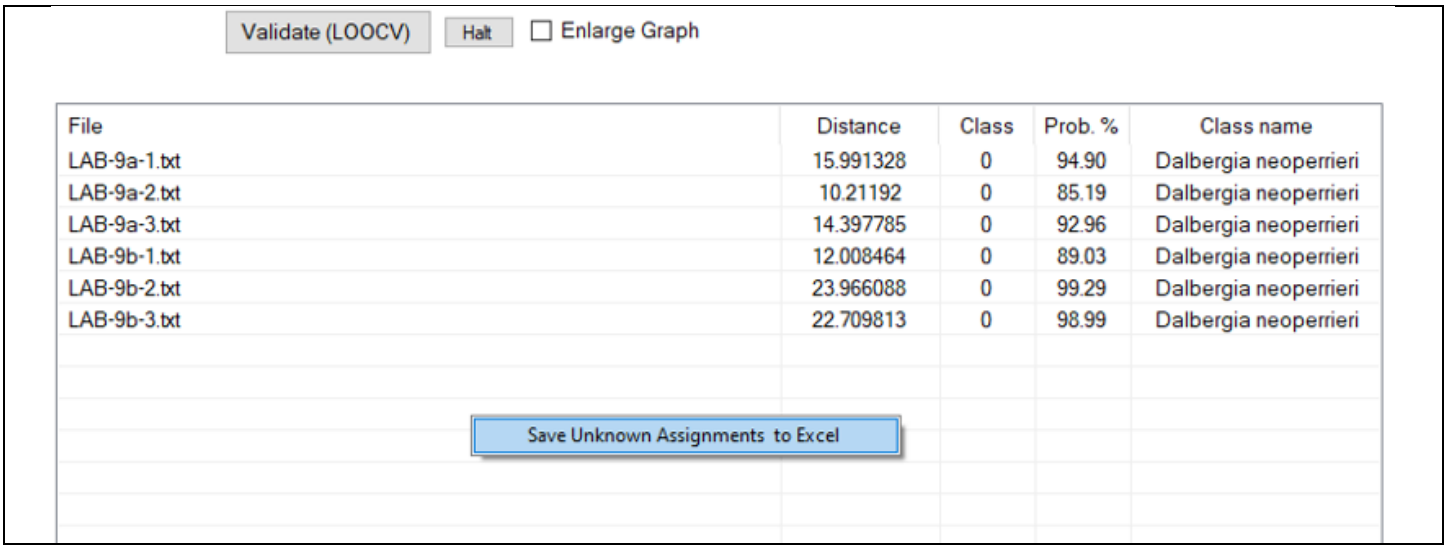


Filename	Class	R.T. (s)	File path
LAB-9a-1.txt	Uncla...	-1	W:\WISC Cases\202
LAB-9a-2.txt	Uncla...	-1	W:\WISC Cases\202
LAB-9a-3.txt	Uncla...	-1	W:\WISC Cases\202
LAB-9b-1.txt	Uncla...	-1	W:\WISC Cases\202
LAB-9b-2.txt	Uncla...	-1	W:\WISC Cases\202
LAB-9b-3.txt	Uncla...	-1	W:\WISC Cases\202

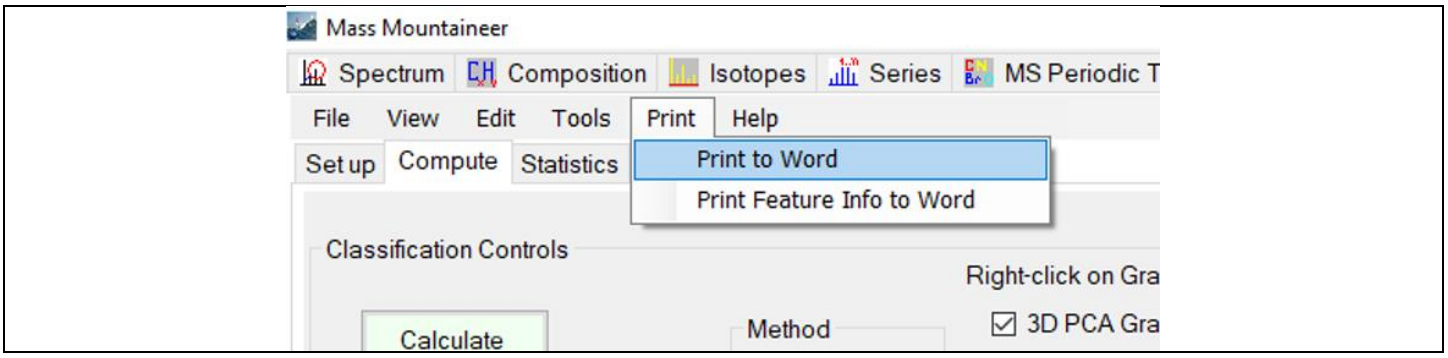
15. Return to the Compute tab and click Calculate.
    i. **NOTE: If you click Validate (LOOCV) after adding spectra to the Unclassified Spectra table they will be deleted and you will have to add them again.**
16. The unclassified spectra will be assigned to the species class that is the best match.

File	Distance	Class	Prob. %	Class name
LAB-9a-1.txt	15.991328	0	94.90	Dalbergia neoperrieri
LAB-9a-2.txt	10.21192	0	85.19	Dalbergia neoperrieri
LAB-9a-3.txt	14.397785	0	92.96	Dalbergia neoperrieri
LAB-9b-1.txt	12.008464	0	89.03	Dalbergia neoperrieri
LAB-9b-2.txt	23.966088	0	99.29	Dalbergia neoperrieri
LAB-9b-3.txt	22.709813	0	98.99	Dalbergia neoperrieri

17. If you wish to keep the results of the classification, right click on the table and select Save Unknown Assignments to Excel.



18. The model parameters can be printed to Word by clicking the Print button and selecting Print to Word. Steps 17 & 18 should always be done when analyzing evidence.



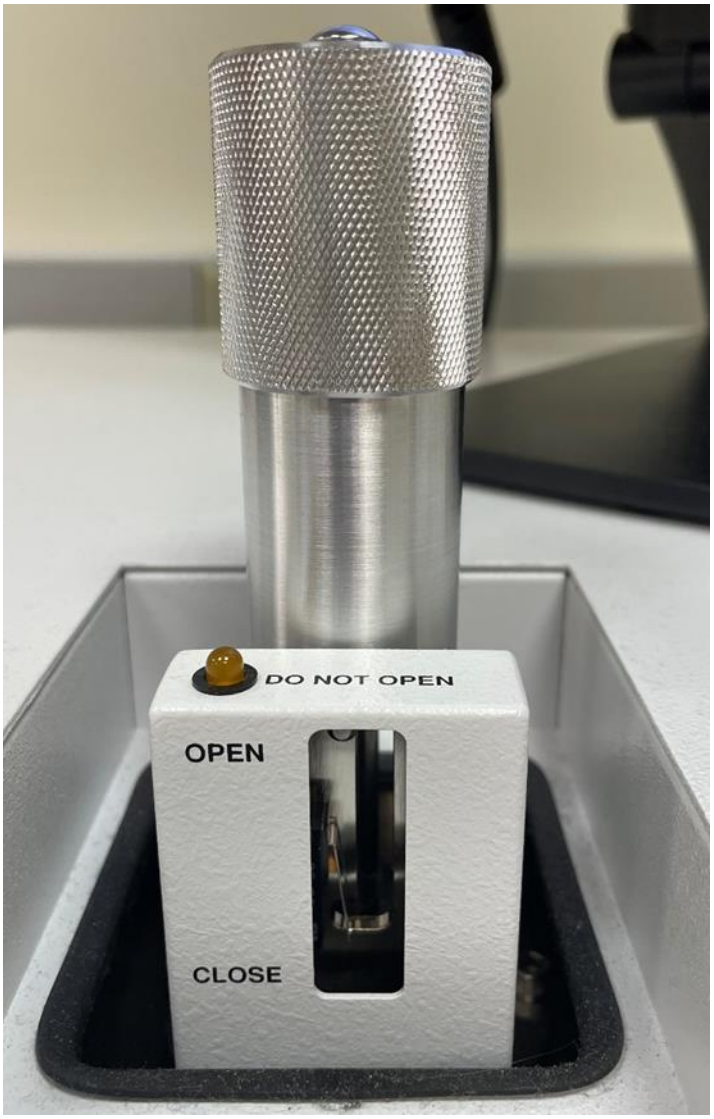
Protocol for Analysis using DART-TOFMS

This document is intended as a general guide, refer to the JEOL manual for complete steps

1. For positive mode operations, dilute the Poly(ethylene glycol) (PEG) calibration standard with a small amount of methanol. If operating in negative mode **do not** dilute the Fomblin calibration standard.

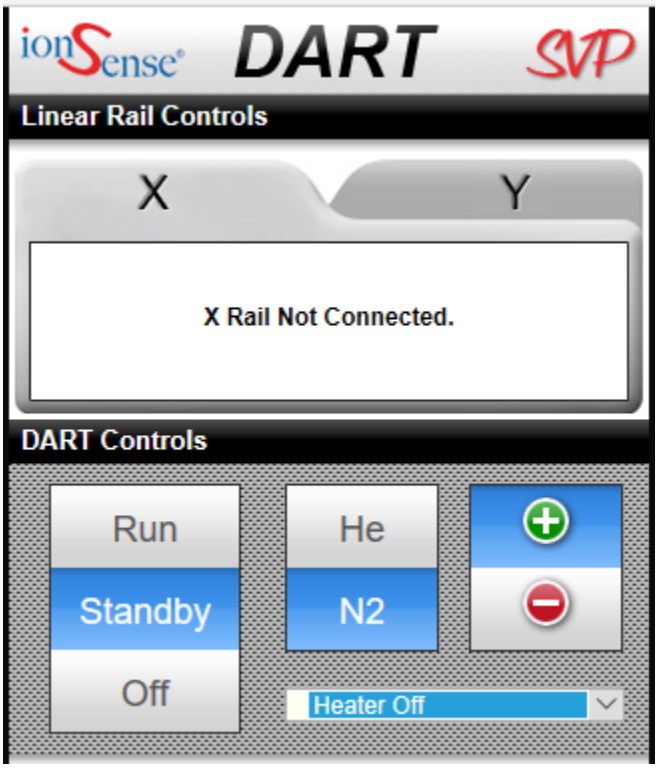
2. Turn on the helium.

3. Locate the Isolation Valve on the top of the Mass Spectrometer. IF LIGHT ON VALVE IS ILLUMINATED DO NOT PROCEED.



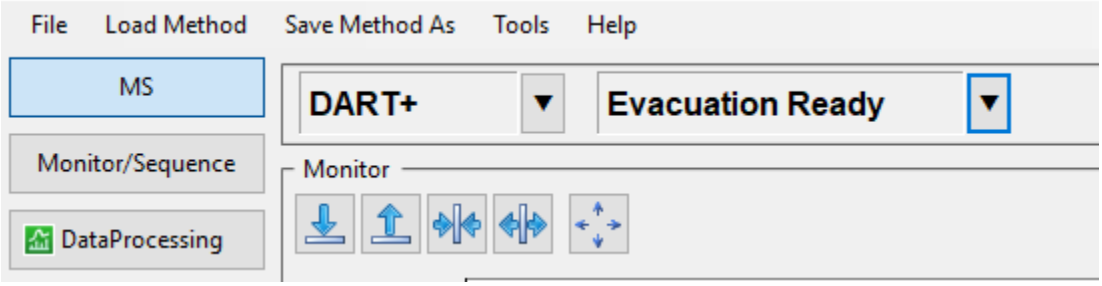
4. Open the Isolation Valve by gently pressing down and to the right, then pull up and turn the valve to the left. There is a small bar that is visible on the valve, look at this to ensure proper placement while opening and closing the valve.

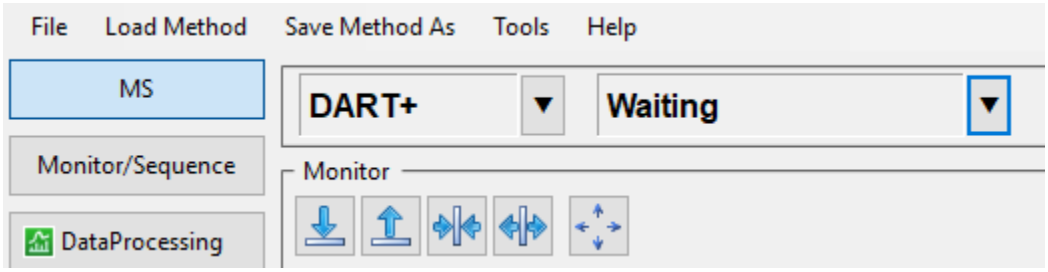
5. Navigate to the DART controller screen by clicking the Internet Explorer icon, the window should be open already. If it is not, enter the following IP address: 192.168.10.111/



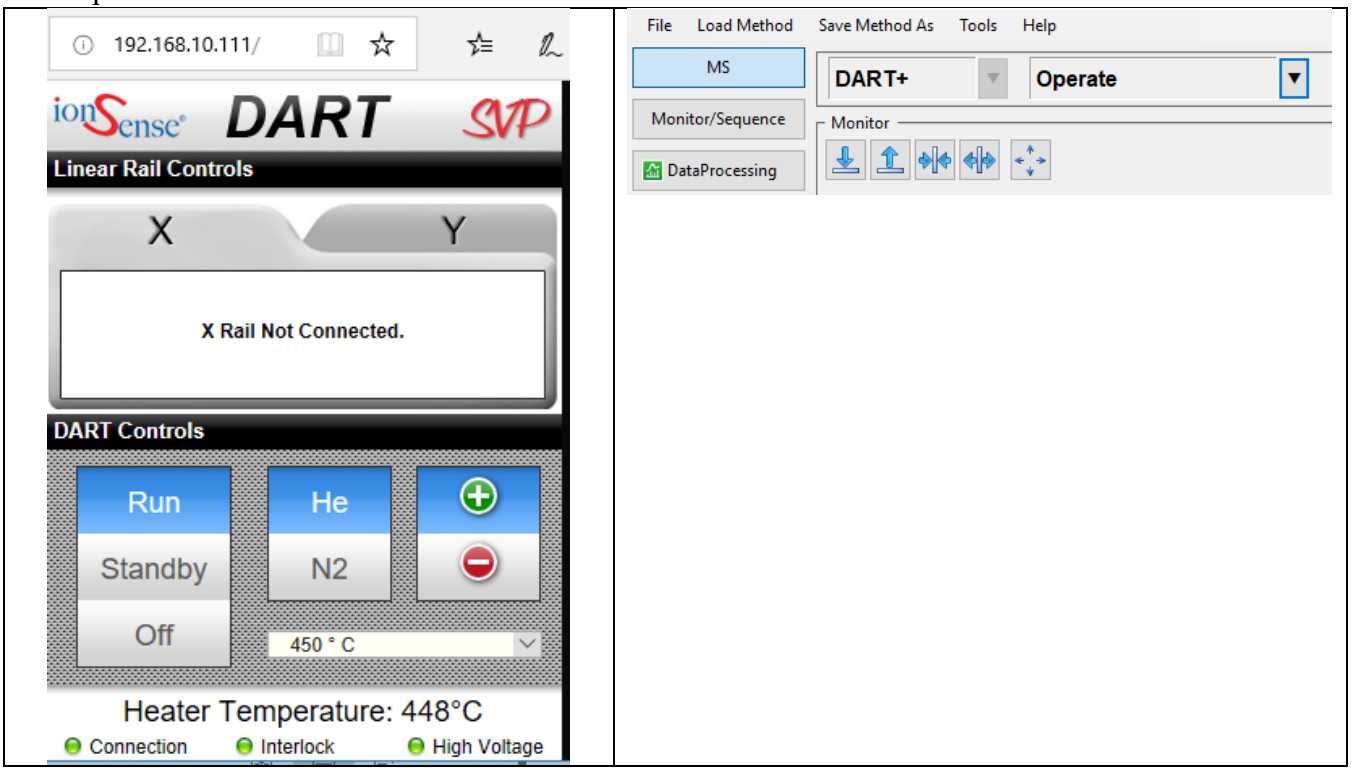
6. Click Standby → turn heater on to desired temperature (350°C or 450°C are typical)

7. While the DART is heating, open msAxel. Make sure that the method matches the method on the DART screen (both should be on DART + **or** DART -). Click the drop-down arrow and switch from Evacuation Ready to Waiting.





8. Once the DART has reached the selected temperature, click Run on the DART screen and Operate from the drop-down arrow on msAxel.



9. Allow the instrument to warm up, intensity and resolution values will fluctuate greatly for the first minute or two, use this time to fill out the calibration sheet. Fill in the highlighted section. The outlined section will be filled out AFTER collecting the new calibration.

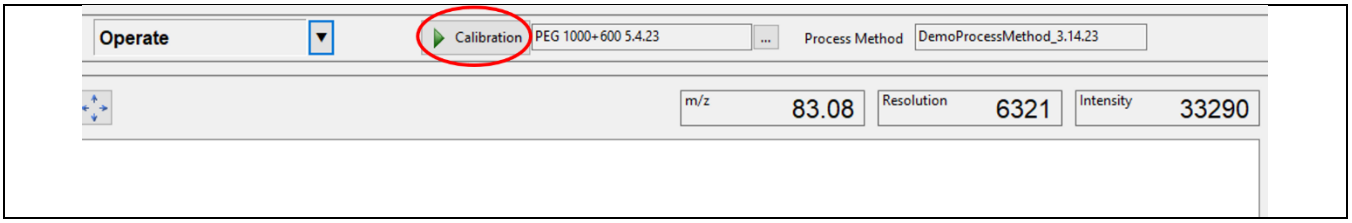
DART TOFMS Daily QA/QC Log

Oil mist filter must be drained every Friday

Date	Initials	Temp °C	Detector Voltage	Ion (~109.10)	Resolution	Intensity	PEG Drift (371.22811)	Caviunin Drift (375.10799)	1-R	File Name /Comments

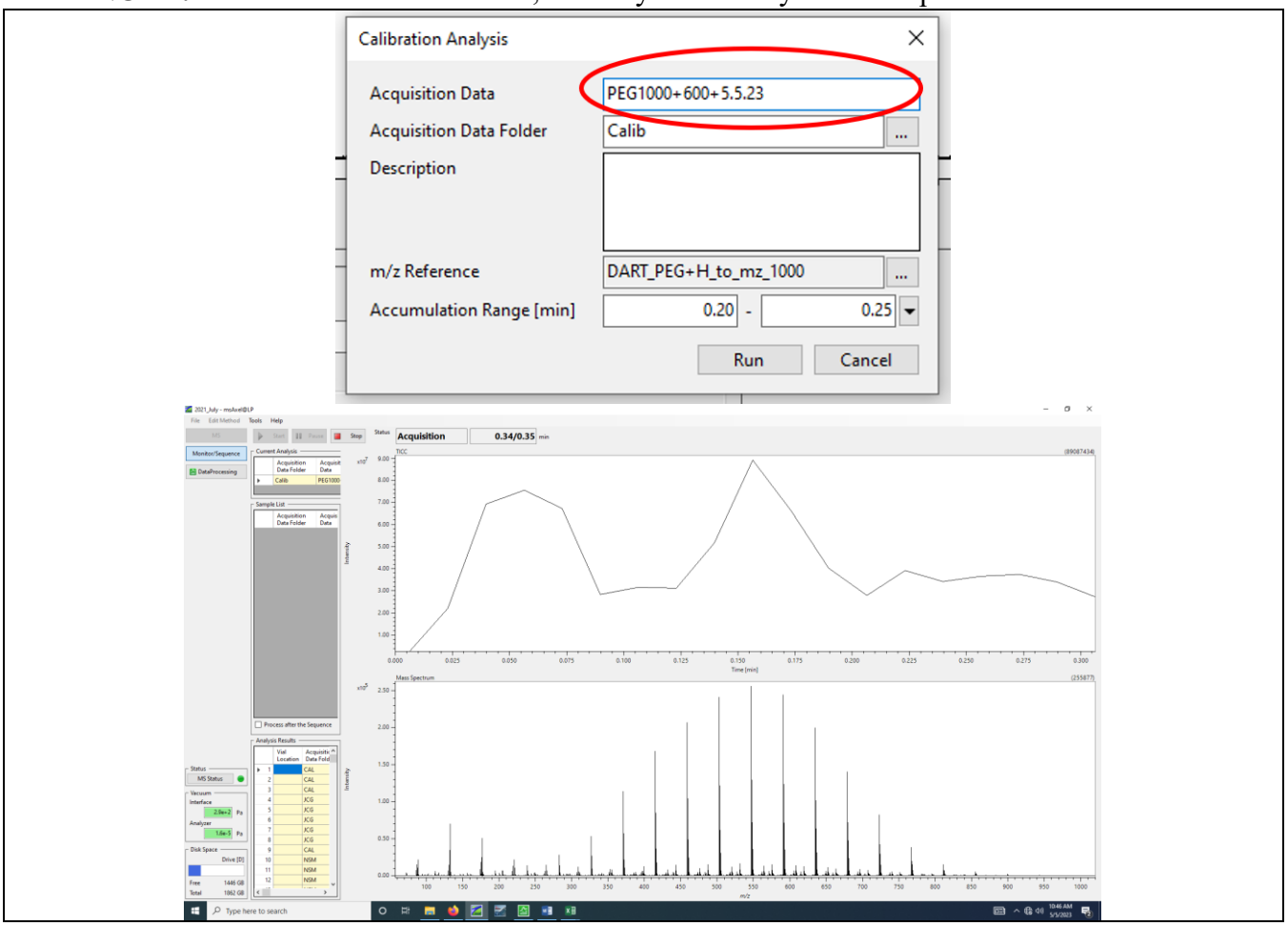
NOTE: Ensure that the instrument is operating correctly by checking the values you see against previous values listed on the calibration sheet, if a discrepancy is noted (e.g., intensity has been previously written as 9000-12000 and is currently fluctuating around 5000 or less) inform personnel. Mass spectra generated while the DART-TOFMS is not operating correctly will need to be re-collected and could damage the instrument.

10. A new calibration file should be made each day before collecting spectra. From the msAxel screen click the Calibration button:



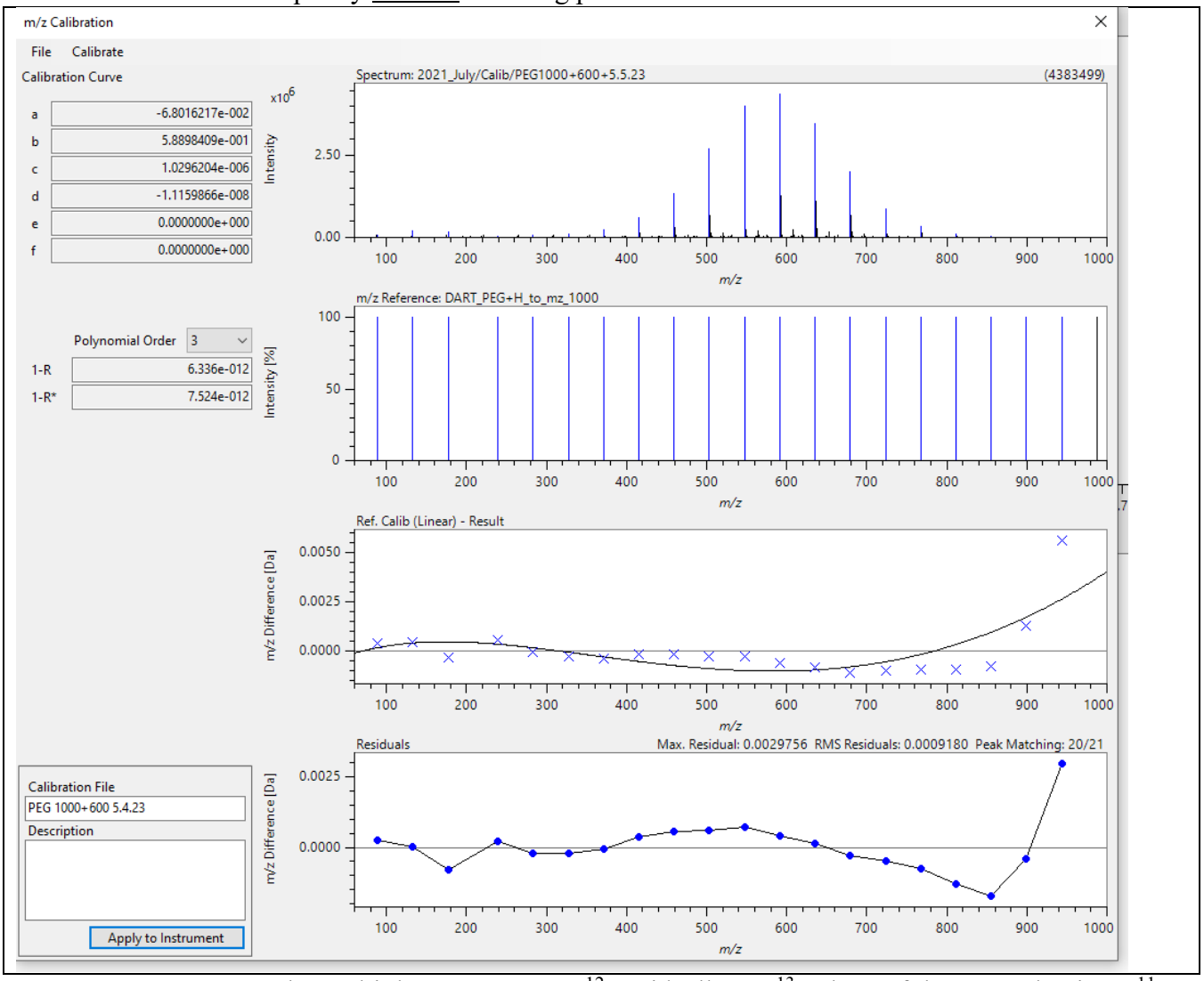
The following screen will appear. Make sure that the m/z Reference file is the same as shown in the picture. Update the Acquisition Data name to the day's date. When ready, click Run and the normal collection window will appear.

NOTE: The collection time is short, be ready to collect your PEG spectrum!



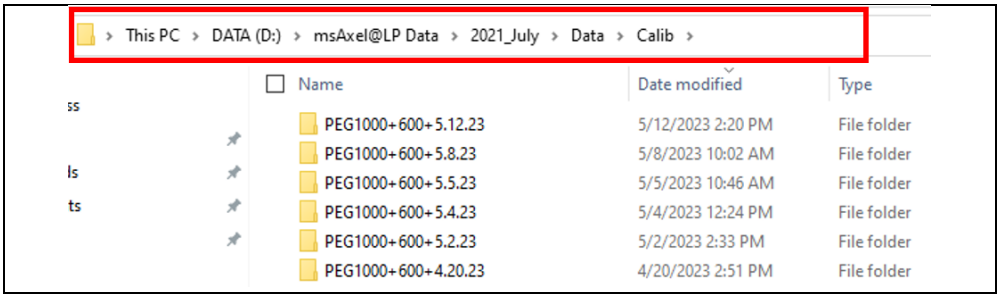
Hold the capillary tube in the sample gap for the entire duration of the run time, this will help the user to collect the full PEG spectrum.

At the end of the collection time, the following screen will automatically appear. Check the 1-R value to be sure it is of sufficient quality without releasing points.

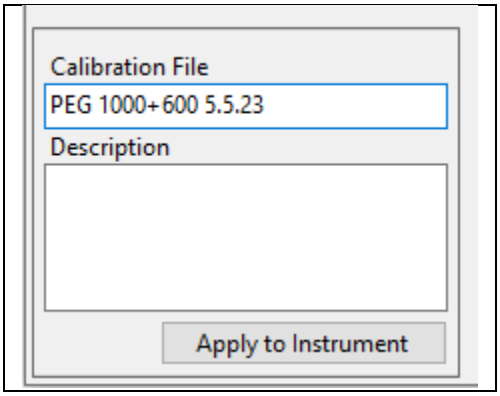


NOTE: It may take multiple runs to get 10^{-12} or, ideally, 10^{-13} values. If the 1-R value is 10^{-11}, we encourage you to find where the file you made was stored, delete it, and collect a new PEG calibration spectrum.

- The location of the calibration files will be similar to the following:



If the 1-R value is acceptable, make sure that the Calibration File name in the bottom left corner is correct, and click Apply to Instrument.



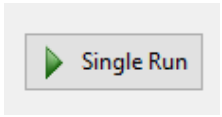
Record the 1-R value in the QA/QC Log and you are now ready to collect mass spectra from wood.

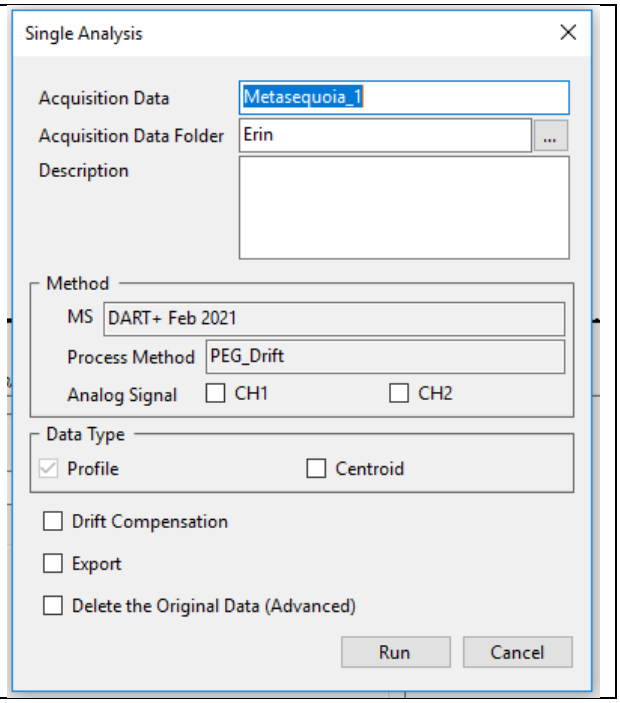
11. We encourage DART TOFMS users to collect a single spectrum from a known timber standard and record a single ion from that standard once per day while operating the DART TOFMS system. WISC & NFWFL use the Caviunin ion from *Dalbergia nigra*, but the standard can be any one of the standards supplied to users who have completed the training in Ashland, OR. Their ions can be found in the Excel spreadsheet "DART-MS Validation", e.g.:
    a. 381.2066 m/z from *Milicia* sp.
    b. 205.1915 m/z from *Entandrophragma* sp.

    Once the reference standard ion is recorded in the QA/QC Log, no more standards need to be collected and samples can be collected as usual.

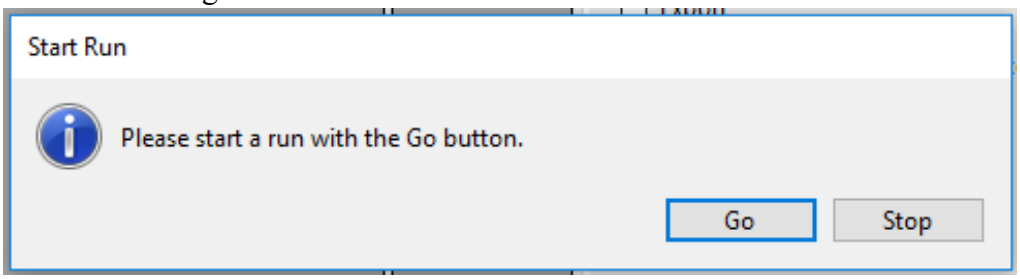
    The following steps have this process added into the flow for clarification, remember that making a new calibration and collecting a single ion from a wood reference standard needs to be done only **ONCE** per day.

12. On the msAxel screen, choose Single Run on the bottom right of the MS screen. Press the **…** button to change the destination folder and type a unique file name that correlates to the set of samples to be analyzed. Create a new destination folder by inputting a file name that is not in use.





13. Press Run and then Go to begin the run.



14. PEG must be run as the first, middle, and last sample. Collecting multiple PEG spectra during a single run is beneficial because it provides multiple spectral options for drift correction and PEG can be used as a marker to assist in identifying sample spectra from a Total Ion Current Chromatogram (see page 7: Data Reduction). E.g.:
    i. PEG
    ii. Positive Control (*Dalbergia nigra* or other species)
    iii. PEG
        1. Sample 1
        2. Sample 2
        3. Sample 3
        4. Sample 4
        5. Sample 5
    iv. PEG
        6. Sample 6
        7. Sample 7
        8. Sample 8
        9. Sample 9
        10. Sample 10
    v. PEG

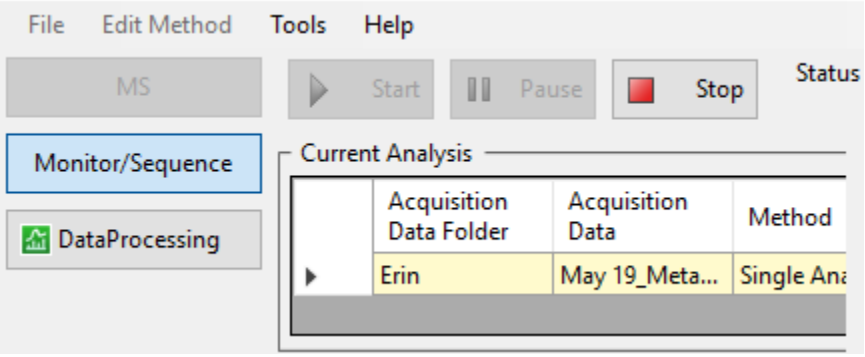
*NOTE: A positive control sample should be run with the first batch of samples in a day AND at the beginning of every batch of evidence samples.

15. Allow time between each sample, there should be sufficient time between chromatograms such that a background can be subtracted from the target chromatogram.

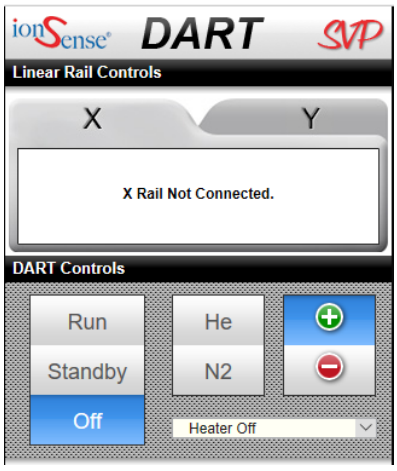
16. Select a sliver and hold it such that the heated helium can flow over the sample. The sample should not be blocking the analytical orifice. Monitor the intensity of the resulting chromatogram, low intensity peaks can be indicative of a poorly placed sample or that there is a blockage in Orifice 1.

    NOTE: If the intensities of a set of samples is persistent and unusually low it is likely that the orifice needs cleaning; contact personnel before taking any steps to solve this problem.

17. Once the selected set of samples is run, press STOP at the upper left of the screen. This will stop the analysis and return the user to the Monitor/Sequence screen. To continue to run additional sample sets, click the MS button and repeat steps 12-17. If you are finished with running samples continue on to step 16.



18. To shut down the DART-TOFMS, click the MS button on msAxel. Click the Internet Explorer icon and turn the DART to Off. Then the drop-down arrow and select Waiting or Evacuation Ready (see step 7 above).



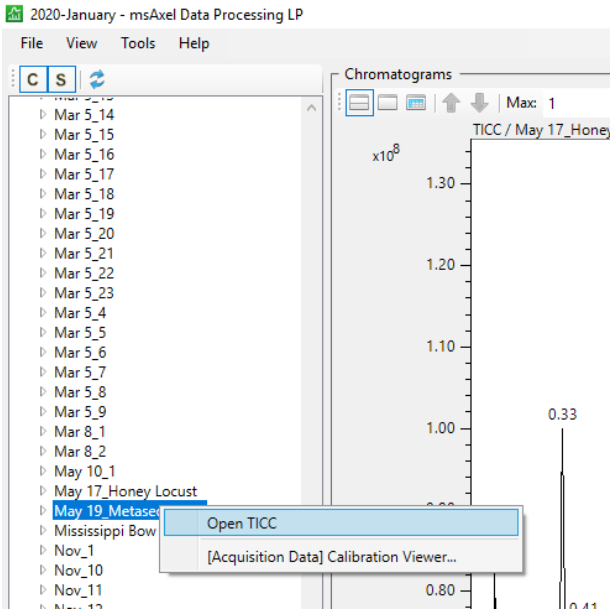
19. **Turn off the helium and close the Isolation Valve!**

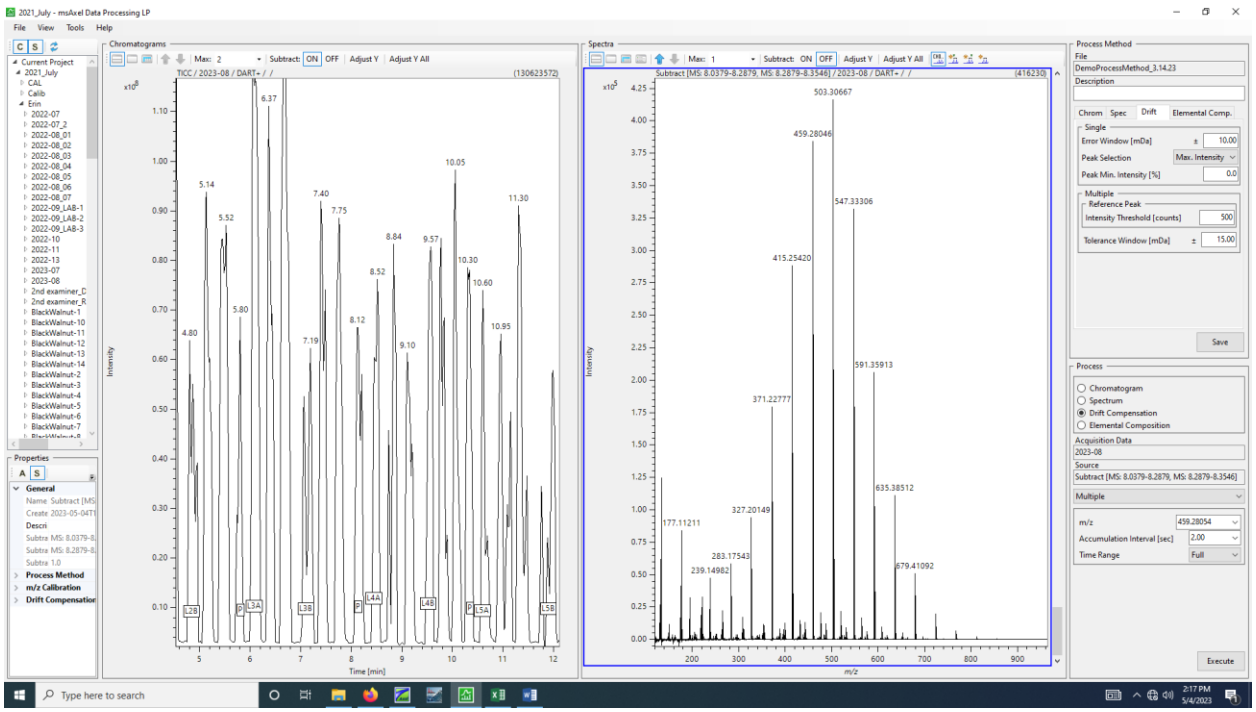
Data Reduction

1. Open msAxel and click DataProcessing

2. Select the folder containing your spectra

3. Right click on the file name and select Open TICC

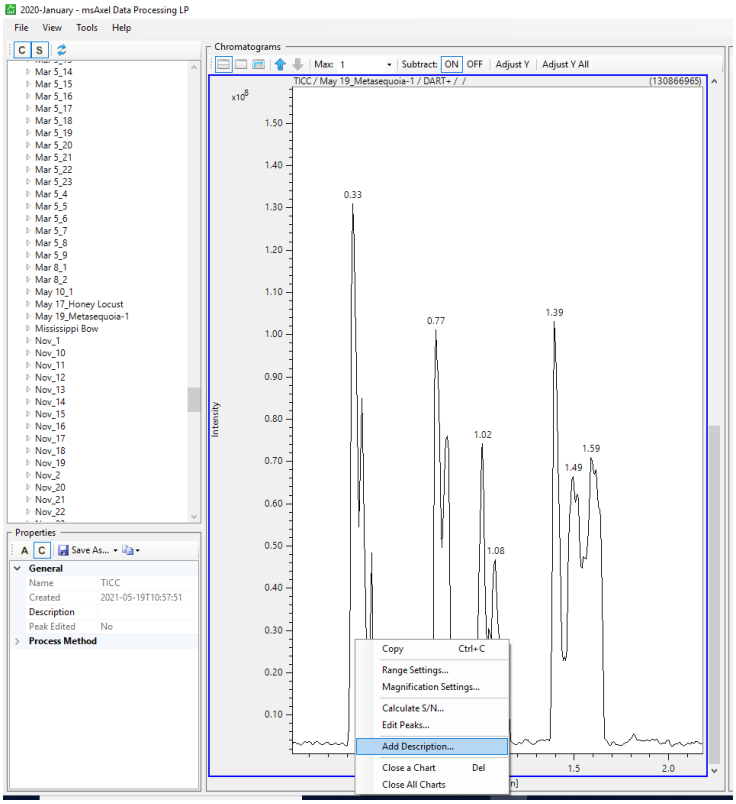




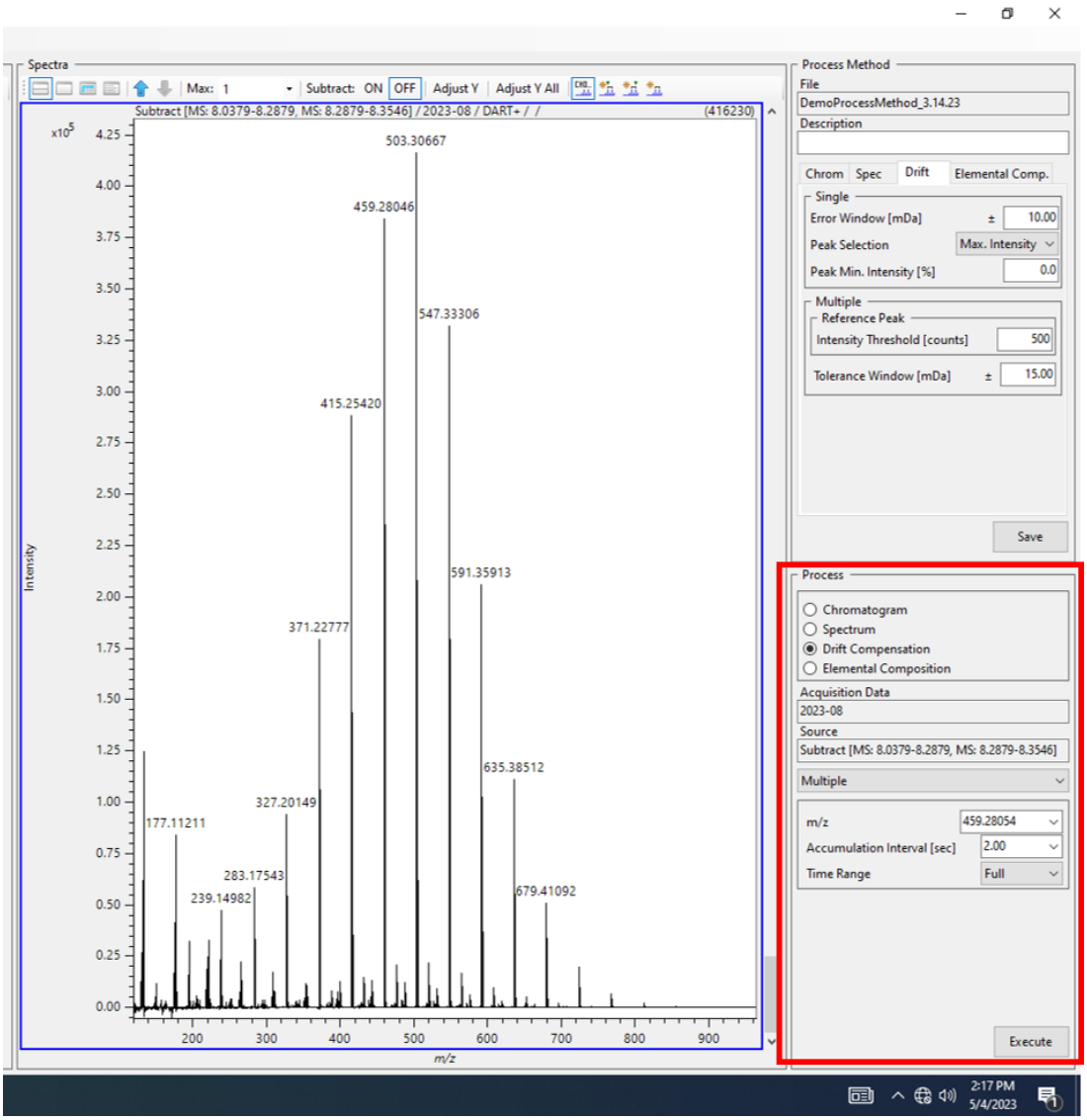
4. In the Chromatograms window on the left, find a calibration standard sample and press both Ctrl and the left mouse button, drag the cursor across the second calibration standard peak. Release the mouse and Ctrl. Press both Shift and the left mouse button and drag the cursor over an

equally sized portion of the background. The collected spectrum will appear in the window to the right.

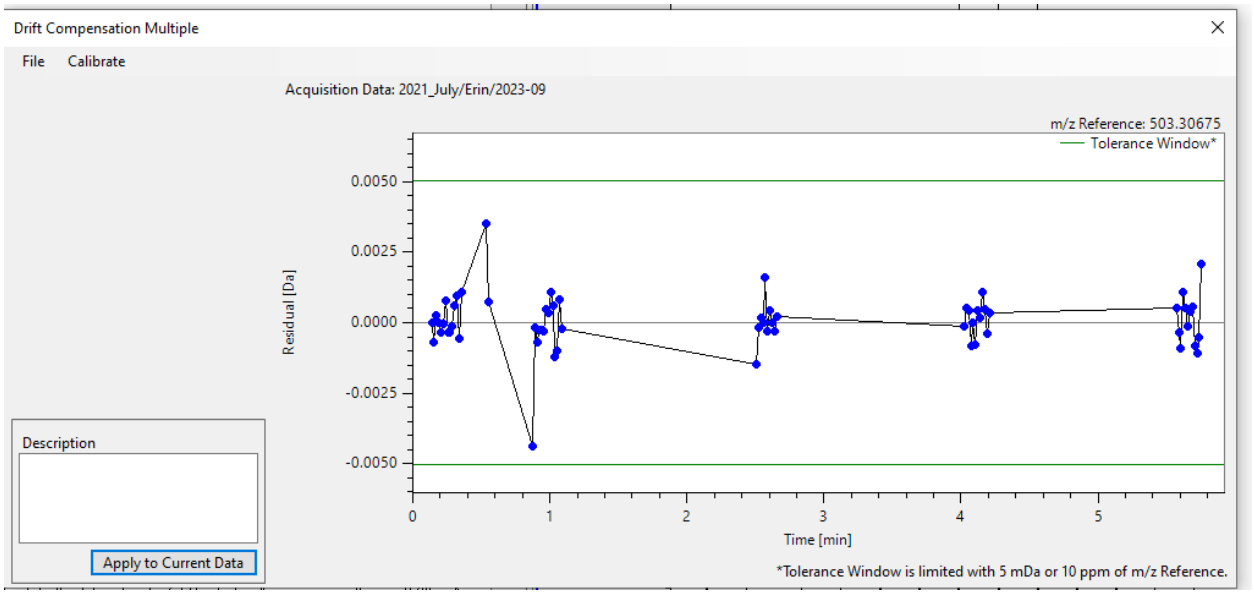
TIP: Label your calibration and/or sample spectra by right clicking on a peak and selecting Add Description.



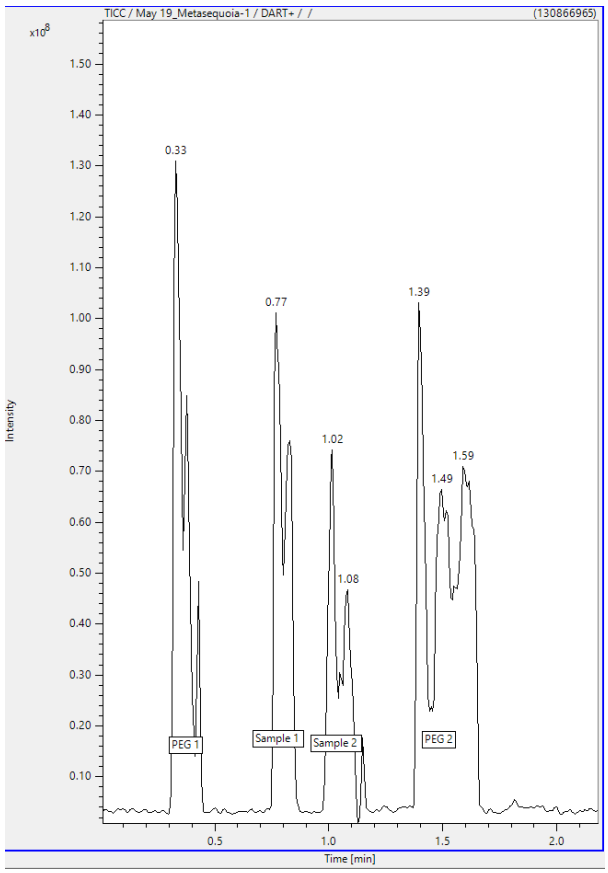
5. Once the calibration spectrum is in view, navigate to the bottom right of the screen and click Drift Compensation, then select Multiple. Use one of the PEG ions to correct the drift in the TICC (Table shown at end of document) and click Execute.



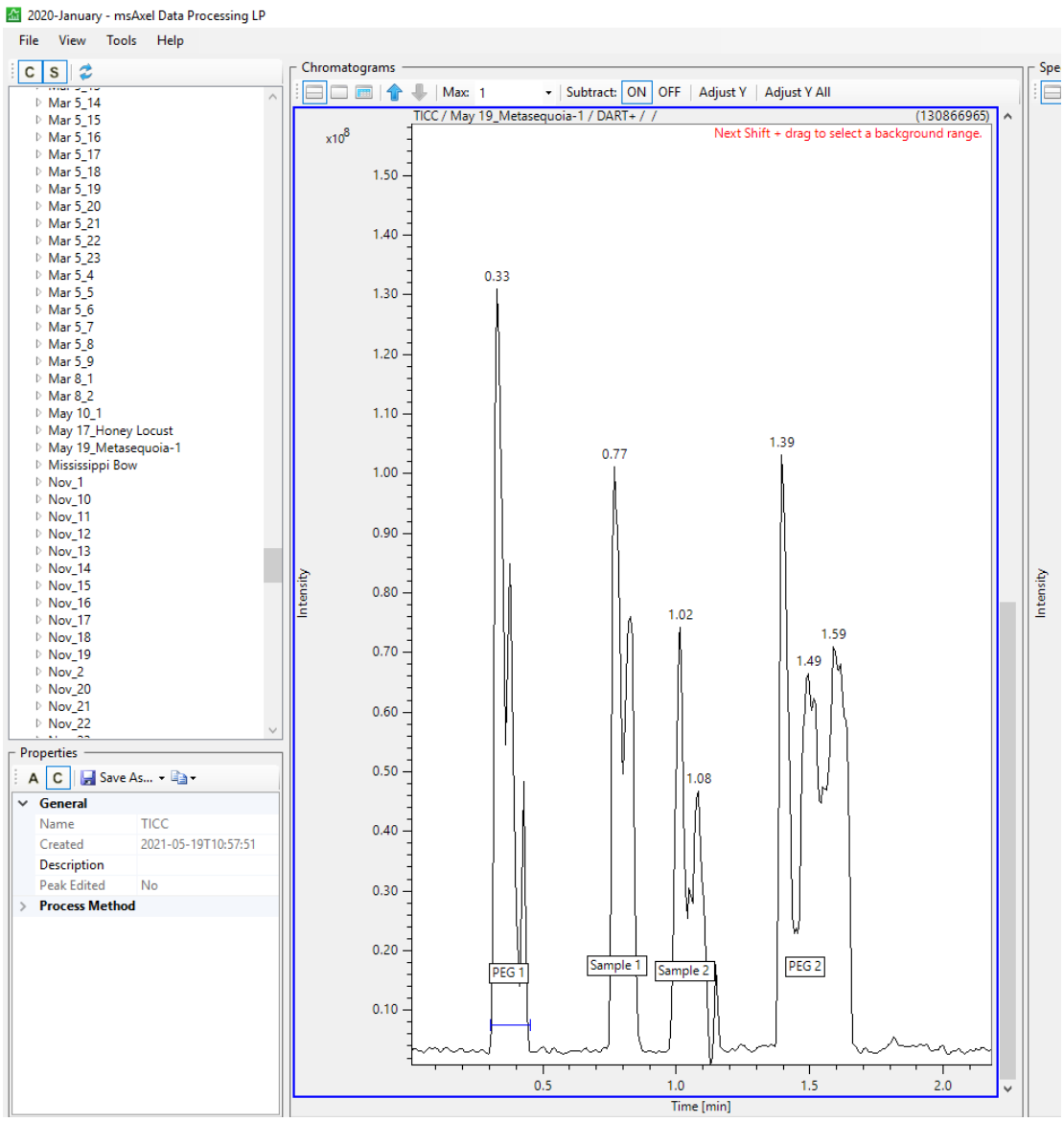
6. A new screen will appear, check that you see clusters of dots at the time where each PEG spectrum was collected. If the PEG dots do not appear, try changing the m/z value shown in the image above to another ion found in PEG.



If the PEG clusters are satisfactory (i.e., there is a cluster of dots at each time that a PEG spectrum was collected), click Apply to Current Data. This process will correct for drift across the entire TICC. Record the actual value of the PEG ion 371.22881 as well as the ion of choice from your wood reference sample in the QA/QC sheet. You are now ready to collect sample mass spectra.

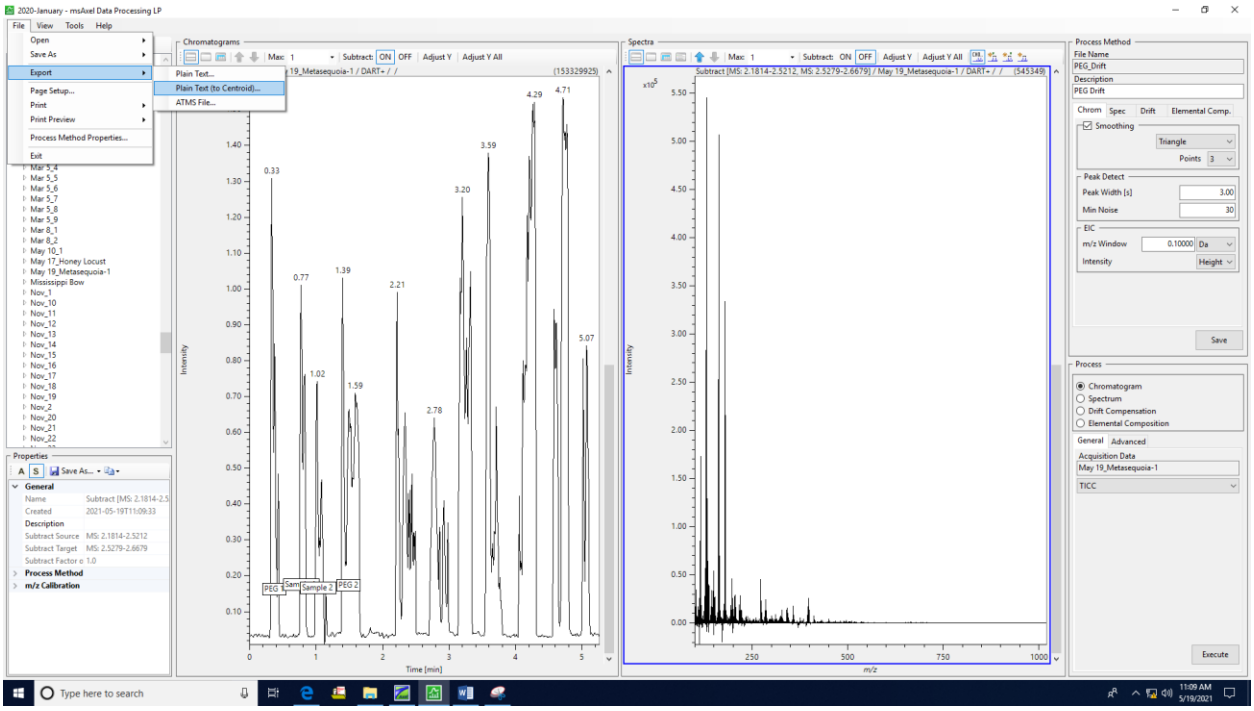


5. Pressing both Ctrl and the left mouse button, drag the cursor across the first sample peak.



Release the mouse and Ctrl. Press both Shift and the left mouse button and drag the cursor over the background space. The sample spectrum will appear in the right Spectra screen.

11. In the Spectra window, right click on empty space in the right window to select the spectrum window, then click File→Export→Plain Text (to Centroid). Be sure that the destination is correct, then save the spectrum under a unique file name.



NOTES: If the mouse is dragged or clicked without holding Ctrl or Shift the screen will zoom in on a chromatogram; simply double click on the left window to return to the larger view.

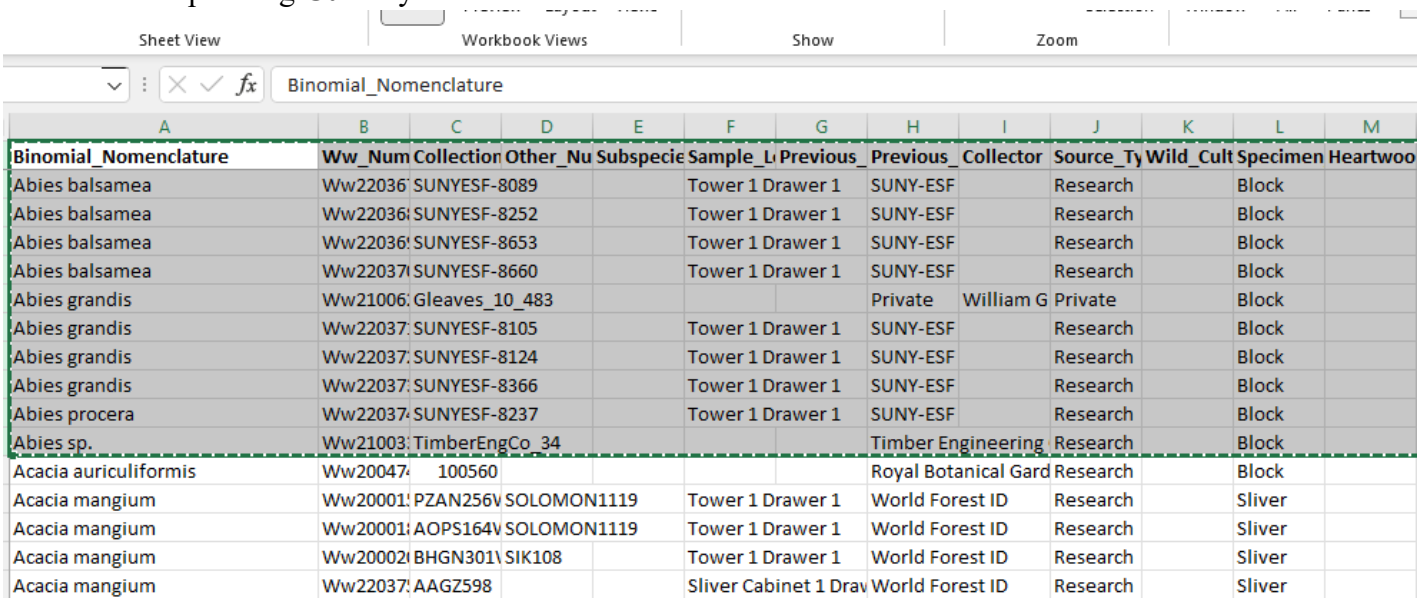
After collecting ~50 spectra, the Spectra screen will not allow additional spectra to be opened. Right click in the right window and select “Close All Charts”. Do this regularly or after each sample set.

Multiple windows can be used by clicking the drop-down arrows at the top of both the left and right screens. If you use this feature, be absolutely certain of which sample spectrum you are collecting, misnamed species spectra can lead to 1) rerunning the sample or 2) unnecessarily removing a spectrum due to lack of consensus.

Protocol for the Creation of File Names

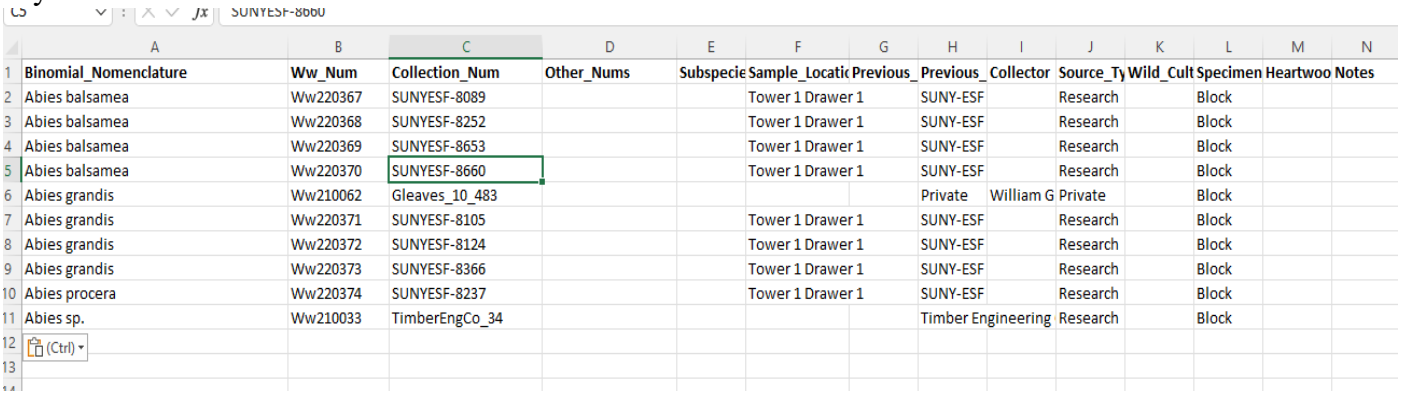
1. To begin, access either the WD or Ww database in Excel.

2. Copy the lines from Excel of the target samples you want to create file names for by highlighting the lines and then pressing **Ctrl** key + c.



Binomial_Nomenclature	Ww_Num	Collection	Other_Nu	Subspecie	Sample_L	Previous_	Previous_	Collector	Source_Ty	Wild_Cult	Specimen	Heartwoo
Abies balsamea	Ww22036	SUNYESF-8089			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies balsamea	Ww22036	SUNYESF-8252			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies balsamea	Ww22036	SUNYESF-8653			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies balsamea	Ww22037	SUNYESF-8660			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies grandis	Ww21006	Gleaves_10_483					Private	William G	Private		Block	
Abies grandis	Ww22037	SUNYESF-8105			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies grandis	Ww22037	SUNYESF-8124			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies grandis	Ww22037	SUNYESF-8366			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies procera	Ww22037	SUNYESF-8237			Tower 1 Drawer 1		SUNY-ESF		Research		Block	
Abies sp.	Ww21003	TimberEngCo_34					Timber Engineering		Research		Block	
Acacia auriculiformis	Ww20047	100560					Royal Botanical Gard		Research		Block	
Acacia mangium	Ww20001	PZAN256V	SOLOMON1119		Tower 1 Drawer 1		World Forest ID		Research		Sliver	
Acacia mangium	Ww20001	AOPS164V	SOLOMON1119		Tower 1 Drawer 1		World Forest ID		Research		Sliver	
Acacia mangium	Ww20002	BHGN301\	SIK108		Tower 1 Drawer 1		World Forest ID		Research		Sliver	
Acacia mangium	Ww22037	AAGZ598			Sliver Cabinet 1 Drav		World Forest ID		Research		Sliver	

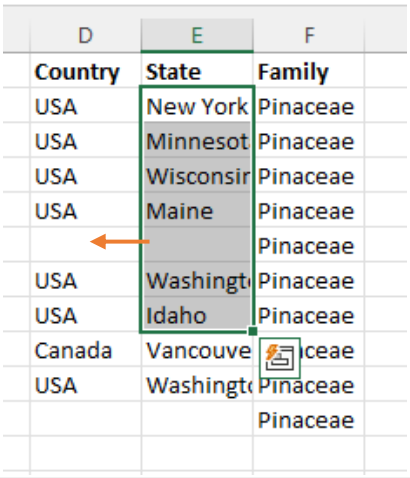
3. Transfer the copied target sample information into a new Excel sheet by pressing the **Ctrl** key and the **V** key.



	Binomial_Nomenclature	Ww_Num	Collection_Num	Other_Nums	Subspecie	Sample_Locatic	Previous_	Previous_	Collector	Source_Ty	Wild_Cult	Specimen	Heartwoo	Notes
2	Abies balsamea	Ww220367	SUNYESF-8089			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
3	Abies balsamea	Ww220368	SUNYESF-8252			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
4	Abies balsamea	Ww220369	SUNYESF-8653			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
5	Abies balsamea	Ww220370	SUNYESF-8660			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
6	Abies grandis	Ww210062	Gleaves_10_483					Private	William G	Private		Block		
7	Abies grandis	Ww220371	SUNYESF-8105			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
8	Abies grandis	Ww220372	SUNYESF-8124			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
9	Abies grandis	Ww220373	SUNYESF-8366			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
10	Abies procera	Ww220374	SUNYESF-8237			Tower 1 Drawer 1		SUNY-ESF		Research		Block		
11	Abies sp.	Ww210033	TimberEngCo_34					Timber Engineering		Research		Block		

4. Keep only the columns that are labeled: **Binomial_Nomenclature**, **Ww_Num** OR **WD_Num**, **Collection_Num**, **Country****, and **Family**.

When the country is USA, the state will be used in the file name instead. To do this, shift the states to the left to replace the country "USA". Finally, the **State column can be deleted.

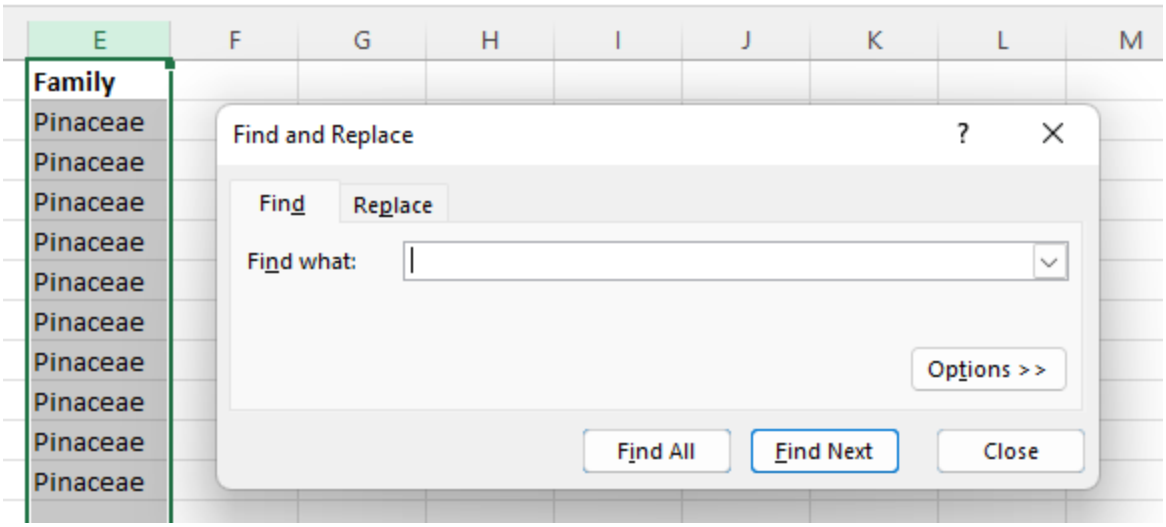


D	E	F
Country	**State**	**Family**
USA	New York	Pinaceae
USA	Minnesot	Pinaceae
USA	Wisconsir	Pinaceae
USA	Maine	Pinaceae
		Pinaceae
USA	Washingt	Pinaceae
USA	Idaho	Pinaceae
Canada	Vancouve	ceae
USA	Washingt	Pinaceae
		Pinaceae

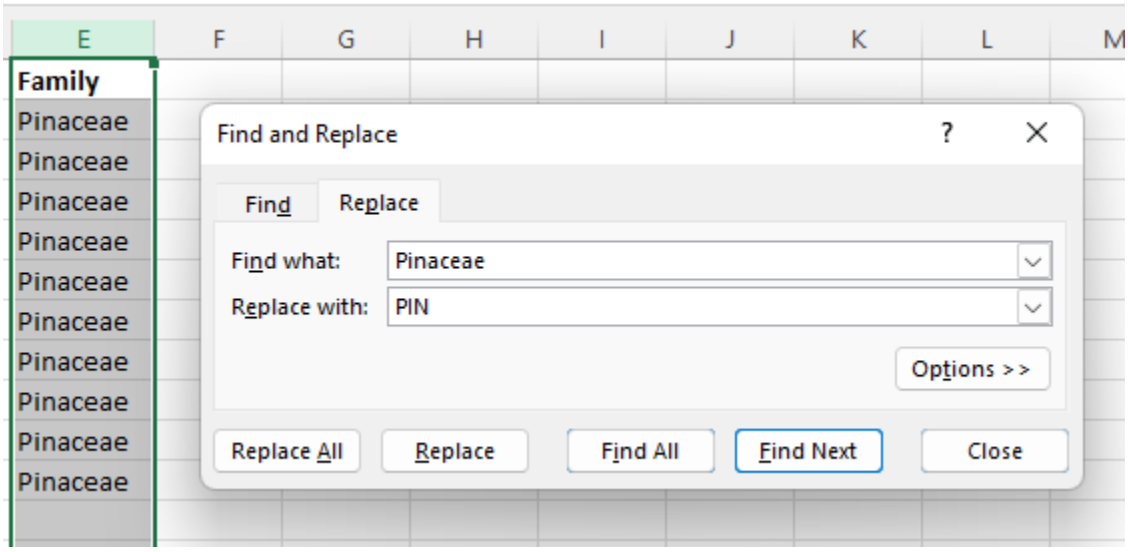
A	B	C	D	E	F
Binomial_Nomenclature	**Ww_Num**	**Collection_Num**	**Country**	**Family**	
Abies balsamea	Ww220367	SUNYESF-8089	New York	Pinaceae	
Abies balsamea	Ww220368	SUNYESF-8252	Minnesota	Pinaceae	
Abies balsamea	Ww220369	SUNYESF-8653	Wisconsin	Pinaceae	
Abies balsamea	Ww220370	SUNYESF-8660	Maine	Pinaceae	
Abies grandis	Ww210062	Gleaves_10_483		Pinaceae	
Abies grandis	Ww220371	SUNYESF-8105	Washington	Pinaceae	
Abies grandis	Ww220372	SUNYESF-8124	Idaho	Pinaceae	
Abies grandis	Ww220373	SUNYESF-8366	Canada	Pinaceae	
Abies procera	Ww220374	SUNYESF-8237	Washington	Pinaceae	
Abies sp.	Ww210033	TimberEngCo_34		Pinaceae	

5. Change the **Family** column to the abbreviated version of the family name in all capitals. For example:
   a. Pinaceae ⟶ PIN
   b. Fabaceae ⟶ FAB
      i. **NOTE: Many families have the same first 3-4 letters, be sure to check what abbreviations have already been used before proceeding.**

This is done by highlighting the **Family** column, then hitting the **Ctrl** key and **F** key. This will bring up the **Find and Replace** box.



Click the **Replace** tab and type the family name in "Find what:" and the abbreviation in "Replace with:", then click "Replace All"



D	E	F
Country	**Family**	
New York	PIN	
Minnesota	PIN	
Wisconsin	PIN	
Maine	PIN	
	PIN	
Washington	PIN	
Idaho	PIN	
Canada	PIN	
Washington	PIN	
	PIN	

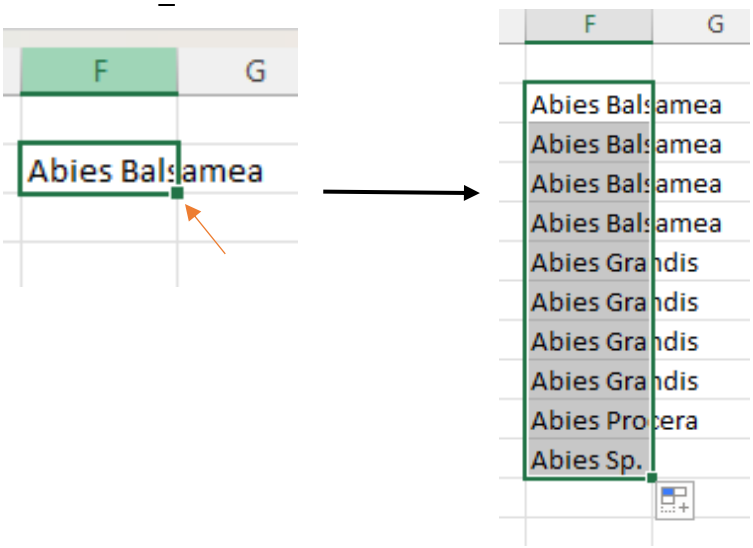
6. Use the same process of "Find" and "Replace" to remove any periods from the names. E.g., "Abies sp." should be "Abies sp"

7. Change the **Binomial_Nomenclature** column to have the genus and species both capitalized: using the adjacent cell in column F, type =PROPER(***then click an empty adjacent column, close the parentheses, and press enter***.)

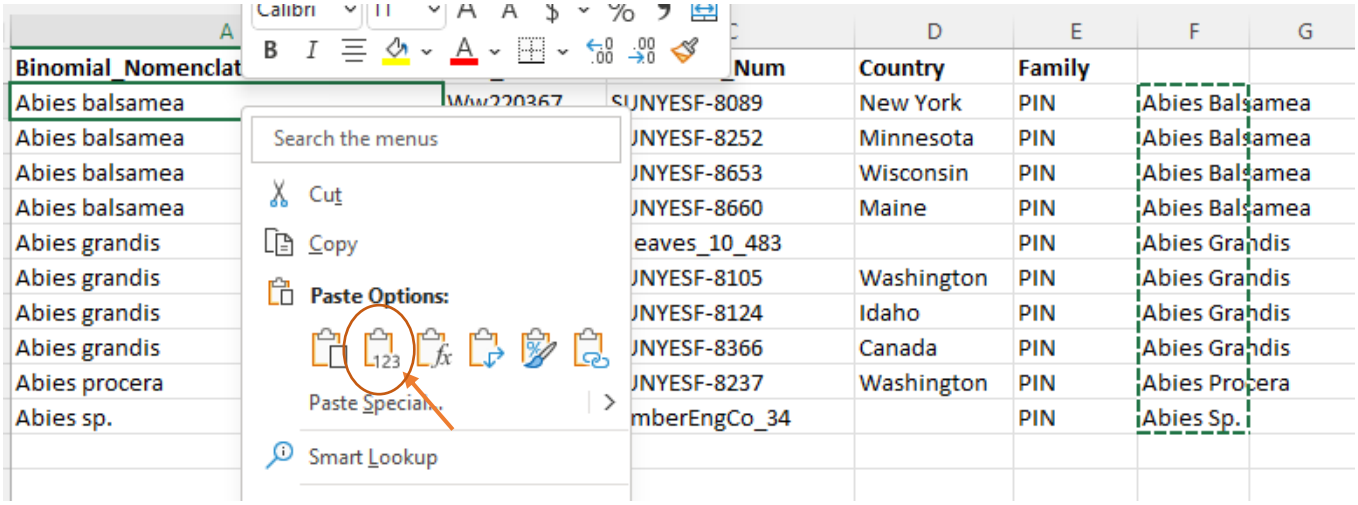
	A	B	C	D	E	F	G
	Binomial_Nomenclature	**Ww_Num**	**Collection_Num**	**Country**	**Family**		
	Abies balsamea	Ww220367	SUNYESF-8089	New York	PIN	=PROPER(A2)	
	Abies balsamea	Ww220368	SUNYESF-8252	Minnesota	PIN		
	Abies balsamea	Ww220369	SUNYESF-8653	Wisconsin	PIN		
	Abies balsamea	Ww220370	SUNYESF-8660	Maine	PIN		
	Abies grandis	Ww210062	Gleaves_10_483		PIN		
	Abies grandis	Ww220371	SUNYESF-8105	Washington	PIN		
	Abies grandis	Ww220372	SUNYESF-8124	Idaho	PIN		
	Abies grandis	Ww220373	SUNYESF-8366	Canada	PIN		
	Abies procera	Ww220374	SUNYESF-8237	Washington	PIN		

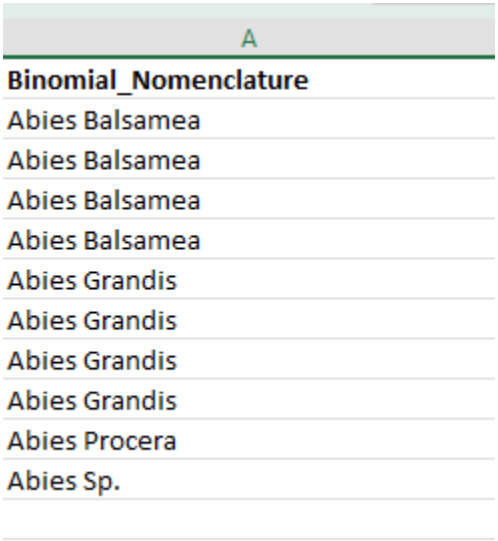
	A	B	C	D	E	F	G
1	**Binomial_Nomenclature**	**Ww_Num**	**Collection_Num**	**Country**	**Family**		
2	Abies balsamea	Ww220367	SUNYESF-8089	New York	PIN	Abies Balsamea	
3	Abies balsamea	Ww220368	SUNYESF-8252	Minnesota	PIN		
4	Abies balsamea	Ww220369	SUNYESF-8653	Wisconsin	PIN		

8. Click and drag the bottom right corner of the highlighted new cell to auto-populate the function for all of the **Binomial_Nomenclature** column.



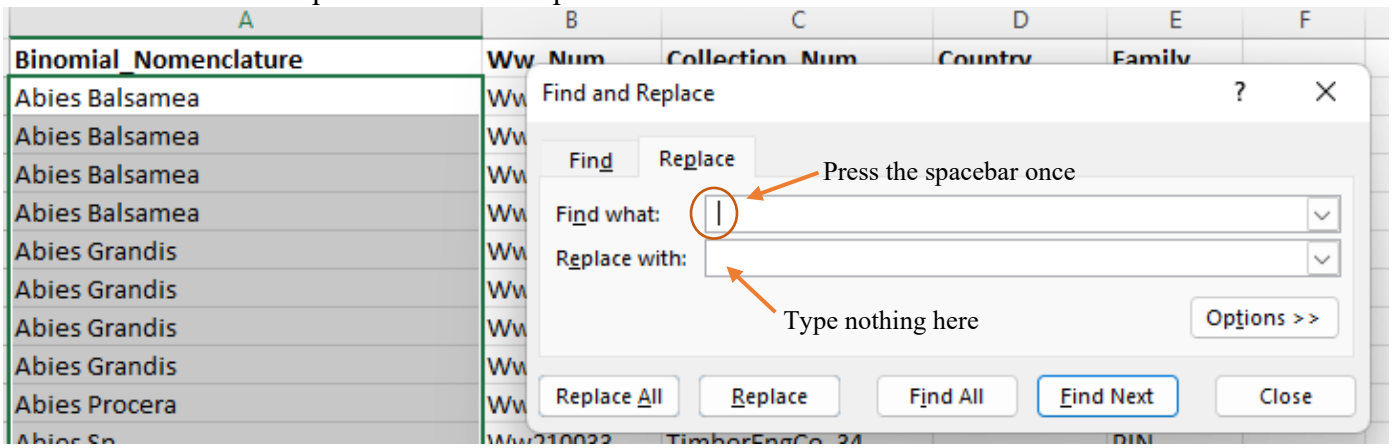
9. Click **Ctrl + c** to copy the highlighted cells and then paste them <u>as values</u> over the original **Binomial_Nomenclature** names. By pasting as values, the function is not copied over, just the names of the species.





The new column used to create the GenusSpecies names can now be deleted.

10. Following the same procedures as step 5, highlight the new, capitalized names in **Binomial_Nomenclature,** then press the space bar once to find a space, " ", and replace with nothing. This will eliminate the space between the species names.

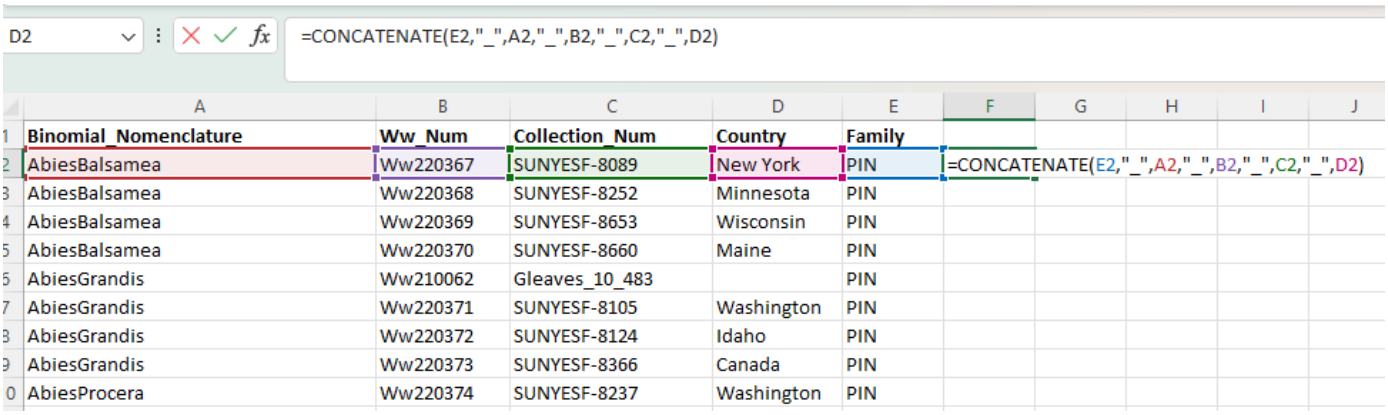


A
Binomial_Nomenclature
AbiesBalsamea
AbiesBalsamea
AbiesBalsamea
AbiesBalsamea
AbiesGrandis
AbiesGrandis
AbiesGrandis
AbiesGrandis
AbiesProcera
AbiesSp.

All the cells are now ready for use as a file name.

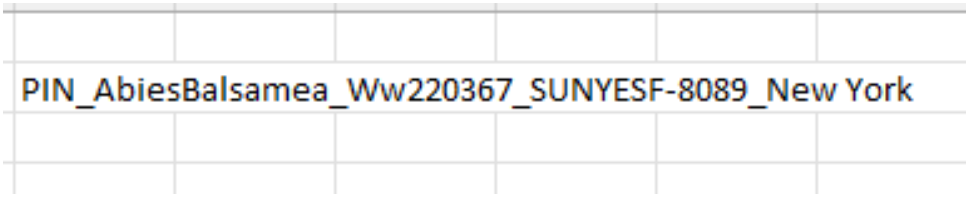
11. In an unused column, type =CONCATENATE(and then follow this format:

=CONCATENATE(**Family**,"_",**Binomial_Nomenclature**,"_",**Ww_Num**,"_",**Collection_Num**,"_",**Country**)



	A	B	C	D	E	F	G	H	I	J
1	**Binomial_Nomenclature**	**Ww_Num**	**Collection_Num**	**Country**	**Family**					
2	AbiesBalsamea	Ww220367	SUNYESF-8089	New York	PIN	=CONCATENATE(E2,"_",A2,"_",B2,"_",C2,"_",D2)				
3	AbiesBalsamea	Ww220368	SUNYESF-8252	Minnesota	PIN					
4	AbiesBalsamea	Ww220369	SUNYESF-8653	Wisconsin	PIN					
5	AbiesBalsamea	Ww220370	SUNYESF-8660	Maine	PIN					
6	AbiesGrandis	Ww210062	Gleaves_10_483		PIN					
7	AbiesGrandis	Ww220371	SUNYESF-8105	Washington	PIN					
8	AbiesGrandis	Ww220372	SUNYESF-8124	Idaho	PIN					
9	AbiesGrandis	Ww220373	SUNYESF-8366	Canada	PIN					
10	AbiesProcera	Ww220374	SUNYESF-8237	Washington	PIN					

The result will look like:



Click and drag the bottom right corner of this new cell the same as in step 7 to copy this function to create all the file names.

PIN_AbiesBalsamea_Ww220367_SUNYESF-8089_New York
PIN_AbiesBalsamea_Ww220368_SUNYESF-8252_Minnesota
PIN_AbiesBalsamea_Ww220369_SUNYESF-8653_Wisconsin
PIN_AbiesBalsamea_Ww220370_SUNYESF-8660_Maine
PIN_AbiesGrandis_Ww210062_Gleaves_10_483_
PIN_AbiesGrandis_Ww220371_SUNYESF-8105_Washington
PIN_AbiesGrandis_Ww220372_SUNYESF-8124_Idaho
PIN_AbiesGrandis_Ww220373_SUNYESF-8366_Canada
PIN_AbiesProcera_Ww220374_SUNYESF-8237_Washington
PIN_AbiesSp._Ww210033_TimberEngCo_34_

12. Use **Shift** to highlight these new file names, press **Ctrl + C** to copy them, and **Paste as Values** (same as step 8) to replace the first column on the sheet.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Binomial_Nomenclature	**Ww_Num**	**Collection_Num**	**Country**	**Family**										
PIN_AbiesBalsamea_Ww220367_SU	Ww220367	SUNYESF-8089	New York	PIN	iPIN_PIN_AbiesBalsamea_Ww220367_SUNYESF-8089_New York_Ww220367_SUNYESF-8089_New York									
PIN_AbiesBalsamea_Ww220368_SU	Ww220368	SUNYESF-8252	Minnesota	PIN	iPIN_PIN_AbiesBalsamea_Ww220368_SUNYESF-8252_Minnesota_Ww220368_SUNYESF-8252_Minnesota									
PIN_AbiesBalsamea_Ww220369_SU	Ww220369	SUNYESF-8653	Wisconsin	PIN	iPIN_PIN_AbiesBalsamea_Ww220369_SUNYESF-8653_Wisconsin_Ww220369_SUNYESF-8653_Wisconsin									
PIN_AbiesBalsamea_Ww220370_SU	Ww220370	SUNYESF-8660	Maine	PIN	iPIN_PIN_AbiesBalsamea_Ww220370_SUNYESF-8660_Maine_Ww220370_SUNYESF-8660_Maine									
PIN_AbiesGrandis_Ww210062_Glea	Ww210062	Gleaves_10_483		PIN	iPIN_PIN_AbiesGrandis_Ww210062_Gleaves_10_483__Ww210062_Gleaves_10_483_									
PIN_AbiesGrandis_Ww220371_SUN	Ww220371	SUNYESF-8105	Washington	PIN	iPIN_PIN_AbiesGrandis_Ww220371_SUNYESF-8105_Washington_Ww220371_SUNYESF-8105_Washington									
PIN_AbiesGrandis_Ww220372_SUN	Ww220372	SUNYESF-8124	Idaho	PIN	iPIN_PIN_AbiesGrandis_Ww220372_SUNYESF-8124_Idaho_Ww220372_SUNYESF-8124_Idaho									
PIN_AbiesGrandis_Ww220373_SUN	Ww220373	SUNYESF-8366	Canada	PIN	iPIN_PIN_AbiesGrandis_Ww220373_SUNYESF-8366_Canada_Ww220373_SUNYESF-8366_Canada									
PIN_AbiesProcera_Ww220374_SUN	Ww220374	SUNYESF-8237	Washington	PIN	iPIN_PIN_AbiesProcera_Ww220374_SUNYESF-8237_Washington_Ww220374_SUNYESF-8237_Washington									
PIN_AbiesSp._Ww210033_TimberEr	Ww210033	TimberEngCo_34		PIN	iPIN_PIN_AbiesSp._Ww210033_TimberEngCo_34__Ww210033_TimberEngCo_34_									

All other data can be deleted except for the newly created file names.

13. If there are any file names that did not have a country listed in the database, there will be an extra "_" at the end of the file name created. These can be deleted to leave the last piece of the file name as the **Collection_Num**.

File Names
PIN_AbiesBalsamea_Ww220367_SUNYESF-8089_New York
PIN_AbiesBalsamea_Ww220368_SUNYESF-8252_Minnesota
PIN_AbiesBalsamea_Ww220369_SUNYESF-8653_Wisconsin
PIN_AbiesBalsamea_Ww220370_SUNYESF-8660_Maine
PIN_AbiesGrandis_Ww210062_Gleaves_10_483_
PIN_AbiesGrandis_Ww220371_SUNYESF-8105_Washington
PIN_AbiesGrandis_Ww220372_SUNYESF-8124_Idaho
PIN_AbiesGrandis_Ww220373_SUNYESF-8366_Canada
PIN_AbiesProcera_Ww220374_SUNYESF-8237_Washington
PIN_AbiesSp._Ww210033_TimberEngCo_34_

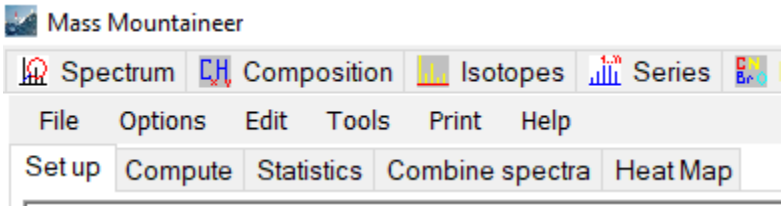
The final file names should look like this:

	File Names
2	PIN_AbiesBalsamea_Ww220367_SUNYESF-8089_New York
3	PIN_AbiesBalsamea_Ww220368_SUNYESF-8252_Minnesota
4	PIN_AbiesBalsamea_Ww220369_SUNYESF-8653_Wisconsin
5	PIN_AbiesBalsamea_Ww220370_SUNYESF-8660_Maine
6	PIN_AbiesGrandis_Ww210062_Gleaves_10_483
7	PIN_AbiesGrandis_Ww220371_SUNYESF-8105_Washington
8	PIN_AbiesGrandis_Ww220372_SUNYESF-8124_Idaho
9	PIN_AbiesGrandis_Ww220373_SUNYESF-8366_Canada
10	PIN_AbiesProcera_Ww220374_SUNYESF-8237_Washington
11	PIN_AbiesSp._Ww210033_TimberEngCo_34

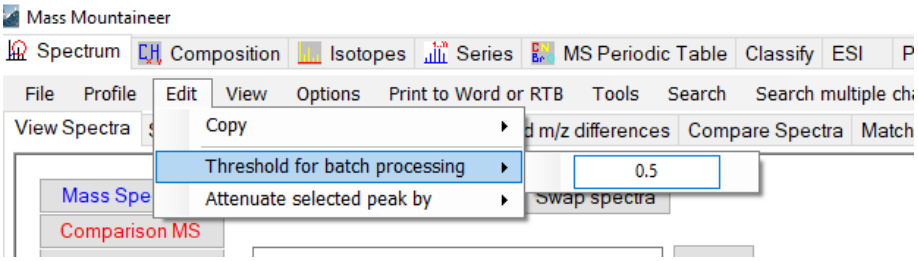
Creating NIST Libraries

Have your spectra folder ready for indexing in a single folder

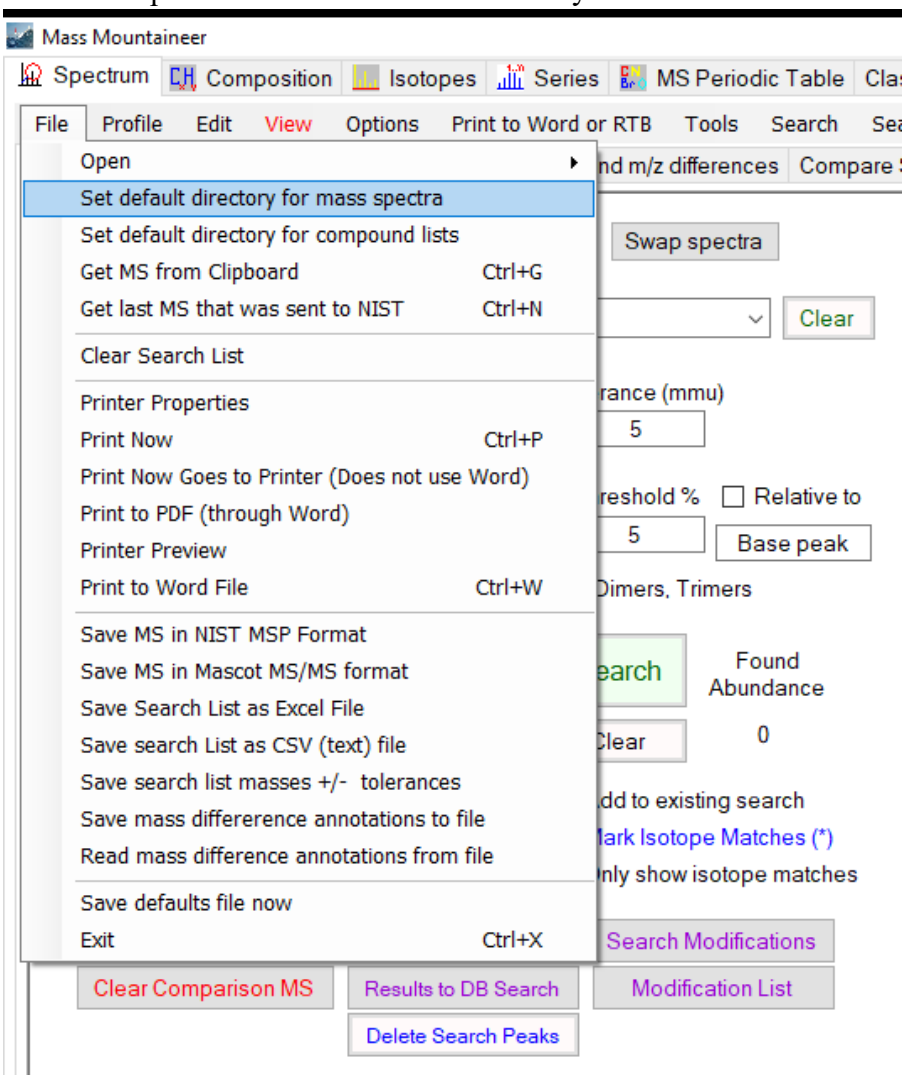
1. Open Mass Mountaineer and Select the Spectrum tab



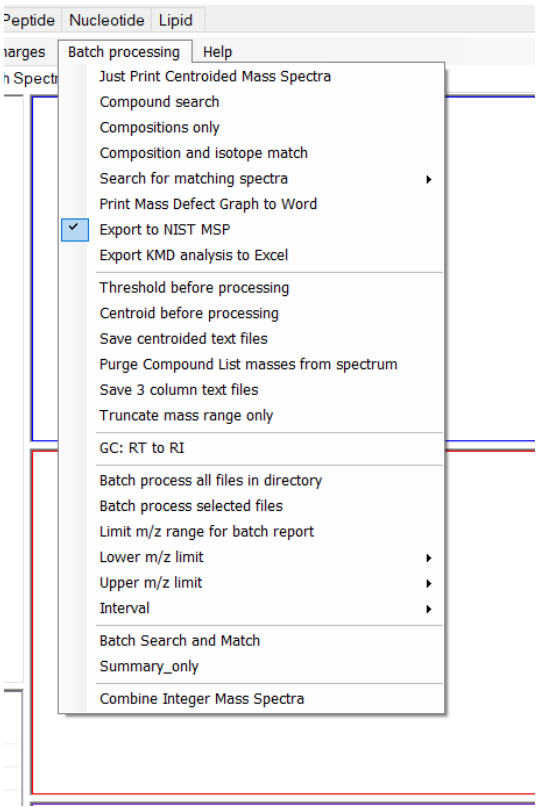
   a. **NOTE**: The default threshold for this process is 0.5, if this requires adjusting the threshold can be changed by navigating to the Edit button in the Spectrum tab.



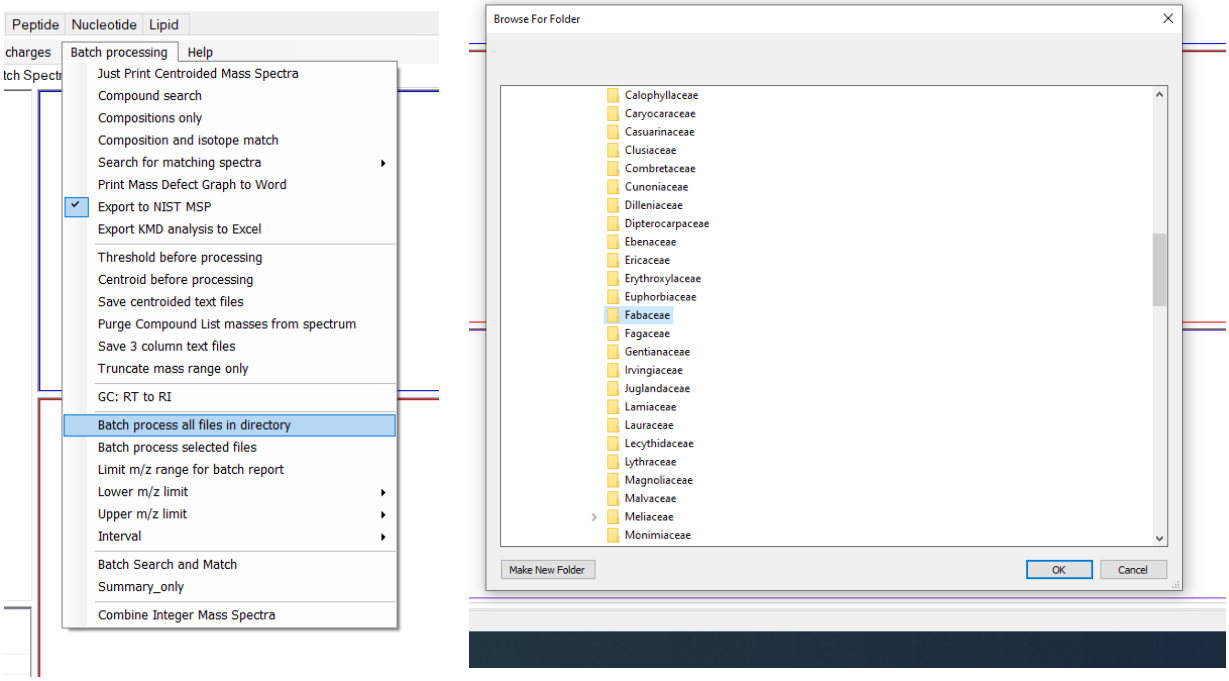
2. Set the path to the folder of spectra that will be used to create your NIST searchable library.



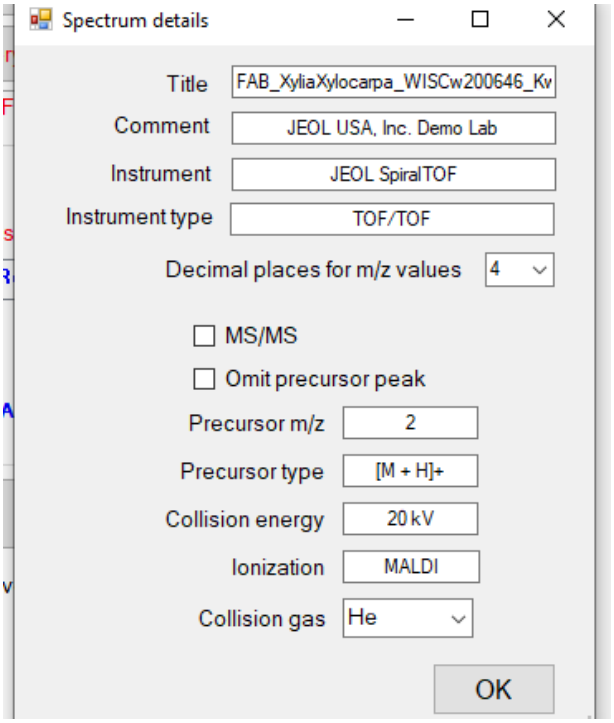
3. Click Batch Processing and select Export to NIST MSP. Deselect any other selected options.

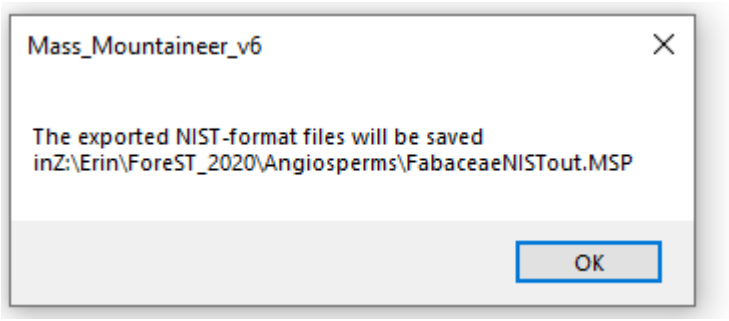


4. Select Batch process all files in directory



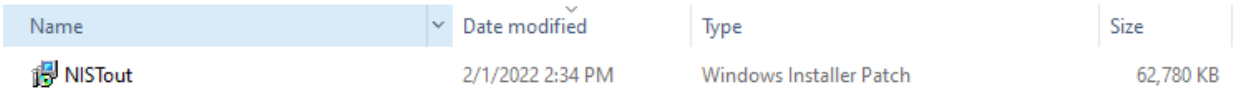
5. Change Decimal places for m/z values to 4. Unselect the MS/MS box. All other settings should match the image below. Click OK and check that your exported files will be saved in the correct path.



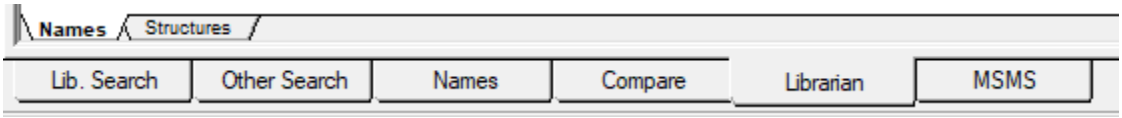


6. Progress can be monitored in the bottom left of the Mass Mountaineer screen. Large datasets will take longer.

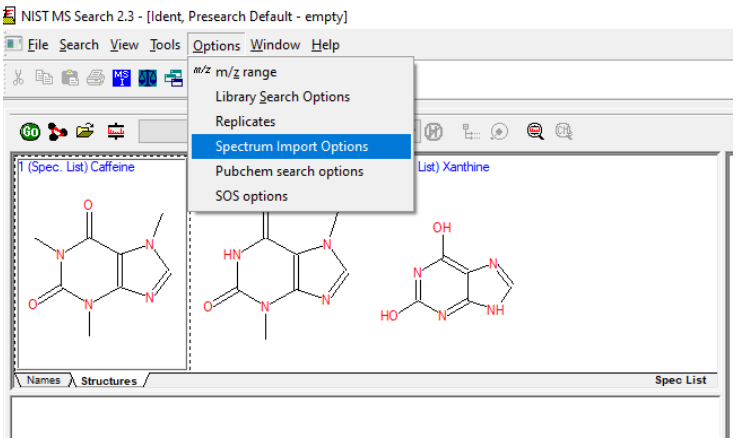
7. Check the spectra folder for the following file:



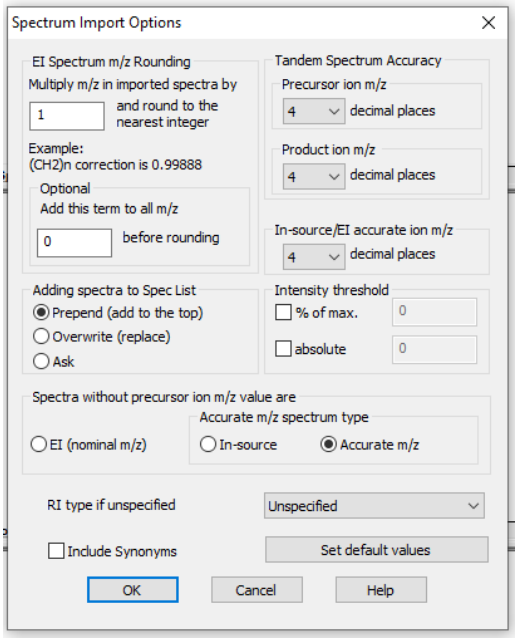
8. Open MS Search  and select the Librarian tab on the bottom left of the screen



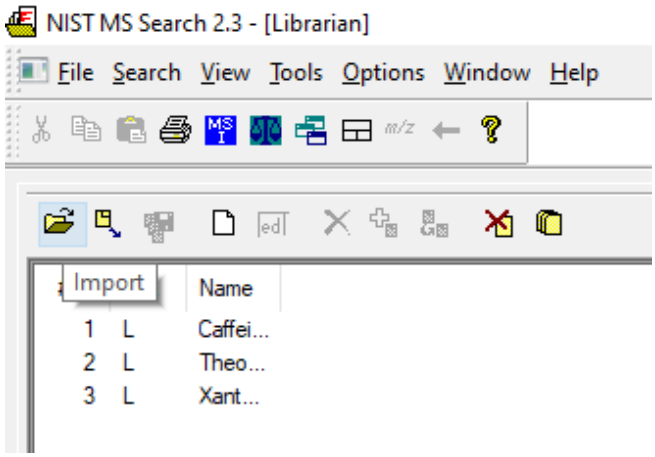
9. Click Options at top of screen and select Spectrum Import Options



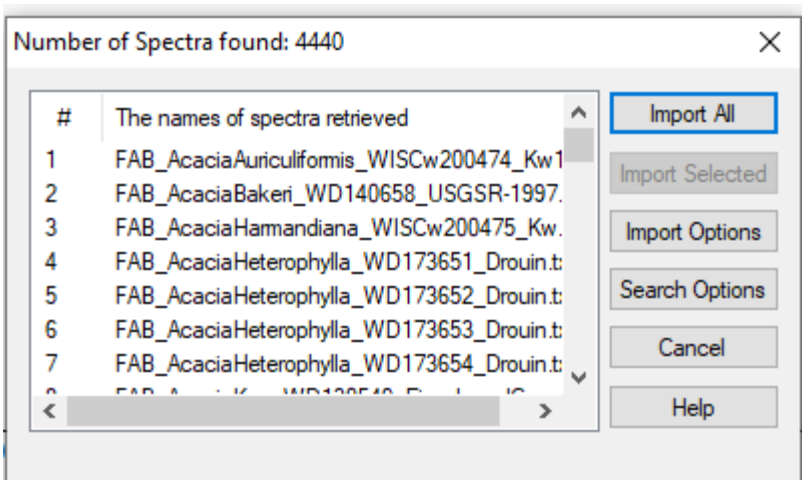
10. Select Accurate m/z. All other settings should be as follows:
    a. **NOTE**: Do not use the nominal m/z setting.

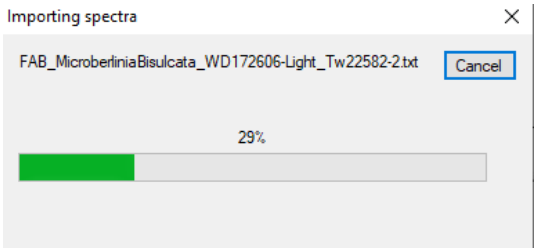


11. Select the Import symbol at the top left of the screen

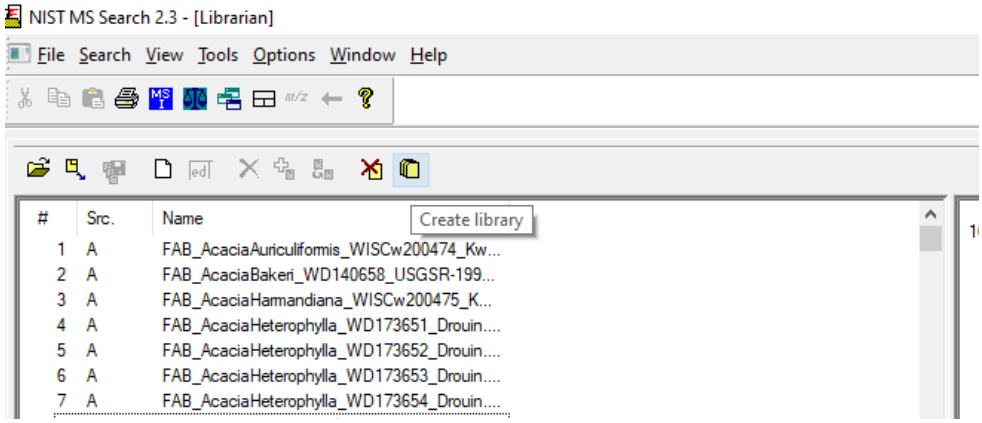


12. Find the NISTout file created in step 6. Click Import All. Progress can be monitored. Cancel the Background search, this is automatic but not needed and will add time to the process.

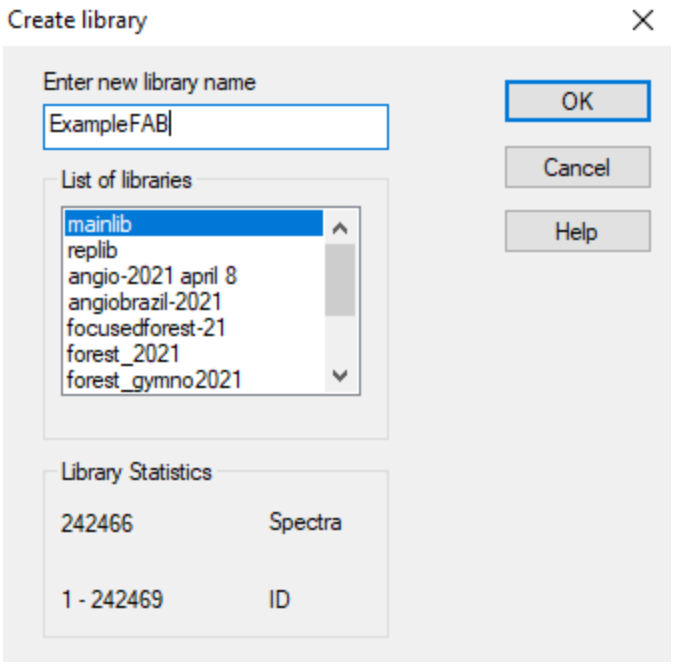




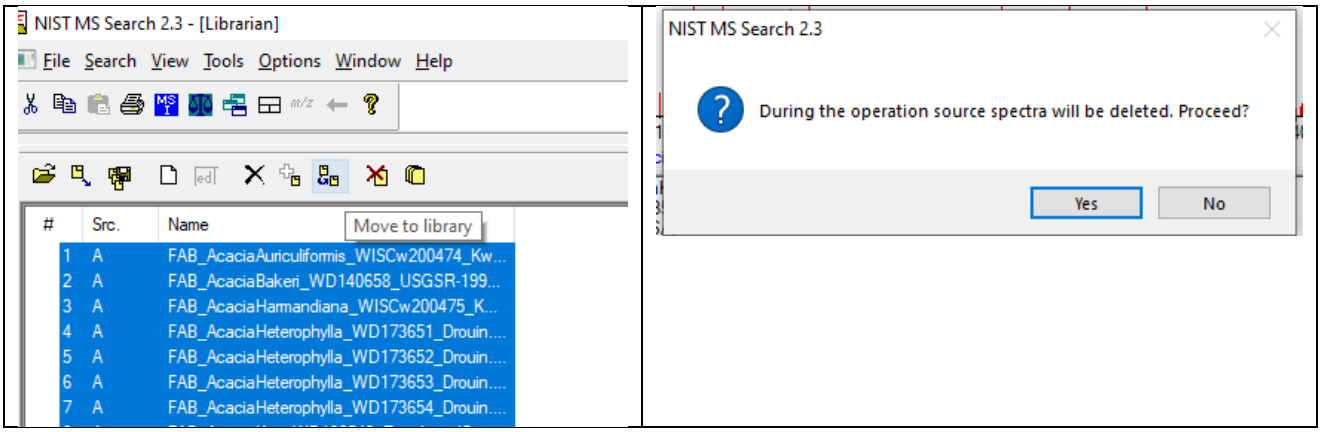
13. Now that spectra files are imported, select Create Library



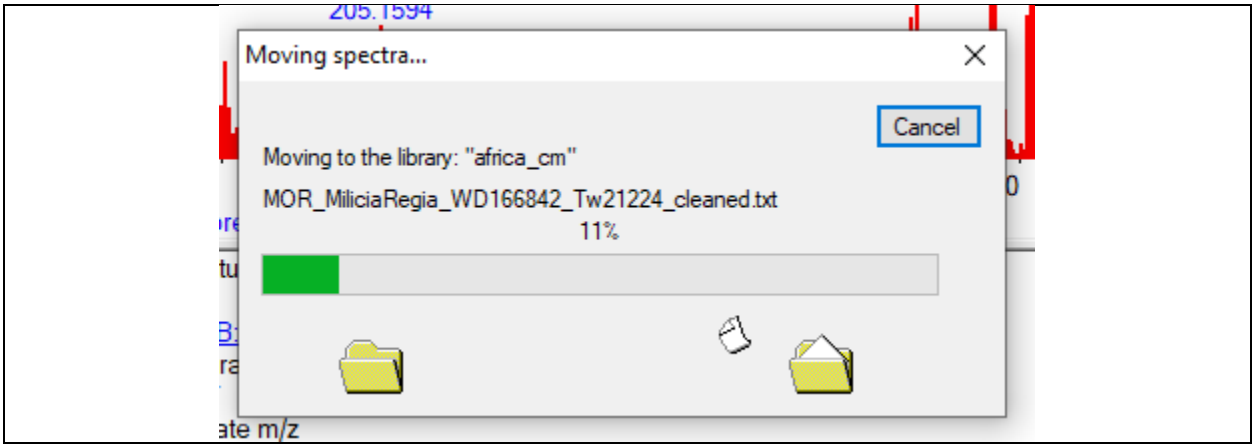
14. Enter a unique and descriptive library name, but keep the name short and simple. Click OK



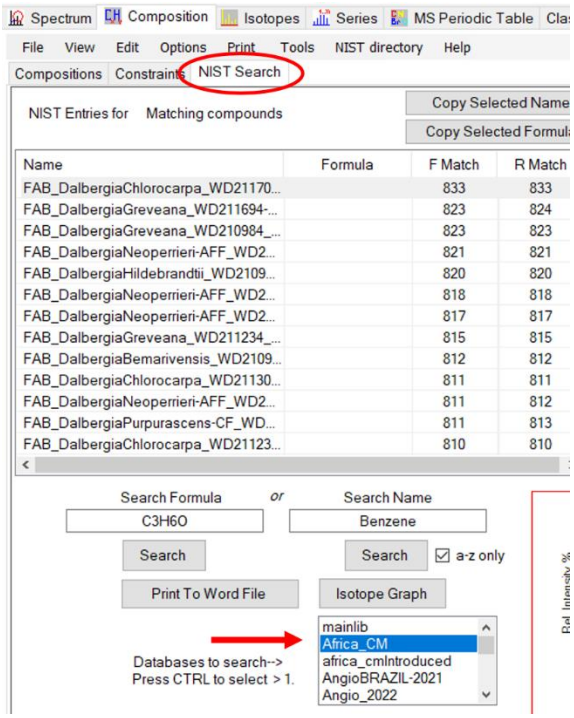
15. Click on any spectrum file in the left pane (Names) and then hold Ctrl + a to select all. Click Move to Library icon. Click OK.



16. Select the unique library name you just created and click OK. The library will begin populating with the spectra.



17. After this final step, the new library is ready for use. Libraries can be changed by navigating to the Composition tab, selecting NIST Search, and selecting the target library from the list.

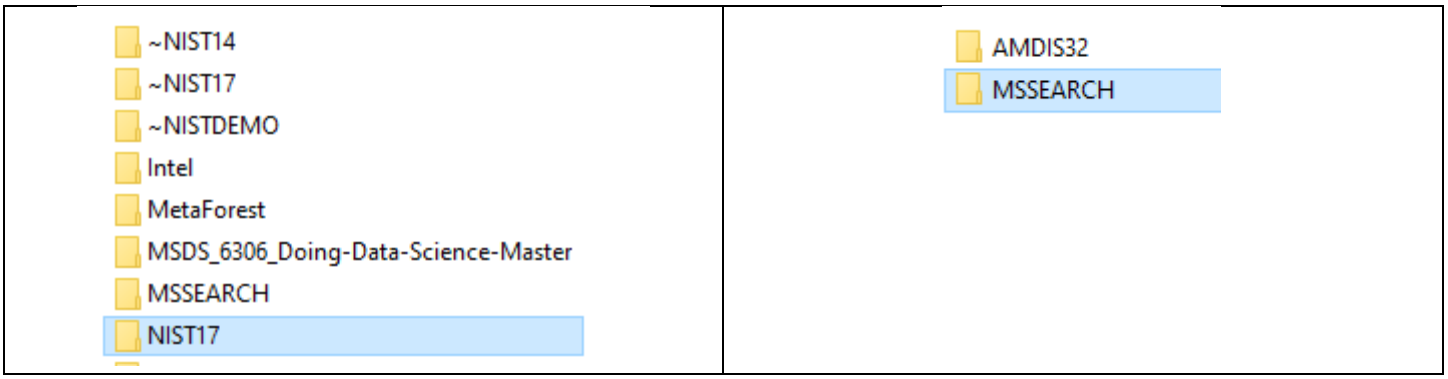


Accurate libraries are used with the “Identity Search” option

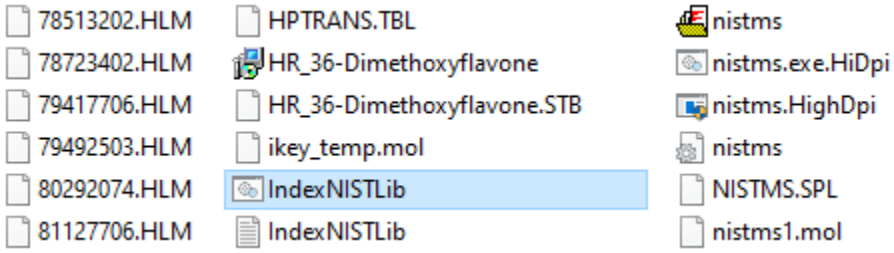
Creating High Resolution Indexed Libraries

****This library type is used with the "In-Source HiRes with PRESEARCH" option in Mass Mountaineer****

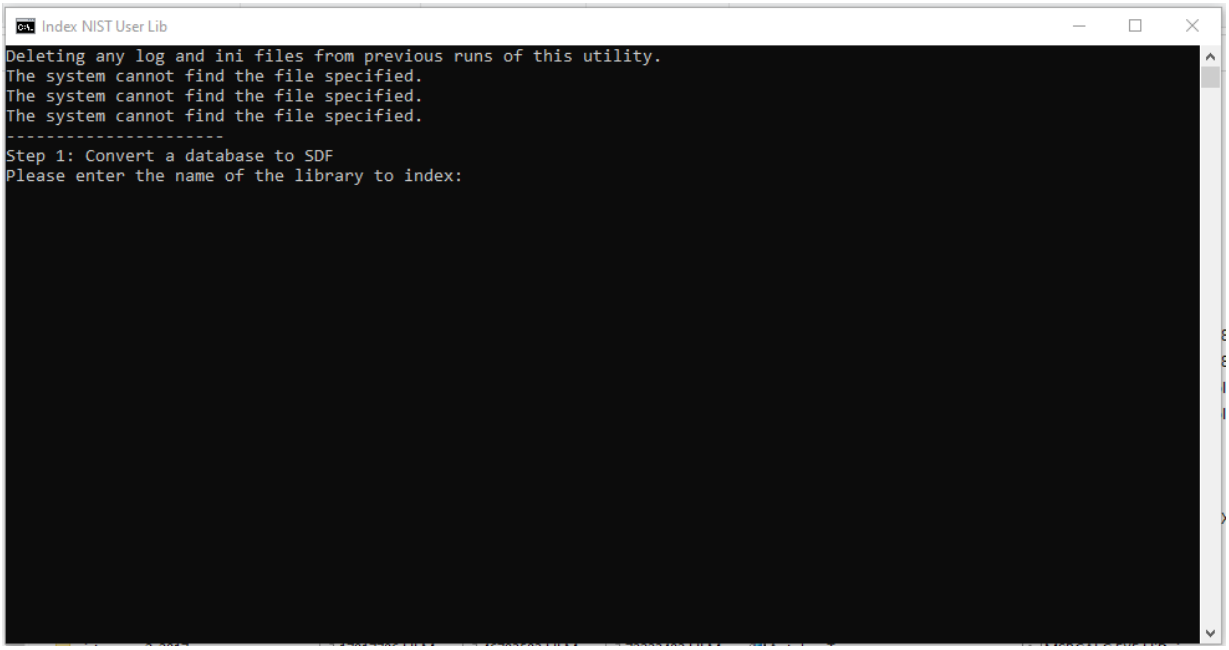
1. Using steps 1-15 listed above, navigate to the NIST folder within the C: drive and open the MSSEARCH folder



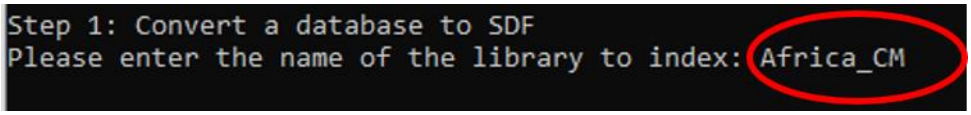
2. Locate the IndexNISTLib.bat file and double click to open it



3. The Index NIST User Lib command screen will open



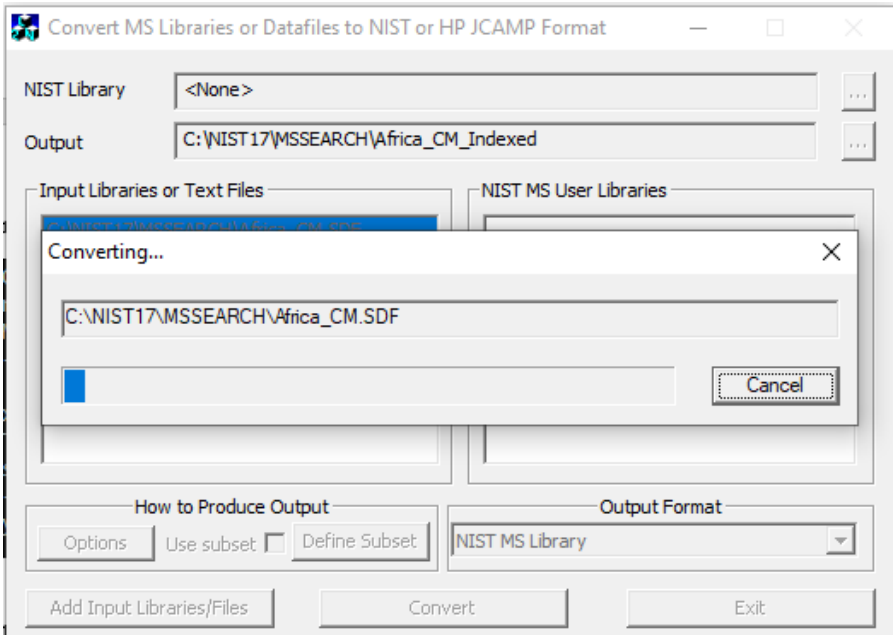
4. Type the *exact* name of the library that is to be converted into an indexed library and hit Enter



5. Press Enter (or any key) again once the prompt “Press any key to continue…” appears, a screen will flash up at this point and immediately close.

```
Step 1: Convert a database to SDF
Please enter the name of the library to index: Africa_CM
Africa_CM was found. Continuing with the indexing procedure.
----------------------
Lib2NIST will start when you press ENTER.
The Indexing will occur in two steps.
IMPORTANT: ALLOW THE STEP 1 CONVERSION TO FINISH BEFORE CONTINUING TO STEP 2.
Then press ENTER to continue.
----------------------
Press any key to continue . . .
```

6. Press Enter (or any key) a third time and the following window will appear, do not click any buttons within the window



7. Once the library is complete the window will disappear. Check the MSSEARCH folder for the new library, it will have the exact name as the original library with _Indexed on the end.