

PHARMACOGNOSY AND PHYTOCHEMISTRY

Dr. Galvina Pereira

Department of Pharmaceutical Sciences and Technology

Institute of Chemical Technology, Mumbai

Week 12

Lecture 57

Week 12: Lecture 57: Natural Product Databases for Phytochemical Search

Thank you. Hello everyone, and welcome back to our Week 12 Session 2 on the application of natural product databases. Now, in the previous session, we learned that we could use some natural product databases to help us gain more information about certain plants. So, if you remember, we saw the Plants of the World Online, JSTOR, Tropicos, as well as the New York Botanical Garden Herbariums. There, we raised a query about certain plants.

And then, what we saw is that whenever you raise a query, the software comes back with herbarium records. They come back with taxonomic information. They come back with geographical information, and some also provide therapeutic applications. Additionally, we saw in the previous session that certain servers help us with traditional medicinal drugs. Now, there are numerous such servers and software available globally.

As I mentioned, some are paid, and some are free, but we explored a few free tools particularly relevant to the Indian traditional system of medicine. This included our TKDL, the Traditional Knowledge Digital Library. It also included IMPPAT, the Indian Medicinal Plants Phytochemistry and Therapeutics Database, as well as the Indian Medicinal Plants Database. Now, among these three, we saw that they not only provided details about the plants and their geography but also shared traditional knowledge, usage locations, systems, methods, formulations, and even the shlokas written about them. In today's session, we will take a step further and explore software that assists with phytochemical searches.

Now, many times it so happens that I have finalized my plant. Now, once I finalize my plant, How do I know what phytoconstituents it contains? Now, many times people resort to traditional techniques. Now, traditional techniques—we also saw some of the quality control techniques.

So you will pick up that plant, say, for example, Ashwagandha, and start analyzing. Does it contain steroids? Does it contain flavonoids? Does it contain triterpenoids and so on? But nowadays, you need not go there.

The reason is, there are certain software servers or web platforms which are created to help you already collect the phytochemical information that is there on the global database. So these servers will collect that information and give you a curated list of phytoconstituents present in a plant. So you already know that if it is Ashwagandha, it has to have this many substances. If it is ginger, it has to have this many substances. Instead of going paper by paper, textbook by textbook, this particular software that we discuss in this session—or these particular web-based servers that we discuss in the session—are going to help us find the phytochemicals and also, to a certain extent, tell us about their properties. So, what are these Softwares? We'll see a few, which includes Dr. Duke's Phytochemical and Ethno botanical Database.

The slide features a green header with the text 'Natural Product Database'. Below the header, on the left, is a green rounded rectangle containing the text 'Phytochemical Search' with a right-pointing arrow. To the right of the arrow is a light blue box containing a list of databases:

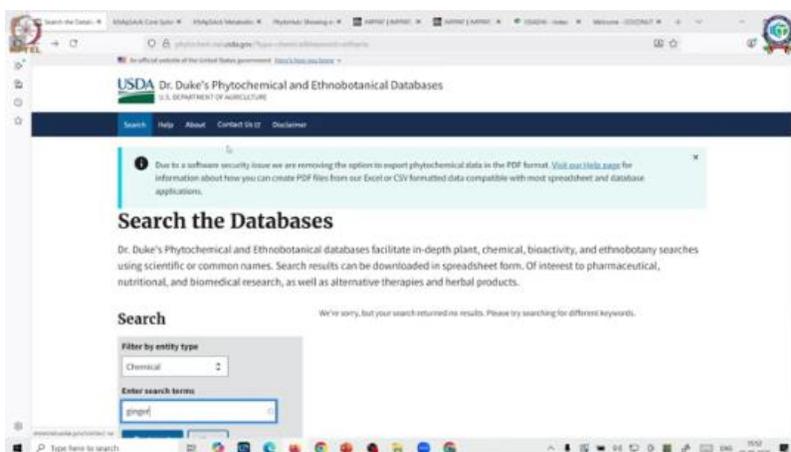
- Dr. Duke's Phytochemical and Ethnobotanical Databases
• <https://phytochem.nsl.usda.gov/>
- KNApSACK: A Comprehensive Species-Metabolite Relationship Database
• <https://www.knapsackfamily.com/KNApSACK/>
- PhytoHub
• <https://phytohub.eu/>
- COCONUT: the COllection of Open NatUral products
• <https://coconut.naturalproducts.net/>
- IHPPAT: Indian Medicinal Plants, Phytochemistry And Therapeutics
• <https://cb.ims.res.in/ihppat/>
- OSADHI - Online Structural and Analytics-based Database for Herbs of India
• <https://neel.ims.res.in/osadhi/>

At the bottom of the slide, it reads: 'Dr. Gayatri Perera, Institute of Chemical Technology, Mumbai'. A small inset image of a woman is visible in the bottom right corner of the slide.

Knapsack is a database of comprehensive species-metabolite relationships. Then you have Phytoherb, which pertains more to food. But many of the plants we are already using are also used in food. So that will also help us find some nutrients. Important plants in that region.

Then you have the Coconut Database. It has nothing to do with coconut. It's just an abbreviated form for the Collection of Open Natural Products. Then again, you have IMPPAT. We saw last time that it also provides details about phytochemicals, plants, and their therapeutic actions.

In addition, the NIST Institute, located in the Northeast, has curated the OSADHI database, which is an online structural analytics-based database, especially for herbs in India. So let's try our hands on a few of the servers and see how each responds with a list of phytoconstituents when we input a plant. Let's take a look at how to access them. The first database we see today is developed by USDA: Dr. Duke's Phytochemical and Ethnobotanical Database. When you click on the link to this database, you will land on a search page.



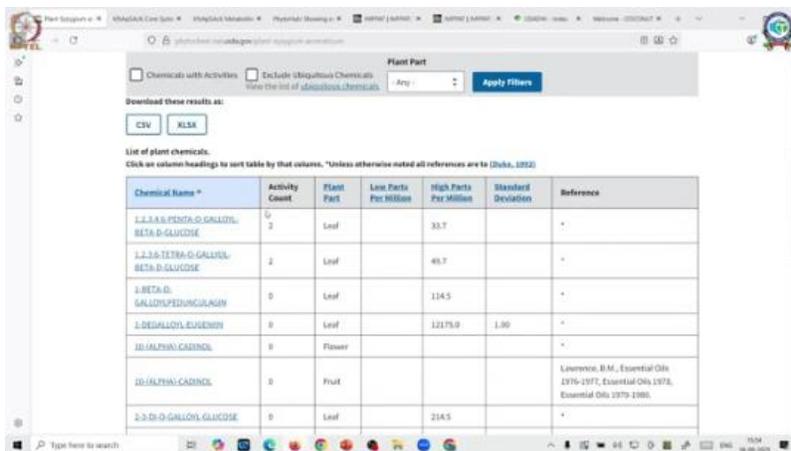
Now this search page you can filter whatever you require. Are you searching for phytochemicals, biological activity, plant or A syndrome or ethno botanical especially you know traditional medicine based activity or a ethno botanical plant. Now just for our convenience let us keep this at any and then let us put any of our query. Say for example I will put a simple case of clover.

Let us put clove for example and then I search. So you will see that when I put clove as a query, now you can see there are certain boxes and in that boxes there are alphabets. So A box indicates biological activity, C indicates the chemical entity, E whether it's a ethno botanical plant or plant depth syndrome or ethno botany. So majority of the queries have returned us with the list of constituents. So you can see clove contains

And you can see here about four compounds that are reported through this database. Then there are plants which are called as clove. So you can see there are so many plants which are called as clove. And then these are kind of ethnobotanical plants. And if you see here, just a plant which is called as.

So these are the traditional ones. These are the modern ones. So I'll just click on *Syzygium aromaticum*. And then when I click on *Syzygium aromaticum*, I am specifically going scientifically to the clove that we want. See, when I put clove, it could be a clove of anything.

And that's why the database was very nonspecific in finding it. Now, when I specifically put *Syzygium aromaticum*, you can see what happens here. So now it has come up with a list of a few compounds. And they are numerous. So you can see the number of pages.



Chemical Name +	Activity Count	Plant Part	Low Parts Per Million	High Parts Per Million	Standard Deviation	Reference
1,2,3,6-PSYTA-D-GALLUOL-BETA-D-GALUCOSE	2	Leaf		33.7		+
1,2,3,6-TETRA-O-GALLUOL-BETA-D-GALUCOSE	2	Leaf		49.7		+
1-BETA-D-GALLUOLPENTACULANIN	0	Leaf		114.5		+
1-DESHALLOYL-EUGENOL	0	Leaf		12175.0	1.00	+
10-ALPHA-CADENOL	0	Flower				+
10-ALPHA-CADENOL	0	Fruit				Lorenson, B.M., Essential Oils 1976-1977, Essential Oils 1978, Essential Oils 1979-1980.
2,3-Di-O-GALLOYL GLUCOSE	0	Leaf		214.5		+

Now, what is this activity count? In literature, how many activities have been reported for this particular compound? Which plant part it occurs in and the reference possibly for the report of this compound. Now you will see that some of them are still Starred, which means there is no clear mention of that. In some cases, you might see the concentrations also being reported.

Many places are where the server is not able to find it. In that case, it is not reported. You can see eugenol and acetyl eugenol. So many of them are terpenoids if you look closely. So I'm just making you run through what all compounds it can contain and see if you

were to manually search this many compounds, how much time you would invest in collecting the list.

Now, one thing about this database is that it also gives you the minerals, the amino acids, and the primary sugars which are present. So you might come across things like aluminum ascorbic acid or certain types of amino acids or proteins which have been characterized from this plant. So just imagine, all your literature search pertaining to phytochemicals present in clove is done within seconds for you. It's an enormous, enormous list. I can download this result as an Excel sheet and then just paste it for my report or any further studies. So, for example, it has come out with around thousands of these.

So these thousands can be nicely downloaded as an Excel sheet and then copied and pasted wherever you want. As you can see here. So the whole of that page that we were seeing. Many of them are repeats also, with spelling mistakes because of which they might be there. And some of them are different plant parts, because of which they are repeated.

So, about say zonarin, you can see the repeat is because of the different flower parts. One is in flower, one is in fruit. So those also, so you come across almost 352 entries here. So I will not save this. Going back.

Similarly, I can search for any other plants and I can even put in a specific query. So, of that, what compounds were in terms of a particular activity, I can do a search there. Now we move on to the next database. The next database is the Knapsack database. Now when you enter the Knapsack database, you come across this as a landing page.



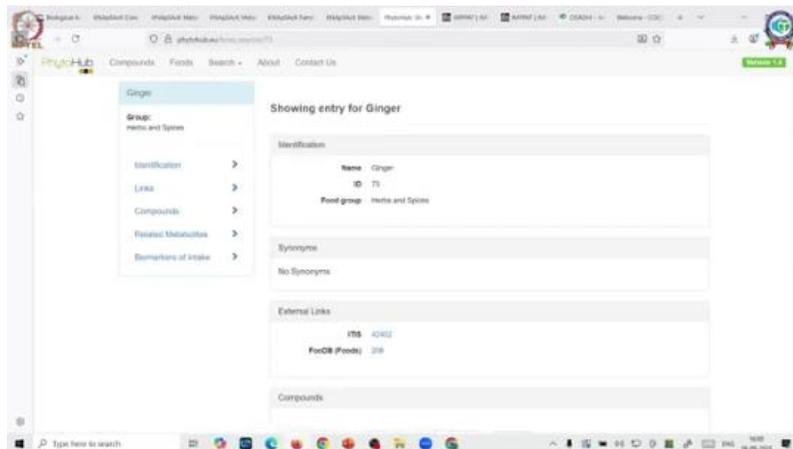
Knapsack actually sources information from many servers. Here are a few of the servers from where it actually picks up the information. Whenever you want to search, you can actually download a version or just do an online search. So we'll just do an online search. We are not looking for anything in particular.



We'll just put a plant and again let's put a clove. So you can see versus your Dr. Duke's database. Here we are looking for a word which is clove. Let me put the query as ginger. And then say list.

So you can see in terms of ginger, it comes back with compounds, molecular formula, CAS ID. But compare that with your Dr. Duke's database. This doesn't give you any information about the plant part. So, this is your knapsack and knapsack specifically gives you information about the phyto constituents in terms of their list and molecular weight. Comparatively, if you see as compared to Dr. Duke's, Dr. Duke's developed by USDA gives you more elaborated list of phytochemicals.

Moving on to the next software, the next software is PhytoHerb. Now, PhytoHerb, if you go to see is a software which will give you a different perspective now this is more pertaining to food so when you click on phytoherb will give you this landing page and phytoherb is a software pertaining to phytochemicals but predominantly occurring in food sources so you can search them by compound search can just see any of the compound and click it so it will give you the properties like it's a terpenoidal family but here like i said more predominantly it gives you a elaborate thing about the food so here you can see in terms of food i can just click on almond say for example And it will come across few compounds that are reported in almond with structure and you can have publications and other details of that.

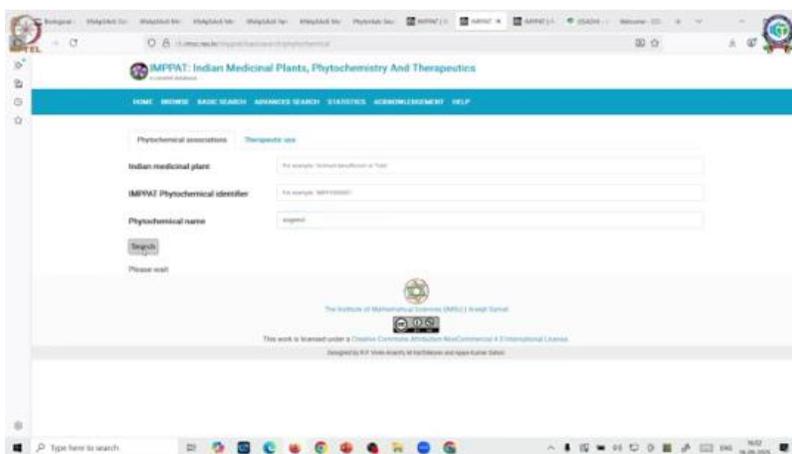


You can also have a set of related metabolites. And if at all, biomarkers of intake. You can also go to the search page and do a basic search there. Let's say, for example, a

formula search. Let me just put, say for example, a glucose formula or something like a carbohydrate formula.

So if I know my compound, I know its molecular formula, whatever compounds are present in this database with the same molecular formula will appear there. So I can do my structure search, LCMS, GCMS search, and so on. Now, the next software is the one we have already seen. You remember the phytochemical search. So in this case, I can just put my phytochemical name.

Or I can put my plant name. So let's see a phytochemical name. And you can see this phytochemical. Eugenol is present in so many plants; there are almost 1,300 entries. So if I have to say quercetin, if I have to say rutin, if I have to say eugenol, and I have to find out which plants contain them, IMPPAT is a good software that will give me this kind of data. Not only does it give me eugenol, but it also gives me the derivatives of eugenol.



So, it can be a name or it can be part of the name also. So, this is where it helps me, and I can get a huge number of entries and come up with the number of plants that will contain eugenol. Now, I can also do a plant search. I'll erase this query. I am just putting 'osimum'.

So, you can see osimum is your tulsi, but osimum has numerous compounds. So, it will give you the list of all compounds present in all osimum species. And that's why the search queries have gone as high as page number 374. I can be more specific. Let's say, for example, I'll put this one: africanum.

So, this time it will end up with only those phyto-constituents present in *osimum africanum*. So, if you see the other osimum, sanctum or tenuiflorum, again you have as high as 60 searches. So, depending on which plant you are searching and how many compounds this database has, it will get you everything here. Now, one more database we can see for phytoconstituents is the OSADHI database, which is developed by NIST. OK, so this gives you a whole, what is called, pan-India database of medicinal plants, which includes traditional knowledge, geographical classification, as well as phytochemical information.

So when you are searching on this, we can do a search on this. See, for example, traditional knowledge. And I just click on one. So it will tell you what its scientific name is, where it is used traditionally. See, for example, dysentery, where it is available geographically and how it is used.

The extract is taken orally. Just going back there. Let's say, for example, if you remember your tannins, acacia catechu. So acacia catechu, again the taxonomy, where in India it is available. What phytochemicals does it contain?

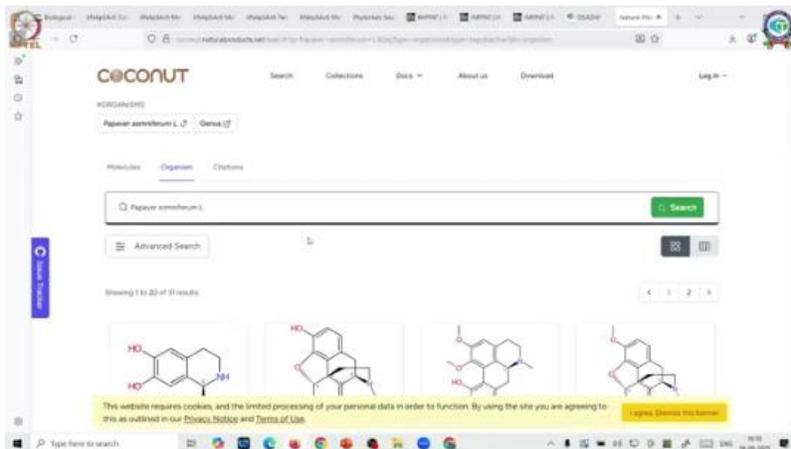
Elements cured by acacia catechu, the plant parts as well as the vernacular names. So I can use this database to find out details about any traditional plant here with respect to most of its aspects. You can just click on aconite. So you can see here Jammu and Kashmir, Madhya Pradesh. So that's where your aconitum is found.

You can have the stem or the roots taken orally and these are all the therapeutic implications of this used in the traditional medicine. Now moving on to the next tab, you can go by geographical classification. Say for example, I want to know what all plants grow in Meghalaya region. So you can see the list of plants which have been found in Meghalaya is updated. So if I want to see geographically also what grows here.

I can just click on that particular plant and again the same details are available but I can search state wise. I can also do a phytochemical database search. So say for example I want to see vitexin. So the database will come up with a solution to that. So you can see vitexin.

This is the plant where it can be found. This is the IUPAC name, the pup chem and the details information. So let's search a common phytochemical. See, for example, kaempferol. So again, *Ficus carica* is where it is found.

This is what is called as the IUPAC name, the 2D version, which is referred to as the canonical smiles, the INCI abbreviation, the PubChem ID, as well as the regular smiles of that. You can also search one more database which gives us the detail about phytoconstituents along with the structure and that is your coconut database. So I am just clicking on this query which is already saved here and that is your *Papaver somniferum*. So if I see *Papaver somniferum*, it will come across what all compounds it contains along with the structure. And I click on this.



I again come up with the details, canonical smiles, synonyms, chemical class, alkaloids, especially the aporphine category. These are your tyrosine alkaloids. You see, you get a lot of details here. As well as the papers in which they are cited. Databases in which they are found.

Molecular properties. Especially how many atoms it contains, how many heavy atoms, how many rotatable bonds, how many amount of rings, what is the charge, does it contain sugar, how much it is more of a natural product likeness. That is how good is it as natural. It is definitely a natural product and hence the score is like more than 1. then total polar surface area log p which gives you a kind of lipophilicity index.

So, you will get different properties of this compound as well from your coconut databases. So, just quickly summarizing this. So now what we saw is, in addition to our previous database, which told us about the plants as well as therapeutic databases, which told us about the pharmacological activity, especially of traditional medicines. In today's session, we saw databases which will give us the list of phytoconstituents. So you can search the plant, you can search the traditional application, and once you are done with that, you can search or curate a list of compounds which are present in it.

This software, especially Dr. Duke's database, Knapsack database, PhytoHerb, Coconut, IMPPAT, as well as OSADHI databases, are going to help you curate that list of phytoconstituents. Many times you might see overlap between the databases, but sometimes some of the databases will come up with additional compounds. So you can consolidate them and come up with a new curated overall list. Still, many of the software programs are at developmental stages, and as a result, you will see many of them incomplete in terms of their information. But yes, as they develop, more and more compounds are being added to them.

So, these databases have been updated on a regular basis, and you can keep on checking them to update your list for the medicinal plant that you are interested in. So, thank you, everyone. Here are a few references that you could check forward, which I'll be sharing. And then, thank you for your time. Keep on searching these queries with the plant that you are interested in and see what information these databases have to offer about the plant of your interest.

Thank you.