The Wired Herbarium #13: Taxonomic Name Resolution Service 3.0
Eric Ribbens

In early June Bradley Boyle (University of Arizona) announced the release of a major upgrade to the Taxonomic Name Resolution Service (TNRS) 3.0. A product of the iPlant collaborative, an NSF-funded entity, the goal of TNRS is to take lists of scientific plant names and compare them to several different commonly used taxonomic resources, including TROPICOS, the USDA Plant Database, BIEN, and the Global Compositae Checklist.

As we all know, taxa get shifted to other names as our taxonomic understanding grows, and there are too many synonyms and names with multiple authorities. I frequently refer to TROPICOS, the USDA Plants Database, and the International Plant Names Index (IPNI), although I have been puzzled by the multiple results that commonly turn up in IPNI, or the differences in authorities between USDA and IPNI or TROPICOS. Furthermore, each of these sites is only usable one species at a time. Therefore, I was quite intrigued by the concept of TNRS and tried it out. I also contacted Dr. Boyle and invited him to respond to my evaluation, which he did. I have inserted his comments as separate paragraphs in the text.

The TNRS homepage is http://tnrs.iplantcollaborative.org/. It features a rotating galley of gorgeous flower pictures that kept distracting me, because I kept trying to guess what the species was. I recommend reading the "Optimize Your Search" and then entering a few names individually, and inspecting the results to get a feel for the website. Annoyingly, the webpage flashes a quick message during processing that I was unable to read, and you cannot simply go back to the input page by hitting the back arrow, but must confirm that you want to "leave the page".

Dr. Boyle: I suspect the quick message you are seeing is the progress bar, which displays what percent of your list of names has been processed. For a small list, the progress bar only displays for a fraction of a second. I'll add a feature request for you that the popup only be displayed if the number of names is above a certain minimum threshold, say, 25 names (we'll have to do some testing). The prompt to confirm leaving the application page is to prevent users from losing their work if they submit a lot of names by typing them in directly.

However, the results are clear and nicely presented. After trying a few individual names, I decided to submit the list of taxa in the database of the herbarium I curate. I copied the genus, species, and subspecies columns from my database into a blank spreadsheet, and sorted it to remove all of the hybrids. Then I pasted the data into a simple ASCII text editor, and ended up with 6139 lines. TNRS accepts a maximum of 5000 lines, so I split the file into two. After saving them as .txt files, I submitted them to TNRS.

Dr. Boyle: The TNRS accepts an unlimited number of names if you use the file upload utility. I've processed nearly 100,000 names at once. The limit only applied to names entered directly in the text box.

The submission process is very straightforward. Upload the txt file, enter your email, and click submit. A message pops up saying the file is being processed, an email confirming the submission appears in your inbox, and four minutes later followup emails reported the results were available.

To retrieve the results, you enter your email and a 30-character submission key (given in the email). The results appear on the screen, and can be downloaded into an excel-ready file. I changed my download settings to detailed results. One complication I encountered was that one of the files I submitted had something in an incorrect format. Annoyingly, the webpage merely reports that there was a problem, but does not indicate which line contained the problem. I never did figure out what the problem was.

The data files are easily uploaded into Excel, and contain a wealth of information. The submitted name, matched name, sources for matched name, authorities, and scores for the strength of the search results are reported, along with several columns of warnings, unmatched names, and a column titled "Selected". I sorted by Selected, and found every row had the value True, so I deleted that column.

Dr. Boyle: If you choose "Best matches only" as your other download option, the TNRS will return only one best match for each name, the the column "Selected" will only contain the value TRUE. If you instead choose "All matches" your results will contain all top matches for a given name; in the column "Selected", one will be marked "TRUE" and the rest "FALSE".

First I sorted by the score. 63 of my taxa had problems. In many cases, it appeared to be a slight spelling difference, but in some cases the genus was recognized but not the species. Some of the nonrecognized taxa puzzled me. For example, *Cucurbita pepo* was not found, and resolved only to genus, although *Cucurbita pepo* is in TROPICOS and USDA, and I had it properly spelled. On the other hand, I had *Potentilla pusilla*, and the webpage matched it with *Potentilla pumila*. That match makes sense to me, although I plan to pull that specimen and check the entry. I created a new column of entries I wanted to evaluate, and marked the 63 entries with problems. Then I sorted by unmatched terms, which turned out to be 43 entries, all also flagged by my earlier search by score.

I then searched by "Warnings", and discovered a number of entries flagged as "Ambiguous match". This made no sense to me, since the scores were 1, the names were identical, and many of the entries were accepted names. Also curiously, the "Warnings" column did not flag the entries for which complete matches were not made. So the *Cucurbita pepo* line did not carry a warning, even though the website had been unable to match the species name. I did turn up six entries that were marked as "Better higher taxonomic match available", which I also marked for further evaluation.

Dr. Boyle: That [the Cucurbita problem] is puzzling. I just checked to confirm the issue. The name is in our database, but as you found, it is not returned by the interface. I'm reporting this issue as a bug. We'll check into it. And the lack of a Warning is definitely a

bug. In general, the parsing application is good at detecting the rank of a name submitted; if you submit a species but the TNRS only finds a genus, the result should be flagged as partial match. Thank you for spotting this issue.

Then I sorted by Taxonomic Status. Most of my entries were accepted or no opinion. I'm not sure what the no opinion means, but the name I entered and the name matched by TNRS were generally identical. Another large set of entries were marked as synonyms. Some of these puzzled me. For example, I entered *Juncus parviflora*. TNRS indicated the correct spelling was *Juncus parviflorus*, and that it was a synonym and the correct name was *Rhynchospora contracta*, and cited TROPICOS as the source. But when I went to TROPICOS and entered *Juncus parviflorus* TROPICOS flagged it as a basionym, and reported a number of alternative combinations, including *Luzula parviflora*. Likewise, USDA places *Juncus parviflora* as a synonym of *Luzula parviflora*. I have no idea where the *Rhynchospora* reference was obtained. Others are areas where the different sites may be in disagreement. For example, USDA lists *Silphium gracile* as a valid name, but TROPICOS places it in *Silphium radula gracile*.

Dr. Boyle: Your question about "No opinion" names is a good one. I'll make sure we add an explanation to our website. A name marked "No opinion" is one for which the taxonomic source provides the name only, without additional information indicating whether they accept that name or treat it as a synonym. Some sources, such as USDA Plants, always provide a taxonomic opinion. Each name is either marked as "Accepted" or "Synonym". Such sources do not have "No opinion" names. Tropicos, perhaps because it is more comprehensive, contains numerous "no opinion" names. These are either names for which no information is currently available, aside from the original publication, or names for which there are conflicting opinions in the literature which have yet to be resolved. The Plant List (http://www.theplantlist.org/) also has many such names. That is the reality of taxonomic knowledge; taxonomists don't always agree.

Nonetheless, I had 647 entries to further investigate. I separated the "problem" entries from the accepted and no opinion results, and sorted by authority. Most results come with authorities, but some did not. Most of the "no authority" entries were subspecies or varieties of one sort or another, but some were just species. For example, *Hieracium diaphanum* did not return an authority, although TROPICOS lists the authority as Fr.

Then I compared authorities, and finally families. I ended up with hundreds of taxa to check, but the decision webpage is cited in the results list, which was extremely helpful in tracking down nomenclatural mistakes. I'm still working through the list of potential mistakes, but this site has fairly quickly helped me identify a large set of taxa to further investigate.

It also repeatedly reminded me that the USDA website fairly frequently has completely different authorities. I have no idea why, but very commonly TROPICOS and other websites will use one set of authorities, and USDA will be different. For example, *Gratiola aurea* is named by Muhl. in IPNI and in TROPICOS, but in USDA is Pursh. However, it is not just USDA that may be different. Both USDA and TROPICOS return Nuttall as the authority for *Solidago speciosa*; the GCC website returns A. Gray. *Quercus robur* is Linnaeus in both USDA and TROPICOS, but

TNRS returns (Ten.) A. DC. In other words, use these results as starting points for investigation, not as final answers.

A few summary comments. First, the user will greatly benefit from some preliminary trials of the search engine, and from reading the information in the "Optimize Your Search" tab. After the search is completed, the "Instructions" tab is very helpful in explaining what the various columns mean.

Second, a significant flaw in my opinion is that when a file that is submitted contains entries that are not in the correct format, the webpage merely returns an error message. It would be much more helpful if the error message included the line # of the entry with the first error.

Third, I think the retrieval process is a bit cumbersome. I don't understand why TNRS doesn't simply email a user-specified file when the job is processed.

Dr. Boyle: We've had a lot of discussion about this, and comments from previous testers. One of the problems is file size. Because there is no limit to the size of the file submitted, sending results by email could blow up some users inboxes. I suppose one solution would be to have a checkbox which would give the user the option of having the results emailed. Another reason for the verification is that the TNRS website does not require a login; the authentication prevents bots and other malicious applications from damaging the website. In general it was decided that this relatively simple authentication at time of access was preferable to requiring users to register and login. But we will definitely take your comments into consideration and see what we can do to streamline the retrieval process.

ERIC: My reply to his comments: I think he is right, and I am wrong about this. I'm not sure about the email size issue, since I'd like to obtain the results in a file anyhow, but I think he has a very good point about the dangers lurking in the background of our web experiences. I do like the option of a checkbox to have the file emailed on.

However, these are all minor complaints. I think this is a useful website, one that I will add to my list of herbarium bookmarks and regularly use. It will not magically make all of your nomenclatural resolution problems disappear, but it is a good tool to efficiently compare your list against several commonly used websites for taxonomic information. It could also be useful if you needed to split a list of taxa into multiple columns and for some reason didn't want to use Excel's text import function, or add authorities to a list in which they are missing.

Well, now I have more work to do with my database! Oh, and Dr. Boyle says that maybe IPNI will be added to TNRS in the future!