Visualizing phylogenetic trees, and linking them to databases

Brent D. Mishler
and
Rebecca L. Shapley
Dept. of Integrative Biology
University and Jepson Herbaria
University of California, Berkeley

Outline:

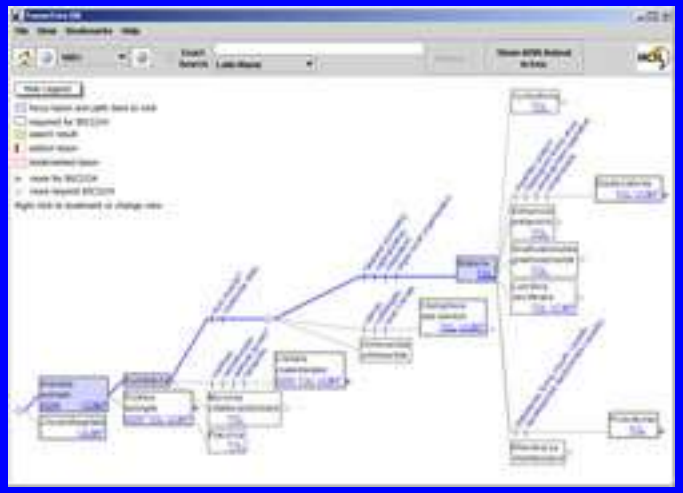
- Representing phylogenetic trees for outreach and understanding. A users' perspective.
- How to incorporate a phylogenetic ontology in bioinformatics?

Representing phylogenetic trees for outreach and understanding

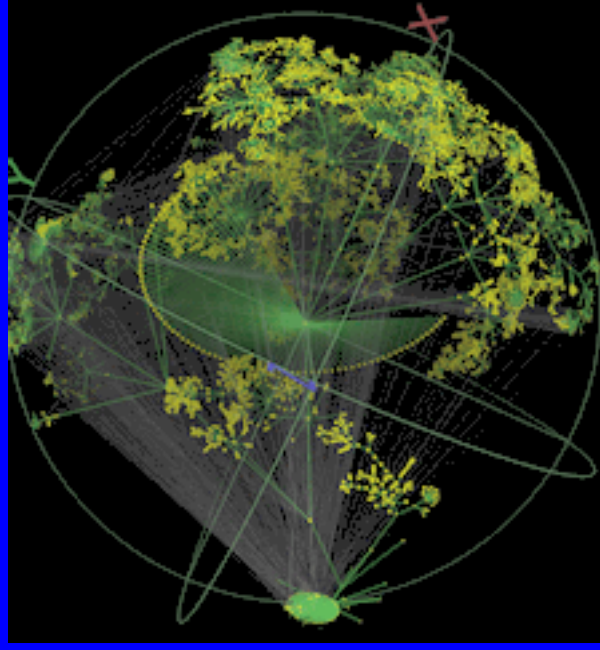
- A users' perspective.
  http://www.sims.berkeley.edu/~rebecca/cipres/compare.htm
- Audiences:
  - middle school
  - high school
  - college students in diversity courses
  - professional users of systematic information
  - systematic biologists

Interim Interview Results, sample of common responses:

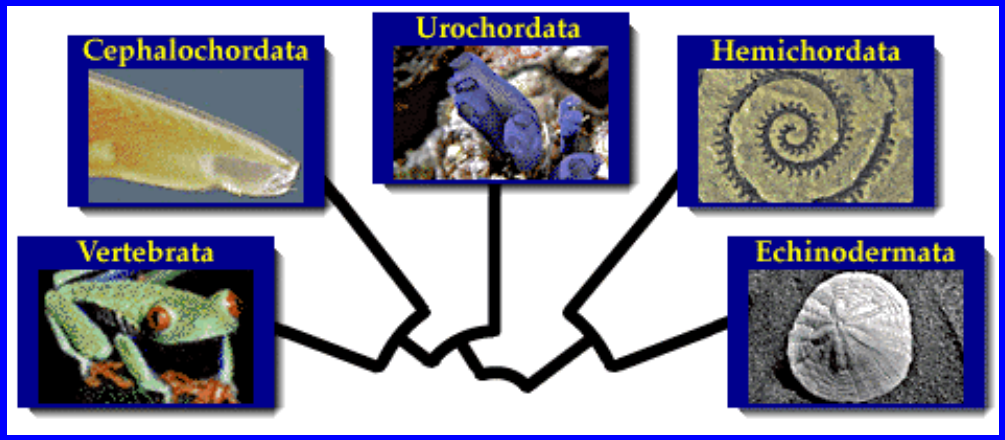
- For middle school, start with where humans fit on the tree. Back up to show context for mammals, vertebrates, animals, eukaryotes, whole tree.
- Won't know how to navigate tree from taxon names...won't recognize them. Therefore, will need search tools and other guidance to find relevant parts.
- Will want to use common names
- Will want to use the tree as a way to access taxon information, not just the classification itself.



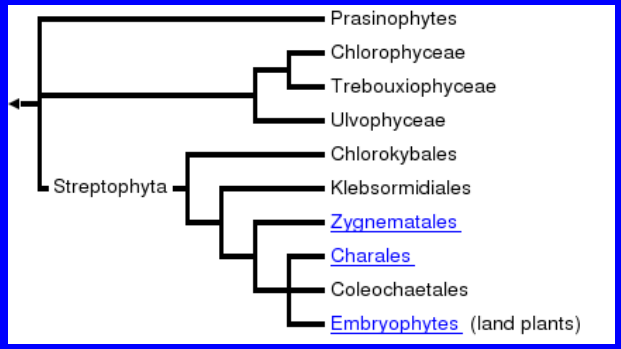
TaxonTree



Walrus -- 3D hyperbolic geometry



UCMP - Evolution site



Tree of Life

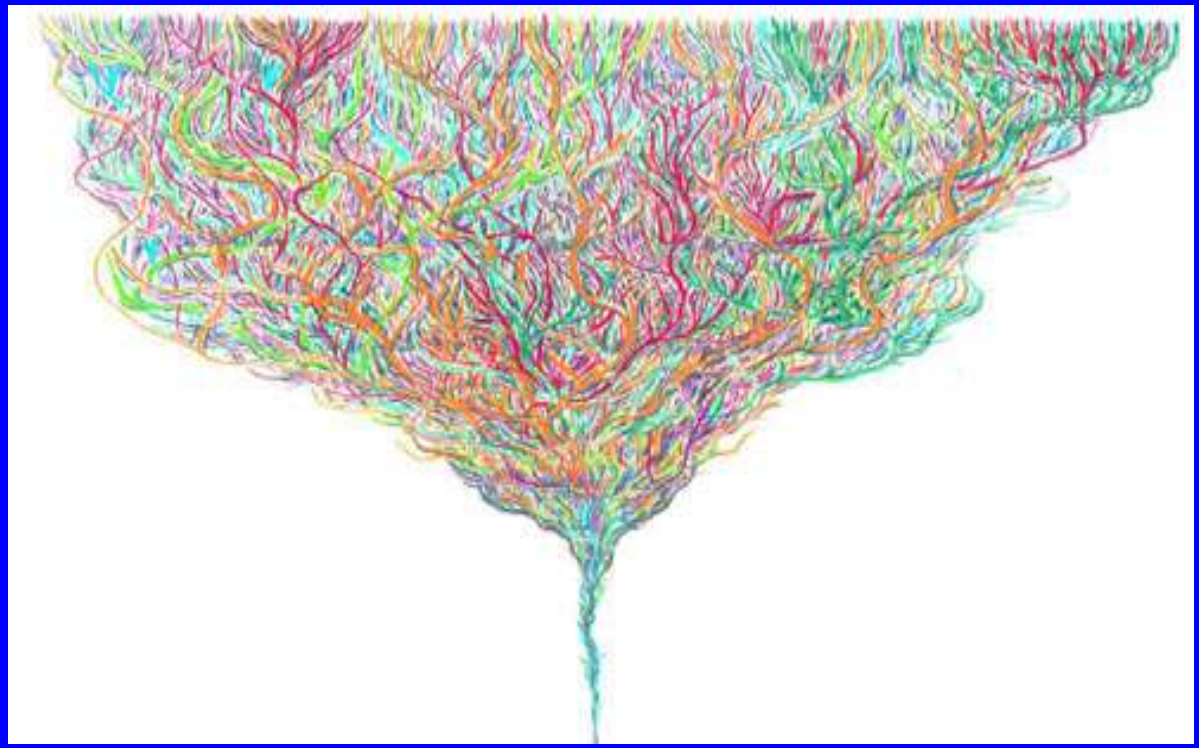
...and my favorite -- the GreenToL hyperbolic tree

Preliminary taxonomy of visualization approaches to nested data structures.

- matrix
- indented text
  - whole tree depth first (like this list)
  - whole tree breadth first (many dichotomous keys)
  - path-plus-breadth lineage browser (BIOT, GenBank, old parts of tol website)
- space-filling
  - linear: set-based display (Prometheus)
  - 2-D: "tree"-map
- "trees" - branching diagrams
  - navigated left-to-right
    - static - (Paup and many published trees)
    - interactive presentation
      - of fixed views (current tol website)
      - of dynamic views (taxonTree)
  - navigated bottom to top
    - static - (MacClade)
    - dynamic
    - hyperbolic (PEG)
  - center-to-outside layout
    - circle (UCMP Evolution website)
    - spiral
    - sphere (WALRUS)

What is the tree of life?

- It is a single, magnificent genealogy connecting all organisms alive today, and that ever lived.
- *Biodiversity is the whole tree of life*, not just the named species.
- In fact, even a complete list of named species would be a very poor representation of biodiversity (contra certain recent writers).



Biodiversity *Isn't* Species

- Biodiversity is the whole tree of life, not just the named species.
- There are lineages smaller and larger than the traditional species level.
- Species are not comparable between lineages in any manner, just an arbitrary cut-off somewhere along a branch in the tree of life.
- Thus only a creationist should think that species are the fundamental units of biodiversity, or that a list of currently named species in some way provides an inventory of biodiversity.

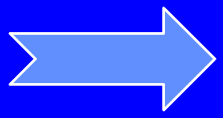
A proliferation of biological databases:

Species-based:

- identification (e.g., DELTA, LucID, MEKA)
- monographs/floras/faunas (e.g., Jepson Flora Project)
- distribution (e.g., BBS, UK Plant Atlas, CalFlora)
- images (e.g., CalPhoto, FishBASE)
- taxonomic management (e.g., IPNI, Tropicos)
- basic and applied ecology (USDA FEIS)

Individual-based:

- specimen-management (e.g., Biota, BRAHMS, SMaSCH)
- observation records (e.g., CalFlora, FishWatcher)
- ecological inventory (e.g., VegBank, SALVIAS)
- gene records (e.g., GenBank)



Very few are **Phylogenetic**, so far those are designed for:

- tree storage (TreeBASE)
- displaying current phylogenetic trees (e.g., Tree of Life, Phylomatic, Apweb).

The Problem:

- Most biological databases remain as essentially a "flat file" with respect to evolution.
- Data are entered with whatever taxon name (usually a species or a genus) happens to be attached to them.
- The only sense of evolutionary relationships is given by a schema of higher-taxon names (say families and phyla) that can be used to group the basic information.
- These higher taxa may or may not be monophyletic, and essentially function as static sorting bins -- there is no way to access or display emergent properties of data at higher evolutionary levels or to discover finer-scale patterns at lower levels.

The Problem, in a nutshell:

- The problem with unification of all these types of databases is that there is low linkage among them, despite high overlap in taxonomic coverage -- the linkage that exists is via a static taxonomic hierarchy.
- This restricts the analytical potential of databases.

The Solution, in a nutshell:

- Use the ubiquitous nature of the tree of life for linking databases.
- All biological data fall *somewhere* on the tree of life, which is the one thing that can unify them all.

- All biological data fall *somewhere* on the tree of life, which is the ***one thing*** that can unify them all.

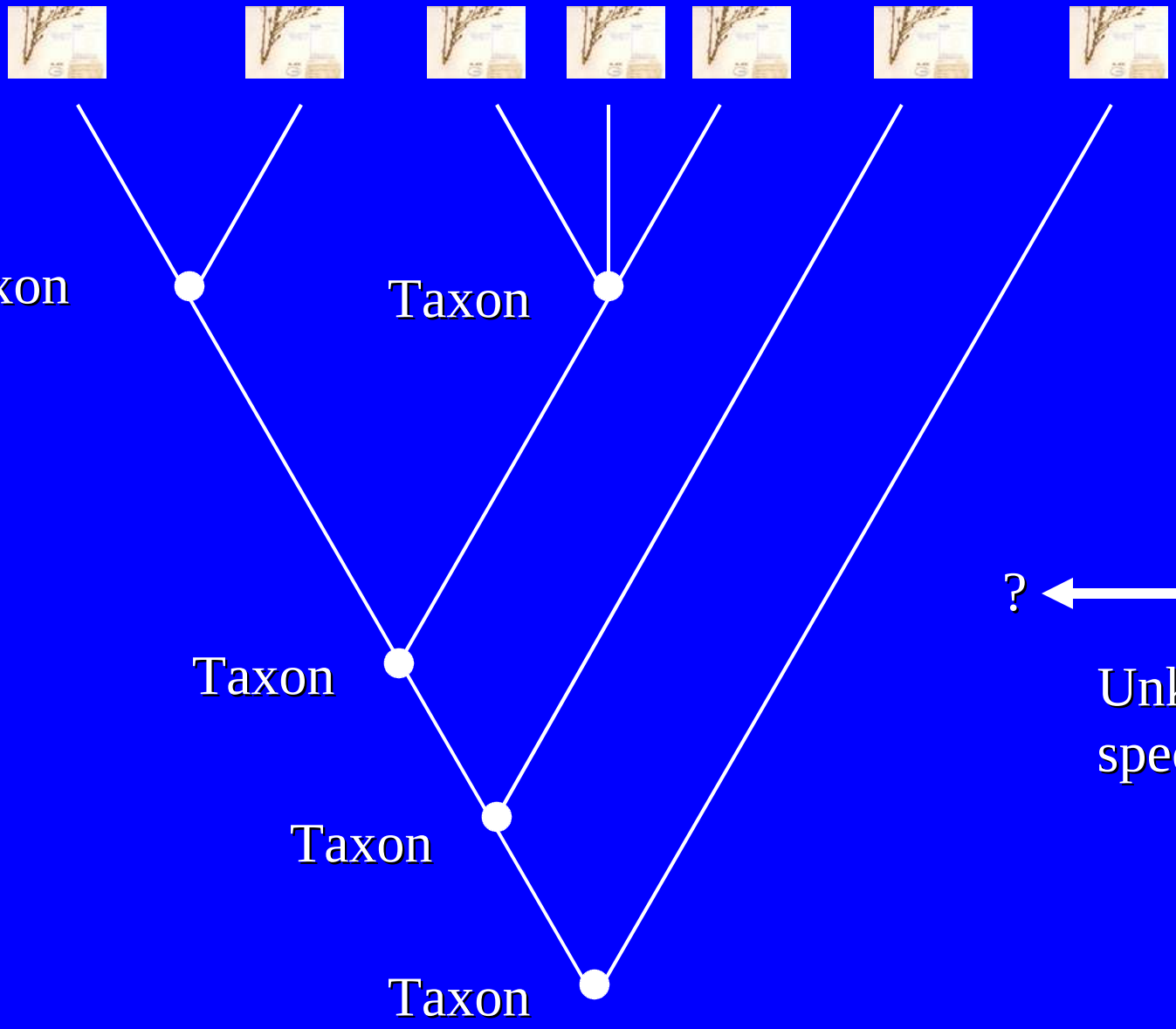
- Seriously, the use of phylogeny-space as the fundamental ontology for biological data, would have the same revolutionary impact as the use of spatial data has had on geography. Instead of Geographic Information Systems (GIS), we would have Phylogenetic Information Systems…

PIS !

The Tree of Life as information-organizer:

- Of all types of biological relationships, phylogeny provides the best *general purpose* classification. In other words, the single most important thing you can know about an organism is what is it related to.
- Phylogeny can be *the* factor that unites together all disparate biological databases.
- The centerpiece of future biological databases will be phylogenetic classification, a deeply nested hierarchy of nodes linked to all available structural and functional data at each level.

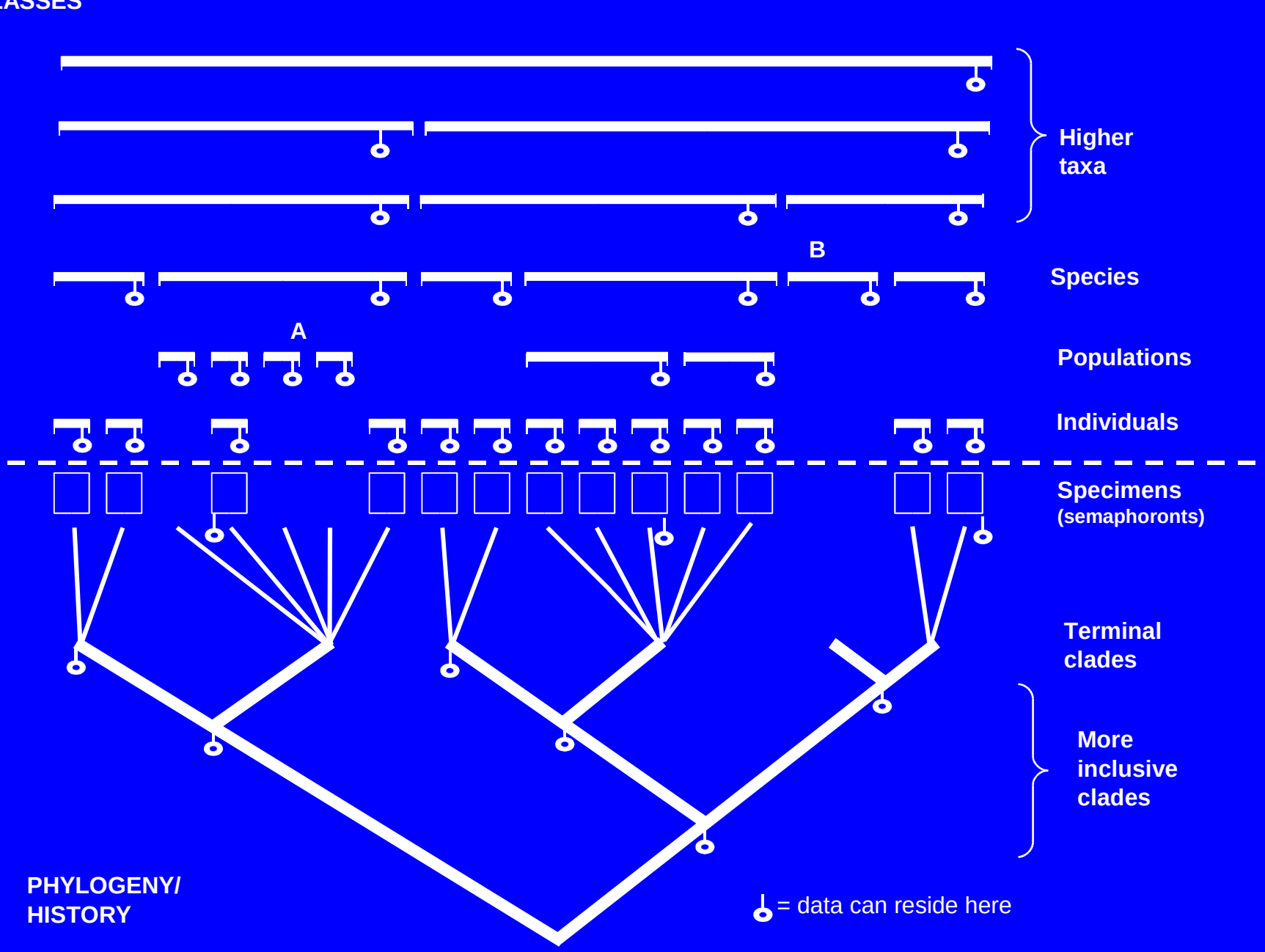
Specimens:



Two distinctions to make:

1 Between the phylogeny (which is an inference about *history*) and the named taxa (which are nested classes of *extant* individuals).

- The ultimate units of phylogenetic analysis are specimens -- semaphoronts in the terms used by Hennig (these include such items as museum specimens, photomicrographs, and tubes of extracted DNA). Never species or other taxa!


CLASSES
Higher taxa
B
Species
A
Populations
Individuals
Specimens
(semaphoronts)
Terminal clades
More inclusive clades
PHYLOGENY/
HISTORY
= data can reside here

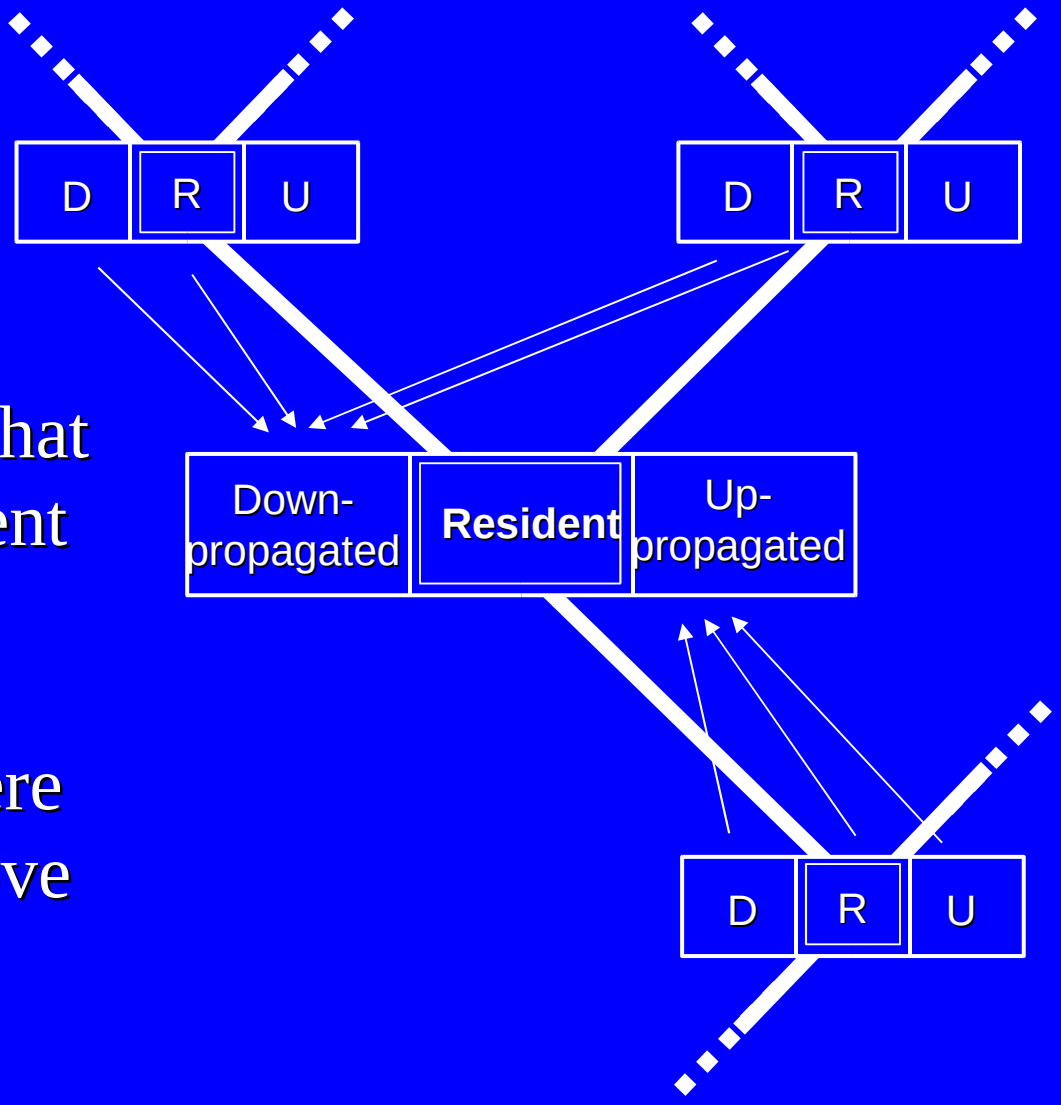
• Semaphoronts are the units bearing the empirical character data, and they are linked together into increasingly more inclusive monophyletic groups, each of which has an hypothesized ancestral node with inferred synapomorphies.

• The important thing to note is that the semaphoronts bear the actual data, and all the higher nodes have *inferred* ancestral states.

• On the other hand, phylogenetic taxa are made up of nested classes of contemporaneous organisms. The taxa can have their own attributes, such as average height and geographic distribution, that are not necessarily the same as the inferred ancestral attribute of the lineage.

Two distinctions to make:

2 Between data that are intrinsically resident at a node versus attributes that are inferred from elsewhere on the tree (either above or below).

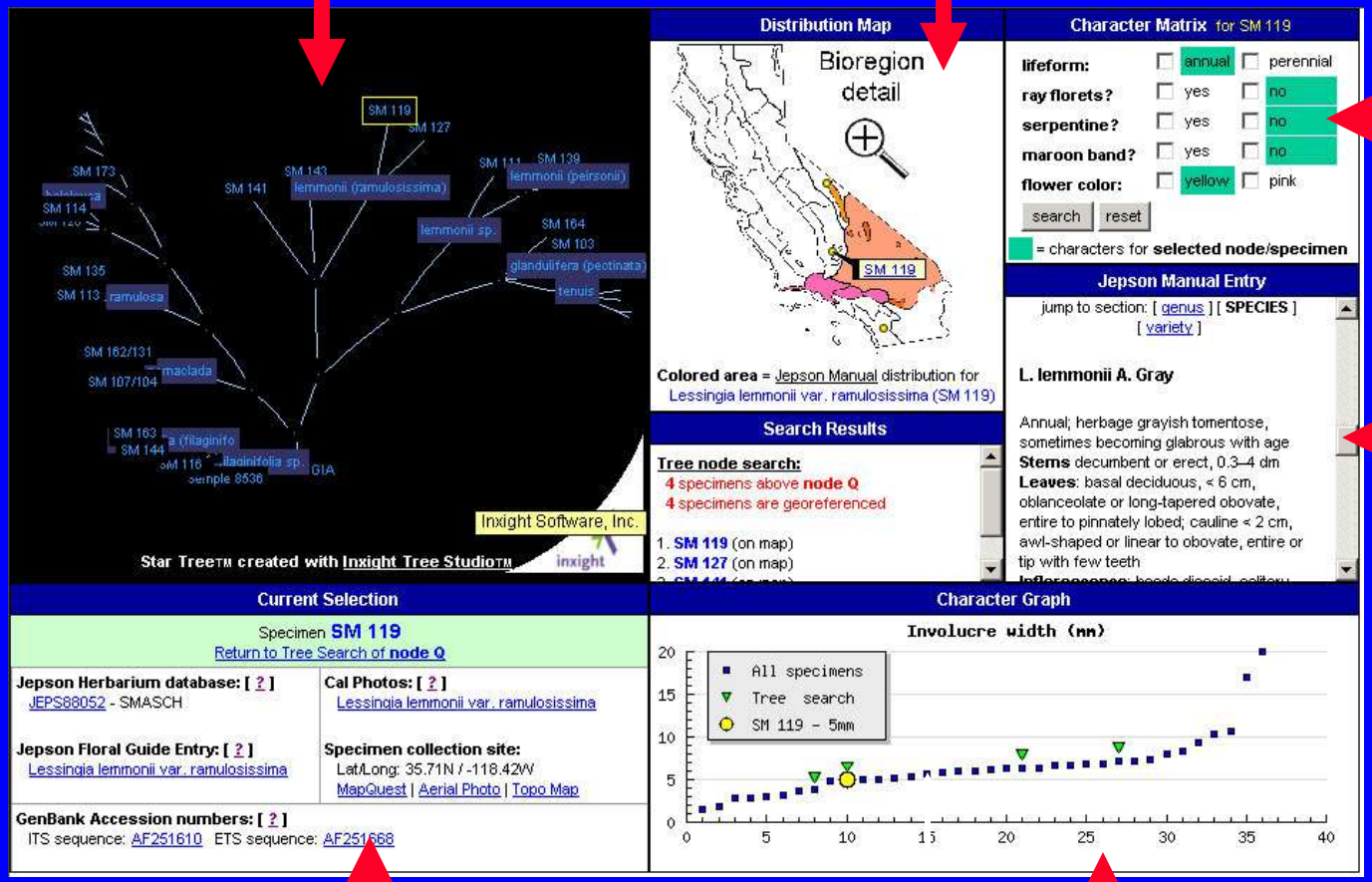


- Conventional statistical analysis focuses on down-propagated data (summary statistics of individual data points).
- Phylogenetic methods, including parsimony and maximum likelihood, utilize up and down-propagated methods in calculation of ancestral states.
- Up-propagated data can also provide predictive insights into organismal traits based on phylogenetic relationships.

PEG: Phylogeny, Ecology, Geography

Tree interface

Map interface



Character Interface

Description from literature

<u>Collaborators:</u>

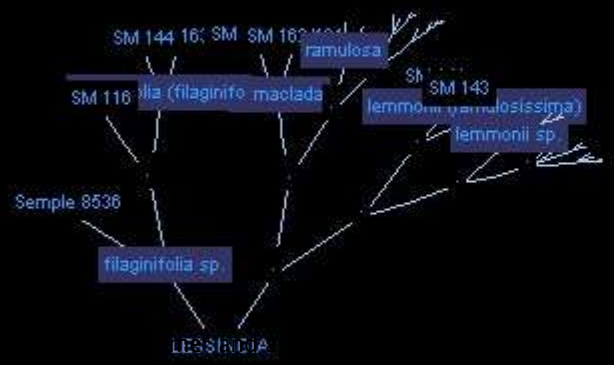
- Brent Mishler
- David Ackerly
- Cam Webb

Collection interface (voucher data; photos; Genebank)

Ecological Interface

Hyperbolic tree software generously provided by Inxight Software, Inc.

Demo at: http://ucjeps.berkeley.edu/upd/



Inxight Software, Inc.

Distribution Map

Character Matrix

lifeform:	annual	perennial
ray florets?	yes	no
serpentine?	yes	no
maroon band?	yes	no
flower color:	yellow	pink

search reset

Search Results

please make a selection

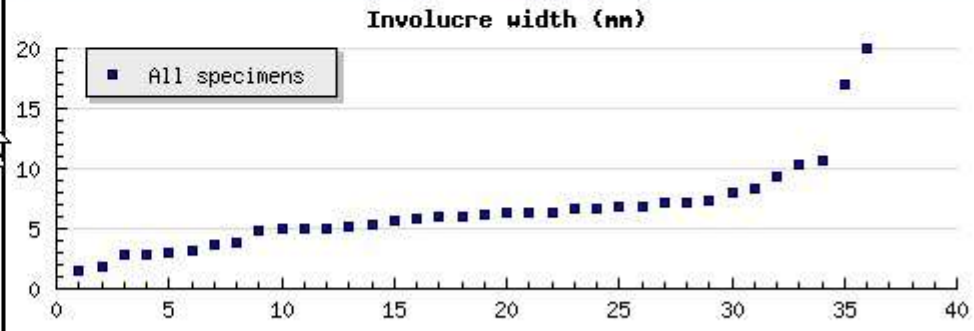
Jepson Manual Entry

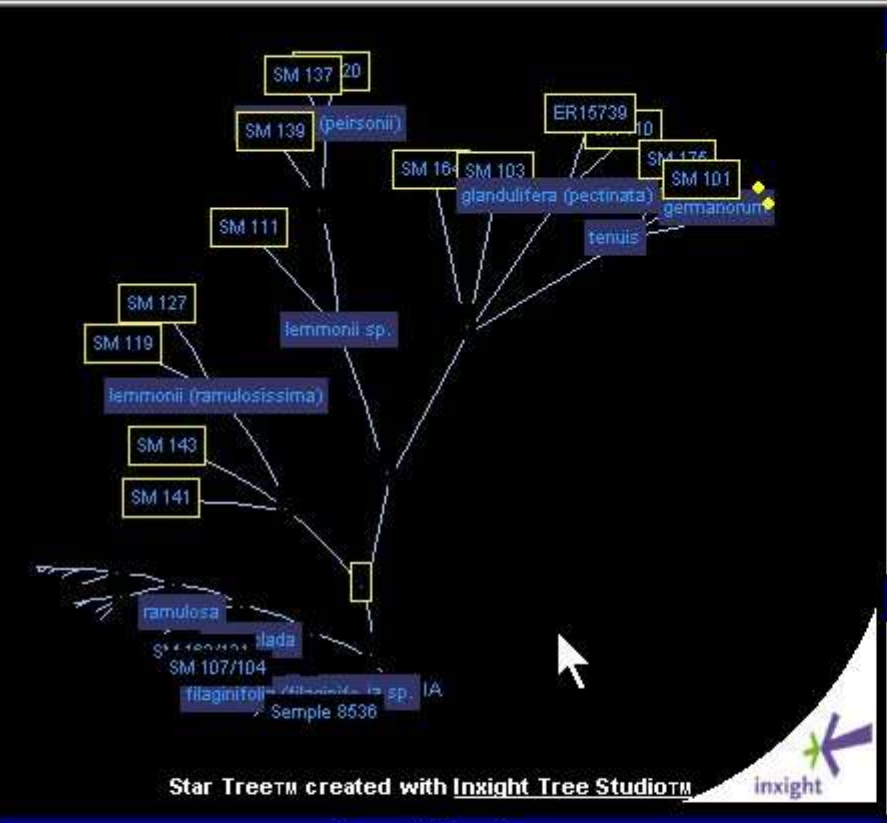
Current Selection

Welcome to the Ultimate Phylogenetics Database.

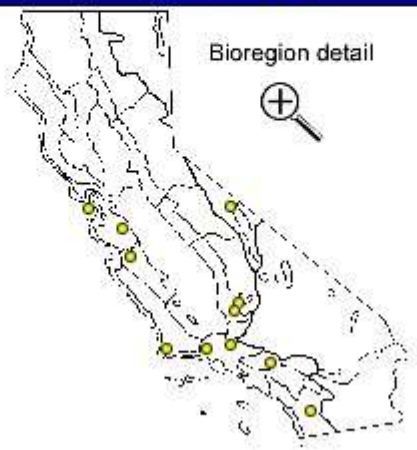
- **Single-click, hold and drag** anywhere on the tree to move it around
- **A Single-click** will center the hyperbolic tree on the node you clicked
- **Double click on a specimen** to load info about that node/specimen, generate a distribution map and show character stats
- **Double click on a node** to see all the specimen above that node on the map in addition to character states for upstream nodes

Character Graph





Distribution Map



Character Matrix for node P

lifeform:	☐	annual	☐	perennial
ray florets?	☐	yes	☐	no
serpentine?	☐	yes	☐	no
maroon band?	☐	yes	☐	no
flower color:	☐	yellow	☐	pink

search reset

= characters for **selected node/specimen**

Jepson Manual Entry

Search Results

Tree node search:

16 specimens above **node P**

12 specimens are georeferenced

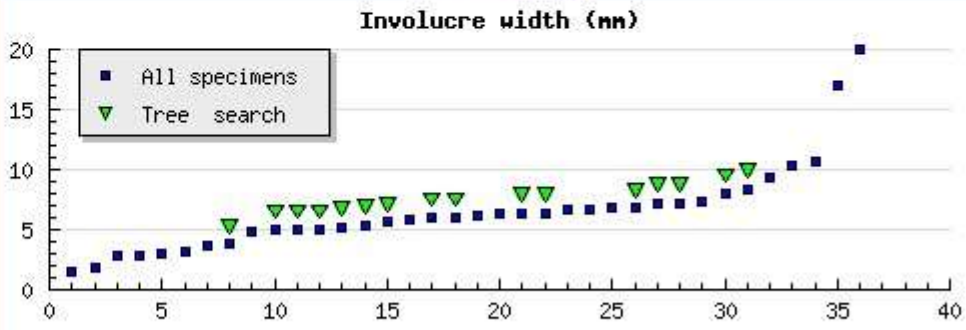
1. **ER15739**
2. **SM 101** (on map)

Current Selection

You performed a Tree search on node P

- (Example) **Click a yellow dot on a map** to see information about that particular specimen. You can then return to this search by clicking **"Return to Character Search"**

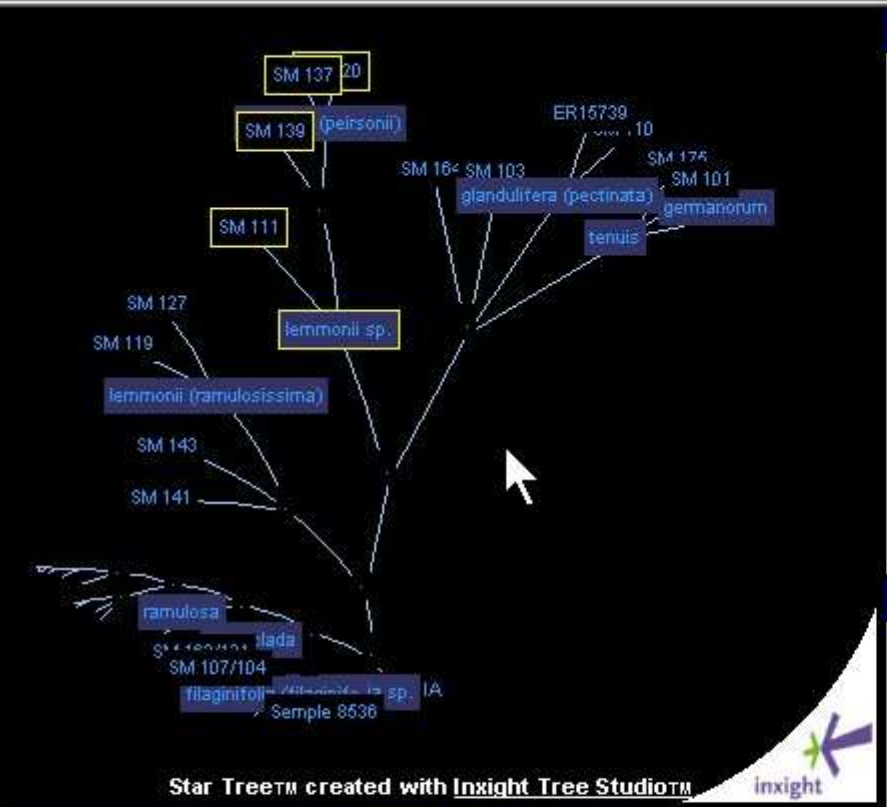
Character Graph



File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://128.32.109.11/upd3/ Go Links



Distribution Map

Colored area = Jepson Manual distribution for *Lessingia lemmonii*

Character Matrix for lemmonii sp

lifeform:	☐	annual	☐	perennial
ray florets?	☐	yes	☐	no
serpentine?	☐	yes	☐	no
maroon band?	☐	yes	☐	no
flower color:	☐	yellow	☐	pink

search reset

= characters for **selected node/specimen**

Jepson Manual Entry

jump to section: [genus] [**SPECIES**] [variety]

L. lemmonii A. Gray

Annual; herbage grayish tomentose, sometimes becoming glabrous with age
Stems decumbent or erect, 0.3–4 dm
Leaves: basal deciduous, < 6 cm, oblanceolate or long-tapered obovate, entire to pinnately lobed; cauline < 2 cm, awl-shaped or linear to obovate, entire or tip with few teeth

Search Results

Tree node search:

4 specimens above **lemmonii sp**
4 specimens are georeferenced

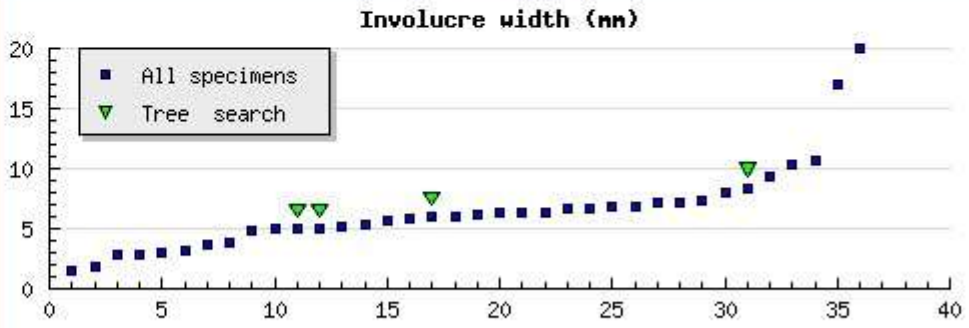
1. **SM 111** (on map)
2. **SM 120** (on map)

Current Selection

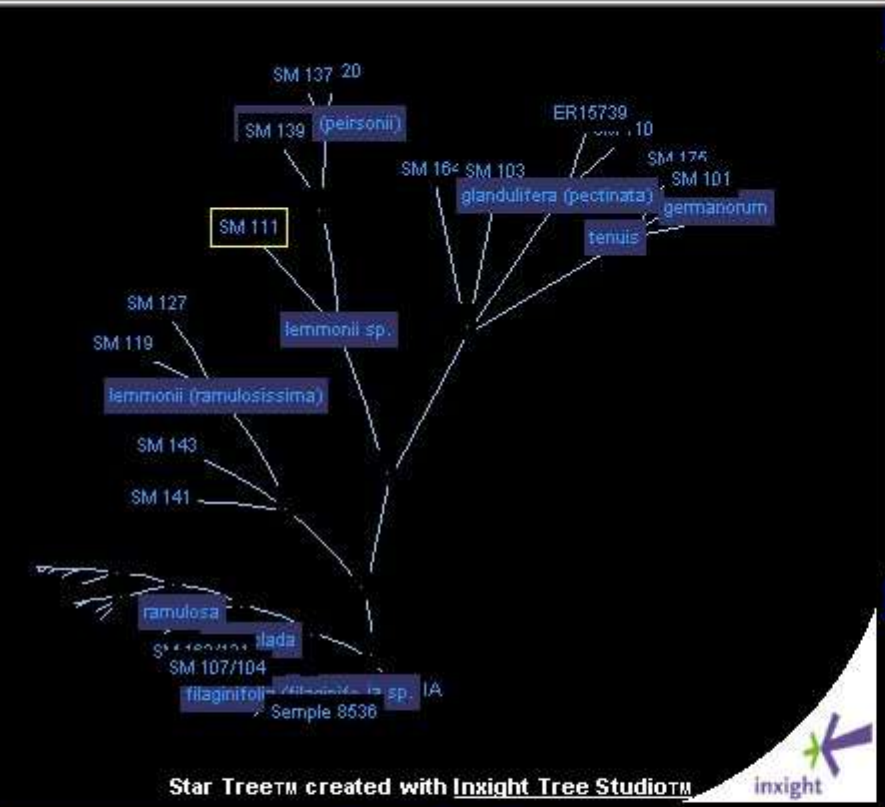
You performed a Tree search on lemmonii sp

- (Example) **Click a yellow dot on a map** to see information about that particular specimen. You can then return to this search by clicking **"Return to Character Search"**

Character Graph



Internet



Distribution Map

Colored area = Jepson Manual distribution for Lessingia lemmonii var. lemmonii (SM 111)

Character Matrix for SM 111

lifeform:	☐	annual	☐	perennial
ray florets?	☐	yes	☐	no
serpentine?	☐	yes	☐	no
maroon band?	☐	yes	☐	no
flower color:	☐	yellow	☐	pink

search reset

= characters for **selected node/specimen**

Jepson Manual Entry

jump to section: [genus] [**SPECIES**] [variety]

L. lemmonii A. Gray

Annual; herbage grayish tomentose, sometimes becoming glabrous with age
Stems decumbent or erect, 0.3–4 dm
Leaves: basal deciduous, < 6 cm, oblanceolate or long-tapered obovate, entire to pinnately lobed; cauline < 2 cm, awl-shaped or linear to obovate, entire or tip with few teeth

Search Results

Tree node search:
4 specimens above **lemmonii sp**
4 specimens are georeferenced

1. **SM 111** (on map)
2. **SM 120** (on map)

Current Selection

Specimen **SM 111**
Return to Tree Search of **lemmonii sp**

Jepson Herbarium database: [?]
JEPS88015 - SMASCH

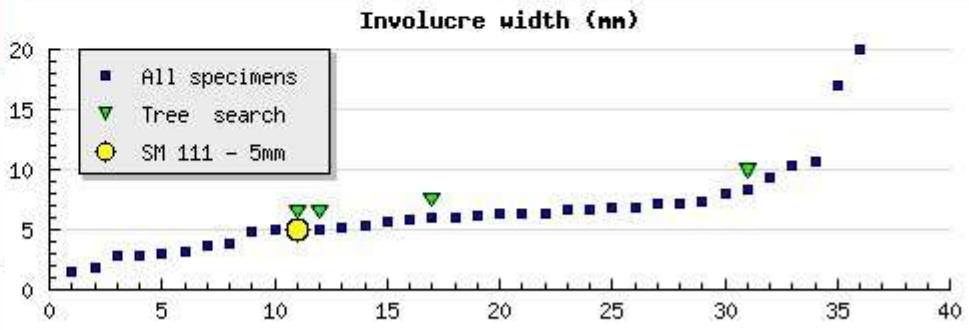
Jepson Floral Guide Entry: [?]
Lessingia lemmonii var. lemmonii

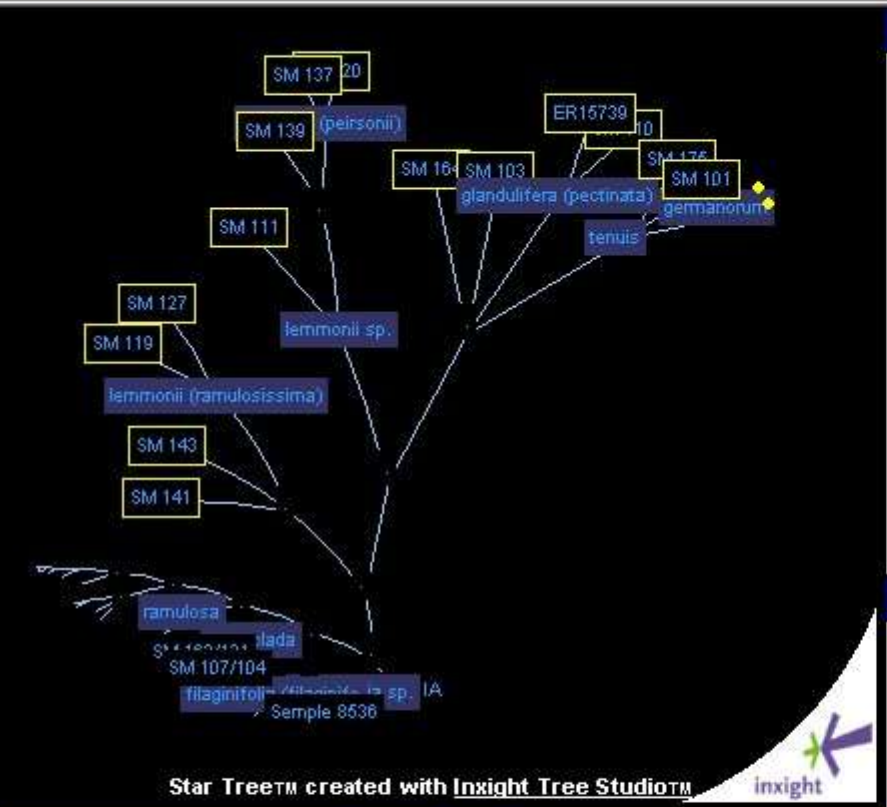
Cal Photos: [?]
Lessingia lemmonii var. lemmonii

Specimen collection site:
Lat/Long: 34.69N / -119.36W
MapQuest | Aerial Photo | Topo Map

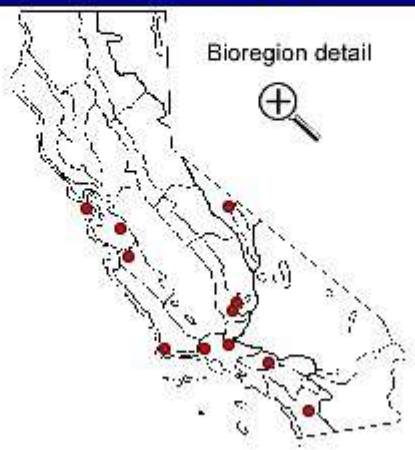
GenBank Accession numbers: [?]
ITS sequence: AF251606 ETS sequence: AF251664

Character Graph





Distribution Map



Character Matrix

lifeform:	☑ annual	☐ perennial
ray florets?	☐ yes	☐ no
serpentine?	☐ yes	☐ no
maroon band?	☐ yes	☐ no
flower color:	☑ yellow	☐ pink

search reset

= characters you selected for **this search**

Jepson Manual Entry

jump to section: [genus] [**SPECIES**] [variety]

L. lemmonii A. Gray

Annual; herbage grayish tomentose, sometimes becoming glabrous with age
Stems decumbent or erect, 0.3–4 dm
Leaves: basal deciduous, < 6 cm, oblanceolate or long-tapered obovate, entire to pinnately lobed; cauline < 2 cm, awl-shaped or linear to obovate, entire or tip with few teeth

Search Results

Character search:

16 total specimens with the characters:
lifeform = annual
color = yellow

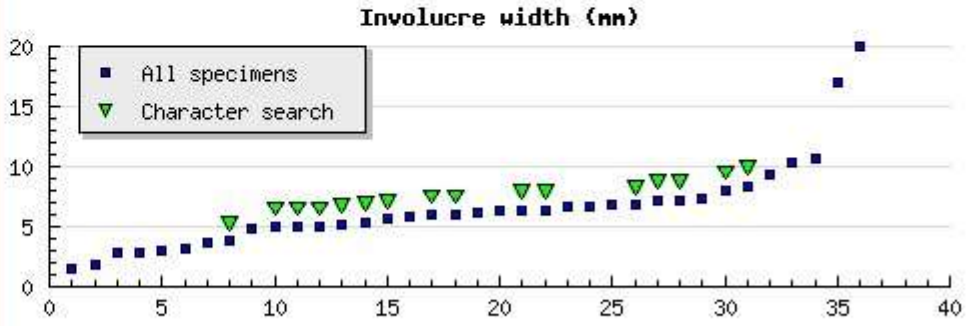
12 specimens are georeferenced

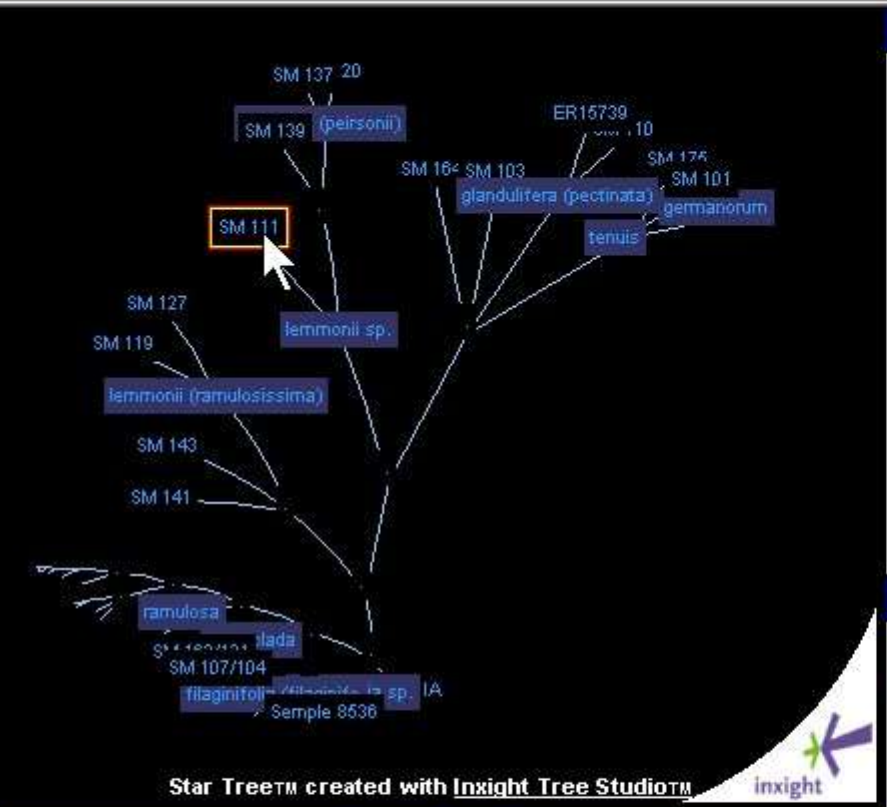
Current Selection

You performed the following Character Search:
lifeform = annual
color = yellow

- (Example) **Click a red dot on a map** to see information about that particular specimen. You can then return to this search by clicking **"Return to Character Search"**

Character Graph





Distribution Map

Colored area = Jepson Manual distribution for Lessingia lemmonii var. lemmonii (SM 111)

Character Matrix for SM 111

lifeform:	☐	annual	☐	perennial
ray florets?	☐	yes	☐	no
serpentine?	☐	yes	☐	no
maroon band?	☐	yes	☐	no
flower color:	☐	yellow	☐	pink

search reset

= characters for **selected node/specimen**

Search Results

Specimen search:

You selected specimen SM 111

Jepson Manual Entry

jump to section: [genus] [**SPECIES**] [variety]

L. lemmonii A. Gray

Annual; herbage grayish tomentose, sometimes becoming glabrous with age
Stems decumbent or erect, 0.3–4 dm
Leaves: basal deciduous, < 6 cm, oblanceolate or long-tapered obovate, entire to pinnately lobed; cauline < 2 cm, awl-shaped or linear to obovate, entire or tip with few teeth

Current Selection

Specimen **SM 111**

Jepson Herbarium database: [?]
JEPS88015 - SMASCH

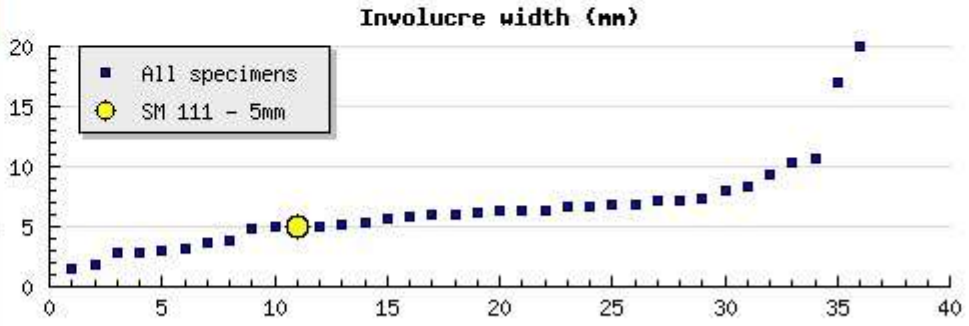
Jepson Floral Guide Entry: [?]
Lessingia lemmonii var. lemmonii

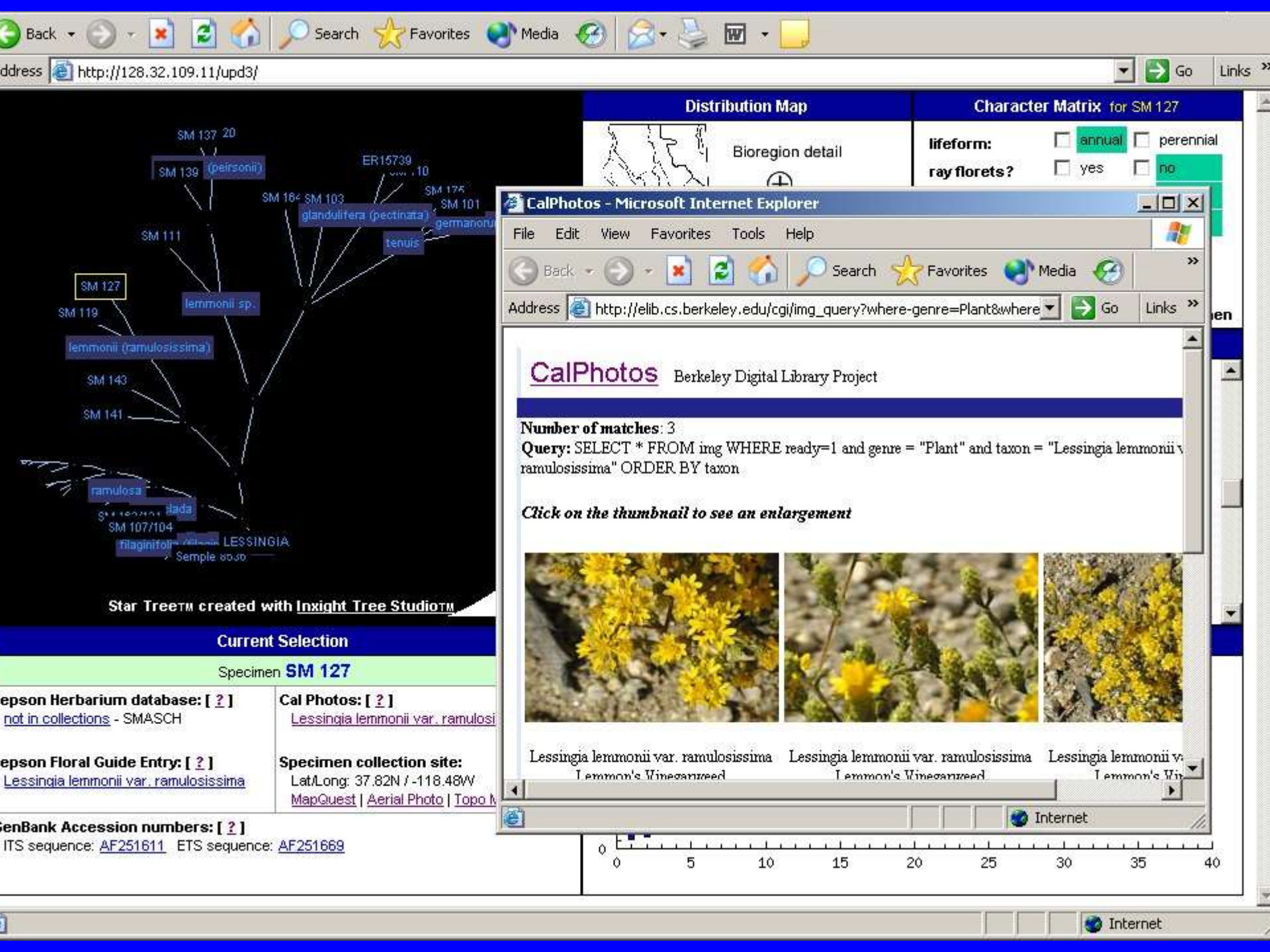
Cal Photos: [?]
Lessingia lemmonii var. lemmonii

Specimen collection site:
Lat/Long: 34.69N / -119.36W
MapQuest | Aerial Photo | Topo Map

GenBank Accession numbers: [?]
ITS sequence: AF251606 ETS sequence: AF251664

Character Graph

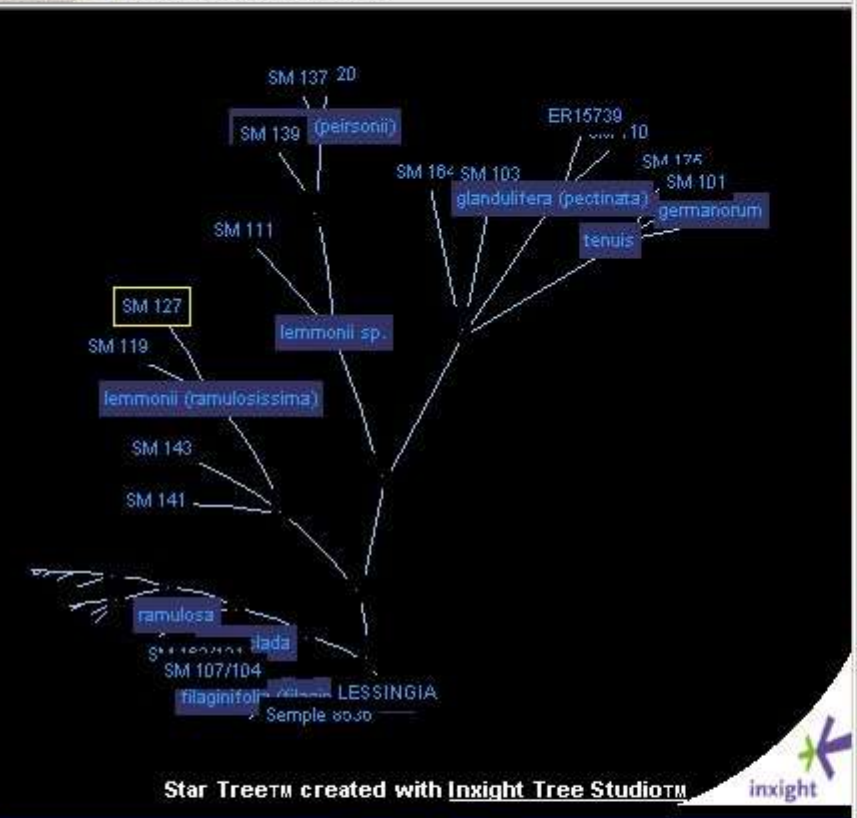



Back
Search
Favorites
Media
Address
http://128.32.109.11/upd3/
Go
Links
Distribution Map
Bioregion detail
Character Matrix for SM 127
lifeform:
annual
perennial
ray florets?
yes
no
SM 137
SM 139
(peirsonii)
ER15739
SM 164
SM 103
glandulifera (pectinata)
SM 101
germanoru
tenuis
SM 111
SM 127
SM 119
lemmonii sp.
lemmonii (ramulosissima)
SM 143
SM 141
ramulosa
SM 107/104
filaginifoli
LESSINGIA
Semple 8030
Star Tree™ created with Inxight Tree Studio™
Current Selection
Specimen SM 127
epson Herbarium database: [?]
not in collections - SMASCH
epson Floral Guide Entry: [?]
Lessingia lemmonii var. ramulosissima
Cal Photos: [?]
Lessingia lemmonii var. ramulosi
Specimen collection site:
Lat/Long: 37.82N / -118.48W
MapQuest | Aerial Photo | Topo M
enBank Accession numbers: [?]
ITS sequence: AF251611 ETS sequence: AF251669
0
0
5
10
15
20
25
30
35
40
Internet
CalPhotos - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://elib.cs.berkeley.edu/cgi/img_query?where-genre=Plant&where
Go
Links
CalPhotos
Berkeley Digital Library Project
Number of matches: 3
Query: SELECT * FROM img WHERE ready=1 and genre = "Plant" and taxon = "Lessingia lemmonii v
ramulosissima" ORDER BY taxon
Click on the thumbnail to see an enlargement
Lessingia lemmonii var. ramulosissima
Lessingia lemmonii var. ramulosissima
Lessingia lemmonii v
Internet
Internet

File Edit View Favorites Tools Help

Back Search Favorites Media

ddress http://128.32.109.11/upd3/



Current Selection

Specimen **SM 127**

epson Herbarium database: [?]
not in collections - SMASCH

epson Floral Guide Entry: [?]
Lessingia lemmonii var. ramulosissima

Cal Photos: [?]
Lessingia lemmonii var. ramulosissima

Specimen collection site:
Lat/Long: 37.82N / -118.48W
MapQuest | Aerial Photo | Topo Map

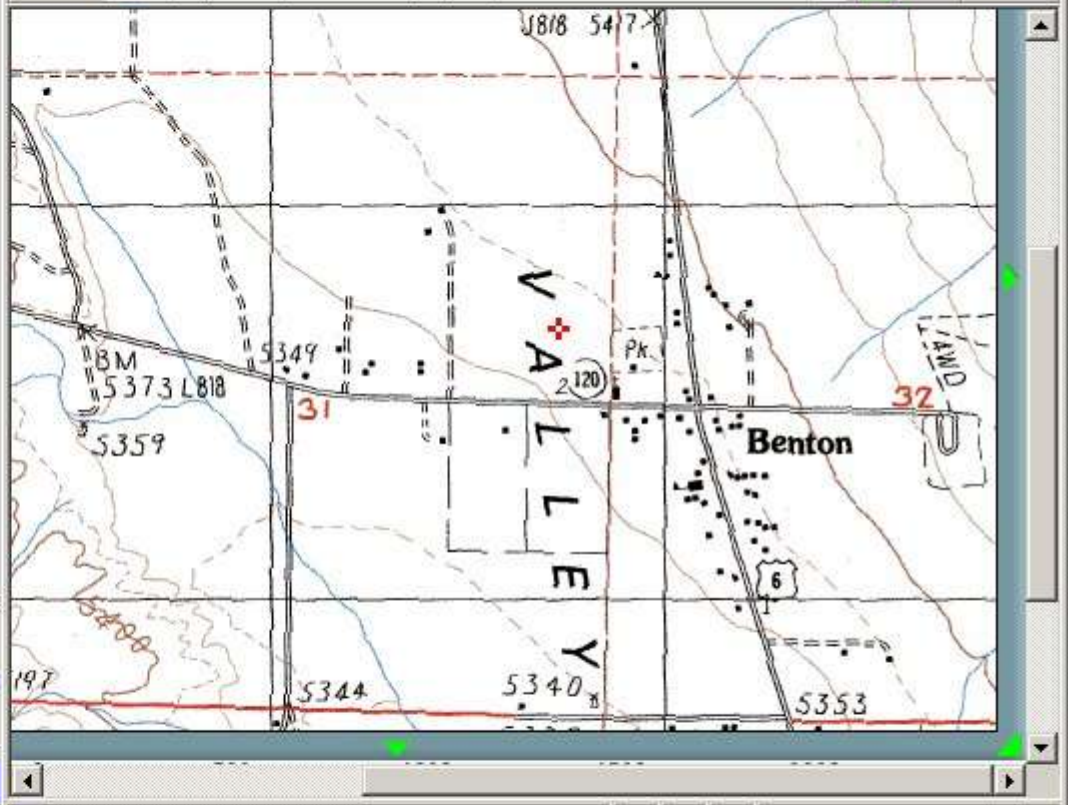
GenBank Accession numbers: [?]
ITS sequence: AF251611 ETS sequence: AF251669

TopoZone - The Web's Topographic Map - Microsoft Internet Explorer

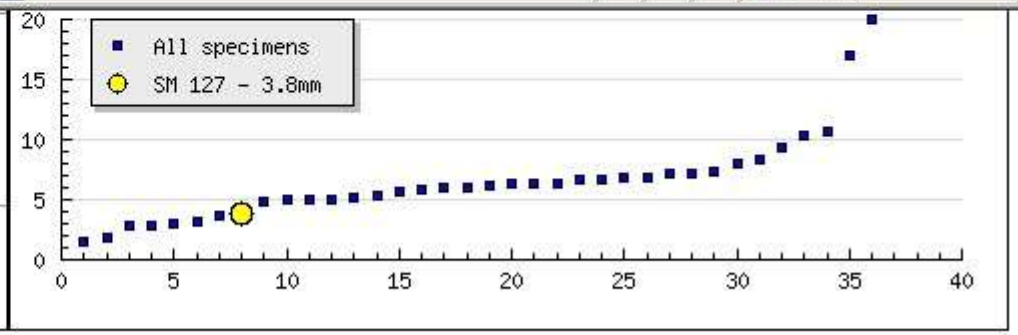
File Edit View Favorites Tools Help

Back Search Favorites Media

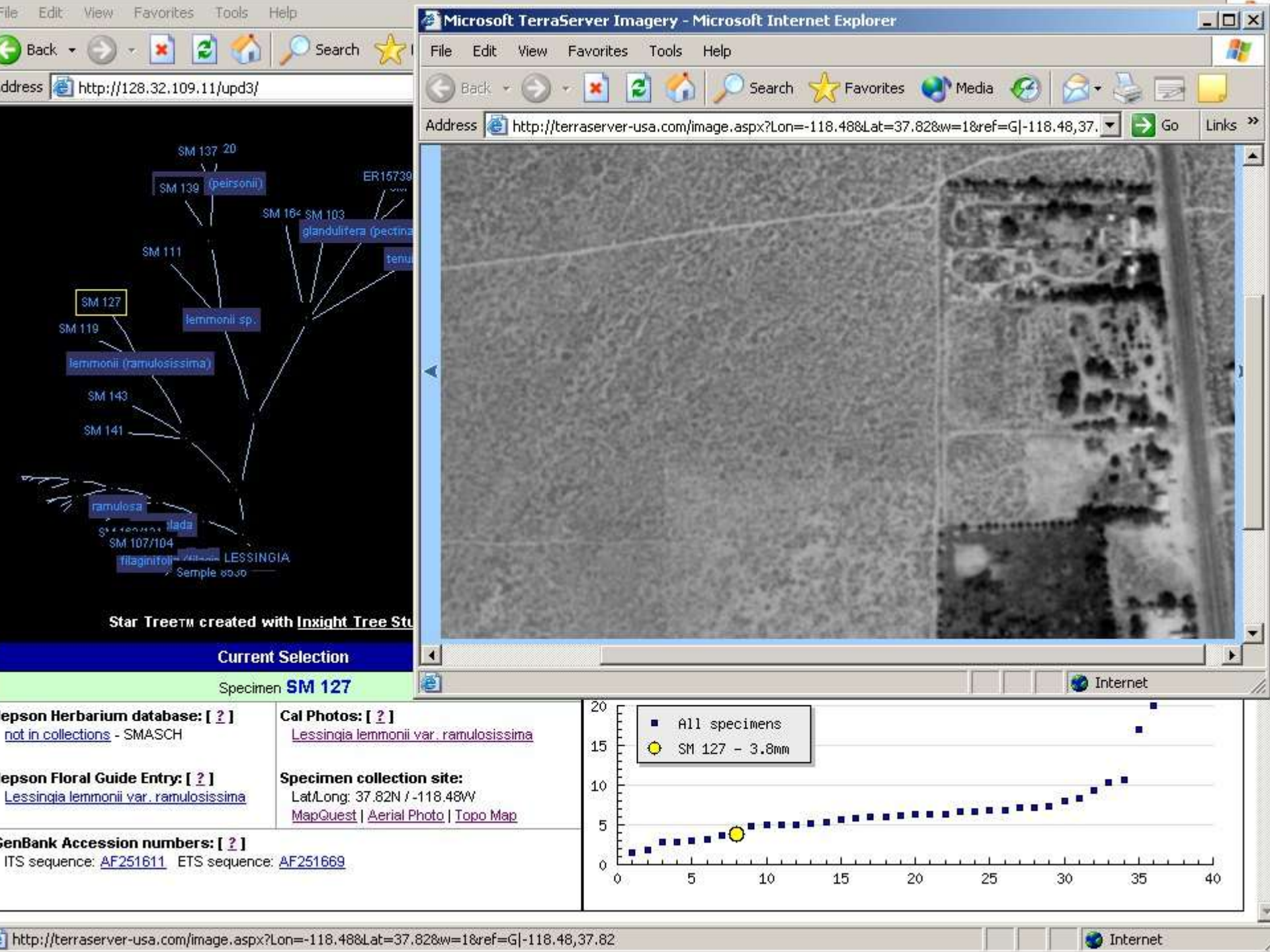
Address http://www.topozone.com/map.asp?s=25&lat=37.828&lon=-118.48 Go Links

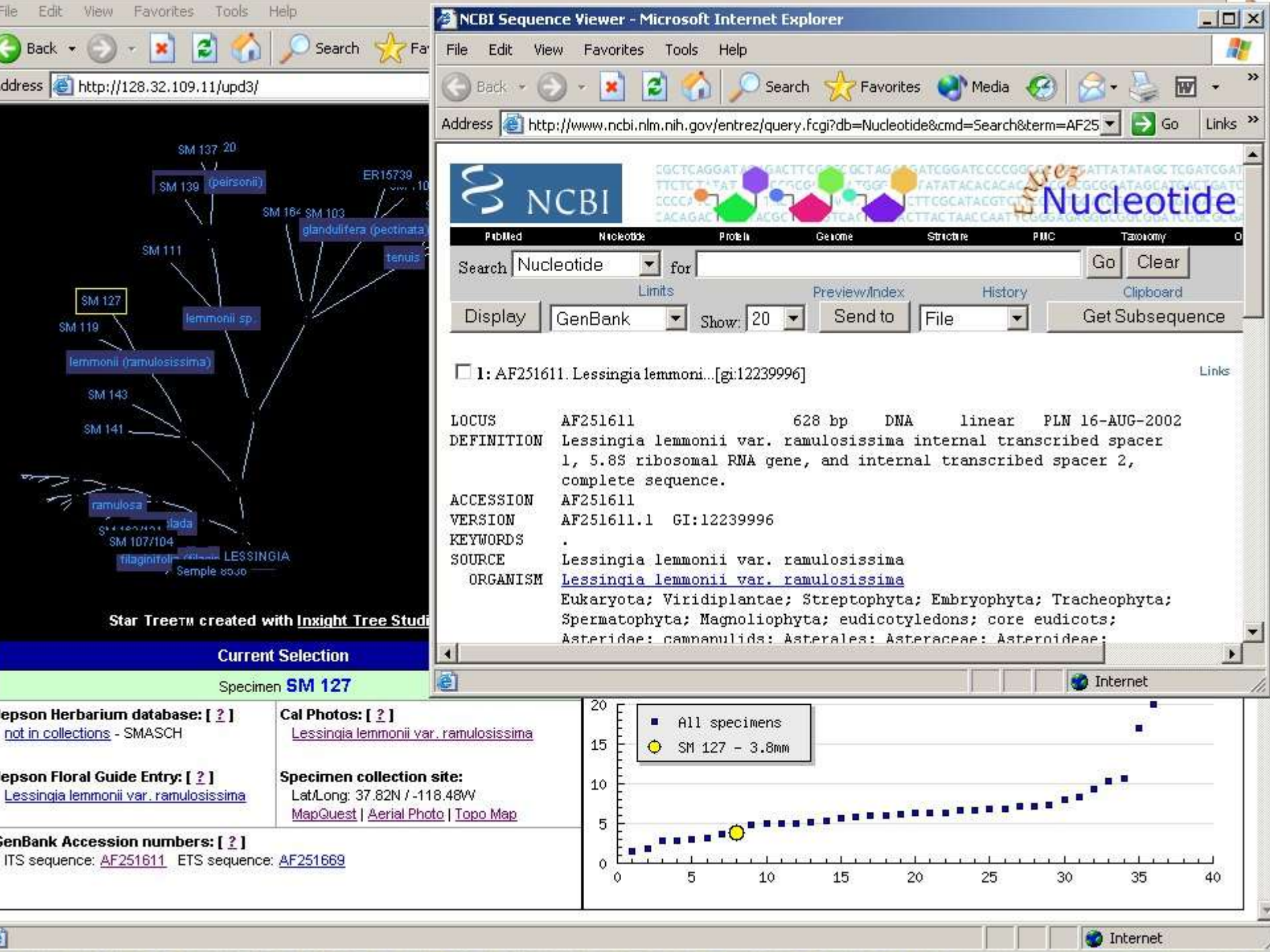


Internet



Internet


Microsoft TerraServer Imagery - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://terraserver-usa.com/image.aspx?Lon=-118.48&Lat=37.82&w=1&ref=G|-118.48,37.
Go Links
Internet
http://128.32.109.11/upd3/
SM 137
SM 139 (peirsonii)
ER15739
SM 103
glandulifera (pectina
SM 111
SM 127
SM 119
lemmonii sp.
lemmonii (ramulosissima)
SM 143
SM 141
ramulosa
SM 107/104
filaginifoli
LESSINGIA
Star Tree™ created with Inxight Tree Stu
Current Selection
Specimen SM 127
epson Herbarium database: [?]
not in collections - SMASCH
epson Floral Guide Entry: [?]
Lessingia lemmonii var. ramulosissima
Cal Photos: [?]
Lessingia lemmonii var. ramulosissima
Specimen collection site:
Lat/Long: 37.82N / -118.48W
MapQuest | Aerial Photo | Topo Map
enBank Accession numbers: [?]
ITS sequence: AF251611 ETS sequence: AF251669
All specimens
SM 127 - 3.8mm
0
5
10
15
20
25
30
35
40
http://terraserver-usa.com/image.aspx?Lon=-118.48&Lat=37.82&w=1&ref=G|-118.48,37.82
Internet


NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide&cmd=Search&term=AF25
NCBI
Entrez Nucleotide
PubMed Nucleotide Protein Genome Structure PMC Taxonomy
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard
Display GenBank Show: 20 Send to File Get Subsequence
1: AF251611. Lessingia lemmoni...[gi:12239996]
Links
LOCUS AF251611 628 bp DNA linear PLN 16-AUG-2002
DEFINITION Lessingia lemmonii var. ramulosissima internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence.
ACCESSION AF251611
VERSION AF251611.1 GI:12239996
KEYWORDS .
SOURCE Lessingia lemmonii var. ramulosissima
ORGANISM Lessingia lemmonii var. ramulosissima
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
Internet
Address http://128.32.109.11/upd3/
SM 137 20
SM 139 (peirsonii)
ER15739
SM 164 SM 103
glandulifera (pectinata)
tenuis
SM 111
SM 127
SM 119
lemmonii sp.
lemmonii (ramulosissima)
SM 143
SM 141
ramulosa
SM 107/104
LESSINGIA
Star Tree™ created with Inxight Tree Studi
Current Selection
Specimen SM 127
Jepson Herbarium database: [?]
not in collections - SMASCH
Jepson Floral Guide Entry: [?]
Lessingia lemmonii var. ramulosissima
Cal Photos: [?]
Lessingia lemmonii var. ramulosissima
Specimen collection site:
Lat/Long: 37.82N / -118.48W
MapQuest | Aerial Photo | Topo Map
GenBank Accession numbers: [?]
ITS sequence: AF251611 ETS sequence: AF251669
All specimens
SM 127 - 3.8mm
0
5
10
15
20
25
30
35
40
Internet

Conclusions

- The **PEG** database will add a **P**hylogenetic ontology to existing resources for **E**cology and **G**eography to allow a brand new way of viewing and understanding biological diversity.
- The PEG database will allow a user to query a unified database to search for the clades that occur in a specified geographic region, the morphological, molecular, or ecological variation in specimens distal to a chosen node, or the subset of clades with a particular combination of character states.
- Such a database can be used for such seemingly disparate tasks as herbarium management, geographic inventories, systematic studies, comparative ecology, macroevolutionary comparisons, conservation planning, or identification of unknown plants.