Botany 2022 — 2022-07-26

Integrating Taxonomic Resources for Alaskan Plants

Campbell O. Webb
Stefanie M. Ickert-Bond

University of Alaska Museum of the North
alaskaflora.org

Version 1.0

Behind a checklist or flora

- ...is a comprehensive **integration of existing data**:
  1. Names: accepted names and synonyms according to various prior opinions;
  2. Distribution: occurrence data from specimens and observations;

  (3. Descriptions: character data)

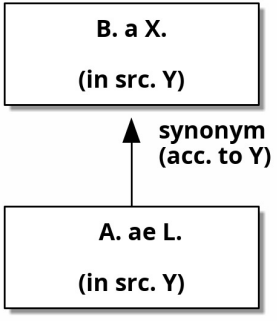
  Creating what we might call the *unchecked-list*.
- Followed by **new taxonomic opinions**

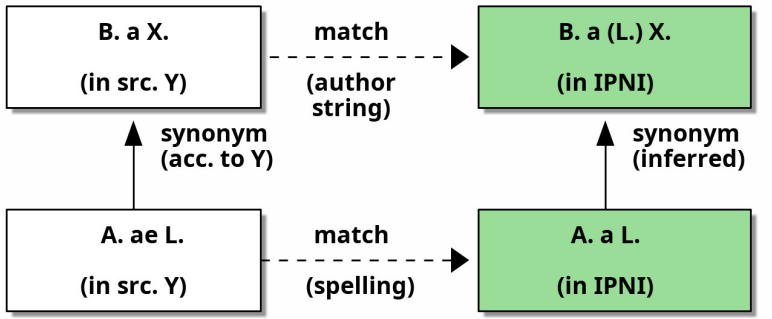
Integration via names

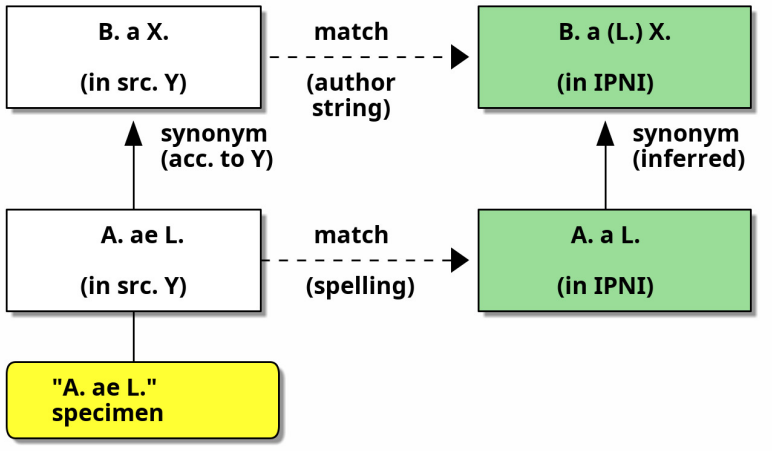
- Patterson et al. (2010): "Names are key to the big new biology" ($\rightarrow$ Global Names Architecture)
- But... **variation in the name string** is common:

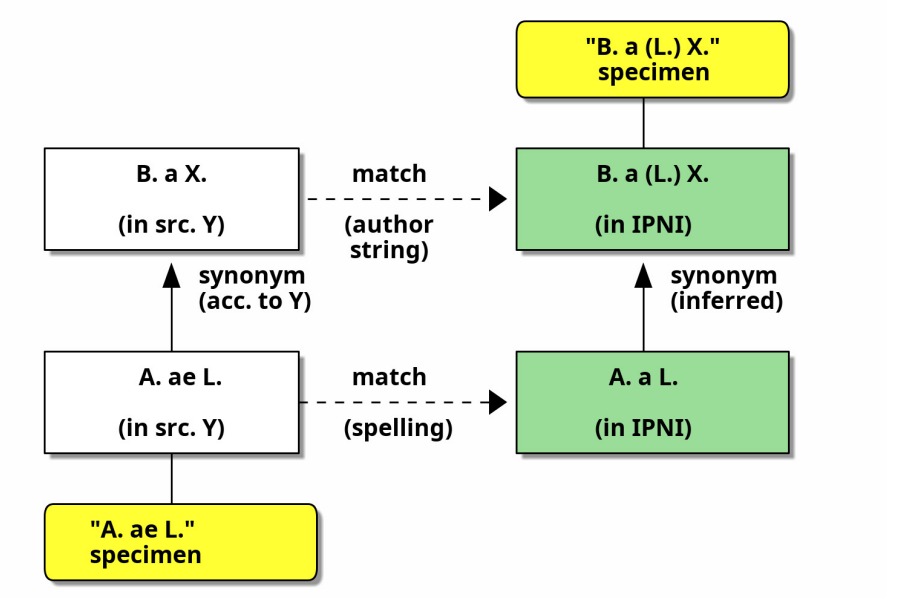
  Ciminalis prostrata (Haenke) Á. Löve & D. Löve
  Ciminalis prostrata (Haenke) A.Love & D.Love
  Ciminalis prostrata Love & Love
  Ciminalis prostratus (Haenke) Á. Löve & D. Löve

- So, **matching** names is also key...

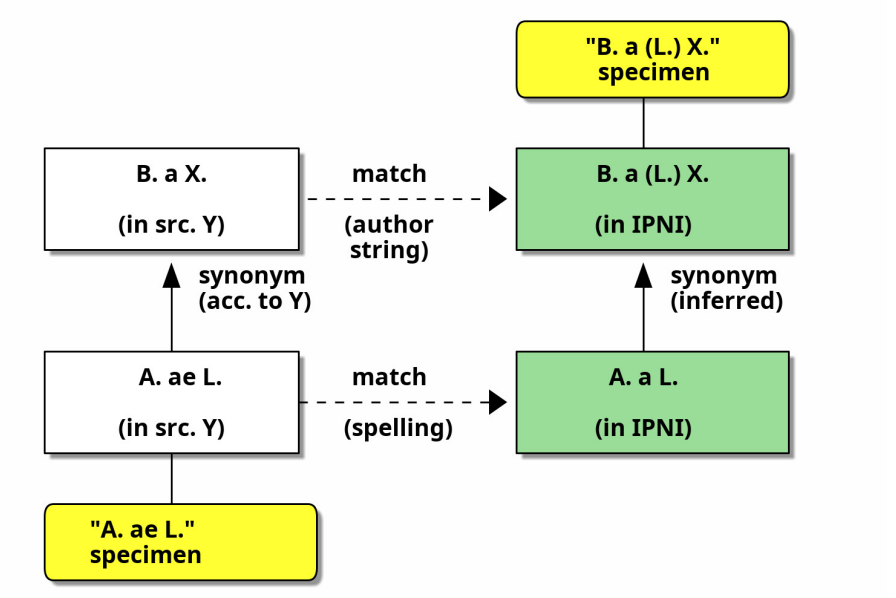

B. a X.
(in src. Y)
synonym
(acc. to Y)
A. ae L.
(in src. Y)


B. a X.
(in src. Y)
match
(author string)
B. a (L.) X.
(in IPNI)
synonym
(acc. to Y)
synonym
(inferred)
A. ae L.
(in src. Y)
match
(spelling)
A. a L.
(in IPNI)


B. a X.
(in src. Y)
match
(author string)
B. a (L.) X.
(in IPNI)
synonym
(acc. to Y)
synonym
(inferred)
A. ae L.
(in src. Y)
match
(spelling)
A. a L.
(in IPNI)
"A. ae L."
specimen


"B. a (L.) X."
specimen
B. a X.
(in src. Y)
match
(author
string)
B. a (L.) X.
(in IPNI)
synonym
(acc. to Y)
synonym
(inferred)
A. ae L.
(in src. Y)
match
(spelling)
A. a L.
(in IPNI)
"A. ae L."
specimen

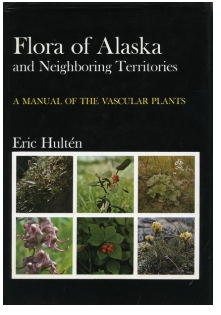
A names *graph*

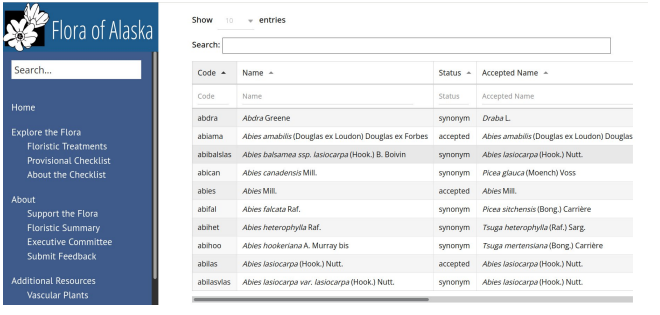


Example graph query: “find all specimens of “B. a (L.) X.”

A new Flora of Alaska

- Time for an update — Hultén's Flora was in 1968
- The (larger) new Flora of Alaska project, https://floraofalaska.org/, since 2018.
- Taxonomic review of checklist almost finished





For our informatics component, thanks to...

- **Funding**: US National Science Foundation, Advances in Biological Informatics Innovation Grant 1759964
- **People**: Flora of Alaska Executive Committee, Nico Franz, Alan Weakley, James Macklin, **Kimberly Cook**
- **Data**: Collectors, flora & monograph authors, iNat observers, data curators, database admins, Ottawa FNA team, Dusty McDonald
- **FOSS**: Linux, Apache, MySQL/MariaDB, Jena, Gawk, ImageMagick, Graphviz, GIMP, ditaa, LaTeX

Data sources

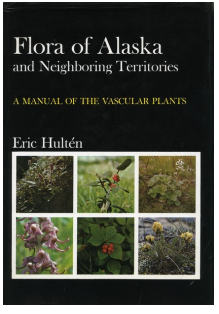
1. External, **"canonical" names**, with GUIDs:
   - Create raw list of all possible names
   - Used Global Names Resolver
   - IPNI > Tropicos > Kew's World Checklist of Selected Plant Families

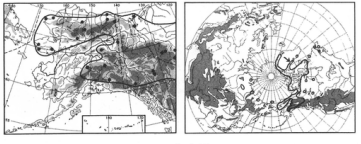
Data sources

1. External, “canonical” names
2. Names in **floras & checklists**
   - Flora of North America (*via* Ottawa team)
   - Previous checklists at ALA, ACCS
   - Panarctic Flora (http://panarcticflora.org/)
   - Hultén’s 1968 Flora of Alaska

Digitizing Hultén's Flora

- Slice up scan with `ImageMagick` (& manual)
- OCR using wordlist including all names
- Distribution dots to long/lat with `ImageMagick`
- View at: https://alaskaflora.org/hulten/





Lesquerella S. Wats.

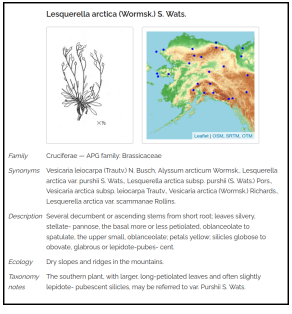
1. Lesquerélla árctica (Wormsk.) S. Wats.

Alyssum arcticum Wormsk.; *Vesicaria arctica* (Wormsk.) Richards.; *V. arctica* subsp. *leiocarpa* Trautv.; *V. leiocarpa* (Trautv.) N. Busch; *Lesquerella arctica* var. *Purshii* S. Wats.; *L. arctica* subsp. *Purshii* (S. Wats.) Pors.; *L. arctica* var. *Scammanae* Rollins.

Several decumbent or ascending stems from short root; leaves silvery, stellate-pannose, the basal more or less petiolated, oblanceolate to spatulate, the upper small, oblanceolate; petals yellow; silicles globose to obovate, glabrous or lepidote-pubescent.

Dry slopes and ridges in the mountains.

The southern plant, with larger, long-petiolated leaves and often slightly lepidote-pubescent silicles, may be referred to var. **Pùrshii** S. Wats.



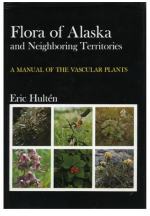
Data sources

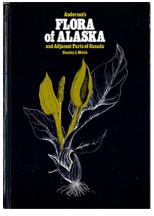
1. External, “canonical” names
2. Names in floras & checklists
3. **Taxonomic concepts**
   - “The problem with names alone”
   - New taxonomic concept mapping work here at the U. Alaska Museum (ALA)

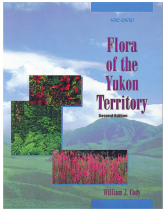
Taxonomic concept mapping

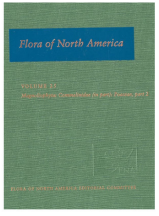
- **Kimberly Cook** aligned concepts in five key Alaskan flora sources: Hultén (1968); Welsh (1974); Cody (2000); FNA Eds. (1993+); Elven et al. (2017)
- TC relationships inferred from taxonomy, morphology, geography (Cook et al., 2021)

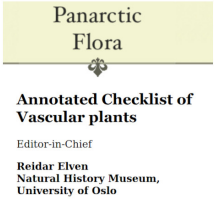
→ 557 taxonomic concepts in 13 genera & 482 taxonomic concept relationships



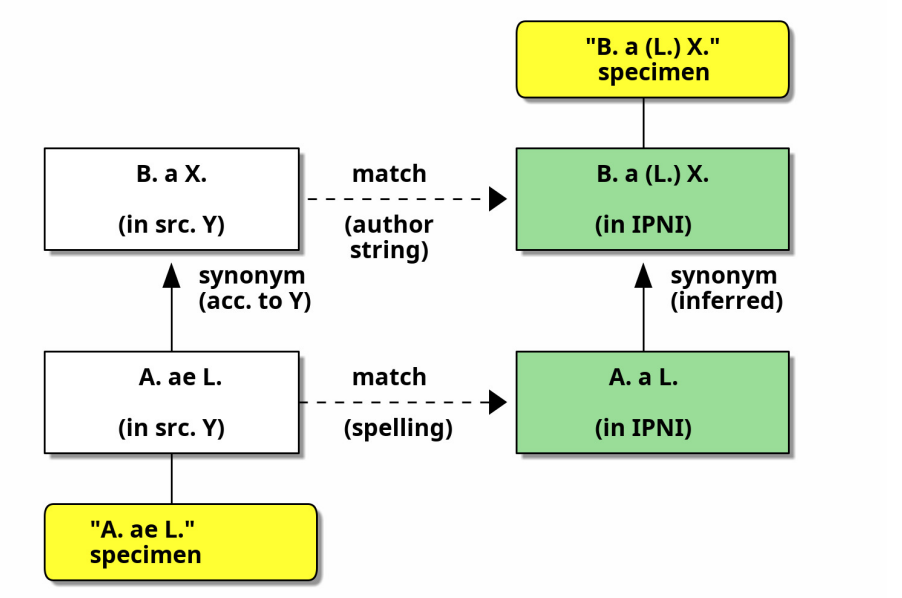




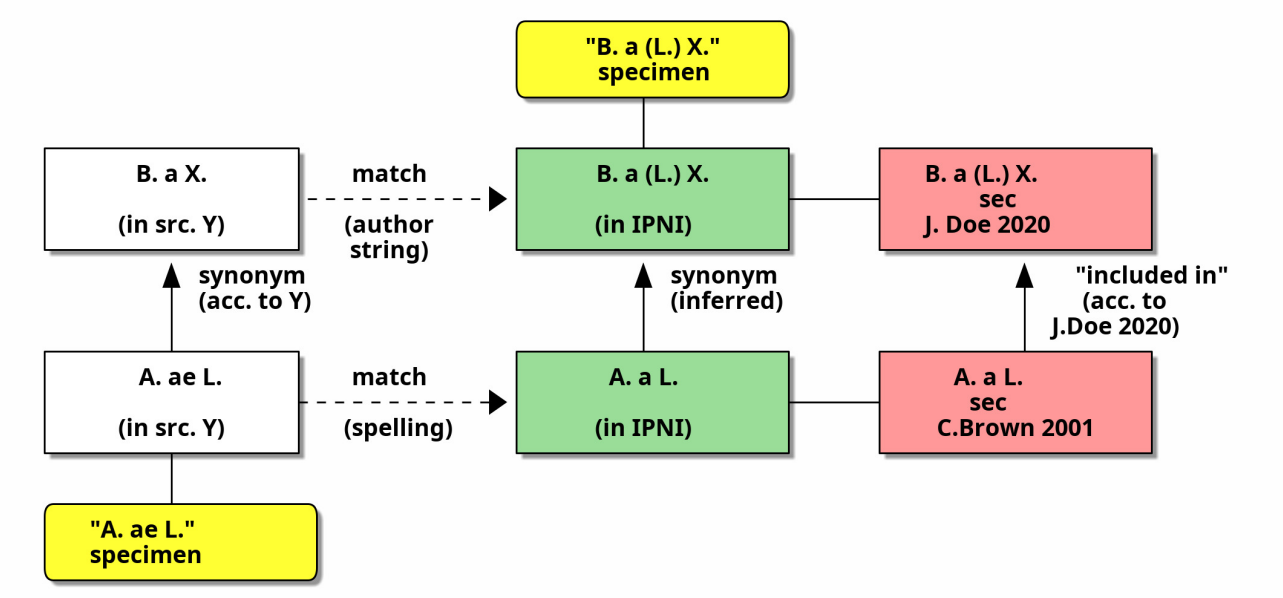




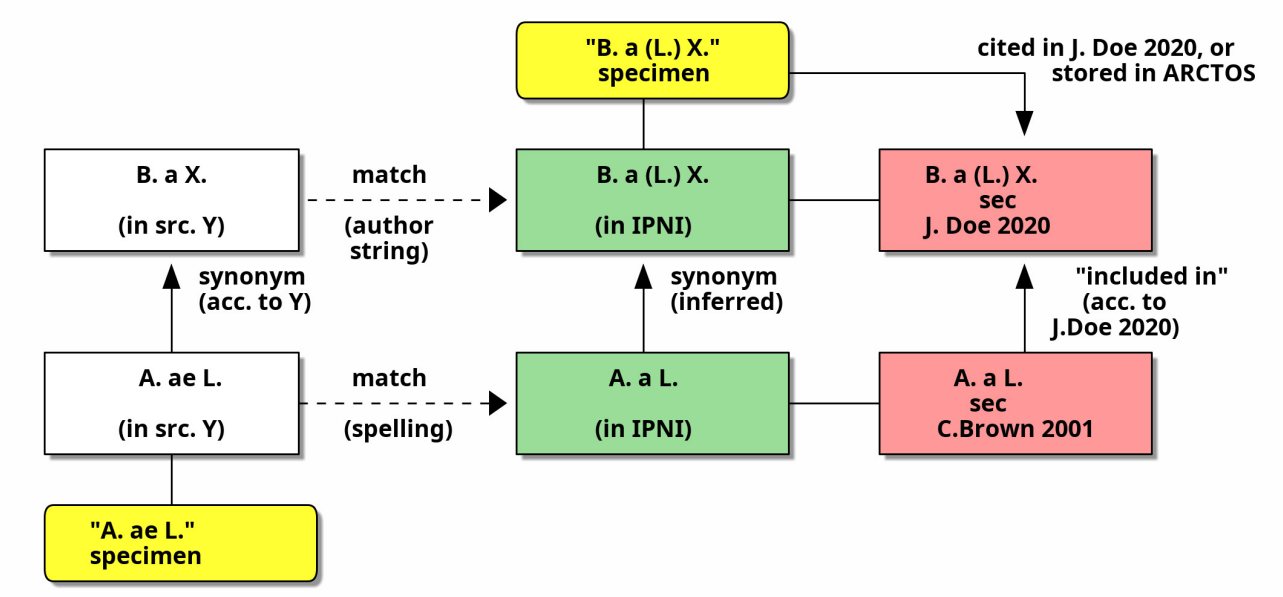
The names graph



The names graph

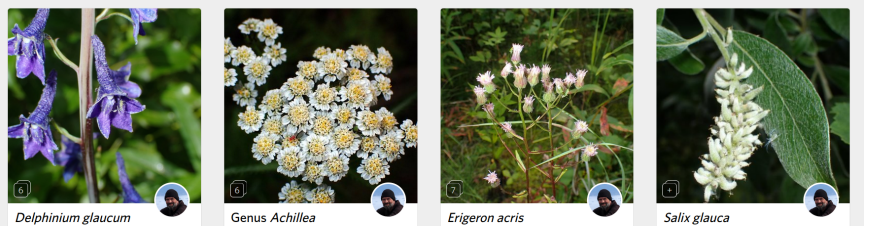


The names graph



Data sources

1. External, “canonical” names
2. Names in floras & checklists
3. Taxonomic concepts
4. **Occurrences**
   - ALA herbarium specimens, GBIF specimens
   - iNaturalist observations



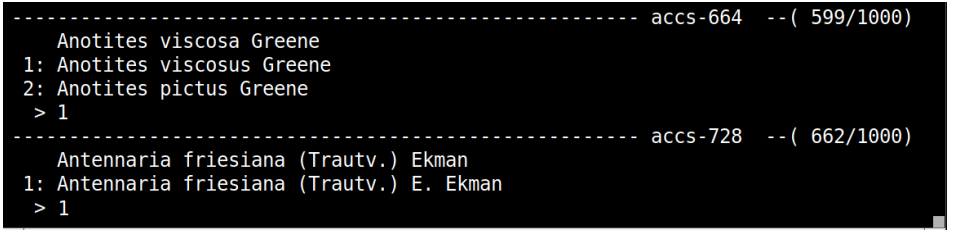
Graph assembly

- Data cleaning, reformatting, filtering scripts
  $\Rightarrow$ Bespoke "data curation"
- All steps recorded in Github and open to review
  https://github.com/camwebb/akflora

Matching names

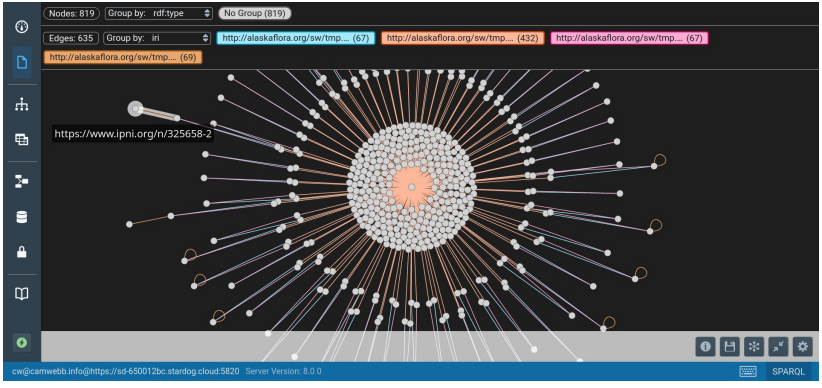
`matchnames` (Webb and Ickert-Bond, 2020)

1. Parse into components
2. Taxonomic logic: missing basionym, *ex*, *in*, etc.
3. De-punctuation matching
4. Fuzzy matching
5. Manual matching:



Graph databasing & query

- Plain text files & Gawk **scripts**
- SQL database (MariaDB) & **SQL**
- Graph databases (Allegrograph, Jena, Stardog) & **SPARQL**



Query via web app

- https://alaskaflora.org/graph/
- Displays name status according to source
- Displays occurrences, both by original name and by synonymized name

Query = *Salix bebbiana*;

Salix bebbiana Sarg. {ipni} ← {accs}, {ala}, {fna}, {gbif}, {hulten}, {paf}
ACCEPTED by ACCS, ALA, FNA, GBIF, PAF
SYNONYM of **Salix depressa subsp. rostrata (Andersson) Hiltonen** {trop} according to HULTEN ← {hulten}

Salix bebbiana var. capreifolia (Fernald) Fernald {trop} ← {accs}, {fna}
SYNONYM of **Salix bebbiana Sarg.** {ipni} according to ACCS ← {accs}, FNA ← {fna}

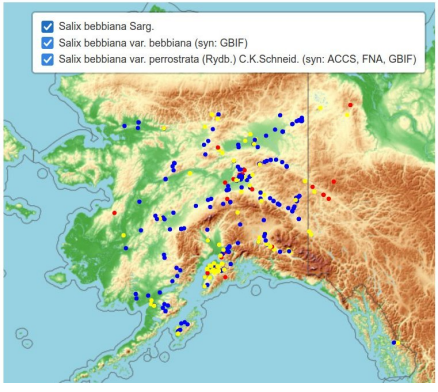
Salix bebbiana var. depilis Raup {ipni} ← {accs}, {fna}
SYNONYM of **Salix bebbiana Sarg.** {ipni} according to ACCS ← {accs}, FNA ← {fna}

Salix bebbiana var. luxurians (Fernald) Fernald {trop} ← {accs}, {fna}
SYNONYM of **Salix bebbiana Sarg.** {ipni} according to ACCS ← {accs}, FNA ← {fna}

Salix bebbiana var. perrostrata (Rydb.) C.K.Schneid. {ipni} ← {accs}, {fna}, {gbif}
SYNONYM of **Salix bebbiana Sarg.** {ipni} according to ACCS ← {accs}, FNA ← {fna}, GBIF ← {gbif}

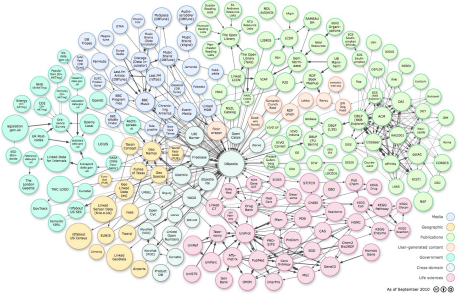
Salix bebbiana var. projecta (Fernald) C.K. Schneid. {trop} ← {accs}, {fna}
SYNONYM of **Salix bebbiana Sarg.** {ipni} according to ACCS ← {accs}, FNA ← {fna}

Salix bebbiana Sarg. {ipni} ← {gbif}



Full graph

- RDF: plain text, XML-based, self-documenting. Directly links into the global Open Data graph.
- 1,243,156 triples: `subject, predicate, object`
- As RDF-XML the file is 78MB (zipped: 5.1 MB)
- Draft version at https://alaskaflora.org, once final also at Zenodo



“Unchecked-lists” are stand-alone resources

- **Integration** of existing taxonomic data (digital and paper) permits complex queries and visualizations
- ... that enable users to see **taxonomic history** and **compare taxonomic opinions**.
- And forms the foundation for new reference checklists and floras.

References

Cody, W. J. 2000. Flora of the Yukon Territory. NRC Research Press, Ottawa.

Cook, K. J., S. M. Ickert-Bond, and C. O. Webb, 2021. Taxonomic concept mapping among historical floras of Alaska: Decision-making and digital implementation. *in* iDigBio 5th Annual Digital Data Conference (virtual). URL https://drive.google.com/file/d/1t1L4-E-0eI-M_fkjOvlZIFVmSJhCWMtS.

Elven, R., D. F. Murray, V. Y. Razzhivin, and B. A. Yurtsev, editors. 2017. Annotated Checklist of the Panarctic Flora (PAF) Vascular plants. URL http://panarcticflora.org/.

FNA Eds., editor. 1993+. Flora of North America North of Mexico. Oxford University Press, New York and Oxford. URL http://www.efloras.org/.

Hultén, E. 1968. Flora of Alaska and Neighboring Territories: a Manual of the Vascular Plants. Stanford University Press, Stanford, CA.

Patterson, D., J. Cooper, P. Kirk, R. Pyle, and D. Remsen. 2010. Names are key to the big new biology. Trends in Ecology & Evolution **25**:686–91. URL https:dx.doi.org/10.1016/j.tree.2010.09.004.

Webb, C. O., and S. Ickert-Bond, 2020. Cyborg matching of taxonomic names, using nomenclatural logic. URL https://camwebb.info/files/webb2020_matchnames.pdf.

Welsh, S. L. 1974. Anderson's Flora of Alaska and adjacent parts of Canada. Brigham Young University Press, Salt Lake City.