On-line Chemical Database for New Crop Seeds

T.P. Abbott*, B.S. Phillips, R.O. Butterfield, T.A. Isbell, and R. Kleiman[1]

New Crops Research, USDA, ARS, NCAUR, Peoria, Illinois 61604

ABSTRACT: Since the 1950s, a database of the chemical composition of seeds, collected throughout the world, has been generated at the National Center for Agricultural Utilization Research. Information contained in the database is primarily selected chemical and physical properties of seed oils. Over more than 38 years, 15,738 accessions of 7,924 species of 2,339 genera from 225 families were collected and analyzed. This database is now accessible on the Internet at http://www.ncaur.usda.gov/nc/ncdb/search.html.ssi. This paper gives an overview of the database, describes the information available, and illustrates how to do searches.
JAOCS 74, 723–726 (1997).

KEY WORDS: Composition, database, germplasm, Internet, oilseed, triglyceride.

The New Crops database on the chemical composition of 15,738 seed accessions has been a resource of the National Center for Agricultural Utilization Research (NCAUR) for a number of years. Although a complete listing of species is not possible, Table 1 shows the 44 families most often represented in the collection and the number of genera in each family in the collection. Much of the data has been published by NCAUR scientists (1–3). Beginning with Fontaine Earle (about 1957), many researchers at the lab have made important contributions to the collection of germplasm for new industrial crops and methods for oilseed analysis. In 1975, the information was incorporated into a database as an inventory and as an archival reference of analytical results, not as a comprehensive source. Inventory information of accessions was entered and those analyses deemed pertinent were added. This feature becomes evident as one searches the database because many sample records in the database do not contain data in all fields. The database is now accessible from the home page of the New Crops Research Unit at NCAUR, Peoria, Illinois, at **http://www.ncaur.usda.gov/nc/ncdb/search.html.ssi.** This is a public access database with no charges. All data were generated by good laboratory practices, and most data were acquired by using AOCS Official Methods. However, as any plant scientist knows, the variation within a species of oil content, protein content, and other components is considerable. The analyses reported in the database are of seeds in our collection, and one can sometimes glean from the database some idea of the variation of oil content and fatty acid composition within a genus or species. Although methods of analysis have changed considerably over the 40 years of analysis, we believe that all values reported are accurate. The gas chromatography (GC) procedures used in the analyses are usually given.

TABLE 1
The 44 Most Frequently Represented Families

Family	No. of genera	Family	No. of genera
Acanthaceae	29	Guttiferae	37
Aceraceae	26	Iridaceae	45
Amaranthaceae	30	Labiatae	319
Amaryllidaceae	27	Leguminosae	1222
Anacardiaceae	53	Liliaceae	167
Apocynaceae	47	Lythraceae	96
Asclepiadaceae	38	Myrtaceae	68
Bignoniaceae	33	Oleaceae	37
Boraginaceae	115	Onagraceae	71
Capparaceae	36	Palmae	38
Caprifoliaceae	37	Polygonaceae	79
Caryophyllaceae	90	Primulaceae	79
Chenopodiaceae	72	Ranunculaceae	120
Compositae	1095	Rosaceae	187
Convolvulaceae	67	Rubiaceae	81
Cruciferae	362	Rutaceae	55
Cucurbitaceae	67	Sapindaceae	50
Cupressaceae	31	Scrophulariaceae	193
Cyperaceae	55	Solanaceae	96
Dipsacaceae	31	Sterculiaceae	38
Euphorbiaceae	181	Umbelliferae	286
Gramineae	354	Verbenaceae	60

SEARCHABLE FIELDS

Field name (underlined) = Definition. Nuno = NCAUR Accession Number assigned when the seed sample was received. Genus = Accession *genus* name. Species = Accession *species* name. Family = Accession family name. Source = Accession source.

Some additional details are needed to clarify the source information: (i) If a country is listed in the source field, the sample was collected in that country, under PL480 grant funds. The number after the country is the accession number from that country, except when PL480 or 480 is present. That prob-

*To whom correspondence should be addressed at New Crops Research, NCAUR, ARS, USDA, 1815 N. University St., Peoria, IL 61604.
E-mail: abbotttp@mail.ncaur.usda.gov.
[1]Retired USDA/ARS. Current address: International Flora Technologies, Ltd., 1151 N. Fiesta Blvd., Gilbert, AZ 85233-2238.

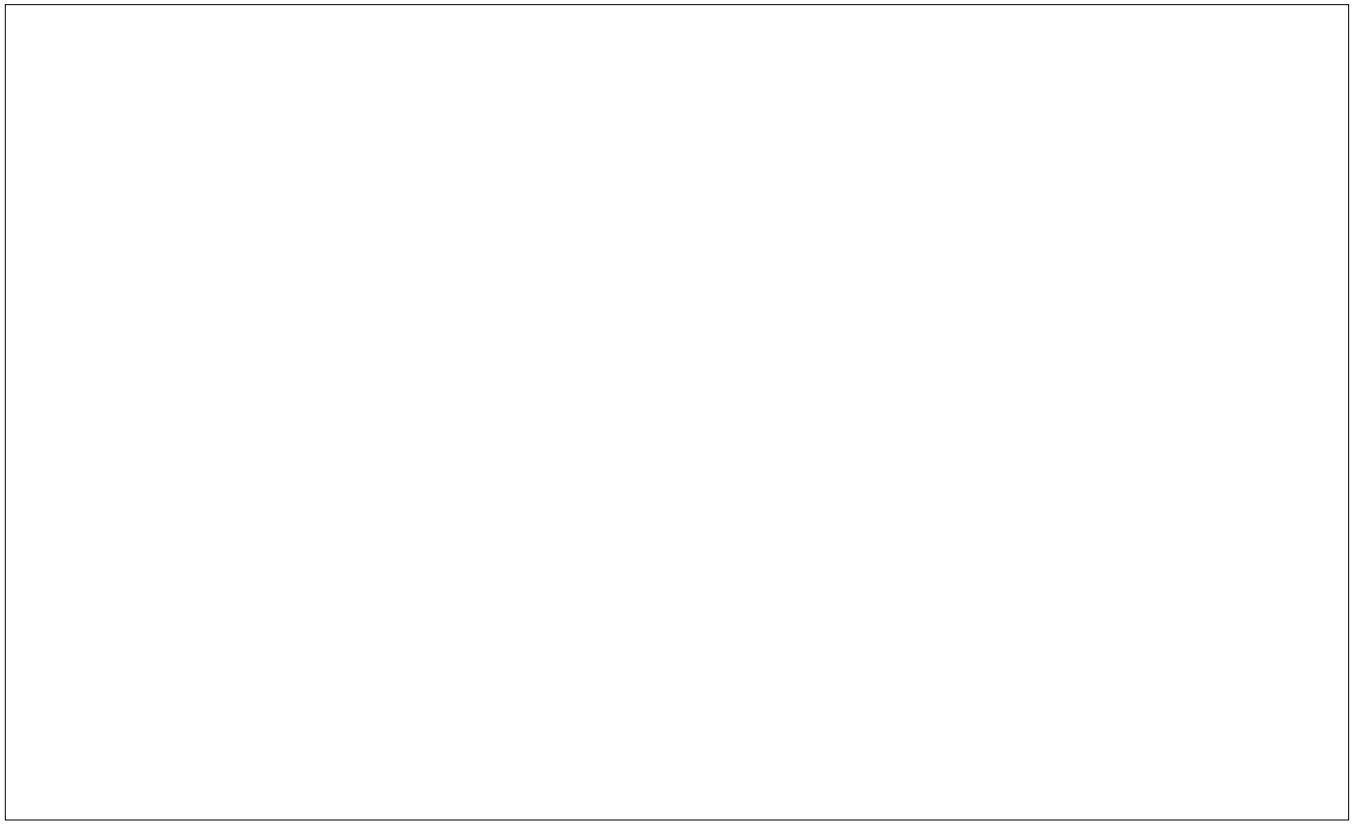
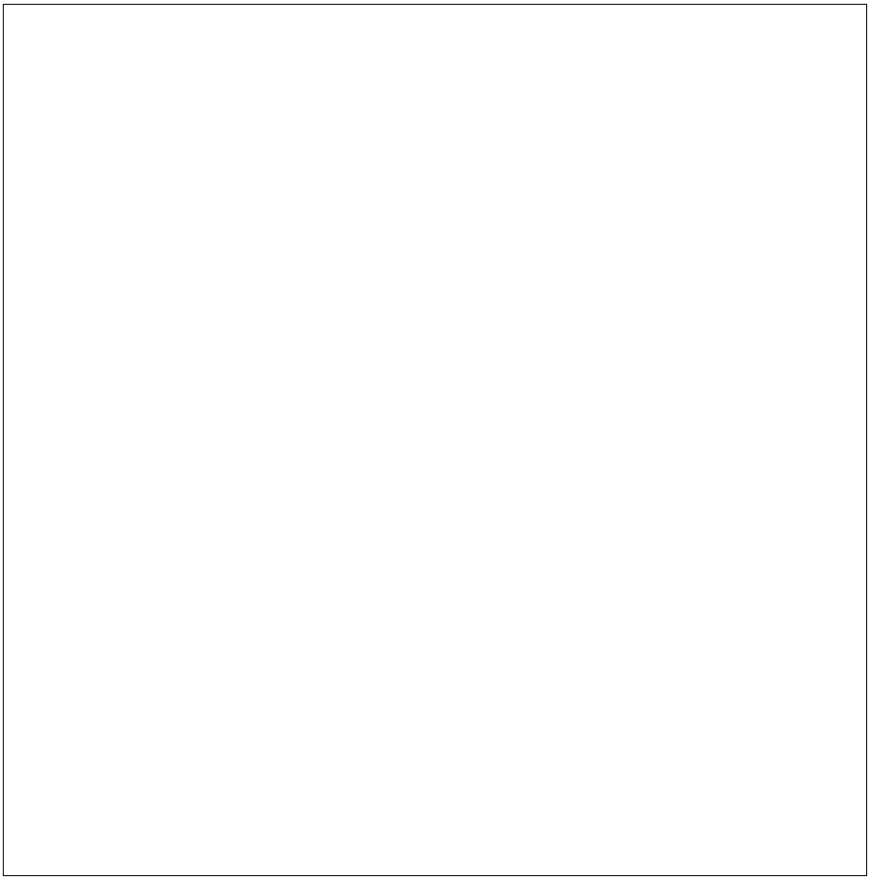


FIG 1. Display resulting from a search of genus: *Limnanthes* and species: *alba.*



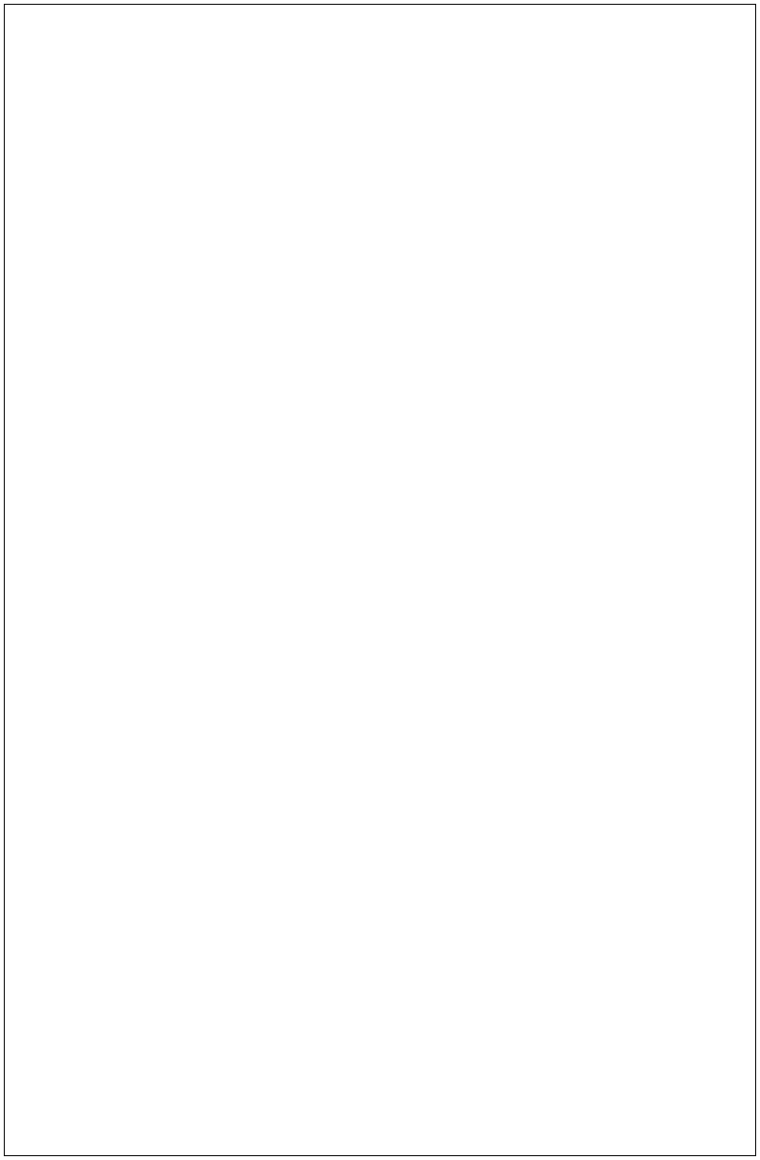


FIG. 2. (A) and (B): Display resulting from selecting accession Nuno 42774 from screen in Figure 1.

ably means "collected under PL480 funding." Hint: A search of "Nuno" records entered into the collection adjacent to the one in question will often reveal a series of samples collected in a group and numbered consecutively, with more complete information about the collection in some records. (ii) If an individual's name is in the source field, the sample was either collected or donated by that individual. The location can mean either where it was collected or the location of that individual. For example, the entry "Gentry" means Howard Scott Gentry, the famous southwestern U.S. collector, and the location provides where the sample was collected; his accession number is usually given. The "Source's Accession #" field can be an aid. Source Accession number entries beginning with the letters "PI" may be searched in the Agricultural Research Service's Germplasm Resources Information Network at www.ars-grin.gov to gain further information or to request seed samples. If it has a PR-#, it was collected by Robert Perdue, a retired USDA employee. He may have further information as to where this material was collected. (iii) Where Irene M. Cull (deceased) is listed as the source, the location where the sample was collected, not where she was located, is shown. Source's Accession # = Identification number assigned to the accession by the source. Comments = General comments about the accession. This field often includes literature citations where the results have been published.

Physical and wet chemical analyses. Wt/1000 = Weight, in grams, per 1000 seeds. % Water = % Moisture in sample. % Nitrogen = Total nitrogen content, %. % Oil = Oil content expressed as a weight percent. % Oil, db = Oil content expressed as a weight percent corrected for moisture (dry basis). Starch = Specific test for presence of starch, positive or negative. HBr = Hydrogen bromide value determined at 55°C. IV Obs. = Iodine value experimentally determined. IV Calc =

Iodine value calculated from GC analysis of the fatty acid methyl esters (FAME). RI = Refractive index of the extracted oil at 40°C.

Infrared and ultraviolet (UV) spectral data. IR code = MB: many bands, not fully interpretable; U: usual bands expected for triglycerides; FA: free fatty acids present. IR1–IR6 = Six significant infrared absorbances of the oil, in microns. UV code = B: high background noise; U: usual bands expected for triglycerides; L2: conjugated diene; L3: conjugated triene; X2: conjugated ene-yne unsaturation.

GC analysis of fatty acid methyl esters of oil. 8:0 to 24:1 = Commonly occurring saturated and unsaturated fatty acids, expressed as a percentage of total chromatographic peak response. The figure 18:2 means the fatty acid has 18 carbons and 2 double bonds, i.e., linoleic acid or other isomers. Isomers are not distinguished unless they are described in the "Ester Comments" field.

Equivalent chainlength of unusual GC peaks. ECL1–ECL 5 = Up to five unidentified GC peaks listed in decreasing order of magnitude with equivalent chainlengths (ECL). ECL are determined for eluting GC peaks on both polar and nonpolar columns by comparison of retention times to saturated FAME retention times as described in a publication by T. Miwa (4). Order of elution is different between polar and nonpolar columns. On polar columns, the nonpolar components elute before the polar components of the same carbon chainlength, whereas the reverse is true for nonpolar columns. Therefore, an ECL must specify whether the value is for a polar or a nonpolar column. The combination of two values, obtained by use of a polar and a nonpolar column, greatly enhances the characterization of most fatty acids. The first ECL1 field was determined on a nonpolar column, and the second ECL1 field was determined on a polar column. The same is true for subsequent ECL# fields. Pct 1 - 5 = Peak magnitude in % of total chromatogramic peak area. Ester Comments = Comments on the structure, location, etc., of the unusual FAME. NPColumn or Pcolumn = Identity of the nonpolar and polar GC capillary columns used in analysis; no entry indicates that data were derived from packed columns with Apiezon L nonpolar coating or LAC-2-R-446 polar coating.

SEARCH INSTRUCTIONS

Entering search criteria. To search the database, click on a field where you would like to enter search criteria, type in the name or number that you would like to search for, and then continue to the end of the fields list to the button labeled "SEARCH." AN =, >, OR < SYMBOL MUST PRECEDE ANY NUMERICAL VALUE ENTERED IN ANY FIELD. Click on the search button, and your search will be processed. Entering capitalized or small letters makes no difference. A good protocol will usually include no more than four search terms. A typical search of *Limnanthes* (Genus) and *alba* (species) resulted in the display shown in Figure 1.

Search results. A list of records from the database that meet the search criteria will display Nuno, Genus, Species, Family, and Source field data (Fig. 1). Clicking on the Nuno of an answer will display the data from all 62 fields for the answer you chose (Fig. 2A and B). If additional information is desired, note the Nuno of your answer and search Nuno's higher and lower for additional information, as described above for the "Source" field. One of the authors, B.S. Phillips, is the database manager and will be in charge of adding new data to the database. He may be reached by e-mail at ncbsp@mail.ncaur.usda.gov.

REFERENCES

1. Earle, F.R., and Q. Jones, Analyses of Seed Samples from 113 Plant Families, *Econ. Botany 16:*440–447 (1962).
2. Kleiman, R., F.R. Earle, I.A. Wolff, and Q. Jones, Search for New Industrial Oils XI. Oils of Boraginaceae, *J. Am. Oil Chem. Soc. 41*:459–460 (1964).
3. Kleiman, R., and G. Spencer, Search for New Industrial Oils XVI. *Umbelliflorae* Seed Oils Rich in Petroselinic Acid, *Ibid. 59*:29–38 (1982).
4. Miwa, T.K., Identification of Peaks in Gas–Liquid Chromatography, *Ibid. 40*:309–313 (1963).

[Received September 10, 1996; accepted March 11, 1997]