P. Bareka, G. P. Mitsainas, T. Constantinidis & G. Kamari

# PhytoKaryon: a karyological database of European and Mediterranean plants

**Abstract**

Bareka, P., Mitsainas, G. P., Constantinidis, Th. & Kamari, G.: PhytoKaryon: a karyological database of European and Mediterranean plants. ─ Flora Mediterranea 18: 109-116. 2008. ─ ISSN 1120-4060.

The remarkable plant diversity that characterizes the Euro-Mediterranean region has been for long the focus of intense scientific interest. Since the data produced daily accumulate at tremendous rates, there emerges the task of organizing them into flexible and widely accessible tools for ongoing research. PhytoKaryon was conceived and built as a database that would contain all available karyological data on plants of the Euro-Mediterranean region. It currently includes more than 45,000 records, based on over 1,200 bibliographical references and relating to about 11,000 plant taxa. This wealth of electronically organized karyological data allows for relevant comparisons among different plant groups and the countries, as here demonstrated by examples of data mining.

*Key words*: karyology, data mining, plant diversity, chromosome number, data entry.

## Introduction

It is generally accepted that the immense knowledge that has been gained on plant diversity of the Euro-Mediterranean region needs to be organized and turned into a flexible and widely accessible tool, to assist ongoing research in this field. Multinational projects have been undertaken in this direction, of which the "Euro+Med Plantbase Project" is one of the most ambitious.

Within the framework of Euro+Med Plantbase, the electronic database, PhytoKaryon, was created, to function as a centralized source of karyological and biosystematic data on plants of the whole Euro-Mediterranean region. In addition, several other karyological databases have been made available during the last years. Most of them focus on specific geographical regions, such as Italy (Garbari 2008), Britain and Ireland (Botanical Society of the British Isles 2008), Catalonia (Cromocat 2008) and Slovakia (Marhold & al. 2008), or on specific plant groups such as the families *Brassicaceae* (Warwick & Al-Shehbaz 2006) and *Asteraceae* (Watanabe 2002) and the genera *Cardamine* (Kučera & al. 2005) and *Hieracium* (Schuhwerk 1996). A world-wide spectrum of plant taxa is covered by the

Index to Plant Chromosome Numbers of the Missouri Botanical Garden, incorporated in the w$^3$TROPICOS database (Missouri Botanical Garden 2008).

The purpose of this work is to present PhytoKaryon, the karyological database that was implemented as part of the Euro+Med PlantBase Project, and to provide examples of the valuable data that can be derived through mining of such a well informed database.

## Database Description

PhytoKaryon was built, using Microsoft® Office Access 2000. It is fed through a single main entry form for karyological data, supported by a series of tables that provide secondary information. For the construction of the database, the model published by Berendsohn & al. (1997) was adhered to. The main data entry form provides for the following data sets, for each karyological record as it appears in the in the original publication:

1. Taxonomy, with 6 data fields, including the name of the taxon as recorded and the name currently accepted as correct, including author citation.
2. Karyology, with 2 data fields: Mitotic or meiotic chromosome number(s), details on karyotype morphology.
3. Specimen information, with 4 data fields: Collector name(s), specimen number, date of collection, and herbarium code according to Holmgren & Holmgren (2007).
4. Information on collecting locality, with 8 data fields: Country name, geographical code (Geocode) and description, locality coordinates, altitude, and habitat description. The geographical codes follow the geographical division of the Euro-Mediterranean region adopted by the Euro+Med PlantBase Project.
5. Reference, with 1 data field: Details of the published source of the record.
6. Notes, with 1 data field: Free text for any additional information (e.g. synonyms, habitat info, etc).

An example of a completed data entry form for a *Bellevalia hyacinthoides* record is given in Fig. 1; the database relationships diagram is shown in Fig. 2.

A special effort has been made to include recorded coordinates of the collecting locality, so that maps showing the distribution of taxa with karyological records can be generated. However, only ca. 20% of the records mention coordinates, which affects negatively the usefulness of such maps, as they are very incomplete.

A variety of published data sources were used, such as peer-reviewed publications, special journal columns reporting karyological data (e.g., Mediterranean Chromosome Number Reports in "Flora Mediterranea"), and chromosome count surveys (e.g., Goldblatt 1981). As a result, PhytoKaryon is currently based on 1,200 bibliographic references and contains ca. 45,000 records that correspond to approximately 11,000 plant taxa of the Euro-Mediterranean region.

## Selected statistical applications of PhytoKaryon

The main advantage of a well-informed database is that it can be instructed to present data in easy-to-understand formats, highlighting trends, indicating research deficits, allow-

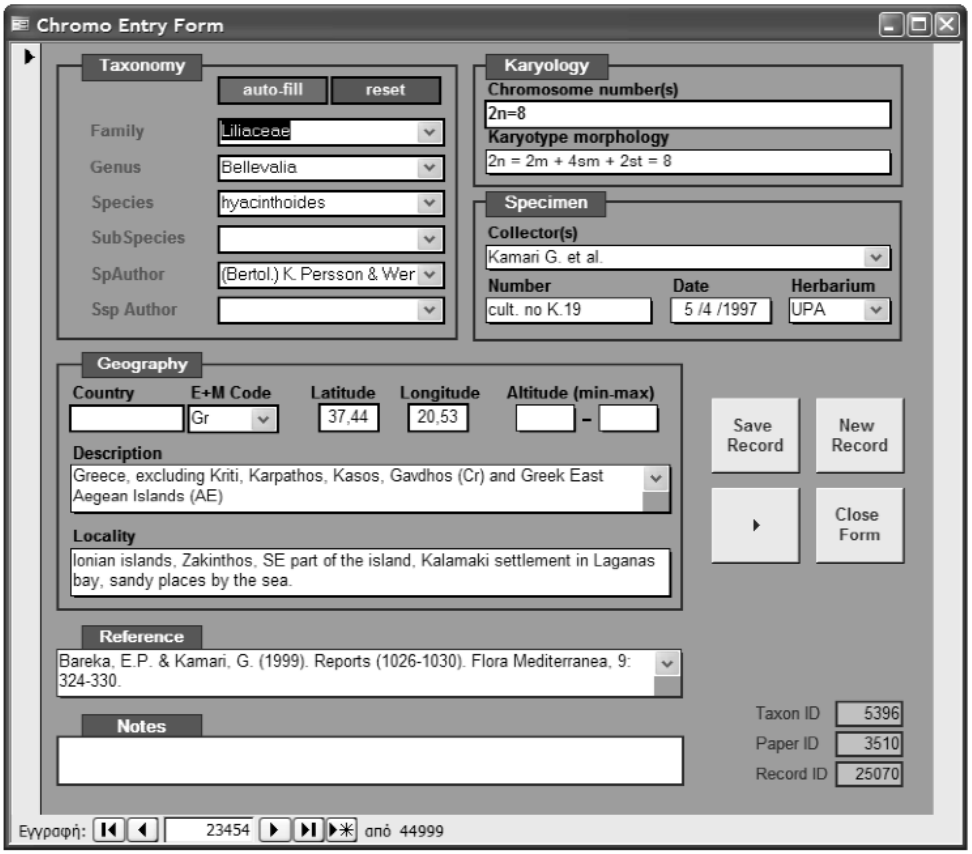


Fig. 1. The PhytoKaryon main data entry form.

ing to draw conclusions, etc. Among other things, PhytoKaryon can demonstrate, through appropriate queries, which plant groups are better studied karyologically than are others, or alternatively, for which Euro-Mediterranean regions or countries the highest number of data has been assembled. Furthermore, a series of correlations can be established, based, for example, on chromosome numbers of specific taxa in different geographical regions.

To judge from the current contents of the database, most karyological studies on plants concern south European countries: Greece, Bulgaria, Italy, Spain, but also some central or extra-European ones: Austria, Turkey, Slovakia (Table 1). Table 2 shows the 20 plant families and 20 plant genera with the highest number of records in PhytoKaryon. *Asteraceae* appears by far as the best studied family, followed by *Liliaceae*, *Fabaceae* and *Brassicaceae*. Among genera, *Allium*, *Campanula* and *Galium* have the highest number of karyological records, for 166, 131 and 71 taxa, respectively.

Using PhytoKaryon to extract data on a family level, one finds that for *Asteraceae* (based on 4,115 records that contain Geocodes), most data concern Austria, followed by

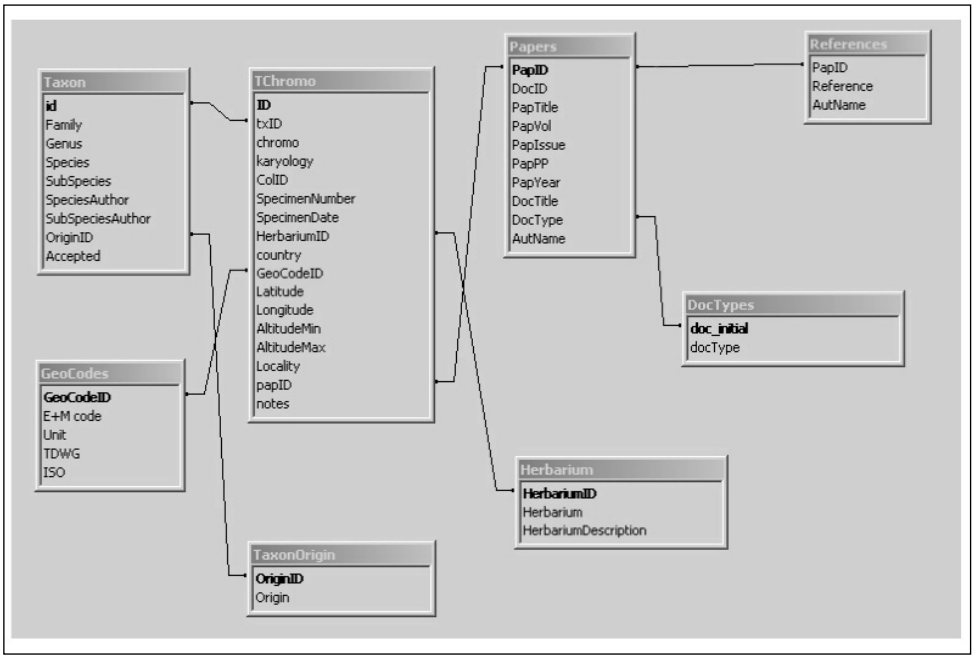


Fig. 2. Diagram showing the relationships among PhytoKaryon's data entry tables.

Table 1. The 20 geographical regions richest in karyological records in PhytoKaryon. Geocode designations correspond to the geographical subdivisions adopted by the Euro+Med Plantbase Project.

| Geocode | Geocode description | No. of records |
|---|---|---|
| Au(A) | Austria | 7,332 |
| Tu(A) + Tu(E) | Asiatic Turkey and Turkey-in-Europe | 3,705 |
| Hs(S) | Peninsular Spain + Mallorca (Balearic Islands) | 3,425 |
| Gr + AE(G) + Cr | Greece, including the Greek East Aegean Islands and Cretan area | 3,281 |
| Bu | Bulgaria | 1,980 |
| It + Si(S) + Sa | Italy, including Sicily and Sardegna | 1,974 |
| Sk | Slovakia | 1,404 |
| Sl | Slovenia | 882 |
| Lu | Portugal | 874 |
| Po | Poland | 816 |
| Ga(F) | Mainland France | 759 |
| Ma | Morocco | 646 |
| Ho | Netherlands | 416 |
| Bl(M) | Balearic Islands: Mallorca | 376 |
| Ct | Croatia | 366 |
| Cs | Czech Republic | 276 |
| Ge | Germany | 252 |
| He | Switzerland (Helvetia) | 220 |
| Cy | Cyprus | 179 |
| Ir | Israel | 171 |

Table 2. The 20 plant families and 20 plant genera richer in records in PhytoKaryon. Percentages are calculated based on the total number of records in the database.

| **Family** | **No. of records (%)** | **Genus** | **No. of records (%)** |
|---|---|---|---|
| *Asteraceae* | 5,865 (13.03%) | *Campanula* | 1,305 (2.90%) |
| *Liliaceae* | 3,525 (7.83%) | *Allium* | 1,117 (2.48%) |
| *Fabaceae* | 3,166 (7.04%) | *Galium* | 922 (2.05%) |
| *Brassicaceae* | 3,135 (6.97%) | *Silene* | 894 (1.99%) |
| *Poaceae* | 2,906 (6.46%) | *Ranunculus* | 844 (1.88%) |
| *Caryophyllaceae* | 2,680 (5.96%) | *Centaurea* | 787 (1.75%) |
| *Lamiaceae* | 2,138 (4.75%) | *Cardamine* | 708 (1.57%) |
| *Ranunculaceae* | 1,557 (3.46%) | *Geranium* | 628 (1.40%) |
| *Apiaceae* | 1,490 (3.31%) | *Ornithogalum* | 563 (1.25%) |
| *Campanulaceae* | 1,488 (3.31%) | *Carex* | 550 (1.22%) |
| *Scrophulariaceae* | 1,299 (2.89%) | *Trifolium* | 488 (1.08%) |
| *Rubiaceae* | 1,102 (2.45%) | *Veronica* | 467 (1.04%) |
| *Boraginaceae* | 1,010 (2.24%) | *Achillea* | 455 (1.01%) |
| *Rosaceae* | 917 (2.04%) | *Vicia* | 438 (0.97%) |
| *Orchidaceae* | 823 (1.83%) | *Viola* | 414 (0.92%) |
| *Geraniaceae* | 769 (1.71%) | *Sedum* | 399 (0.89%) |
| *Cyperaceae* | 699 (1.55%) | *Crocus* | 392 (0.87%) |
| *Iridaceae* | 632 (1.40%) | *Euphorbia* | 382 (0.85%) |
| *Chenopodiaceae* | 533 (1.18%) | *Asphodelus* | 371 (0.82%) |
| *Dipsacaceae* | 510 (1.13%) | *Festuca* | 367 (0.82%) |
| **Total** | **36,244 (80.54%)** | **Total** | **12,491 (27.26%)** |

Portugal, Spain, Bulgaria and Greece (Fig. 3a). For *Fabaceae* (based on 1,926 records with Geocodes) data from Turkey, Spain, Bulgaria and Portugal are most numerous (Fig. 3b). Similar information can be extracted on a genus level, as here shown for *Centaurea* and *Silene*. The former shows particular taxonomic and karyological interest, especially for countries such as Greece, Bulgaria, Croatia, Italy and Spain, where an impressive 75% of the data belong (Fig. 3c), for a total of 200 different taxa represented in the database. With regard to *Silene*, records exist for 156 taxa, and more than 50% of them relate to Greece and Spain (Fig. 3d).

Other kinds of conclusions that can be drawn from the PhytoKaryon data are exemplified by *Crepis* and *Fritillaria*. In *Crepis*, as shown in Fig. 3e, the most common diploid chromosome number is $2n = 8$, recorded for 45 taxa, followed by $2n = 10$ (19 taxa) and $2n = 12$ (11 taxa). Most karyological studies for this genus were made in Greece and Bulgaria. In *Fritillaria*, most of the data come from Greece. The genus shows remarkable stability in chromosome number: in 87% of the records, for 33 taxa, the chromosome number is the same, $2n = 24$, with the occasional appearance of up to 5 B-chromosomes (Fig. 3f).

The above examples show the considerable potential and varied uses of PhytoKaryon, by means of simple or more complex queries yielding diverse results and conclusions. It should be stressed, nevertheless, that the effectiveness and usefulness of any database depend on the amount and quality of its data, when data entry, the most important step, is also the most time consuming and arduous one.

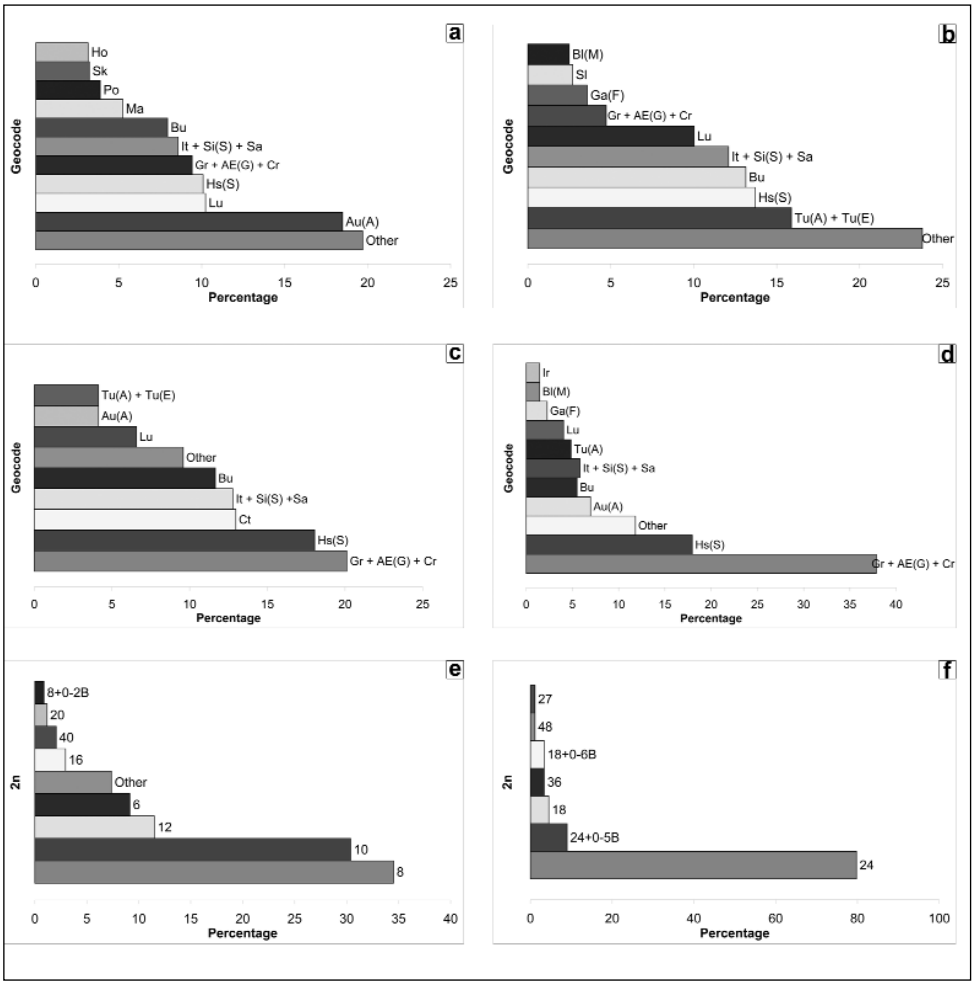


Fig. 3. Diagrams showing the comparative degree of available karyological data on the families *Asteraceae* (a) and *Fabaceae* (b), and on the genera *Centaurea* (c) and *Silene* (d), in the Euro-Mediterranean region, as well as the prevailing diploid chromosome numbers for the genera *Crepis* (e) and *Fritillaria* (f). - Geocode symbols as in Table 1.

## Prospects

One of the main prerequisites of databases, especially larger ones, is that they do not remain static but rather be continuous enlarged and improved in their capabilities. Therefore, several measures are under way, to allow PhytoKaryon to cope with new challenges and satisfy the ever-increasing demands of users. Meanwhile, data entry continues and new relevant karyological information is included as it is published. In addition, the interface for online consultation of the database is being thoroughly redesigned and will

soon be accessible at 'www.PhytoKaryon.biology.upatras.gr'. Through simple queries, it will provide open access to the basic data (taxon, chromosome number, country, reference) of all database records, and controlled access to additional stored information. Also, it will allow accredited editors to remotely update the database by adding new published or otherwise available chromosome data.

In view of the low percentage of records in PhytoKaryon that include geographical coordinates, the possibility to map karyological information at different levels of resolution is being explored. After extraction of the requested data from the database, appropriate software with specific functions, e.g. MATLAB, can be used to produce such maps. Authors of karyological data are strongly encouraged to include coordinates with the indication of locality of origin for their material. At present, only ca. 20% of the karyological records in PhytoKaryon include coordinate information, so that the majority of the data cannot be represented on maps. To overcome this shortcoming, we consider adding coordinates manually for those records that contain detailed indications of collecting locality, but this is a tedious and time-consuming process.

Among our immediate priorities is the addition of visual information to the features that can be accessed online by the users of PhytoKaryon: published photographs of karyograms, idiograms or even metaphase spreads, subject to permission from the copyright holders.

**Acknowledgments**

This work was partially funded by the "Project "EURO+MED Plantbase", (Co EVR1-2000-40004 EURO+MED) and by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Project "New graduate programs of the University of Patras".

**References**

Anonymous 2008: Cromocat (part of the general Catalan biodiversity database BIOCAT). – http://biodiver.bio.ub.es/biocat/homepage.html

— [Botanical Society of the British Isles] 2008: BSBI database – Cytology. – http://rbg-web2.rbge.org.uk/BSBI

— [Missouri Botanical Garden] 2008: w$^3$TROPICOS – Index to plant chromosome numbers data base. – http://mobot.mobot.org/W3T/Search/ipcn.html

Berendsohn, W. G., Greilhuber, J., Anagnostopoulos, A., Bedini, G., Jakupović, J., Nimis, P. L. & Valdés, B. 1997: A comprehensive datamodel for karyological databases. – Pl. Syst. Evol. **205:** 85-98.

Garbari, F. (ed.) 2008: Chromosome numbers for the Italian flora. – http://www. biologia.unipi.it/chrobase

Goldblatt, P. 1981: Index to plant chromosome numbers 1975-1978. – Missouri Botanical Garden.

Holmgren, P. K. & Holmgren, N. H. 2007: Index herbariorum. – http://www.sciweb. nybg.org/science2/IndexHerbariorum.asp

Kučera, J., Valko, . & Marhold, K. 2005: On-line database of the chromosome numbers of the genus *Cardamime (Brassicaceae).* – Biologia, Bratislava **60:** 473-476.

Marhold, K., Mereďa, P. Mráz, P., Hodálová, I., Kolník, M., Kučera, J., Lihová, J., Mártonfi, P., Mrázová, V. & Perný, M. (ed.) 2008: Karyological database of ferns and flowering plants of Slovakia. – http://147.213.100.121/webapp/

Schuhwerk, F. 1996: Published chromosome counts in *Hieracium*. – http://www.botanischestaatssammlung.de/projects/chrzlit.html

Warwick, S. I. & Al-Shehbaz, I. A. 2006: *Brassicaceae*: Chromosome number index and database on CD-Rom. – Pl. Syst. Evol. **259:** 237-248.

Watanabe, W. 2002: Index to chromosome numbers in *Asteraceae*. – http://www-asteraceae.cla.kobe-u.ac.jp/index.html

Address of the authors:
Pepy Bareka & Georgia Kamari, Botanical Institute, Section of Plant Biology, Department of Biology, University of Patras, 265 00 Patras, Greece (Hellas).
E-mails: bareka@upatras.gr; kamari@upatras.gr
George P. Mitsainas, Laboratory of Zoology, Section of Animal Biology, Department of Biology, University of Patras, 265 00 Patras, Greece (Hellas).
E-mail: mitsain@upatras.gr
TTheophanis Constantinidis, Laboratory of Systematic Botany, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece (Hellas). E-mail: const-gr@aua.gr