Joan Simon, Marta Margelí & Cèsar Blanché

CROMOCAT: the Chromosome Database of the Catalan Countries

Abstract

Simon, J., Margelí, M. & Blanché, C.: CROMOCAT: the Chromosome Database of the Catalan Countries. – Bocconea 13: 281-297. 2001. – ISSN 1120-4060.

The structure, contents and current progress of the Chromosome Database of the Catalan Countries are presented. This database, named CROMOCAT, covers a flora of 4,360 taxa of vascular plants on an area of nearly 70,000 square kilometres. The total number of records are 23,610 for the table named CRO-OUT and 2,332 for the table named CRO-IN; the territories with the highest number of chromosome reports are the Balearic Islands and the Pyrenaean provinces.

Introduction

The chromosome atlas of a given country, linked to the general flora of its territory is a very useful tool for documenting infraspecific variation, cytobiogeography, endemism classification, and origin of the flora. As demonstrated on many occasions, such chromosome surveys are helpful for solving ordinary taxonomic problems in the light of biosystematics. In the near future, chromosome databases could be considered as the embryo of information systems documenting genetic biodiversity either as genetic resources of a given region or more precisely identifiable diversity units for conservation policies.

Although many countries have already finished their chromosome atlas (in coordination, or not, with their flora), the Catalan Countries lack this tool. We tried to fill the gap following the project proposed by the OPTIMA Karyosystematics Commission at the Meeting held in Borovec in 1993 (Kamari, 1996). This project provides a common basis for building a network of chromosome databases leading to a future shared Mediterranean Chromosome Database. As stated by Berendsohn & al. (1997), only sporadic efforts have previously been made to make karyological information available in electronic form; for this reason, CROMOCAT should be considered only as a first attempt to implement a karyological database. Opinions, comments and information on similar projects will be appreciated by the authors.

History of Catalan chromosome information - The first report dates back to 1926 (2n=16 for *Diplotaxis erucoides* from Barcelona, in a PhD Thesis of J. Homedes, see Báez 1933); the second one from a review of chromosome numbers of Spanish *Cruciferae* by A. Báez, published in the old prestigious catalan botanical journal *Cavanillesia*, also from *D. erucoides*, collected from Girona (Báez 1993).

The current development of karyology in our country began under the influence of the school of Prof. C. Favarger in Neuchâtel, through the Contributions of A. M. Cauwet (Perpinyà), M. A. Cardona (Menorca) and the visits of Ph. Küpfer to the Pyrenees between the late 60's and the beginning of the 70's. The visit to Barcelona of Contandripoulos (Marseille) at the end of the 70's and the courses by Prof. C. Favarger, Prof. M. A. Cardona and Prof. J. Molero started the formation of new cytotaxonomists who obtained chromosome data for their PhD Thesis. Today, researchers from the whole Iberian Peninsula and from all the world continue to increase the chromosome knowledge of our flora and some new techniques, as chromosome banding or in-situ DNA hybridization are being developed in our laboratories.

Scope and goals

Our chromosome database covers the taxa of vascular plants of the Catalan Countries, including the Regions of Valencia and Catalonia, the Balearic Islands, Andorra and the Northern (French) Catalonia, corresponding to the OPTIMA territories of Hs, Bl and Ga, respectively.

A first, but very important decision, was the taxonomic scheme to be used. The only published complete list of taxa from the Catalan Countries is the "Flora Manual dels Països Catalans" by Bolòs & al. (1993); it was thereby selected for CROMOCAT. This flora was also chosen as a basis for the Chorologic Database (Font 1996), and other projected databases, such as the Ethnobotanical Database designed by J. Vallès (com. pers.) will follow the same model, so that a further integrated system of Plant Information Databases could be implemented in the future.

This Catalan System of Plant Information is embraced in the planning of the Catalan Strategy for Biodiversity Conservation (following the Rio Convention). It will also be at the core of a future Documentation Centre of Plant Biodiversity, which will be organized by the University of Barcelona to coordinate plant information held by Libraries, Herbaria and Databases, and to provide consistent information to researchers, decision makers, administrators and the General Public.

From this starting point, the following goals were designed for CROMOCAT:

- To include chromosome numbers and associated chromosomal data (karyotypes, chromosome formulae, photographs, banding, etc.), i.e. textual and image data.
- To include both complete records of chromosome studies based on populations inside the Catalan Countries, and reference records for our taxa from other areas.
- To provide access to the original publication of the data, as a large amount of complementary information cannot be included in the general cards and to allow for verification of the transcription process.
- To design interfaces for information input directly from the microscope, from CD-ROM, remote databases and on-line sources.
- To provide for further extension of CROMOCAT towards other data such as isozymes or DNA markers.
- To produce a data access interface to the Internet and allow linkage to other databases in a networked system of plant information.

Setup, structure and design

Hardware.- The equipment used consists of a Pentium-class computer connected to a LAN and the Internet, a Hewlett Packard ScanJet 4c/T scanner, two laser printers, an Axiolab E Zeiss microscope, a Hitachi VideoDeck VT-S80E video recorder equipped with an Averkey Plus system and a Sony TV monitor, and a CD-ROM duplicator-recorder Philips CDD-2000 IPW.

Software.- The database was implemented using Microsoft Access 2.0; for image capture and processing the images Corel Photo-Paint 5.0 and Visioner Paper-Port 3.0 were used.

Tables, fields and structure.- As a relational database system, Access allows the building of a system of interrelated tables, which helps to avoid duplicate entries and can be used to facilitate the completion of records in individual tables (Fig. 1).

There are two main and six complementary tables, comprised of major fields defined by the OPTIMA Commission for Karyosystematics (included there in the tables named CHRODATA, CHROTAXON and CHROBIBLIO, Kamari 1996).

A particular effort to avoid mistakes and misinterpretations has been made with the aim to provide an appropriate degree of reliability of the information facilitated by CROMOCAT. Following a recommendation given by A. Strid during the discussion of the OPTIMA database project five years ago, all data have been carefully checked and markers and comments have been included to indicate remaining doubts. With the exception of very few records, all information in CRO-IN have been verified by direct consultation of the original source. In the following, details are given for the fields of the two main tables (CRO-IN and CRO-OUT) as well as for the six complementary ones.

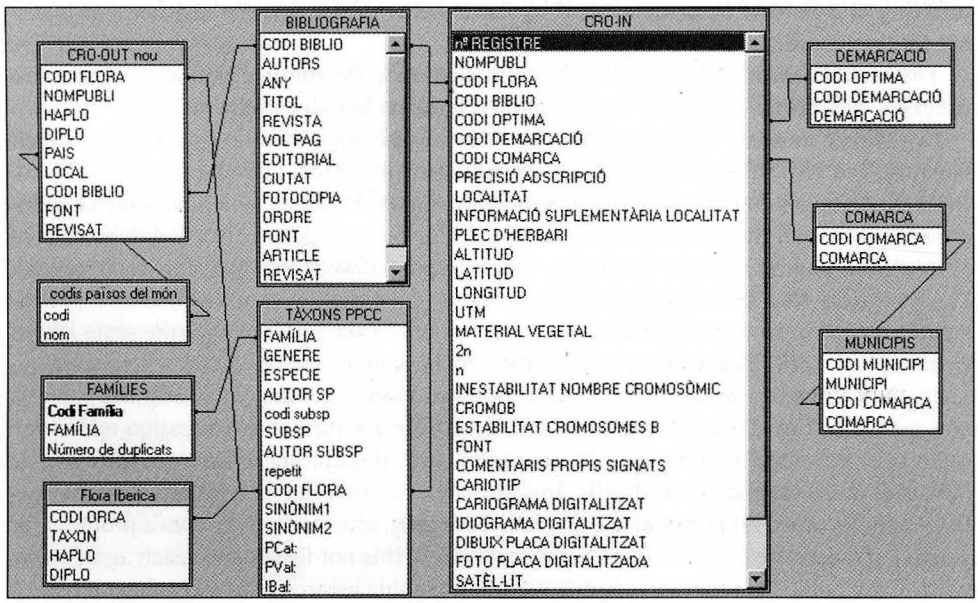

Figure 1.- Table relationships in CROMOCAT.

CRO-IN - This is the longest file, comprising 30 fields of information of any chromosome data from the Catalan Countries (Table 1). All records created from a bibliographic source are linked to the digitalized original document via the table BIBLIOGRAFIA. The registration number (N. registre) is marked by a trailing parenthesis when the original source could not be obtained. The published name (NomPubli) has been corrected for orthography, author abbreviation or other very evident errors, following the recommendations of Uotila & Pellinen (1985). Otherwise, the name is maintained as published. The field CODI FLORA contains the taxon number according to the Flora dels Països Catalans, which is the basis for links to other Catalan databases. To ensure the accurate allocation of names even in case of taxonomic disagreements, some supplementary fields for synonyms have been added to the table "Taxons PPCC"; the names coming from other standard works such as Med-Checklist (Greuter & al. 1984-1989), Flora Iberica (Castroviejo & al. 1986-1997) and Flora Europaea (Tutin & al. 1964-1980). This was of particular importance as a guide for the young people helping us in to enter records. For unresolved problems, a set of markers is available.

With respect to the origin of the populations analyzed, the OPTIMA territories, administrative units (Spanish provinces and French departments) and the natural districts (comarca) are given; doubtful or uncertain ascriptions are indicated by a question mark. The locality has been translated to Catalan, doubtful records again being marked. For the Balearic Islands, the name of each particular island is given in the first place. Fields for ecological details, altitude and geographic coordinates, complete the localization data. A field for a herbarium voucher citation is present, however, this important data item is missing from 32 % of the CRO-IN chromosome reports.

The chromosome number is given, under n or 2n followed by fields for indications of B chromosomes, the chromosome formula, etc., each of which may be marked by an "i" referring to a field of comments typed by the member of the Secretariat who is responsible for the record.

The code given in the source field is removed once the original publication has been checked; it indicates the source in which the report has initially been detected.

Digitalized images.- The graphical data elements are imperative in karyology and their handling is a critical point in the database management system (Berendsohn & al. 1997). In the first version of CROMOCAT, the individual graphics were shown on screen together with the textual items in the database (Fig. 2). To do so, separate files had to be created to store karyograms, plate drawings, photographs, idiograms, etc. for each record. However, this lead to a series of problems of technical nature (a lot of memory is consumed even if links were used instead of inserts) and of scientific nature (the scale length is not always provided with the publication, some errors are produced when scanning images). It forced us to use the currently followed procedure: the entire original paper is now scanned (Fig. 3). This makes the data accessible for further confirmation and at the same time provides the photographs, idiograms, etc. through the PaperPort software, as indicated in the respective field cells. In some countries this can be considered an alternative to photocopies for personal use and thus be legally admitted but it poses a problem for general and open distribution due to copyright rules. It is not feasible to reach agreements with all the publishers for more than 5,300 bibliographic records over more than 50 years; so this problem remains to be solved.

CRO-OUT.- This 9-field table holds chromosome reports on Catalan taxa which have been obtained from populations outside the study area (Table 2). An effort to standardize has been made in accordance with Berendsohn (1998): country abbreviations follow the ISO 3166 standard and, as in the CRO-IN data, BPH is used for periodical abbreviations, Levan & al. (1964) for chromosome formulae, Index Herbariorum (Holmgren & al. 1990) for Herbaria, Grid data follow UTM. CRO-OUT is used as a table of reference for the

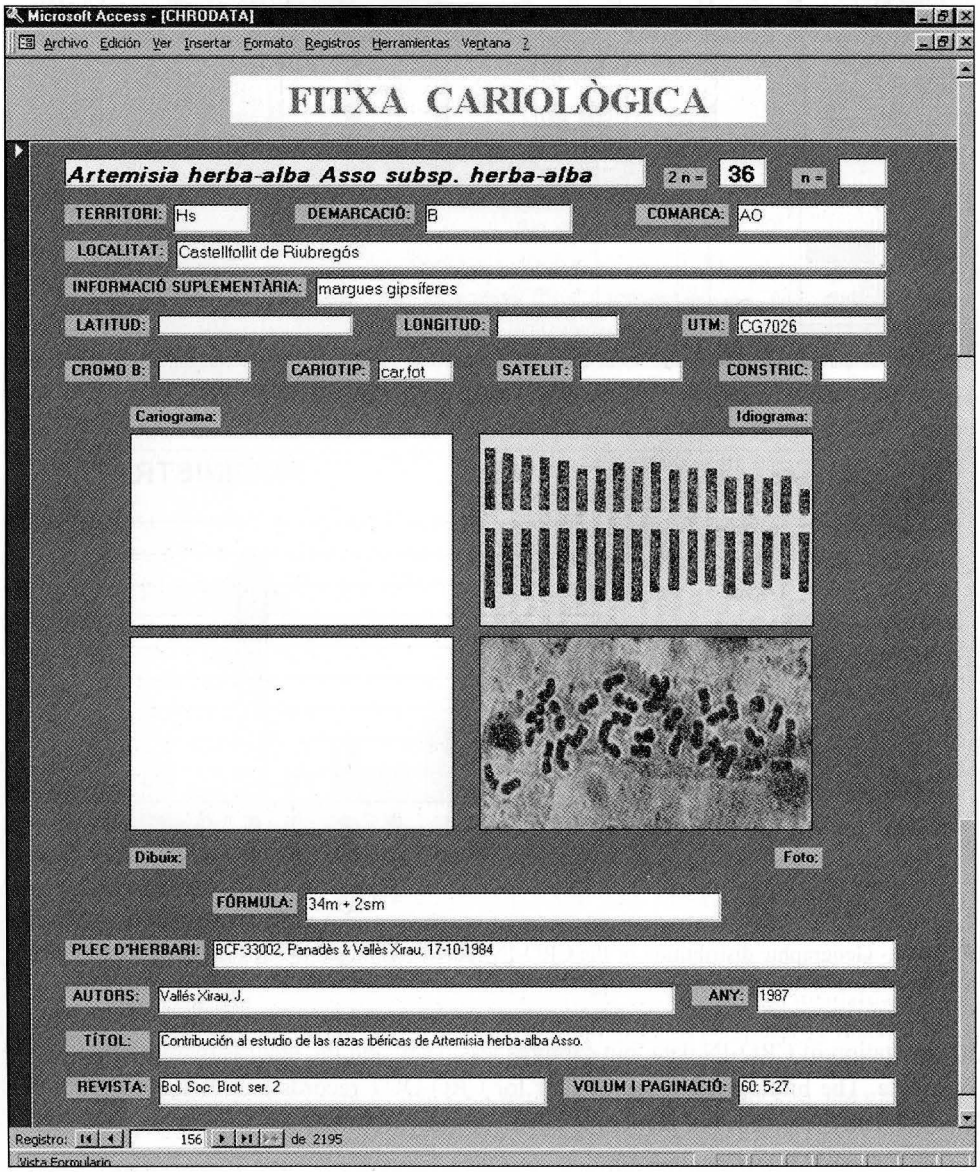

Figure 2.- CROMOCAT output in 1996.

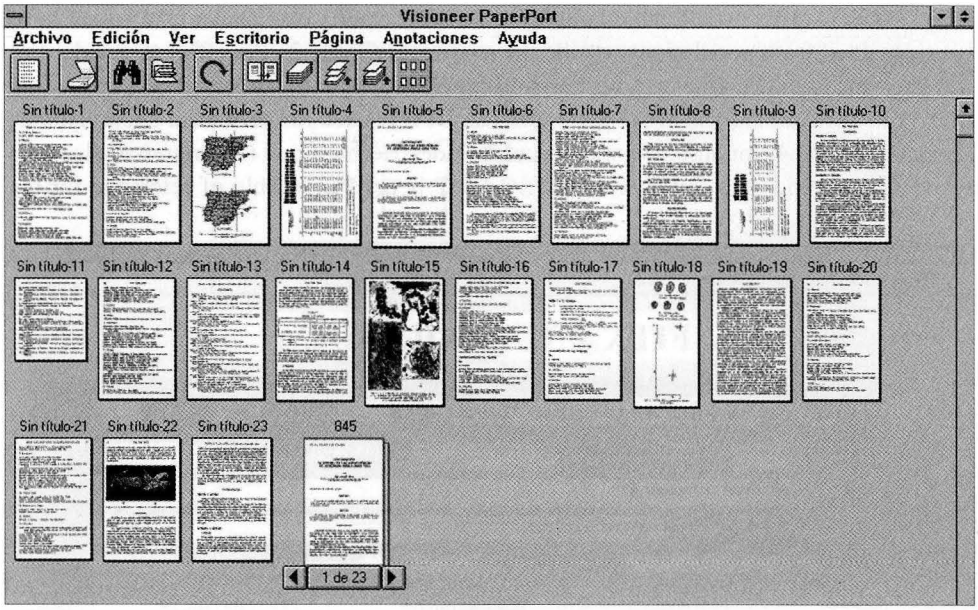

Figure 3.- Example display for CRO-IN digitalized bibliography.

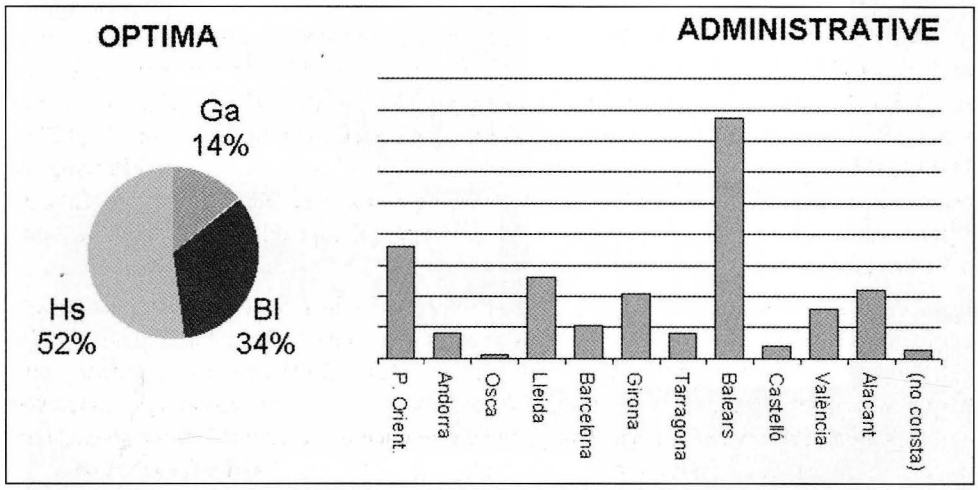

Figure 4.- Geographic distribution of the CRO-IN records, arranged by OPTIMA territories (left) and by administrative demarcations (right).

interpretation of CRO-IN data, but captures less detail and the records themselves are less complete. The bibliographic sources her for CRO-OUT records are photocopied but not electronically scanned. For records in CRO-IN and CRO-OUT, the person responsible for the data entry is noted.

The table BIBLIOGRAFIA. holds standardized bibliographic citations; it corresponds

Table 1.- CRO-IN table structure.

Nombre del campo	Tipo de datos	Descripción
nº REGISTRE	Texto	Número únic que identifica cada fitxa cromosòmica
NOMPUBLI	Texto	Nom del tàxon tal com consta a la publicació original
CODI FLORA	Texto	Codi numèric corresponent al número de tàxon de la Flora Manual dels PPCC
CODI BIBLIO	Texto	Codi que identifica de manera única cada fitxa bibliogràfica
CODI OPTIMA	Texto	Codi Optima segons Med-Checklist i MCNR
CODI DEMARCACIÓ	Texto	Codi demarcació segons delimitació administrativa
CODI COMARCA	Texto	Codi comarca segons tesaure ORCA
PRECISIÓ ADSCRIPCIÓ	Texto	Precisió de l'adscripció comarcal i/o de la demarcació
LOCALITAT	Texto	Transcrita de la publicació, amb ortografia corregida al català
INFORMACIÓ SUPLEMENTÀRIA LOCALITAT	Texto	Altra informació sobre substrat, orientació, localització, etc, citada a la publicació
PLEC D'HERBARI	Texto	Abreviatura Index Herbariorum (si en té) - núm. de plec- col.lector/s - (si consta) -data
ALTITUD	Numérico	Extreta de la publicació en m s.m. Si s'indica una franja, s'ha transcrit la mínima
LATITUD	Texto	Tal com consta a la publicació original
LONGITUD	Texto	Tal com consta a la publicació original
UTM	Texto	Tal com consta a la publicació original
MATERIAL VEGETAL	Texto	Tal com consta a la publicació original
2n	Texto	Nombre cromosòmic somàtic
n	Texto	Nombre cromosòmic gamètic
INESTABILITAT NOMBRE CROMOSÒMIC	Sí/No	Si la publicació ho reporta
CROMOB	Texto	Presència de cromosomes B

Table 2.- CRO-OUT table structure.

Nombre del càmpo	Tipo de datos	Descripción
CODI FLORA	Texto	Codi numèric corresponent al número de tàxon de la Flora Manual dels PPCC
NOMPUBLI	Texto	Nom del tàxon tal com consta a la publicació original
HAPLO	Texto	Nombre cromosòmic gamètic
DIPLO	Texto	Nombre cromosòmic somàtic
PAIS	Texto	Codi del país segons la norma ISO 3616
LOCAL	Texto	Transcrita de la publicació a una unitat administrativa més petita
CODI BIBLIO	Texto	Codi que identifica de manera única cada fitxa bibliogràfica
FONT	Texto	Font de la referència bibliogràfica si aquesta s'ha obtingut indirectament (codi intern a
REVISAT	Texto	Signatura de qui ha complimentat la fitxa cromosòmica

to the OPTIMA table CHROBIBLIO. It also provides the link to the digitalized copy of each paper. TAXONS PPCC provides the taxonomic data ascribed to each record, following Bolòs & al. (1993), and including the main synonyms. FAMÍLIES is linked to the previous table contains family names and their codes according to Bolòs & al. (1993). MUNICIPIS contains a thesaurus of municipalities of the Catalan Countries, following the codification adopted by Font (1996) in his Chorologic Database. Finally, DEMARCACIÓ includes the codes for the administrative and geographical units allowing for queries using different unit schemes.

ISO-IN is a table created recently as a test for the inclusion of isozyme data, which appear increasingly in the botanical literature (Table 3). The table includes fields for the most commonly used diversity parameters (Hamrick & al., 1991), such as: *A* (mean number of alleles per locus), *P* (mean number of polymorphic loci) and *Ho* and *He*, observed and expected heterozygosity respectively, and a reference if the scanned paper gives tables of allellic frequences, zimogrammes, etc. Analogous tables are planned for DNA data.

Current state of the system

Although the CROMOCAT system is open to direct microscope image inputs, as well as to Internet or other electronic connections, actually the chromosome data in the system has the following sources:

- Published papers available in the Libraries of Barcelona
- PhD. Thesis and Master Thesis

Table 3.- ISO-IN table designed to include isozyme records.

n°	NOMPUBLI	CODI	COL	CC	CO	COI	ALTI	UTM	A	P	Ho	He	TAU	ZIM
1	Delphinium montanum DC.	0118a	1647	Ga	P.	CAY	1980	31TDH20	2,5	100	0,263	0,307	Si	No
2	Delphinium montanum DC.	0118a	1647	Hs	L	BY	2350	31TCG98	2,8	87,5	0,291	0,389	Si	Si
3	Delphinium bolosii C. Blanché et J.	0119b	1647	Hs	L	NO	290	31TCG34	1,8	61,5	0,141	0,157	Si	No
4	Delphinium bolosii C. Blanché et J.	0119b	1647	Hs	T	PR	600	31TCF27	1,7	53,8	0,173	0,183	Si	Si
5	Delphinium verdunense Balb.	0115a	1647	Hs	B	AD	196	31TDF09	1,8	36,4	0,147	0,172	Si	No
6	Delphinium verdunense Balb.	0115a	1647	Hs	L	AB	1100	31TCH10	2,5	45,5	0,161	0,238	Si	No
7	Delphinium gracile DC.	0115b	1647	Hs	Os	BI	300	31TBF69	1,7	45,5	0,212	0,162	Si	No
8	Delphinium gracile DC.	0115b	1647	Hs	A	MA	100	31SBC48	1,7	36,4	0,182	0,147	Si	No
9	Delphinium gracile DC.	0115b	1647	Hs	A	MA	200	31SBC48	1,9	54,5	0,201	0,168	Si	No

- Books, Chromosome Atlases, etc, paper format
- Chromosome Atlases in electronic format (Jardín Botánico de Madrid, Universidad de Sevilla)
- Chromosome Files of M. Àngels Cardona (Universitat Autònoma de Barcelona)
- Unpublished data (from some PhD Thesis they are not cited and need the author's permission to be consulted)

Outputs from CROMOCAT is currently obtained by way of queries in MS Access 2.0 or 97. A Internet version of CROMOCAT has not yet been produced, but e-mail consultations to the author's addresses will be answered. Further development of CROMOCAT can be followed at our web site (http://www.ub.es/botanica/greb/cromocat.htm).

Data summary.- From the present degree of development of CROMOCAT, some remarks and data on its contents could be given.

Summary of information contents.- The Catalan flora contains 149 families including 935 genera and 3,580 species, and including subspecies, the number of taxa of the vascular flora is 4,360. The total surface of the area covered by CROMOCAT is nearly 70,000 square kilometres.

The total number of records is 2,332 for CRO-IN and 23,610 for CRO-OUT. The number of bibliographic records stored at present is 5,348. The number of taxa of the Flora of the Catalan Countries from which CRO-OUT gives information is 3,413. This means that 64.5% of the total taxa have been studied elsewhere. Currently, the main sources of data for CRO-OUT are from Mediterranean countries and from eastern Europe where extensive chromosome checklists exist.

The 2,332 chromosome cards of CRO-IN cover only 18.4 % of the total flora and belong to 43.6 % of the total number of families. To convey an idea of the completeness of the data it should be noted that 24.7 % of the reports include plate drawings or photographs and only 4.1% give the chromosome formula. This means that the majority of records still gives only the chromosome number. 22.9 % of CRO-IN cards are n reports whereas 77.1% are 2n reports.

A first geographical analysis of the CRO-IN data from CROMOCAT shows that 52% of the total records come from the OPTIMA territory of Hs, 34% from Bl and 14% from Gal. Grouping these data by administrative demarcations, the Balearic Islands show the highest number of chromosome reports, followed by the Pyrénées Orientales and the Lleida province. There is a small number of populations of doubtful adscription, usually

located in the boundary between two demarcations (Fig. 4). By comparing the number of chromosome reports by comarques (districts), there are six comarques at the bottom, with only 1 record, all six being inland or plain areas, whereas the 5 top comarques are headed by the Mallorca and Menorca Islands, and the 3 Pyrenaean ones. This indicates that the efforts of cytotaxonomists have stressed on such high diversity areas, and that the traces of the intense job of some OPTIMA members, particularly of the late and esteemed M. Àngels Cardona are mirrored in CROMOCAT data.

Development of Cromocat

After a first phase of design, the phase of database implementation started in 1996, and a Secretariat composed of Maria Bigordà, Marta Margelí, Míriam Galisteo began to introduce chromosome data, mainly from literature. This was helped by the indexes produced by the Real Jardín Botánico de Madrid and the Universitay of Sevilla, to which we are indebted. The Secretariat is based at the Faculty of Pharmacy, University of Barcelona, and incorporated in the Research Group on Plant Biodiversity and Biosystematics (GReB). A third phase of implementation is currently in progress, including the introduction of data from selected chromosome databases, e.g. the data gathered by Joan Vallès and Montse Torrell from the genus *Artemisia* or the personal files of M.Àngels Cardona, deposited at the Universitat Autònoma de Barcelona and gathered by Isabel Moya.

A next step to be fasten is the organization of a Scientific Committee to ensure the quality of the information contained in the database and to guide further development, such as making the information available through the Internet, linking to other information systems, designing chromosome research projects in those taxonomic groups or regions in which a low level of cytotaxonomic knowledge is present, and, last but not least, obtaining a solid base of financial support the relatively low pace of CROMOCAT growth can be explained by its dependence on private initiative.

Acknowledgements:

To A. M. Cauwet, X. Font, J. Molero, J. Vallès and C. Benedí, for their technical advice and suggestions. We also thank the Real Jardín Botánico de Madrid and G. Nieto Feliner, as well as the Departamento de Botánica of the Universidad de Sevilla and J. Pastor, respectively, for kindly allowing us to use their computer files as reference indexes for our work. To the anonymous reviewer of the manuscript who highly improved its content. This work has been partly supported by Grant AMB- 97.0375 from CICYT, Spain.

References

Báez, A. 1933: Estudio cariológico de algunas crucíferas y su interpretación en la sistemática. — Cavanillesia **6**: 59-103.

Berendsohn, W. G. 1998: Standards, Information Models, and Data Dictionaries for Biological Collections. International Working Group on Taxonomic Databases and BIOCISE. In: http://www.bgbm.fu-berlin.de/TDWG/acc/Reference.htm

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

Bolòs, O. de, Vigo, J., Masalles, R.M. & Ninot, J. 1993.: Flora Manual dels Països Catalans (2nd. ed.) — Barcelona.

Castroviejo, S., Laínz, M., López González, G., Montserrat, P., Muñoz Garmendía, F., Paiva, J. & Villar, L. (ed.) 1986-1997: Flora Iberica, ***1-7*** — Madrid.

Font, X. 1996: Els bancs de dades de la Flora i la Vegetació de Catalunya - IVth Conference on Plant Taxonomy Abstracts Book: 60. — Barcelona.

Greuter, W., Burdet, H. M. & Long, G. (ed.) 1984-1989: Med-Checklist, **1-2, 4**. — Genève & Berlin

Hamrick, J. L., Godt, M. J. W., Murawski, D. A., & Loveless, M. D. 1991: Correlations between species traits and allozyme diversity: implications for conservation biology.- Pp: 75-86 in: Falk, D.A. & Holsinger, K. E. (ed.), Genetics and Conservation of rare plants. — Oxford.

Kamari, G. 1996: Report of the OPTIMA Comission for Karyosystematics. In: http://www.bgbm.fu-berlin.de/OPTIMA/op27in.htm

Levan, A., Fredga, K. & Sandberg, A. 1964: Nomenclature for centromeric position on chromosomes. — Hereditas **52**: 201-220

Tutin, T. G., Heywood, V. H., Burges, N. A., Valentine, D. H., Walters, J. M. & Webb, D. A. (ed.): 1964-1980: Flora Europaea, **1-5** — Cambridge.

Uotila, P. & Pellinen, K. 1985: Chromosome numbers in vascular plants from Finland — Acta Bot. Fenn. **130**: 1-37

Address of the authors:
Joan Simon, Marta Margelí & Cèsar Blanché, GReB, Laboratori de Botànica, Facultat de Farmàcia, Universitat de Barcelona. Av. Joan XXIII s/n, E-08028 Barcelona, Catalonia (Spain). E-mail: simon@farmacia.far.ub.es, blanche@farmacia.far.ub.es & mmargeli@pie.xtec.es