VI Simposio SEBOT
Mieres - Asturias 28-30 Noviembre 2025

Universidad de Oviedo

SEBOT
Sociedad Botánica Española

Cytogenetic and cytogenomic databases in plants: practical utility in the genomics era

Why evolutionary genomics, taxonomy, systematics… still need chromosomes?

Sònia Garcia

institut botànic de Barcelona
Centre mixt
CSIC
museu de ciències naturals de Barcelona

Why evolutionary genomics needs cytogenomics

The gap in evolutionary genomics

Genome sequences are exploding in number, but:

- **Repetitive regions often poorly assembled** or collapsed (rDNA, heterochromatin. TE-rich)
- **Sex chromosomes and supernumerary elements** are frequently under-characterized

Many comparative genomic analyses ignore chromosome-level traits

To interpret genome evolution we need standardized, curated cytogenetic metadata

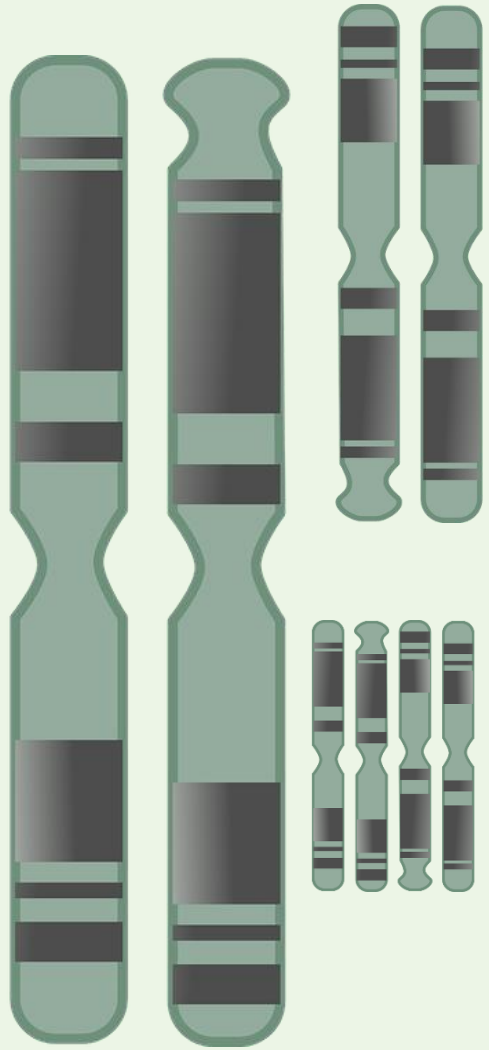
Cytogenetic traits as evolutionary characters

Examples of **cytogenetic traits:**

- Number and chromosomal position of 35S/5S rDNA
- Presence, frequency and type of B chromosomes
- Sex chromosome type (XY, ZW, UV, homomorphic vs heteromorphic)

<u>These traits:</u>

- **Are easily mappable on phylogenies**
- **Can be correlated with life-history, ecology, genome size, polyploidy…**
- **Provide independent axes of genomic variation beyond sequence**



VI SIMPOSIO SEBOT
MIERES ASTURIAS 28-30 NOVIEMBRE 2025

SEBOT
SOCIEDAD BOTÁNICA ESPAÑOLA

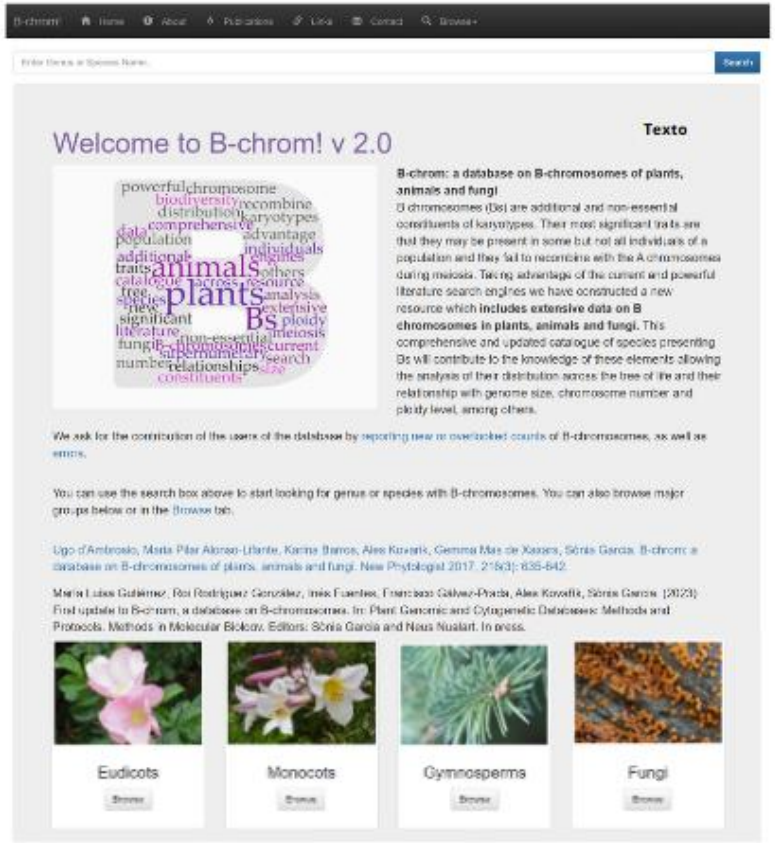
Our three databases: three complementary resources

1 Plant rDNA database

https://www.plantrdnadatabase.com

2 B-chrom

https://www.bchrom.csic.es

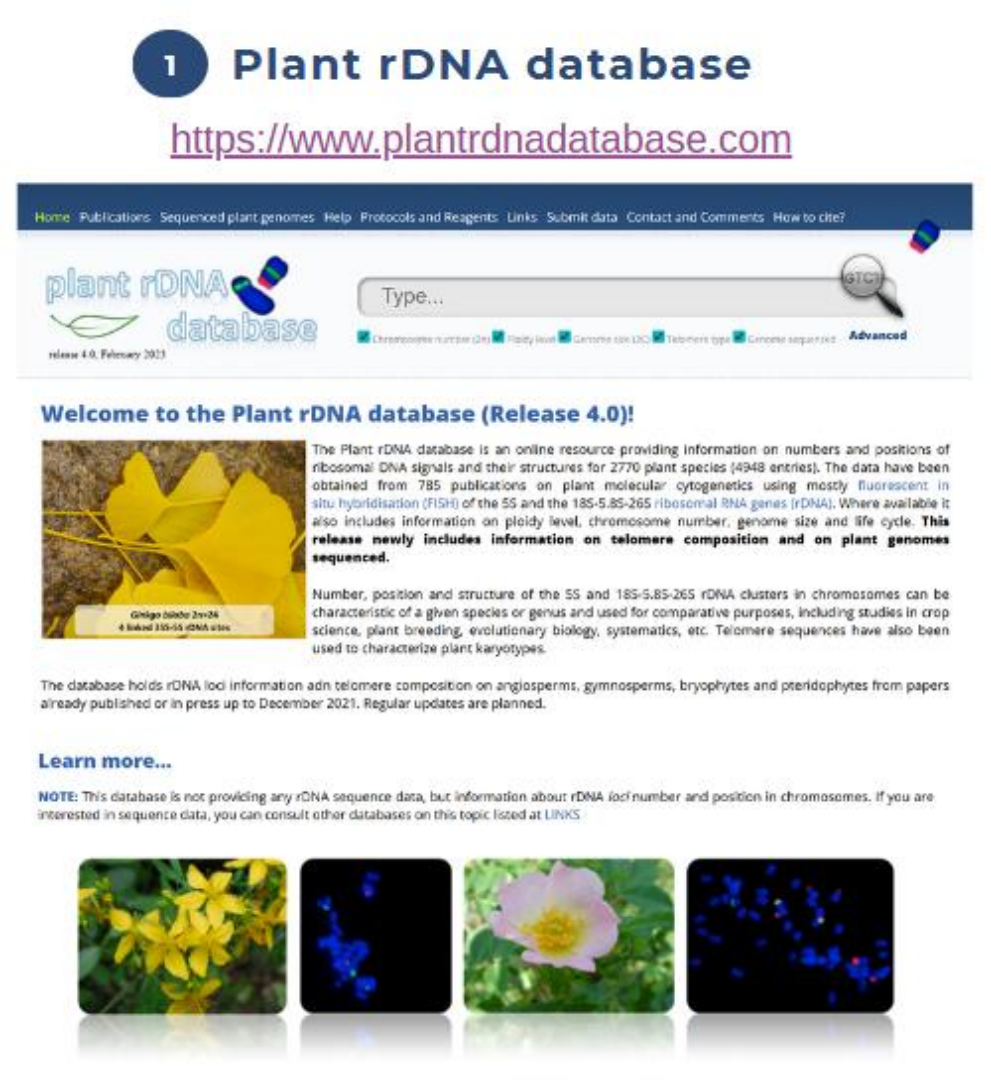


3 Sex-chrom

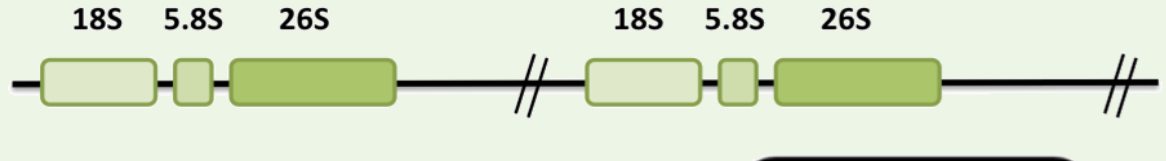
https://www.sexchrom.csic.es

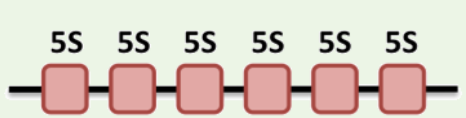
Together they cover major aspects of genome architecture and sexual systems in plants

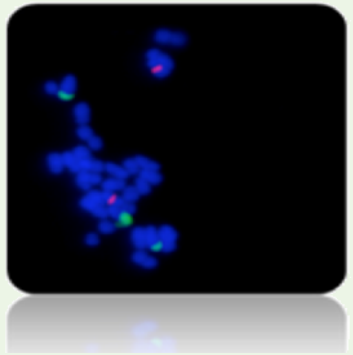
The plant rDNA database: why it exists



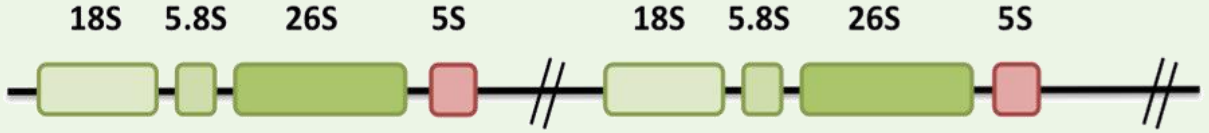
- **Original motivation**
  - Looking for unusual configurations of rDNA (L-type and S-type):

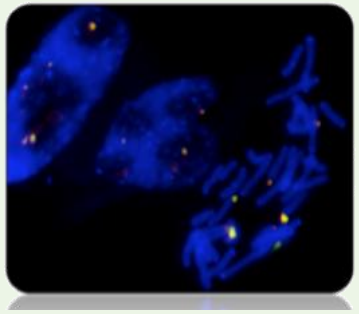






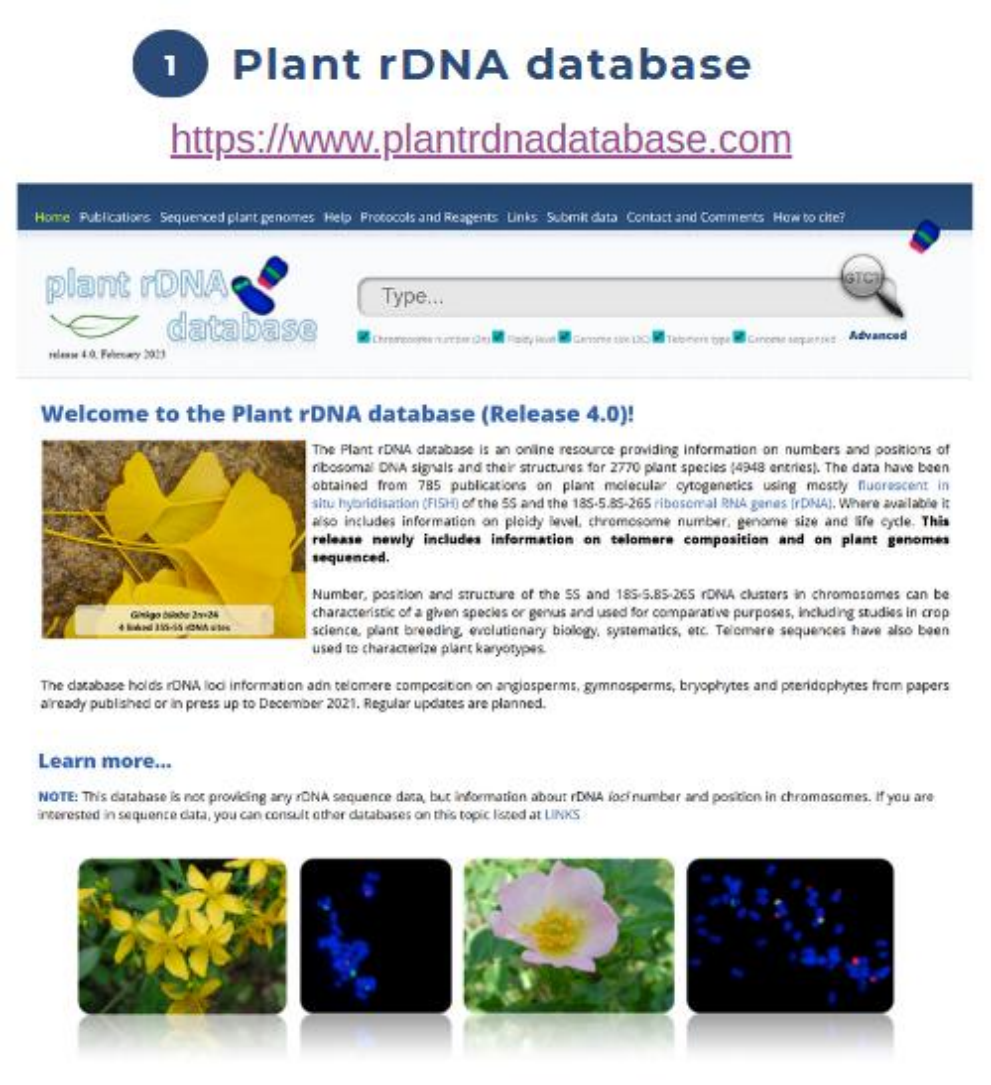
1. Separated rDNA configuration (S-type)





2. Linked rDNA configuration (L-type)

The plant rDNA database: contents, scope, evolutionary questions...



Plant rDNA database: content & scope

- Variables:
  - Species, 2n, ploidy level, genome size
  - Number of 35S and 5S loci
  - Physical position (telomeric, subtelomeric, interstitial, pericentromeric)
  - Co-localisation vs separate loci
- Taxonomic coverage across major plant clades

Evolutionary questions using rDNA data

- Macroevolution of genome architecture:
  - Gains/losses of rDNA loci along phylogenies
  - Shifts in rDNA position (e.g. telomeric → interstitial) and chromosomal rearrangements
- Polyploidy and diploidisation:
  - Retention vs loss of parental rDNA loci in allopolyploids
- Genome size and repeat dynamics:
  - Correlations between rDNA CNV and genome size, TE content, recombination landscapes

VI SIMPOSIO SEBOT HUESCA 26-30 NOVIEMBRE 2025

SEBOT SOCIEDAD BOTÁNICA ESPAÑOLA

The plant rDNA database: contents, scope, evolutionary questions…

Use as a resource

- Rapid accumulation of rDNA-FISH data in the literature
- No other resource giving standardized locus number and position information across plants

Scope

- Angiosperms, gymnosperms, bryophytes and pteridophytes (land plants + early diverging lineages)
- A parallel resource for animals (www.animalrdnatabase.com) was also developed few years later

For each taxon / cytogenetic record, the database stores:

- **rDNA locus data**
  - Number of 5S and 35S loci per 2n set
  - Chromosomal position simplified into three categories:
    - (Peri)centromeric, interstitial, (sub)terminal
  - Mutual arrangement of 5S and 35S:
    - **L-type** (linked in a single unit 18S–5.8S–26S–5S)
    - **S-type** (separate 5S and 35S arrays)
- **Others:** number of chromosomes with both 5S and 35S (separate arms or same arm), genome size, ploidy level, life cycle, telomere composition

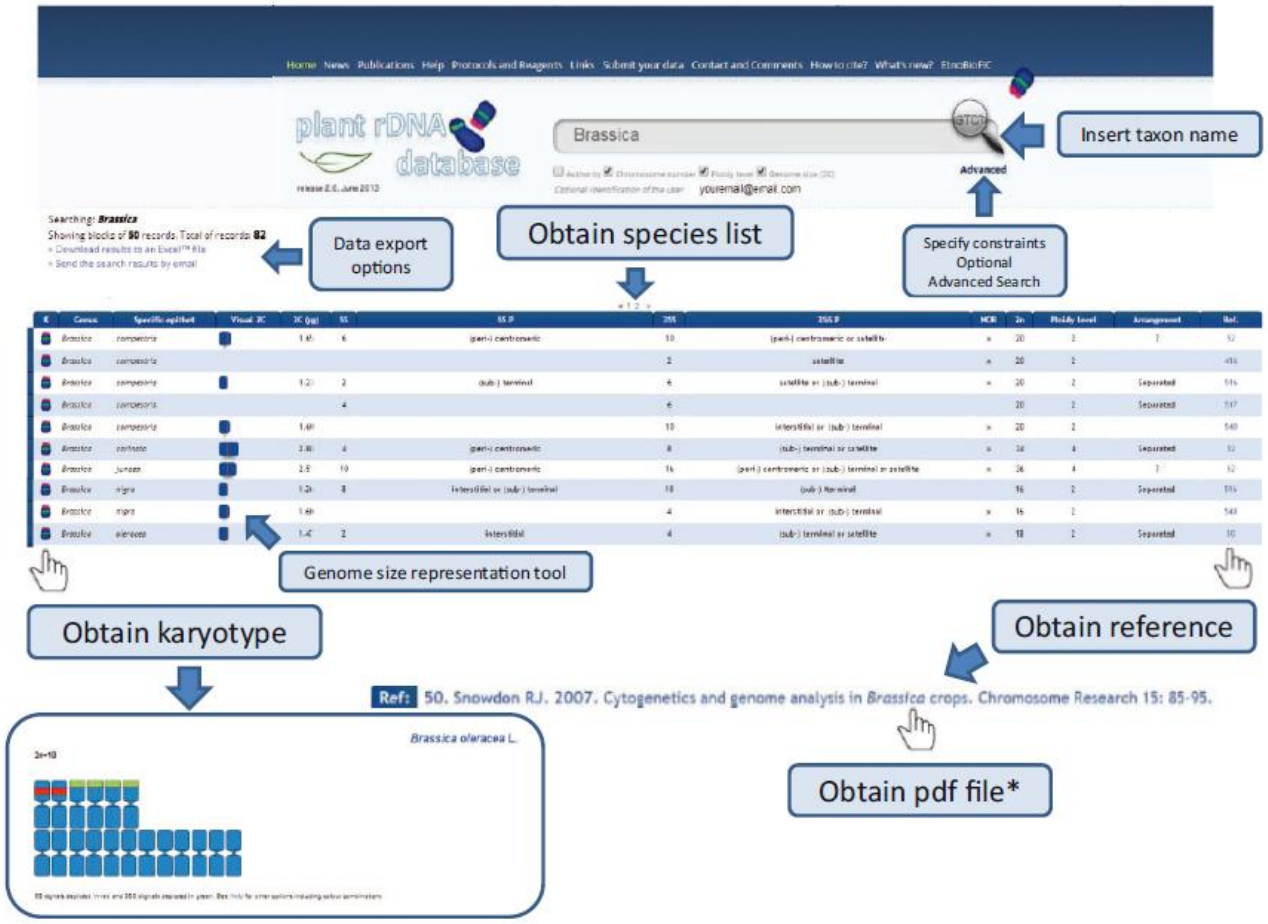
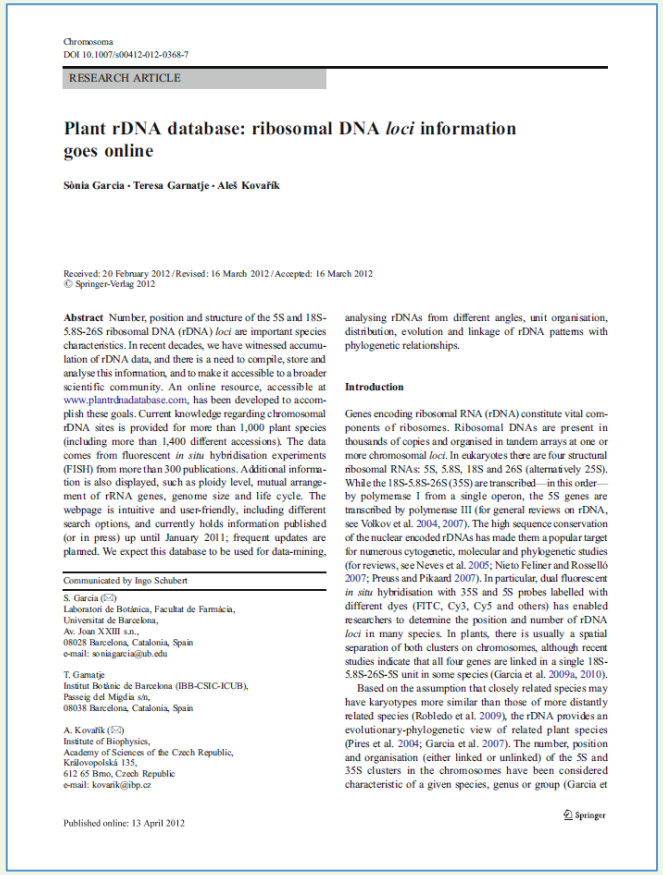


Figure 4. Flowchart showing the possible database outputs and options for a given search. The karyotype appears as a pop-up window, in which the 5S rDNAs are depicted in red and the 35S in green.

Garcia et al. 2014

The plant rDNA database: growth across releases (1→4)

- **Release 1.0 (Garcia et al. 2012)**
  - 1,033 species; 1,461 records
  - 319 publications; 60 families, 245 genera
  - Focus on entries with both 5S and 35S data



Chromosoma
DOI 10.1007/s00412-012-0368-7

RESEARCH ARTICLE

Plant rDNA database: ribosomal DNA *loci* information goes online

Sònia Garcia · Teresa Garnatje · Aleš Kovařík

Received: 20 February 2012 / Revised: 16 March 2012 / Accepted: 16 March 2012

Abstract Number, position and structure of the 5S and 18S-5.8S-26S ribosomal DNA (rDNA) *loci* are important species characteristics. In recent decades, we have witnessed accumulation of rDNA data, and there is a need to compile, store and analyse this information, and to make it accessible to a broader scientific community. An online resource, accessible at www.plantrdnadatabase.com, has been developed to accomplish these goals. Current knowledge regarding chromosomal rDNA sites is provided for more than 1,000 plant species (including more than 1,400 different accessions). The data comes from fluorescent *in situ* hybridisation experiments (FISH) from more than 300 publications. Additional information is also displayed, such as ploidy level, mutual arrangement of rRNA genes, genome size and life cycle. The webpage is intuitive and user-friendly, including different search options, and currently holds information published (or in press) up until January 2011; frequent updates are planned. We expect this database to be used for data-mining, analysing rDNAs from different angles, unit organisation, distribution, evolution and linkage of rDNA patterns with phylogenetic relationships.

Communicated by Ingo Schubert

S. Garcia (✉)
Laboratori de Botànica, Facultat de Farmàcia,
Universitat de Barcelona,
Av. Joan XXIII s.n.,
08028 Barcelona, Catalonia, Spain
e-mail: soniagarcia@ub.edu

T. Garnatje
Institut Botànic de Barcelona (IBB-CSIC-ICUB),
Passeig del Migdia s/n,
08038 Barcelona, Catalonia, Spain

A. Kovařík (✉)
Institute of Biophysics,
Academy of Sciences of the Czech Republic,
Královopolská 135,
612 65 Brno, Czech Republic
e-mail: kovarik@ibp.cz

Introduction

Genes encoding ribosomal RNA (rDNA) constitute vital components of ribosomes. Ribosomal DNAs are present in thousands of copies and organised in tandem arrays at one or more chromosomal *loci*. In eukaryotes there are four structural ribosomal RNAs: 5S, 5.8S, 18S and 26S (alternatively 25S). While the 18S-5.8S-26S (35S) are transcribed—in this order—by polymerase I from a single operon, the 5S genes are transcribed by polymerase III (for general reviews on rDNA, see Volkov et al. 2004, 2007). The high sequence conservation of the nuclear encoded rDNAs has made them a popular target for numerous cytogenetic, molecular and phylogenetic studies (for reviews, see Neves et al. 2005; Nieto Feliner and Rosselló 2007; Preuss and Pikaard 2007). In particular, dual fluorescent *in situ* hybridisation with 35S and 5S probes labelled with different dyes (FITC, Cy3, Cy5 and others) has enabled researchers to determine the position and number of rDNA *loci* in many species. In plants, there is usually a spatial separation of both clusters on chromosomes, although recent studies indicate that all four genes are linked in a single 18S-5.8S-26S-5S unit in some species (Garcia et al. 2009a, 2010).

Based on the assumption that closely related species may have karyotypes more similar than those of more distantly related species (Robledo et al. 2009), the rDNA provides an evolutionary-phylogenetic view of related plant species (Pires et al. 2004; Garcia et al. 2007). The number, position and organisation (either linked or unlinked) of the 5S and 35S clusters in the chromosomes have been considered characteristic of a given species, genus or group (Garcia et

Published online: 13 April 2012

- **Release 2.0 (Garcia et al. 2014)**
  - 1,609 species; 2,839 records; >600 publications
  - Now includes single-locus papers (5S-only or 35S-only)
  - First inclusion of pteridophytes; clear growth especially in monocots

DATABASE
Database, 2014, 1–7
doi: 10.1093/database/bau063
Database update

Database update

Plant rDNA database: update and new features

Sònia Garcia[1,*], Francisco Gálvez[2], Airy Gras[1], Aleš Kovařík[3] and Teresa Garnatje[4]

[1]Laboratori de Botànica-Unitat associada CSIC, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, 08028 Catalonia, Spain, [2]BioScripts – Centro de Investigación y Desarrollo de Recursos Científicos, Sevilla, 41012 Andalusia, Spain, [3]Institute of Biophysics, Academy of Sciences of the Czech Republic. Brno, CZ-612 65, Czech Republic and [4]Institut Botànic de Barcelona (IBB-CSIC-ICUB). Barcelona, 08038 Catalonia, Spain

*Corresponding author: Tel: +34 934 024 490; Fax: +34 934 025 879; Email: soniagarcia@ub.edu or sphaeromeria@gmail.com

Present address: Sònia Garcia, Laboratori de Botànica-Unitat associada CSIC, Facultat de Farmàcia, Universitat de Barcelona. Av. Joan XXIII s/n 08028 Barcelona, Catalonia, Spain

Citation details: Garcia,S., Gálvez,F., Gras,A. *et al.* Plant rDNA database: update and new features. *Database* (2014) Vol. 2014: article ID bau063; doi:10.1093/database/bau063

Received 26 February 2014; Revised 16 May 2014; Accepted 4 June 2014

Abstract

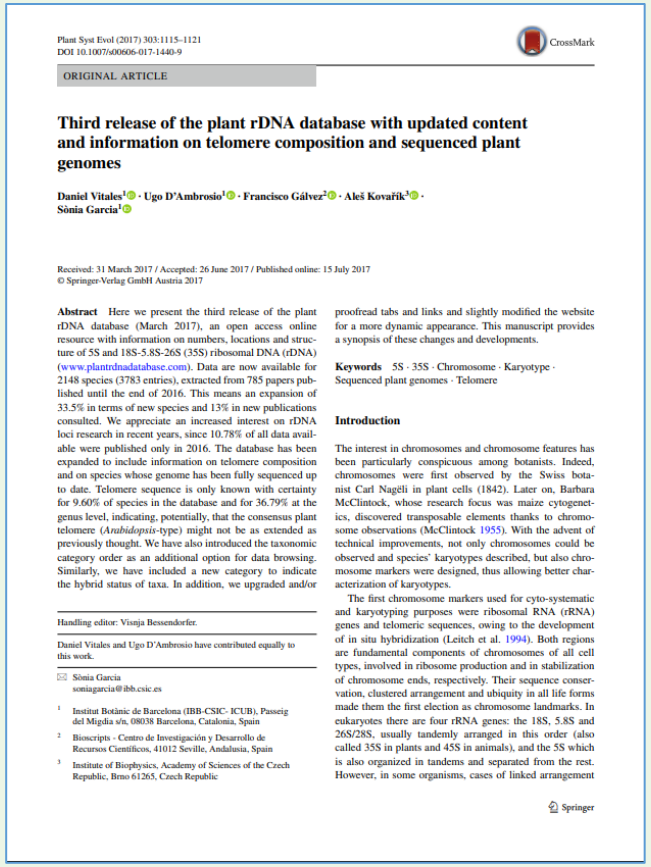
The Plant rDNA database (www.plantrdnadatabase.com) is an open access online resource providing detailed information on numbers, structures and positions of 5S and 18S-5.8S-26S (35S) ribosomal DNA *loci*. The data have been obtained from >600 publications on plant molecular cytogenetics, mostly based on fluorescent *in situ* hybridization (FISH). This edition of the database contains information on 1609 species derived from 2839 records, which means an expansion of 55.76 and 94.45%, respectively. It holds the data for angiosperms, gymnosperms, bryophytes and pteridophytes available as of June 2013. Information from publications reporting data for a single rDNA (either 5S or 35S alone) and annotation regarding transcriptional activity of 35S *loci* now appears in the database. Preliminary analyses suggest greater variability in the number of rDNA *loci* in gymnosperms than in angiosperms. New applications provide ideograms of the species showing the positions of rDNA *loci* as well as a visual representation of their genome sizes. We have also introduced other features to boost the usability of the Web interface, such as an application for convenient data export and a new section with rDNA–FISH-related information (mostly detailing protocols and reagents). In addition, we upgraded and/or proofread tabs and links and modified the website for a more dynamic appearance. This manuscript provides a synopsis of these changes and developments.

Database URL: http://www.plantrdnadatabase.com

VI SIMPOSIO SEBOT
MERES RETURNS 28–30 NOVIEMBRE 2025
SEBOT
SOCIEDAD BOTÁNICA ESPAÑOLA

The plant rDNA database: growth across releases (1→4)

- **Release 3.0 (Vitales et al. 2017)**
  - 3,783 entries; 2,148 species; 785 papers (data to end 2016)
  - New: telomere composition; list of species with sequenced genomes; order-level browsing; flag for hybrids



Plant Syst Evol (2017) 303:1115–1121
DOI 10.1007/s00606-017-1440-9

ORIGINAL ARTICLE

Third release of the plant rDNA database with updated content and information on telomere composition and sequenced plant genomes

Daniel Vitales[1] · Ugo D'Ambrosio[1] · Francisco Gálvez[2] · Aleš Kovařík[3] · Sònia Garcia[1]

Received: 31 March 2017 / Accepted: 26 June 2017 / Published online: 15 July 2017

Abstract Here we present the third release of the plant rDNA database (March 2017), an open access online resource with information on numbers, locations and structure of 5S and 18S-5.8S-26S (35S) ribosomal DNA (rDNA) (www.plantrdnadatabase.com). Data are now available for 2148 species (3783 entries), extracted from 785 papers published until the end of 2016. This means an expansion of 33.5% in terms of new species and 13% in new publications consulted. We appreciate an increased interest on rDNA loci research in recent years, since 10.78% of all data available were published only in 2016. The database has been expanded to include information on telomere composition and on species whose genome has been fully sequenced up to date. Telomere sequence is only known with certainty for 9.60% of species in the database and for 36.79% at the genus level, indicating, potentially, that the consensus plant telomere (*Arabidopsis*-type) might not be as extended as previously thought. We have also introduced the taxonomic category order as an additional option for data browsing. Similarly, we have included a new category to indicate the hybrid status of taxa. In addition, we upgraded and/or proofread tabs and links and slightly modified the website for a more dynamic appearance. This manuscript provides a synopsis of these changes and developments.

Keywords 5S · 35S · Chromosome · Karyotype · Sequenced plant genomes · Telomere

Handling editor: Visnja Bessendorfer.

Daniel Vitales and Ugo D'Ambrosio have contributed equally to this work.

✉ Sònia Garcia
soniagarcia@ibb.csic.es

[1] Institut Botànic de Barcelona (IBB-CSIC- ICUB), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain

[2] Bioscripts - Centro de Investigación y Desarrollo de Recursos Científicos, 41012 Seville, Andalusia, Spain

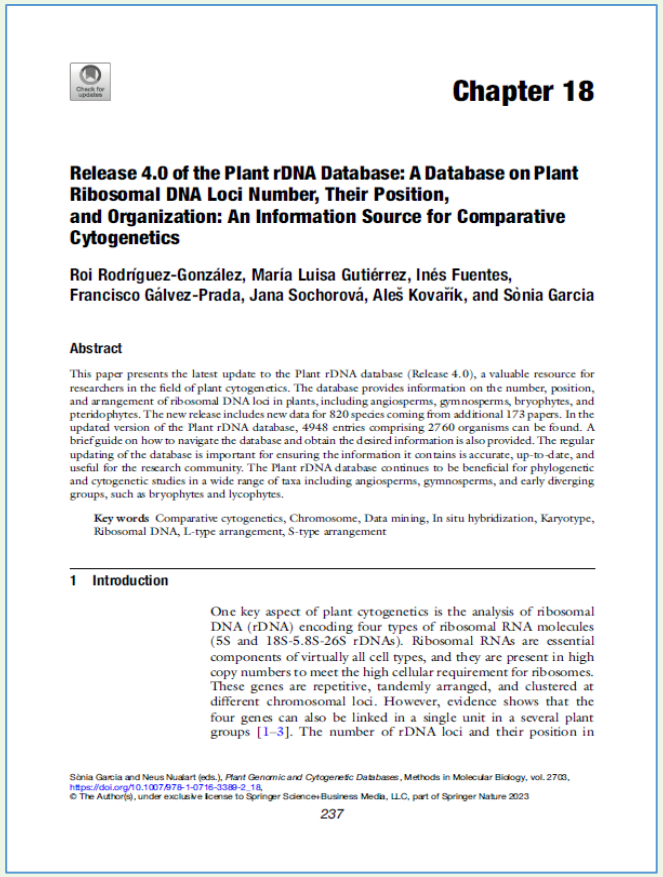
[3] Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno 61265, Czech Republic

Introduction

The interest in chromosomes and chromosome features has been particularly conspicuous among botanists. Indeed, chromosomes were first observed by the Swiss botanist Carl Nageli in plant cells (1842). Later on, Barbara McClintock, whose research focus was maize cytogenetics, discovered transposable elements thanks to chromosome observations (McClintock 1955). With the advent of technical improvements, not only chromosomes could be observed and species' karyotypes described, but also chromosome markers were designed, thus allowing better characterization of karyotypes.

The first chromosome markers used for cyto-systematic and karyotyping purposes were ribosomal RNA (rRNA) genes and telomeric sequences, owing to the development of in situ hybridization (Leitch et al. 1994). Both regions are fundamental components of chromosomes of all cell types, involved in ribosome production and in stabilization of chromosome ends, respectively. Their sequence conservation, clustered arrangement and ubiquity in all life forms made them the first election as chromosome landmarks. In eukaryotes there are four rRNA genes: the 18S, 5.8S and 26S/28S, usually tandemly arranged in this order (also called 35S in plants and 45S in animals), and the 5S which is also organized in tandems and separated from the rest. However, in some organisms, cases of linked arrangement

- **Release 4.0 (Rodríguez-Gonzalez et al. 2023)**
  - 4,948 entries; 2770species; 951 papers (data to end 2020)
  - 127 families, 566 genera; strongly dominated by angiosperms (~96% of entries)
  - New data for 820 species from 173 additional publications; interface and search functions refined



Chapter 18

Release 4.0 of the Plant rDNA Database: A Database on Plant Ribosomal DNA Loci Number, Their Position, and Organization: An Information Source for Comparative Cytogenetics

Roi Rodríguez-González, María Luisa Gutiérrez, Inés Fuentes, Francisco Gálvez-Prada, Jana Sochorová, Aleš Kovařík, and Sònia Garcia

Abstract

This paper presents the latest update to the Plant rDNA database (Release 4.0), a valuable resource for researchers in the field of plant cytogenetics. The database provides information on the number, position, and arrangement of ribosomal DNA loci in plants, including angiosperms, gymnosperms, bryophytes, and pteridophytes. The new release includes new data for 820 species coming from additional 173 papers. In the updated version of the Plant rDNA database, 4948 entries comprising 2760 organisms can be found. A brief guide on how to navigate the database and obtain the desired information is also provided. The regular updating of the database is important for ensuring the information it contains is accurate, up-to-date, and useful for the research community. The Plant rDNA database continues to be beneficial for phylogenetic and cytogenetic studies in a wide range of taxa including angiosperms, gymnosperms, and early diverging groups, such as bryophytes and lycophytes.

Key words Comparative cytogenetics, Chromosome, Data mining, In situ hybridization, Karyotype, Ribosomal DNA, L-type arrangement, S-type arrangement

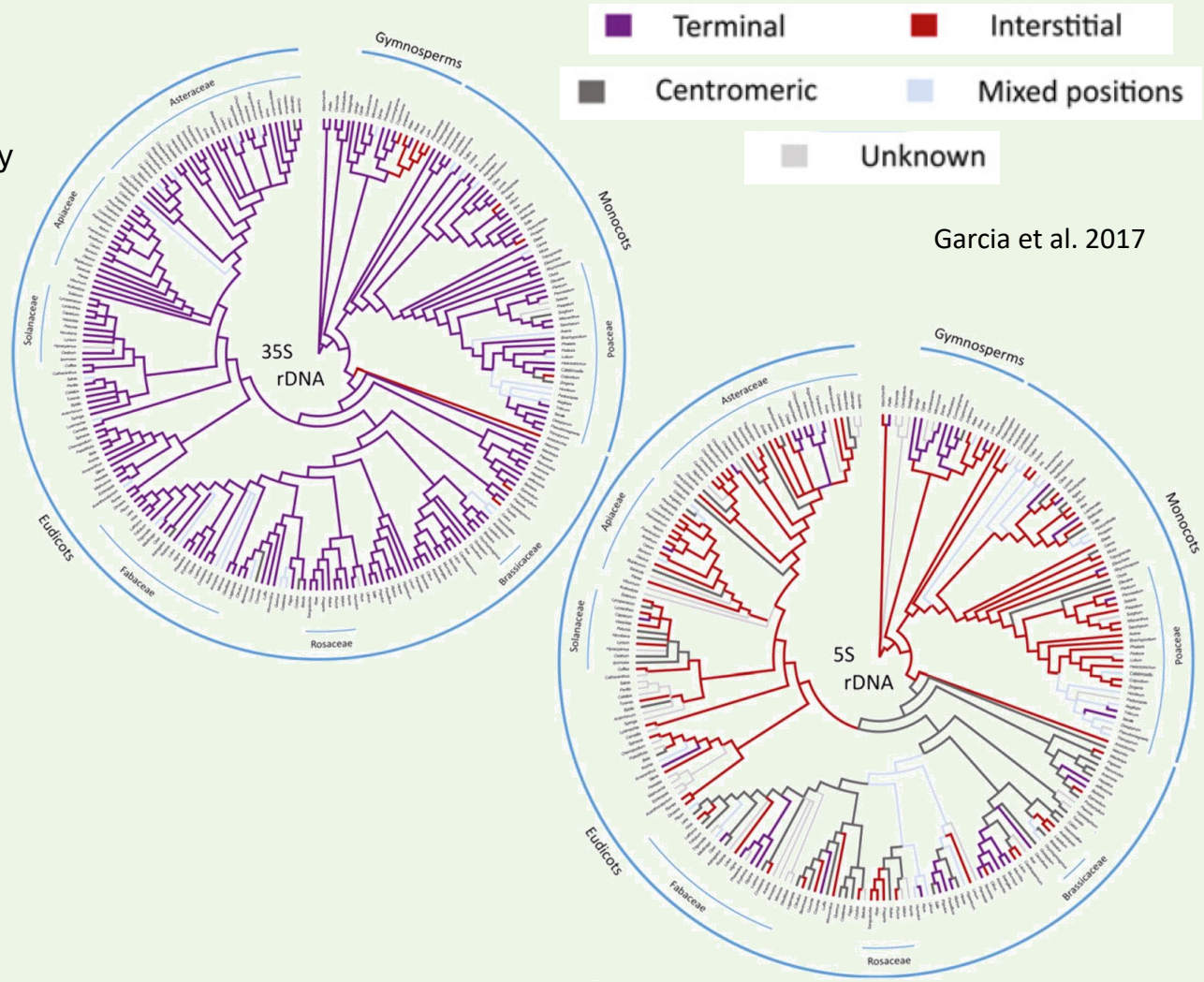
1 Introduction

One key aspect of plant cytogenetics is the analysis of ribosomal DNA (rDNA) encoding four types of ribosomal RNA molecules (5S and 18S-5.8S-26S rDNAs). Ribosomal RNAs are essential components of virtually all cell types, and they are present in high copy numbers to meet the high cellular requirement for ribosomes. These genes are repetitive, tandemly arranged, and clustered at different chromosomal loci. However, evidence shows that the four genes can also be linked in a single unit in a several plant groups [1–3]. The number of rDNA loci and their position in

Sònia Garcia and Neus Nualart (eds.), *Plant Genomic and Cytogenetic Databases*, Methods in Molecular Biology, vol. 2703,
https://doi.org/10.1007/978-1-0716-3389-2_18

The plant rDNA database: What the global data tell us

- **Low locus numbers are the rule**
  - Most species **have few rDNA loci**; single-locus karyotypes are very common, even in polyploids
  - Interpreted as **locus loss after polyploidy**, part of the diploidization process
- **Contrasting dynamics in major groups**
  - Angiosperms vs gymnosperms: both usually have low numbers of 5S/35S sites, but **35S in gymnosperms shows much more heterogeneous distributions**
- **Positional patterns**
  - 5S loci: roughly **balanced** across pericentromeric, interstitial, and terminal positions
  - 35S loci: **strongly biased to terminal/subterminal regions** (~90% of loci), fewer centromeric/interstitial sites
- **Genomic organization (L vs S)**
  - Only **~5% of entries show L-type arrangement** (linked 18S–5.8S–26S–5S); ~81% are S-type (separate arrays)



Garcia et al. 2017

The plant rDNA database: What the global data tell us

Extra layers added in later releases

- **Telomere composition** (Release 3.0 onward)
  - Database records telomere motif type (*Arabidopsis*-type, human-type, *Cestrum*-type, *Allium*-type, etc.)
  - Highlight: **contributed to discovering human-type telomeres in *Zostera marina* and underpins reviews of telomere diversity in plants**

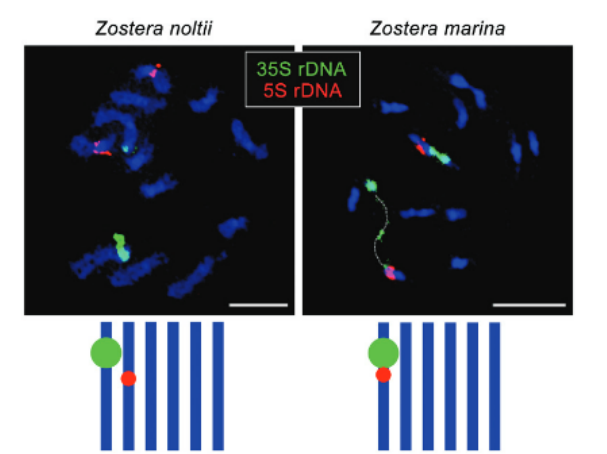


Fig. 2. Karyotypes of *Zostera noltii* and *Z. marina*. Both species bear a single terminally located 35S rDNA locus and a single interstitial 5S rDNA locus. While 35S and 5S are situated on different chromosomes in *Z. noltii*, they occupy the same chromosome in *Z. marina*. Note the highly decondensed 35S rDNA site in *Z. marina* (dotted line). Scale bars=10 μm. (This figure is available in colour at *JXB* online.)

Peska et al. 2020

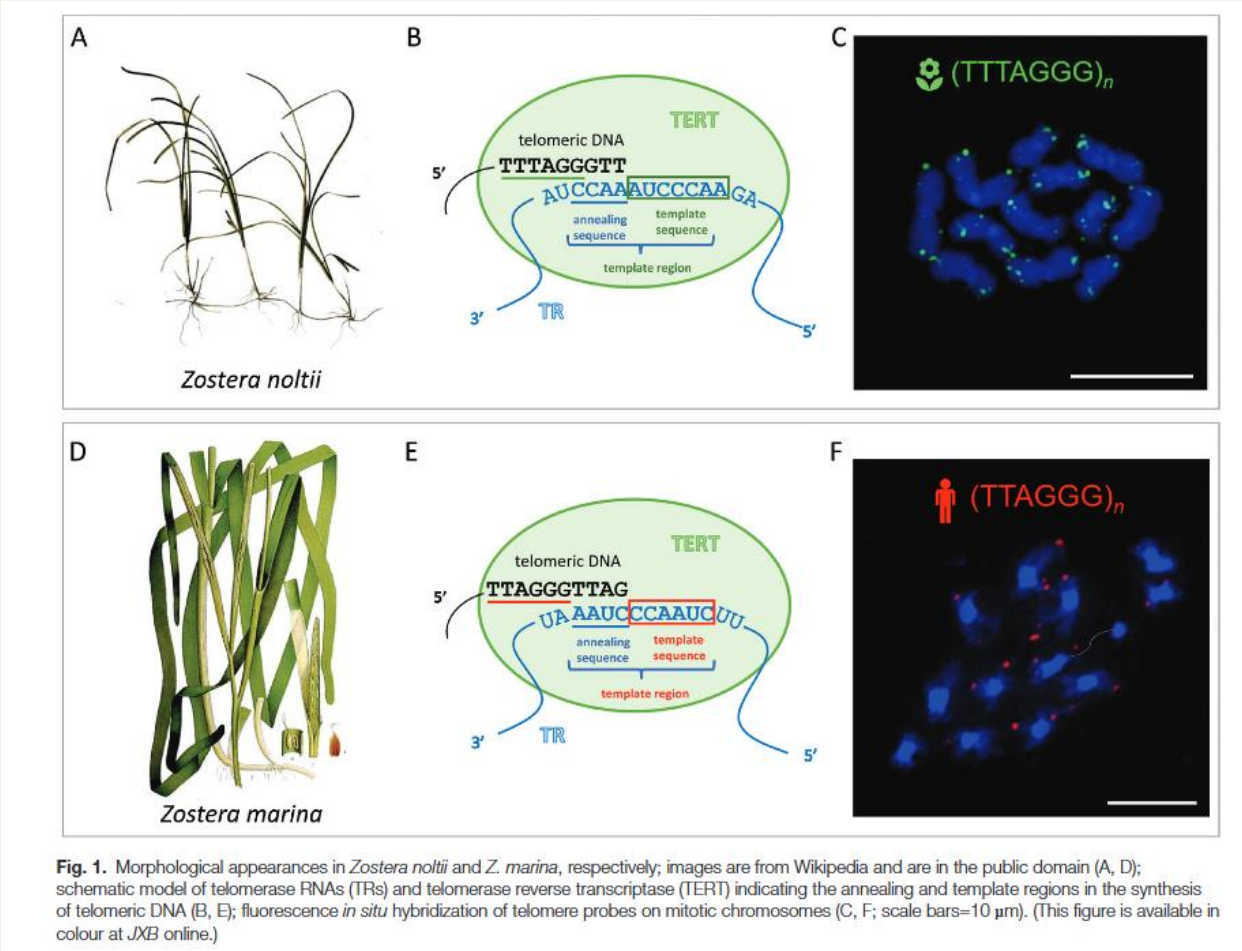
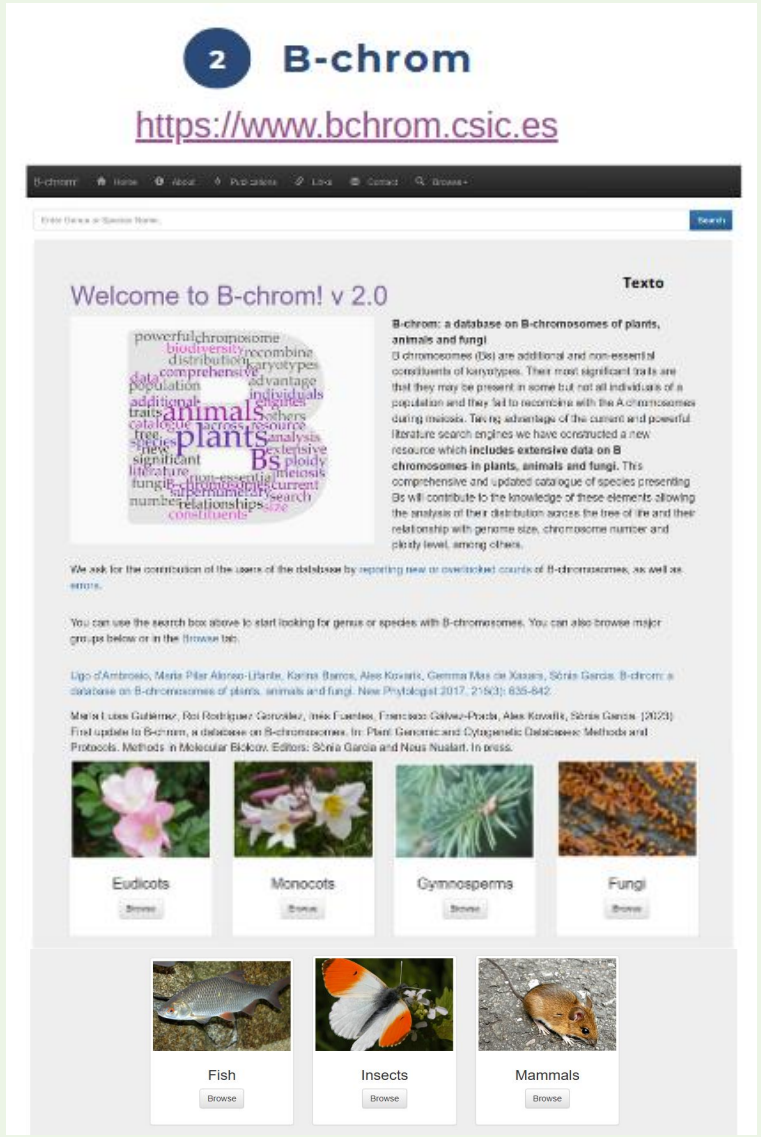


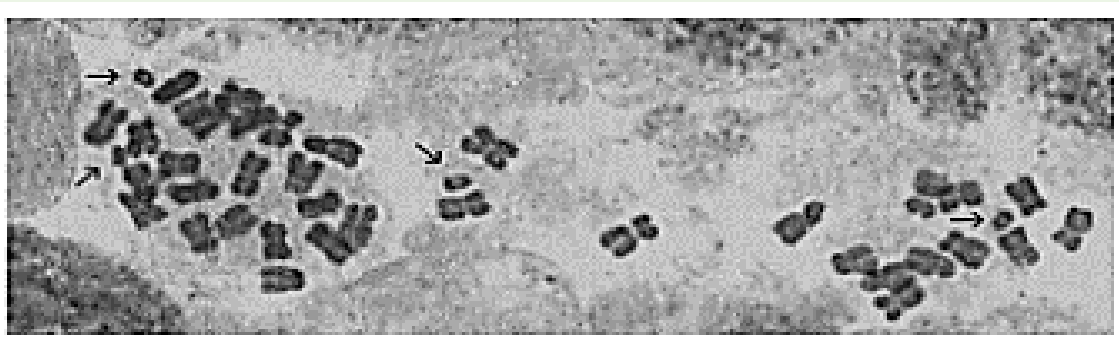
Fig. 1. Morphological appearances in *Zostera noltii* and *Z. marina*, respectively; images are from Wikipedia and are in the public domain (A, D); schematic model of telomerase RNAs (TRs) and telomerase reverse transcriptase (TERT) indicating the annealing and template regions in the synthesis of telomeric DNA (B, E); fluorescence *in situ* hybridization of telomere probes on mitotic chromosomes (C, F; scale bars=10 μm). (This figure is available in colour at *JXB* online.)

Peska et al. 2020

B-chrom database: why it exists



- **B-chrom (www.bchrom.csic.es)**: online database on B-chromosomes (Bs) in **plants, animals and fungi**
- Motivation: >100 years of scattered literature on Bs (since 1906) with no up-to-date, searchable global synthesis
- **Bs = supernumerary, mostly dispensable chromosomes:**
  - Present only in some individuals or tissues
  - Do not recombine with A chromosomes; often show **drive (non-Mendelian transmission)**
- Bs are relevant because they:
  - Occur in **~15% of eukaryotes**
  - Contribute to **intraspecific genome size variation**
  - Can have **neutral, deleterious or beneficial** effects (e.g. vigour, fertility, resistance, adaptation)



Garcia et al. 2006

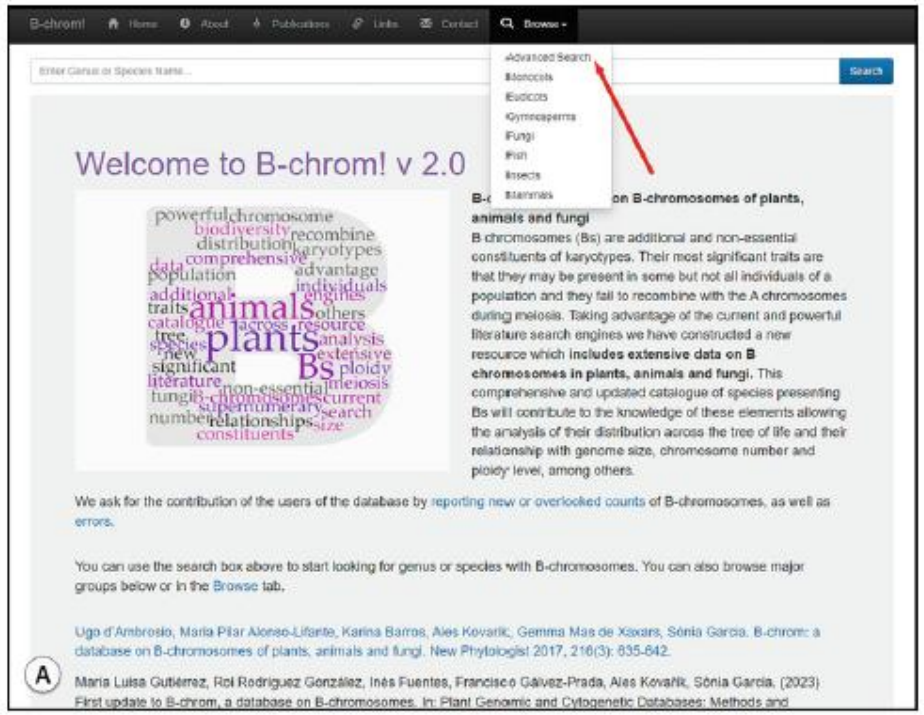
B-chrom database

B-chrom database: content & scope

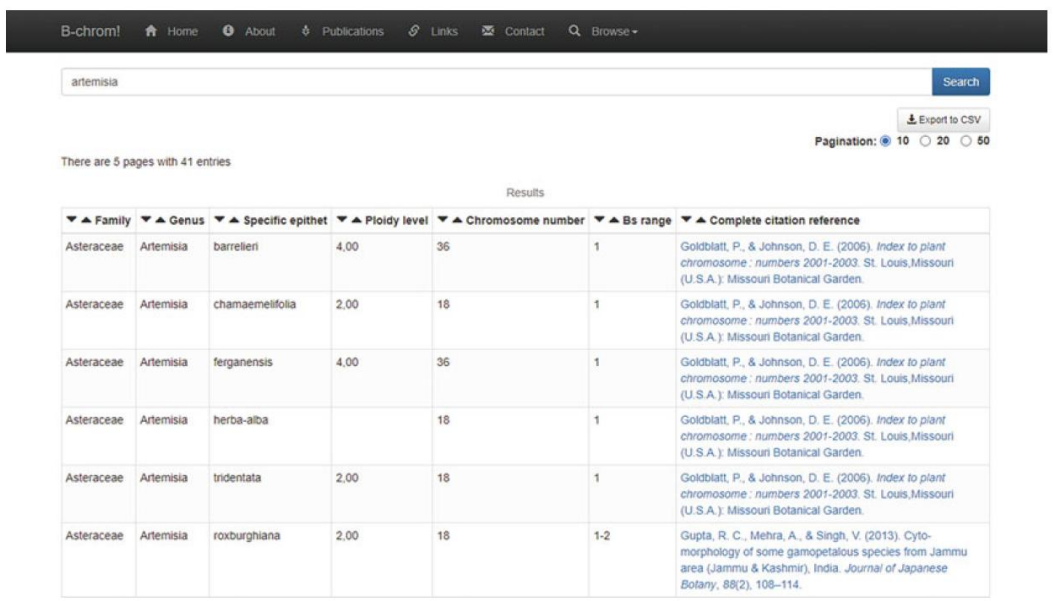
- Variables:
  - **Presence** of B chromosomes
  - Number and/or range
- Includes plants, animals and fungi → comparative, cross-kingdom perspective

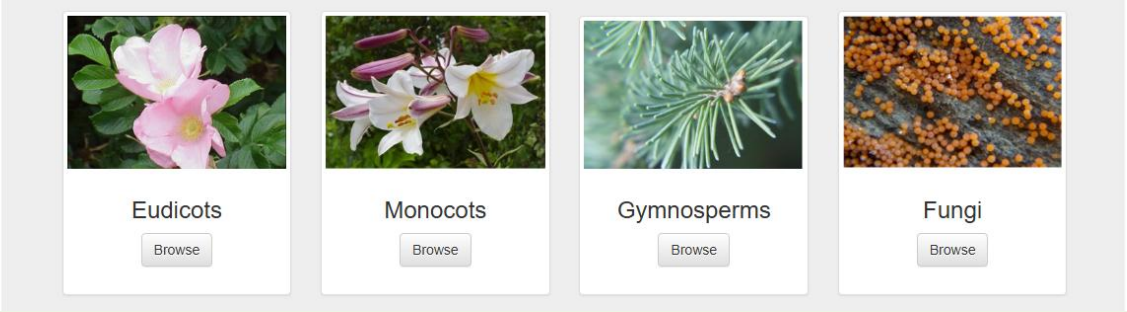
Evolutionary questions using B-chrom data

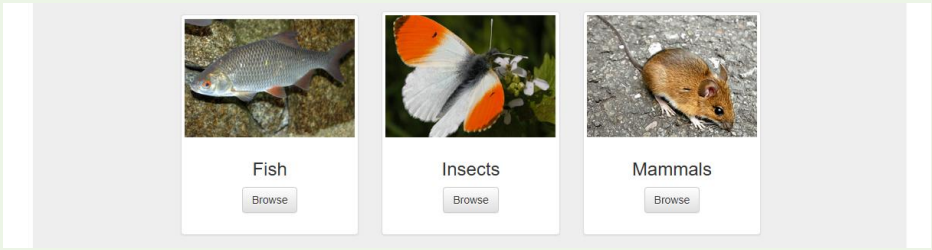
- Genomic conflict and selfish elements:
  - Comparative analyses of where B chromosomes **arise and persist**
  - **Association with life history** (annual vs perennial, mating system)
- Macro-evolution:
  - Is B-chromosome presence **phylogenetically clustered** or **recurrent**?
  - Links to **polyploidy, genome size, and karyotype instability**



B-chrom database







What an entry in B-chrom contains:

- **Taxonomy and group info**
  - Family, genus, specific epithet
  - Search for specific “large” group
- **Karyological data**
  - Ploidy level
  - Somatic chromosome number (2n).
- **B-chromosome data**
  - Presence of Bs and the **number or range** of Bs (e.g. 0–4, 2–6)
  - Codes indicating context:
    - “Bs” = presence
    - suffixes such as “G” (gametic), “n” (haploid), “m/f” (male/female), “PRS” for paternal sex ratio Bs in arthropods
- **Bibliographic metadata**
  - Full citation (APA style) and DOI/URL
  - Each entry linked to a specific source and has its own ID

- **Results can be exported as a csv file**

B-chrom database: growth across releases

1. **Initial release (2017)** – D'Ambrosio et al (2017)

   5760 entries; **2828 eukaryotic species** reported with Bs

   - Composition of species:
     - 73.6% plants (2087 spp; ~53% monocots, ~47% eudicots)
     - 26.0% animals (736 spp)
     - 0.5% fungi (14 spp)
   - 311 families (119 plant, 185 animal, 7 fungal) and 1095 genera (635 plant, 450 animal, 10 fungal)
   - Data from **3041 references**, combining Jones & Díez (2004) (1906–1994, 3484 records) plus new literature (1995–2016, 2276 new records)

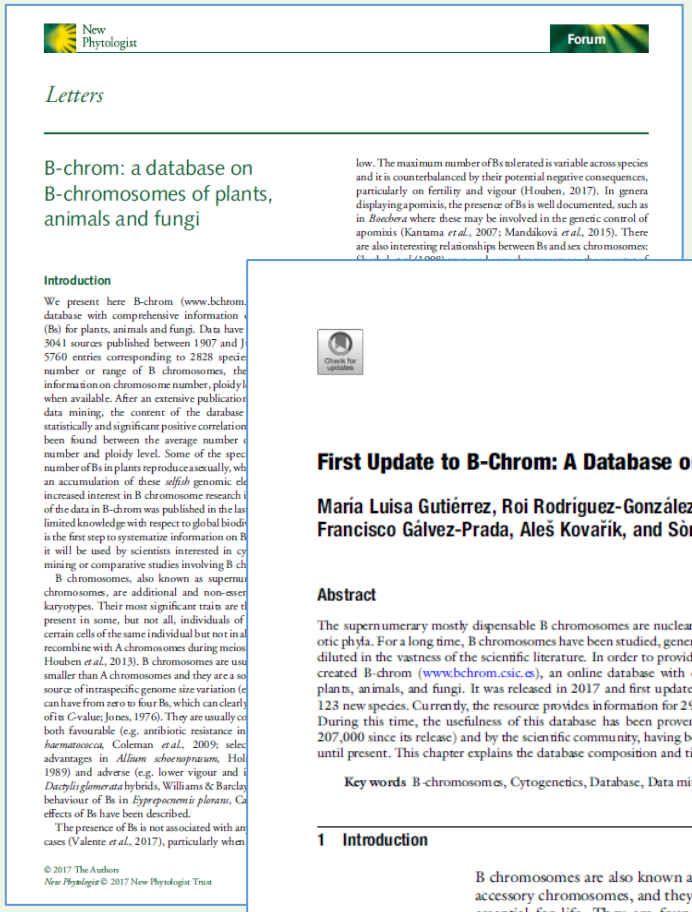
     +65% entries and +61.5% species

2. **First update (2021)** – Gutierrez et al. (2021)

   2. Added **334 entries and 123 new species**
   - Current content: **6089 entries**, **2951 species**, from **3292 sources** (articles, books, proceedings)

Search strategy: using Scopus and Web of Science with a broad set of search terms (B chromosome, accessory, supernumerary, selfish chromosome, etc.)

D'Ambrosio et al. 2017



New Phytologist — Forum

Letters

B-chrom: a database on B-chromosomes of plants, animals and fungi

Introduction

We present here B-chrom (www.bchrom.
database with comprehensive information
(Bs) for plants, animals and fungi. Data have
3041 sources published between 1907 and J
5760 entries corresponding to 2828 specie
number or range of B chromosomes, the
information on chromosome number, ploidy l
when available. After an extensive publicatior
data mining, the content of the database
statistically and significant positive correlation
been found between the average number
number and ploidy level. Some of the spec
number of Bs in plants reproduce asexually, wh
an accumulation of these *selfish* genomic ele
increased interest in B chromosome research i
of the data in B-chrom was published in the las
limited knowledge with respect to global biodi
is the first step to systematize information on B
it will be used by scientists interested in cy
mining or comparative studies involving B ch
B chromosomes, also known as supernu
chromosomes, are additional and non-esse
karyotypes. Their most significant traits are t
present in some, but not all, individuals of
certain cells of the same individual but not in al
recombine with A chromosomes during meios
Houben *et al.*, 2013). B chromosomes are usu
smaller than A chromosomes and they are a so
source of intraspecific genome size variation (e
can have from zero to four Bs, which can clearly
of its *C*-value; Jones, 1976). They are usually co
both favourable (e.g. antibiotic resistance in
haematococca, Coleman *et al.*, 2009; selec
advantages in *Allium schoenoprasum*, Hol
1989) and adverse (e.g. lower vigour and i
Dactylis glomerata hybrids, Williams & Barcla
behaviour of Bs in *Eyprepocnemis plorans*, Ca
effects of Bs have been described.
The presence of Bs is not associated with an
cases (Valente *et al.*, 2017), particularly when

low. The maximum number of Bs tolerated is variable across species and it is counterbalanced by their potential negative consequences, particularly on fertility and vigour (Houben, 2017). In genera displaying apomixis, the presence of Bs is well documented, such as in *Boechera* where these may be involved in the genetic control of apomixis (Kantama *et al.*, 2007; Mandáková *et al.*, 2015). There are also interesting relationships between Bs and sex chromosomes:

Gutiérrez et al. 2023

Chapter 17

First Update to B-Chrom: A Database on B-Chromosomes

María Luisa Gutiérrez, Roi Rodríguez-González, Inés Fuentes, Francisco Gálvez-Prada, Aleš Kovařík, and Sònia Garcia

Abstract

The supernumerary mostly dispensable B chromosomes are nuclear components of about 15% of eukaryotic phyla. For a long time, B chromosomes have been studied, generating an enormous bulk of knowledge, diluted in the vastness of the scientific literature. In order to provide better access to this information, we created B-chrom (www.bchrom.csic.es), an online database with comprehensive information on Bs for plants, animals, and fungi. It was released in 2017 and first updated in 2021, by adding 334 entries and 123 new species. Currently, the resource provides information for 2951 species coming from 3292 sources. During this time, the usefulness of this database has been proven by the number of visits (more than 207,000 since its release) and by the scientific community, having been cited in more than 60 publications until present. This chapter explains the database composition and tips on how to use it.

Key words B-chromosomes, Cytogenetics, Database, Data mining, Evolution, Karyology, Karyotype

1 Introduction

B chromosomes are also known as supernumerary, dispensable, or accessory chromosomes, and they are additional, by definition, not essential for life. They are found only in some individuals of a species, and in some cases, only in some tissues or some cells. They are often found by chance in routine karyological analyses. One of the most prominent features of B chromosomes is that they do not recombine with standard (A) chromosomes during meiosis. They are preferentially inherited deviating from usual Mendelian segregation, leading to a transmission advantage for B chromosomes, a mechanism called "chromosome drive" [1].

B-chromosome's evolutionary origin is still poorly understood, and as with many evolutionary phenomena, it is likely that B chromosomes have originated repeatedly and independently on several occasions (reviewed in [2]). The most accepted idea is that they come from A-chromosomes, for example, after asymmetric

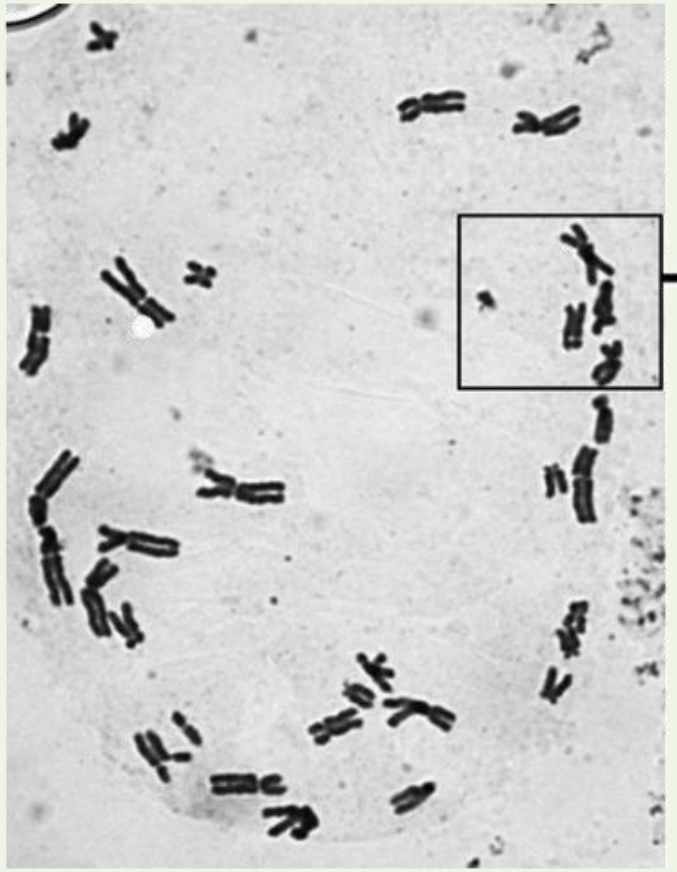
Sònia Garcia and Neus Nualart (eds.), *Plant Genomic and Cytogenetic Databases*, Methods in Molecular Biology, vol. 2703,
https://doi.org/10.1007/978-1-0716-3389-2_17,

VI SIMPOSIO SEBOT — MIERES, ASTURIAS 26–30 NOVIEMBRE 2025

SEBOT SOCIEDAD BOTÁNICA ESPAÑOLA

B-chrom database: what the global data tell us about B-chromosomes

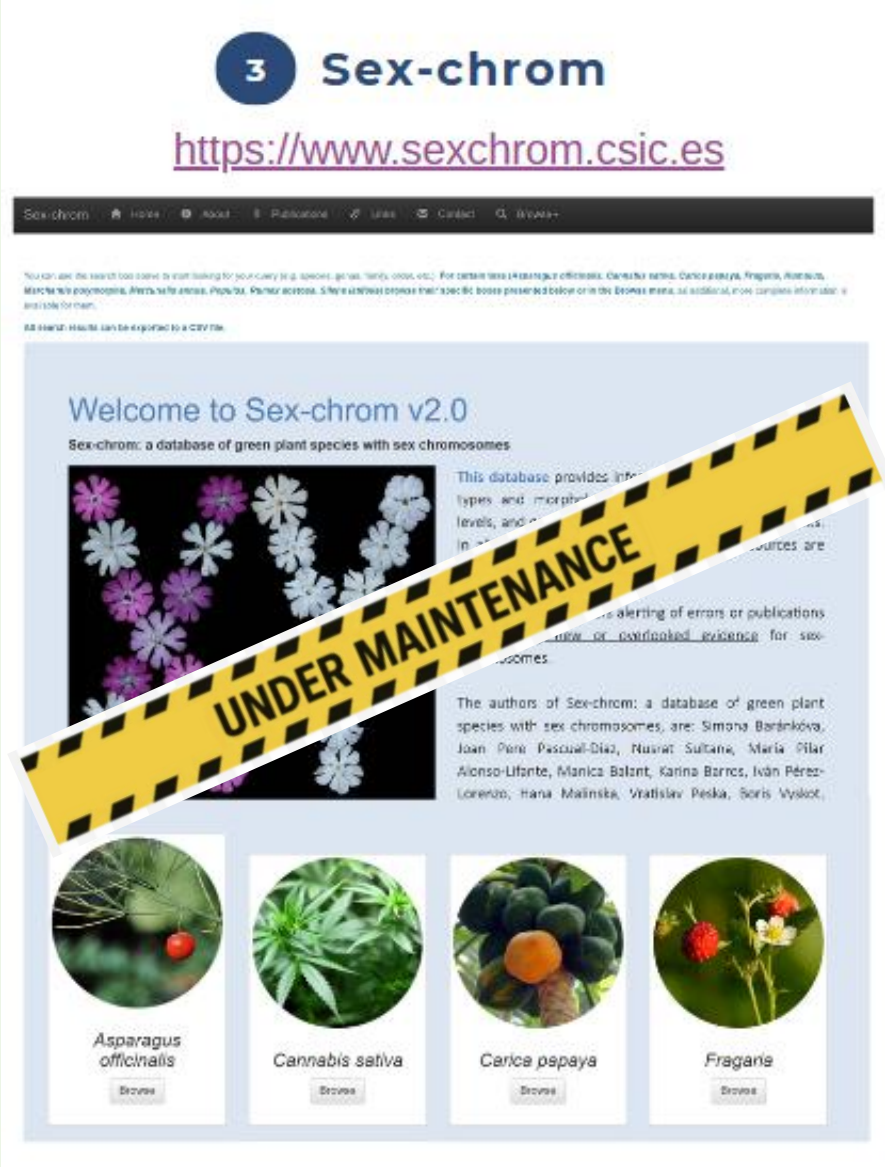
- **Bs are widespread but still under-documented**
  - In plants, by combining B-chrom with CCDB, we estimate that around **3% of species with known chromosome number** have Bs, consistent with earlier estimates (3–4% of angiosperms, higher in monocots)
  - This is a **lower bound**, given strong taxonomic and geographic biases in cytogenetic sampling
- **Weak but significant correlations**
  - Statistically significant, though weak, **positive correlations** between:
    - average number of Bs and **chromosome number (2n)**
    - average number of Bs and **ploidy level**
  - No strong correlation with genome size overall
- **Life-history and reproduction**
  - Some plant species with **very high B numbers** reproduce **asexually (e.g. apomixis)** suggesting that **reduced sexual reproduction may favour accumulation of Bs**
- **Evolutionary links to sex chromosomes and selfish elements**
  - Multiple independent origins from A chromosomes (e.g. rye Bs composed of A fragments, cichlid Bs from rearranged A segments)
  - Documented relationships between Bs and sex chromosomes (e.g. Bs derived from or ancestral to sex chromosomes in frogs and fish)



Caltha palustris

(pictures: Llez & Jelena V Blagojević)

Sex-chrom: a database on plant sex chromosomes



Aim: provide an accessible, organized source of information on plant sex chromosomes for researchers and general users

- For each species:
  – full taxonomy, chromosome number (2n), genome size (2C), ploidy, sexual system
  – **sex-determination** mechanism (XY / ZW / UV / other) and sexual system
  – presence of **homomorphic vs heteromorphic** sex chromosomes
  – **literature source and method used to infer sex chromosomes**
- Focus: bridge **cytology, molecular genetics and genomics**

Evolutionary questions using sex-chrom data

- Independent origins of dioecy and sex chromosomes:
  - **Mapping sexual systems** and sex chromosomes on phylogenies
  - Testing whether sex chromosome evolution is associated with **genome size, polyploidy, life history**
- Tempo and mode of sex chromosome differentiation:
  - Contrasts between **homomorphic vs heteromorphic systems**

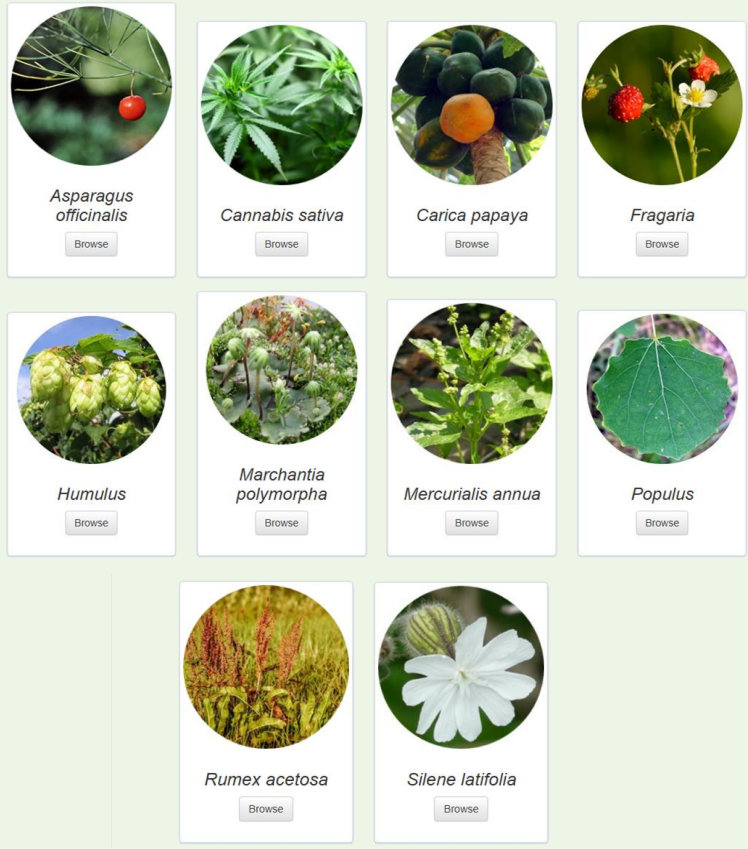
VI SIMPOSIO SEBOT
SEBOT
SOCIEDAD BOTÁNICA ESPAÑOLA

Sex-chrom: a database on plant sex chromosomes

What an entry contains (data model): Information is extracted manually and standardised.

- **Taxonomy and name checking**
  - Clade (upper/lower), order, family, genus, species, infraspecific category
  - Names verified and standardised with **TNRS** (Taxonomic Name Resolution Service)
- **Basic genomic / karyotype information**
  - Chromosome number (2n), ploidy level.
  - Genome size (2C), taken from Kew C-values when not given in source.
- **Sex-system and sex-chromosome information**
  - Sexual system: dioecious, androdioecious, gynodioecious, etc.
  - **Sex-determination mechanism: XY (male heterogamety), ZW (female heterogamety), UV (haploid phase), balance systems (X:A), etc.**
  - Sex chromosomes in males and females (e.g. XX/XY, ZZ/ZW, UV)
  - Coded as **homomorphic / heteromorphic / uncertain** based on cytological or molecular evidence
- **Method and extra genomics**
  - **Method for primary characterisation** (cytology, FISH, genetic mapping, NGS-based approaches, etc.)
  - **Miscellaneous genomic info**

Additional sheets for 10 model systems (*Asparagus, Cannabis, Carica, Fragaria, Humulus, Marchantia, Mercurialis, Populus, Rumex, Silene*) store detailed data on sex-linked genes, sex-biased genes, tandem repeats, transposons, etc.



Sex-chrom: coverage and growth

Barankova et al. 2020

Forum

New Phytologist

Letters

Sex-chrom, a database on plant sex chromosomes

Introduction

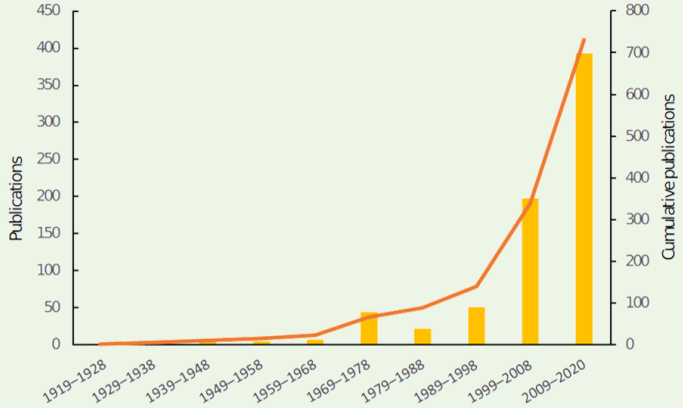
Sex determination and sex chromosomes have been studied extensively owing to their importance in evolutionary biology. Species of dioecious plants, which may have sex chromosomes, offer us unique insights into the history of this phenomenon across the Tree of Life. Here, we present Sex-chrom: a database on plant sex chromosomes (www.sexchrom.csic.es), which aims to provide an easily accessible and organized information source for scientists and a general audience interested in this field. Basic data such as complete taxonomic classification of the species, chromosome number, genome size (2C), ploidy level, sex determination mechanism, presence of homomorphic or heteromorphic sex chromosomes and their corresponding sources are readily available for 178 species of 84 genera and 65 families. Besides, the database contains specific sections for 10 selected model systems in plant sex chromosome research such as *Silene latifolia* and *Rumex acetosa*. In these sections, more detailed information also is available, including data on sex-linked genes, transposable elements or tandemly repeated DNA present in sex chromosomes. Data have been extracted from 431 sources published between 1919 and 2020.

Although most plant species are hermaphroditic, carrying both female and male reproductive organs, some are monoecious in which some flowers are either female or male in the same individual, and only *c.* 6% of angiosperms are dioecious, characterized by the appearance of individuals having separate sexes (Renner, 2014). The sex determination pathways differ significantly among plant species with various breeding systems (Fig. 1), and in dioecious species, they are presumed to be under control of nuclear gene(s), usually located in the sex chromosomes. Among species with sex chromosomes, only a few have heteromorphic sex chromosomes, in a similar way to most animals. Although in plants dioecy is uncommon, it has usually emerged from hermaphroditism: it is likely that the 15 600 known dioecious species would have appeared in 900–5000 independent transitions from hermaphrodite ancestors (Renner, 2014), usually through an intermediate (Barrett, 2002). Also, their sex chromosomes have emerged from a pair of autosomes (Charlesworth, 2015) in which there is initially partial lack of recombination. It can be assumed that sex-determining genes often are located in the sex chromosomes, albeit sex determination has been so far studied in a minority of dioecious plant species and in many cases, the nonrecombining region of the sex chromosomes is too small to be easily revealed (Hobza *et al.*, 2018). The relatively rare plant species with known sex chromosomes have become the subject of wide research. The main topics deal with the process of the Y/W chromosome degeneration, searching for sex-linked genes underlying the sex determination, or with mechanisms of epigenetic control of the dosage compensation (reviewed by Charlesworth, 1996).

Types of plant sex chromosomes, sex determination systems and their diversity

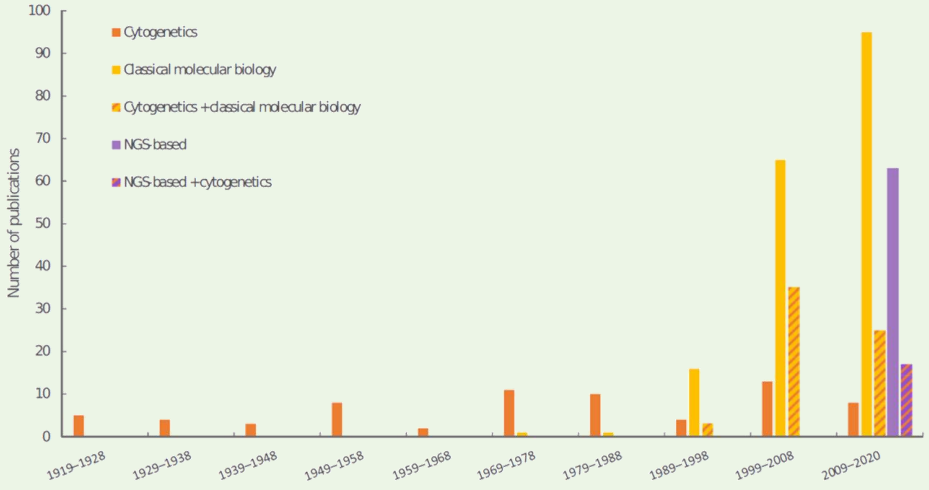
Most dioecious plants possess morphologically indistinguishable (homomorphic) sex chromosomes, whereas only a minority have conspicuously morphologically different (heteromorphic) sex chromosomes (Fig. 2). There are different opinions on this classification, because some chromosomes may seem homomorphic when observed by light microscopy but their heteromorphism is further revealed via other cytogenetic methods such as fluorescent *in situ* hybridization (FISH), or it is only apparent in meiotic chromosomes. In this report, we consider heteromorphism of the sex chromosomes if there is any karyological or molecular cytogenetics evidence.

We can distinguish between several types of sex determination, which could be very diverse across the Tree of Life. The best-characterized and most common in the Plant Kingdom is the one in which males are the heterogametic sex (XY) and females are the homogametic sex (XX) (Westergaard, 1958), which is the same system as mammals and some insects. Some models are derived from the XY, as the XX/XY_1Y_2 system – reported in some species, in which males possess two Y chromosomes (e.g. *Humulus japonicus* or *R. acetosa*) – and the $X_1X_1X_2X_2/X_1X_2Y$ – where there are four and two X chromosomes in females and males, respectively, as in some *Podocarpus* species (Zhou & Gu, 2001). Another mechanism observed in plants, although much less frequently, is the ZZ/ZW system: in this case, females are the heterogametic sex (ZW), whereas males are homogametic (ZZ) (Ming *et al.*, 2011). This system has been found in genera *Fragaria*, *Salix*, *Silene* and a few other genera (reviewed by Slancarova *et al.*, 2013, and Balounova *et al.*, 2019); it also is the sex determination of birds, and of some reptiles and insects. Besides, when sex is determined genetically in organisms with haploid-phase sex determination systems (such as green algae, bryophytes and brown algae), the chromosomes containing the sex-determining region (SDR) are referred to as U (female) and V (male) sex chromosomes (Bachtrog *et al.*, 2011; Coelho *et al.*, 2018). The sex determination systems vary largely amongst plant species, and we can observe this variability even within one genus or species. The last systems known in plants are the cytoplasmic male sterility, which is encoded by mitochondria, and the nuclear-encoded male fertility restorer genes (Saumitou-Laprade *et al.*, 1994; Caruso *et al.*, 2012). In some cases, sex expression is labile and is controlled by environmental conditions (Korpelainen, 1998). Often an interplay between genetic, epigenetic and physiological factors (e.g. the role



Version 1.0 (Barankova et al. 2020 New Phytologist)

- Data from **431 sources** (1919–2020)
- **178 species**, **84 genera**, **65 families**, 258 entries
- Composition (entries):
  - **87.6% angiosperms; within these: 92.9% eudicots, 6.6% monocots, 0.4% ANA grade**
  - **5.0% gymnosperms, 5.0% bryophytes, 2.3% green/red/brown algae**
- Sex-determination mechanisms:
  - **69.0% XY, 16.3% ZW, 5.4% UV; remainder other/uncertain**
- Morphology:
  - **48.5% homomorphic sex chromosomes, 31.0% heteromorphic, 20.5% not clearly assigned**



VI SIMPOSIO SEBOT
MERIDA RETURNS 24–30 NOVIEMBRE 2025

SEBOT
SOCIEDAD BOTÁNICA ESPAÑOLA

Sex-chrom: coverage and growth

Garcia et al. 2023

Chromosoma (2023) 132:55–58
https://doi.org/10.1007/s00412-023-00786-7

COMMENT

Sex-chrom v. 2.0: a database of green plant species with sex chromosomes

Sònia Garcia[1] · Bohuslav Janousek[2] · Joan Pere Pascual-Díaz[1] · Susanne S. Renner[3]

Received: 14 July 2022 / Revised: 20 January 2023 / Accepted: 23 January 2023 / Published online: 2 February 2023

Introduction

Sex chromosomes primarily determine the sex of an individual and ensure a stable sex ratio at conception in animals or at fertilization in plants. Their sex-specific specialization and evolutionary stability vary greatly among species, defying generalizations, but usually a pair of sex chromosomes will carry sex-determining genes (Kamiya et al. 2012; Renner and Müller 2021, 2022 for reviews). In some cases, chromosomal sex determination instead involves a balance mechanism that "counts" the relative numbers of X chromosomes versus autosomes and triggers downstream pathways of male or female development as, for example, in *Rumex* species with X/0 sex chromosome systems (Zuk 1963; Navajas-Pérez et al. 2005). Research on the mechanisms of sex determination is rapidly expanding, and we here present version 2 of a database first released in 2019 (Baránková et al. 2020). The database summarizes studies for green plants published between 1919 and March 2022.

Before the onset of molecular-cytogenetic methods and genomic data, studies of plant sex chromosomes focused on microscopically distinct, heteromorphic chromosomes that were unique to either males or females in separate-sexed gametophytes or sporophytes (e.g., Westergaard 1940; Newton 1984; Ramsay and Berrie 1982). In species with microscopically homomorphic chromosomes, such as *Asparagus*, *Ecballium*, or *Populus*, chromosomal sex determination was inferred from sex ratios in family pedigrees and other evidence, such as differences in staining ability, different fluorescent in situ hybridization (FISH) signals of repetitive elements confined to a single chromosome in one of the two sexes, or viable crosses with close relatives that have highly heteromorphic sex chromosomes. Such approaches have been used, for example, to infer sex chromosomes in date palm, *Phoenix dactylifera*, using differential chromomycin staining in multiple individuals of each sex (Siljak-Yakovlev et al. 1996), in the liverwort *Frullania dilatata* using sex-specific distribution of FISH signals in the two sexes (Sousa et al. 2021), and in *Silene heuffelii*, where heteromorphic sex chromosomes have been inferred because the species crosses freely with *S. latifolia*, which has strongly dimorphic sex chromosomes (Prentice 1978). In *Coccinia grandis*, the eukaryote species with the largest extent of sex chromosomes heteromorphism known so far, Naudin (1859, 1862) successfully hybridized it with *C. schimperi*, suggesting that the latter species might also have heteromorphic sex chromosomes, something recently confirmed by cytogenetic work (Janousek et al. 2022).

Information on sex chromosome morphology can be difficult to access because cytogenetic studies have often been published in highly specialized journals, not all of which are completely digitized (back to the early twentieth century). This makes answering new research questions about plant sex chromosomes unnecessarily difficult. For example, differences in chromosome size and morphology between sexes may relate to differences in organismal longevity (see Marais and Lemaître 2022), but obtaining the relevant data on either homomorphic or heteromorphic chromosomes in XY, ZW, and UV systems or plant longevity is fraught with difficulties.

Materials and methods

The first author's collection of article reprints and pdfs of studies in plant sex chromosomes from 1919 onwards was the basis for building version 1.0 of this database (Baránková et al. 2020). We have now revised and updated the data up to March 2022.

✉ Sònia Garcia
soniagarcia@ibb.csic.es

[1] Institut Botànic de Barcelona (IBB-CSIC), Passeig del Migdia S/N, 08038 Barcelona, Catalonia, Spain

[2] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic

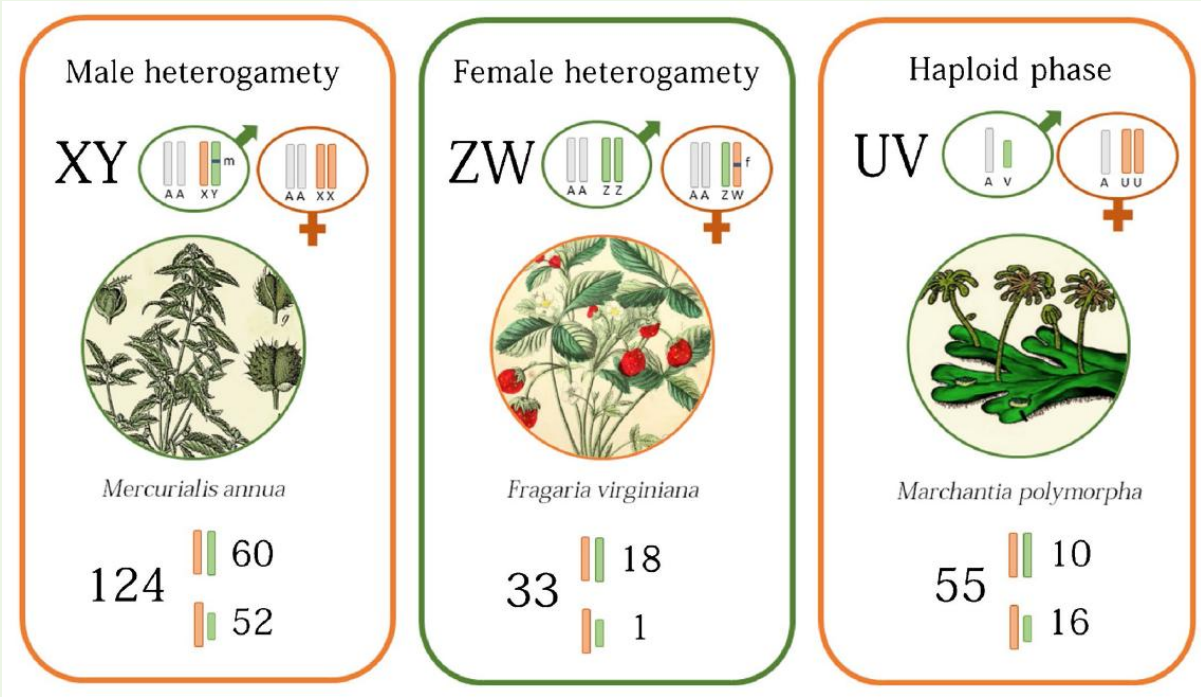
[3] Department of Biology, Washington University, Saint Louis, MO 63130, USA

Springer

Version 2.0 (Garcia et al. 20203 Chromosoma)

Sex-chrom v2

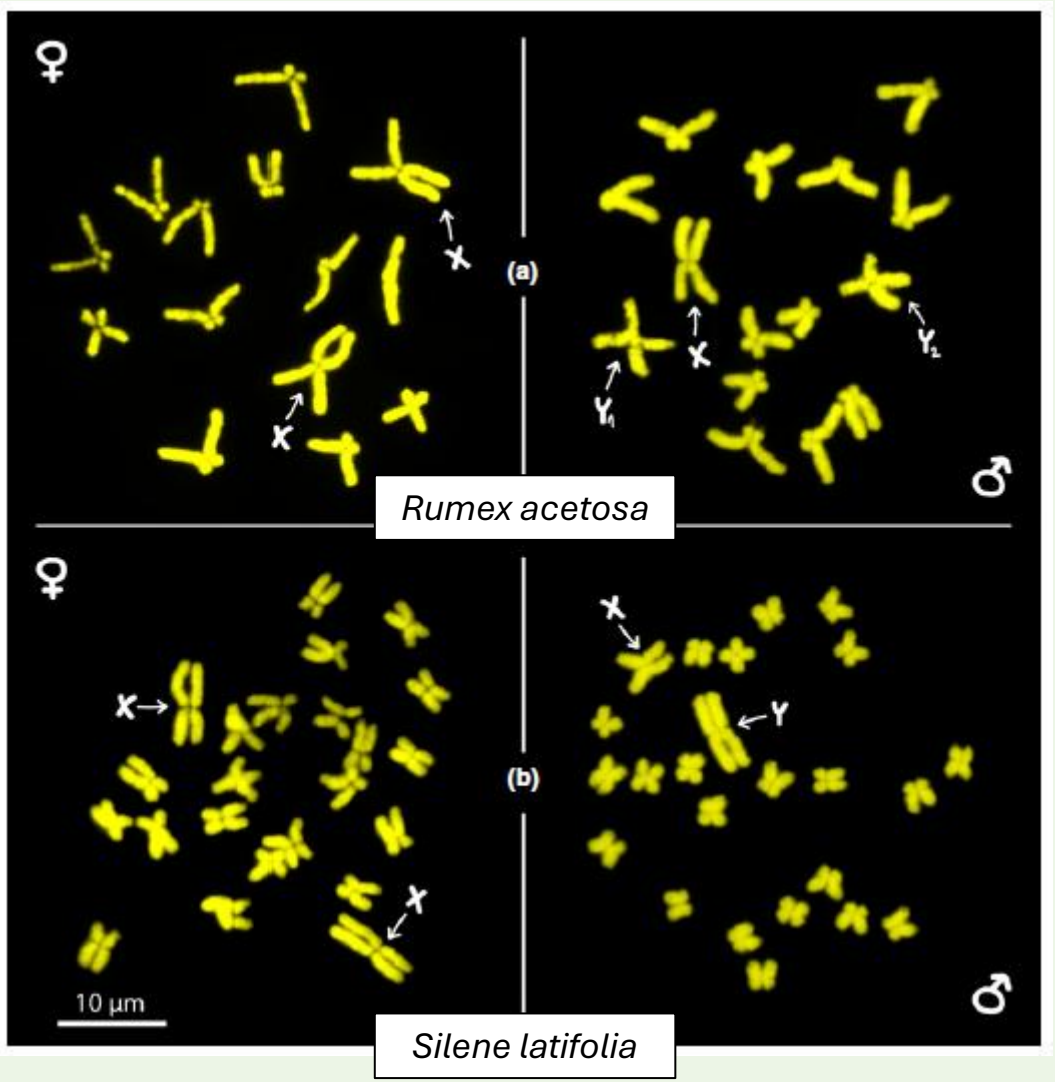
- Now **229 green plant species** with sex chromosomes



- Genome size data available for 160 species; **seven show male–female differences in 2C values, indicating strong chromosome heteromorphy**
- Morphology:
  - **88 species with homomorphic sex chromosomes**
  - **42 additional ones with suspected homomorphy**

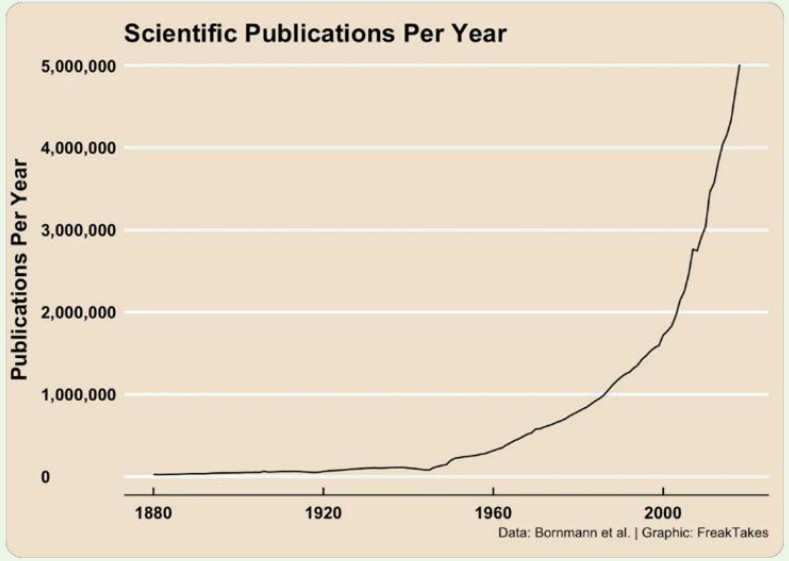
Sex-chrom: what the global data tell us

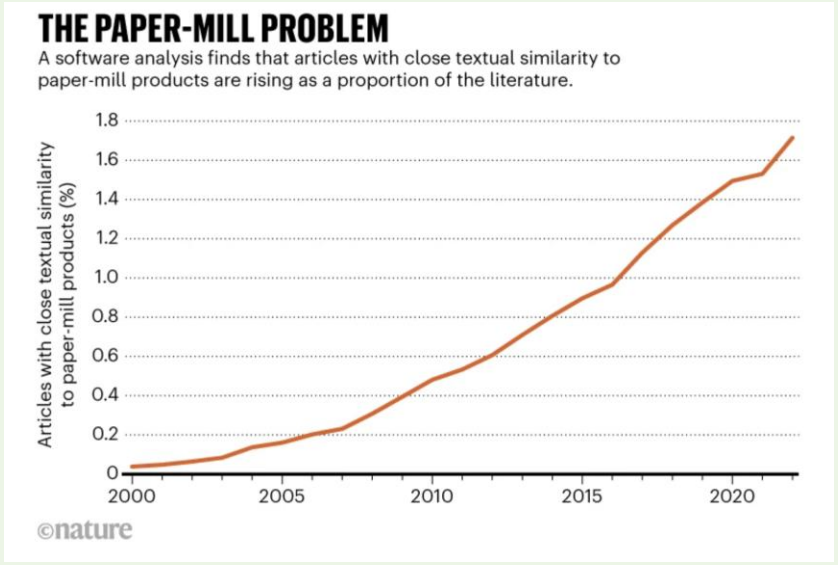
- **Sex systems are diverse and repeatedly evolved**
  - XY (male heterogamety) is most frequent, but ZW and UV are not rare
  - ZW systems are more common than previously thought, once molecular tools are applied
- **Morphology: many systems remain homomorphic**
  - Roughly half of known systems have homomorphic sex chromosomes; many others are suspected homomorphic but not yet fully resolved
  - Heteromorphy is common in XY systems (e.g. *Rumex, Silene, Coccinia*) but rare in ZW systems so far
- **Genome size and heteromorphy**
  - For several species, males and females differ in 2C value, reflecting size differences of sex chromosomes (e.g. *Cannabis sativa, Rumex acetosa, Viscum album*)
- **Strong methodological shift**
  - **Early decades dominated by classical cytology; since ~2010 clear move towards NGS-based approaches**
  - **Very few studies integrate both genomics and cytogenetics** in the same system, highlighting the gap Sex-chrom is designed to help bridge



Challenges: correctness, maintenance and update of the databases

1. **Keeping up with the literature explosion & fraud:** literature **volume grows faster** than curator time; each **new release is a big one-off effort** rather than a smooth, continuous process; growing **fraud allegations**
2. **Manual data extraction from heterogeneous papers:** extremely **time-consuming,** and difficult to automate without losing accuracy, especially for **nuanced cytogenetic traits**
3. **Terminology, inclusion criteria and "fuzzy edges":** defining consistent inclusion/exclusion **rules for keywords** and trait categories, and then applying them consistently over a century of very heterogeneous literature





Challenges: correctness, maintenance and update of the databases

1. **Taxonomic instability and name standardization:** every update requires revisiting and sometimes re-coding older records when taxonomy changes

1. **Quality control and error handling:** large portions of the curation effort are not data entry, but error detection and reconciliation of conflicting or ambiguous reports

Challenges: correctness, maintenance and update of the databases

1. **Technical and interface maintenance:** these are small, grant-dependent academic projects, not large funded infrastructures, so web-app maintenance competes with scientific curation for limited time and money
2. **Funding and human resources:** these databases depend on small teams, Msc students, predoc and postdoc contracts and project grants; long-term maintenance is structurally fragile

Possible solution: AI? Pilot test with the Plant rDNA database

The next update for the Plant rDNA database will be closed by December 2025 (time frame: 2021-2025)

Update workflow:

1. Literature search
2. Data extraction
3. Data cleaning and validation (quality control, error detection)
4. Web release

...

5. Data analysis

(together with a parallel resource: **the Animal rDNA database**)

Literature searches were performed with:

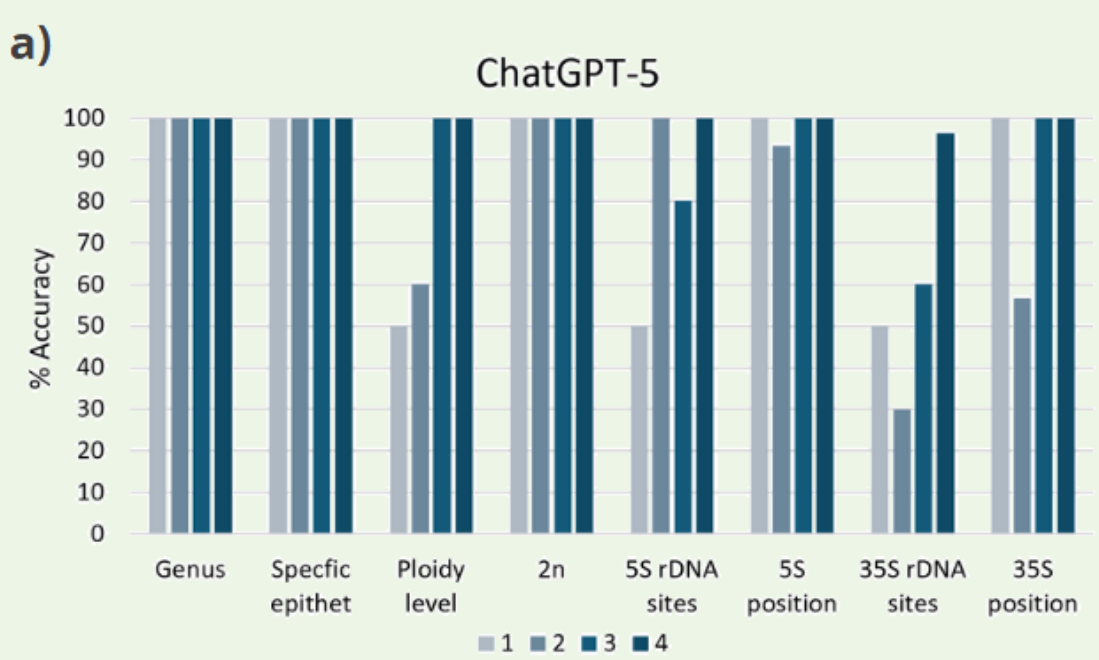
- Scopus
- Web of Science

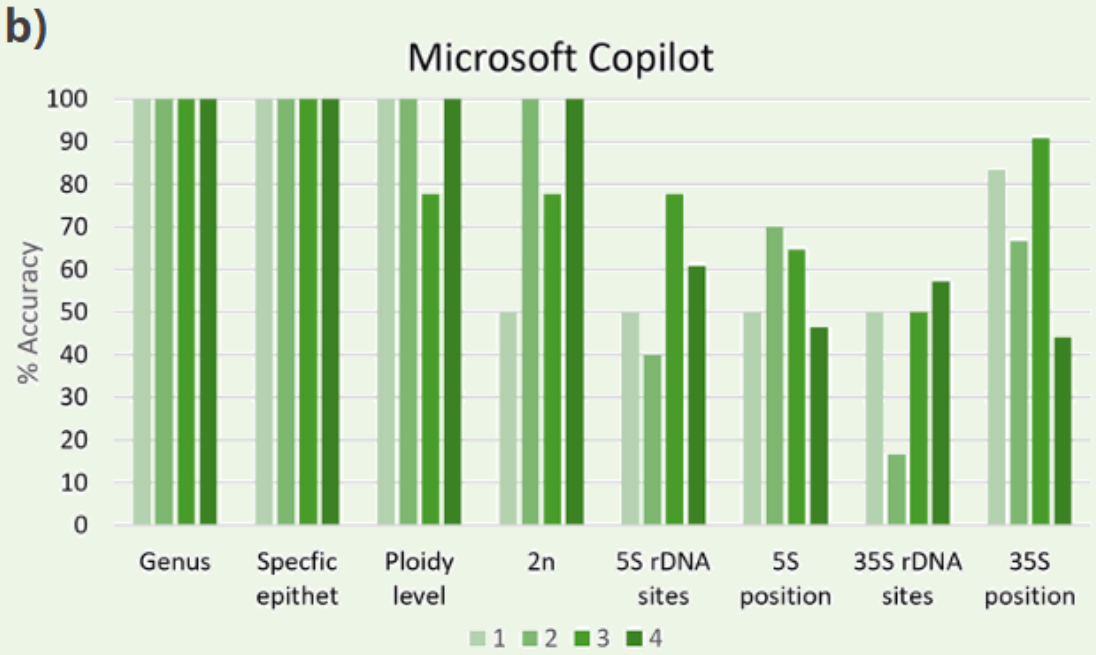
Using targeted keyword combinations (important step!)

A total of **340 new & relevant publications** were identified.

Possible solution: AI? Pilot test with the Plant rDNA database

- We established a set of criteria to assess the accessibility of information within each publication, assigning a score from **1 (least accessible)** to **4 (most accessible)**
- To assess AI reliability, **three publications for each category (1 to 4) were selected**, and data were extracted manually to serve as a reference for comparison
- An initial screening was conducted different publicly available AI tools, including ChatGPT, Gemini, and SciSpace, along with premium versions of ChatGPT-5 and Microsoft Copilot





Possible solution: AI? Pilot test with the Plant rDNA database

- ChatGPT-5 outperformed Microsoft Copilot, achieving higher accuracy across all categories and nearly reaching 100% accuracy for the most accessible papers

ChatGPT-5 > Microsoft Copilot

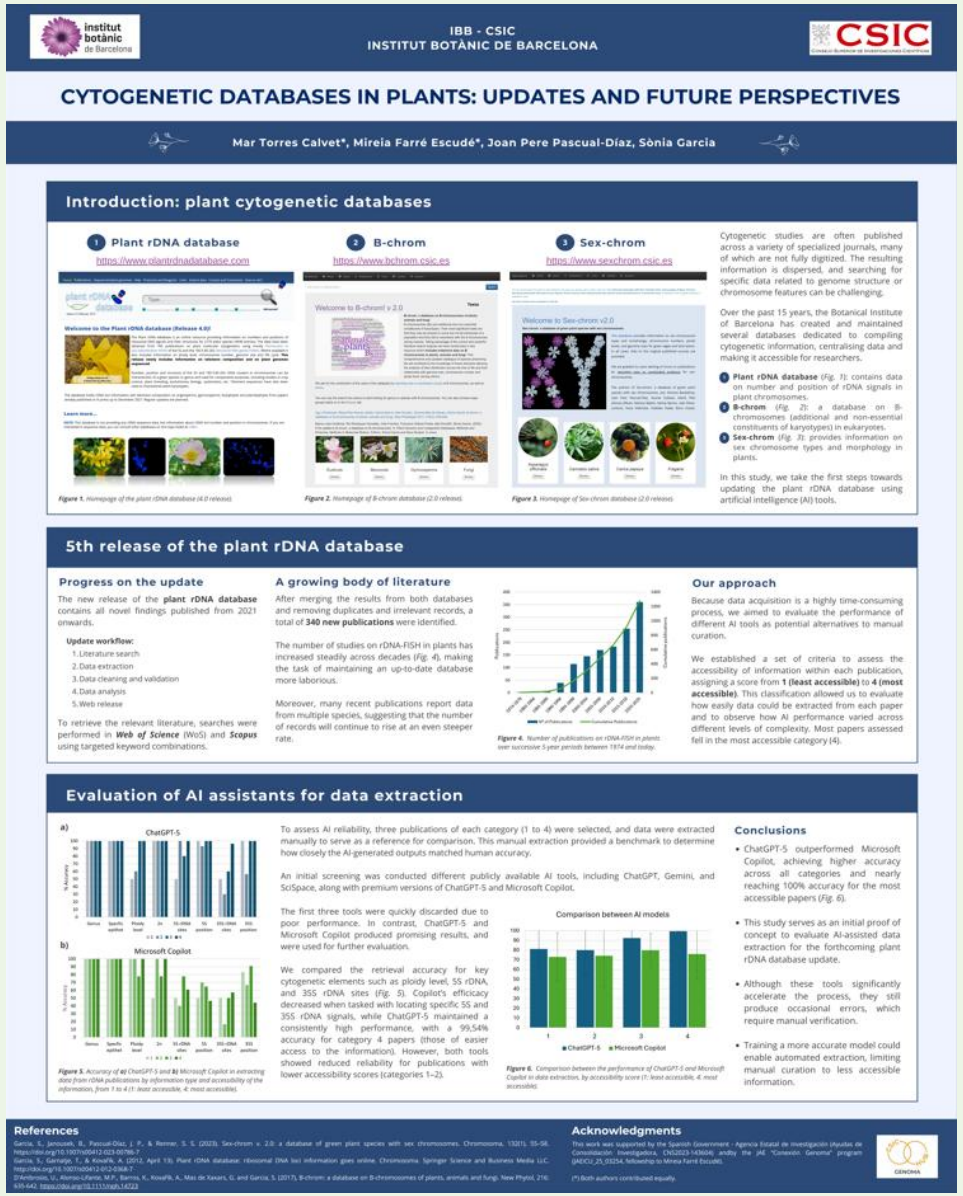
- This study serves as an **initial proof of concept** to evaluate AI-assisted data extraction for the forthcoming plant rDNA database update

- They still produce occasional errors, so **manual verification is still required**

- Training a more accurate model could enable automated extraction, **limiting manual curation to less accessible information**

Do you want to learn more? Ask Mar and Mireia!

POSTER
N. 37



IBB - CSIC
INSTITUT BOTÀNIC DE BARCELONA

CYTOGENETIC DATABASES IN PLANTS: UPDATES AND FUTURE PERSPECTIVES

Mar Torres Calvet*, Mireia Farré Escudé*, Joan Pere Pascual-Díaz, Sònia Garcia

Introduction: plant cytogenetic databases

1 Plant rDNA database — https://www.plantrdnadatabase.com
2 B-chrom — https://www.bchrom.csic.es
3 Sex-chrom — https://www.sexchrom.csic.es

Cytogenetic studies are often published across a variety of specialized journals, many of which are not fully digitized. The resulting information is dispersed, and searching for specific data related to genome structure or chromosome features can be challenging.

Over the past 15 years, the Botanical Institute of Barcelona has created and maintained several databases dedicated to compiling cytogenetic information, centralising data and making it accessible for researchers.

1. **Plant rDNA database** (*Fig. 1*): contains data on number and position of rDNA signals in plant chromosomes.
2. **B-chrom** (*Fig. 2*): a database on B-chromosomes (additional and non-essential constituents of karyotypes) in eukaryotes.
3. **Sex-chrom** (*Fig. 3*): provides information on sex chromosome types and morphology in plants.

In this study, we take the first steps towards updating the plant rDNA database using artificial intelligence (AI) tools.

5th release of the plant rDNA database

Progress on the update

The new release of the **plant rDNA database** contains all novel findings published from 2021 onwards.

Update workflow:
1. Literature search
2. Data extraction
3. Data cleaning and validation
4. Data analysis
5. Web release

To retrieve the relevant literature, searches were performed in ***Web of Science*** (WoS) and ***Scopus*** using targeted keyword combinations.

A growing body of literature

After merging the results from both databases and removing duplicates and irrelevant records, a total of **340 new publications** were identified.

The number of studies on rDNA-FISH in plants has increased steadily across decades (*Fig. 4*), making the task of maintaining an up-to-date database more laborious.

Moreover, many recent publications report data from multiple species, suggesting that the number of records will continue to rise at an even steeper rate.

Our approach

Because data acquisition is a highly time-consuming process, we aimed to evaluate the performance of different AI tools as potential alternatives to manual curation.

We established a set of criteria to assess the accessibility of information within each publication, assigning a score from **1 (least accessible)** to **4 (most accessible)**. This classification allowed us to evaluate how easily data could be extracted from each paper and to observe how AI performance varied across different levels of complexity. Most papers assessed fell in the most accessible category (4).

Evaluation of AI assistants for data extraction

To assess AI reliability, three publications of each category (1 to 4) were selected, and data were extracted manually to serve as a reference for comparison. This manual extraction provided a benchmark to determine how closely the AI-generated outputs matched human accuracy.

An initial screening was conducted different publicly available AI tools, including ChatGPT, Gemini, and SciSpace, along with premium versions of ChatGPT-5 and Microsoft Copilot.

The first three tools were quickly discarded due to poor performance. In contrast, ChatGPT-5 and Microsoft Copilot produced promising results, and were used for further evaluation.

We compared the retrieval accuracy for key cytogenetic elements such as ploidy level, 5S rDNA, and 35S rDNA sites (*Fig. 5*). Copilot's efficacy decreased when tasked with locating specific 5S and 35S rDNA signals, while ChatGPT-5 maintained a consistently high performance, with a 99,54% accuracy for category 4 papers (those of easier access to the information). However, both tools showed reduced reliability for publications with lower accessibility scores (categories 1–2).

Conclusions

- ChatGPT-5 outperformed Microsoft Copilot, achieving higher accuracy across all categories and nearly reaching 100% accuracy for the most accessible papers (*Fig. 6*).
- This study serves as an initial proof of concept to evaluate AI-assisted data extraction for the forthcoming plant rDNA database update.
- Although these tools significantly accelerate the process, they still produce occasional errors, which require manual verification.
- Training a more accurate model could enable automated extraction, limiting manual curation to less accessible information.

References

Acknowledgments

Mar Torres Calvet

Mireia Farré Escudé

ACKNOWLEDGEMENTS

Ales Kovarik, Mar Torres, Mireia Farré, Joan Pere Pascual-Díaz, Daniel Vitales, Teresa Garnatje, ,Ugo D'Ambrosio, Simona Barankova, Maria Luisa Gutiérrez, Roi Rodríguez, Suzanne Renner, Karina Barros, M. Pilar Alonso-Lifante, Jana Sochorova, Bohuslav Janousek, Andrew Leitch

institut botànic de Barcelona
Centre mixt CSIC museu de ciències naturals de Barcelona

CSIC
Consejo Superior de Investigaciones Científicas

Ayudas Consolidación Investigadora
CNS2023-143604
Ayudas Plan Estatal Investigación
PID2020-119163GB-I00

IBP

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro
AGENCIA ESTATAL DE INVESTIGACIÓN

Universidad de Oviedo

SEBOT
SOCIEDAD BOTÁNICA ESPAÑOLA

Generalitat de Catalunya

GReB
Grup de Recerca en Biosistemàtica i Biodiversitat Vegetal

Ajuts Suport Grups Consolidats
2021SGR00315