



## SyntenyViewer: a comparative genomics-driven translational research tool

Raphael Flores, Cécile Huneau, Laura Burlot, Mathilde Lainé, Erik Kimmel, Cyril Pommier, Michael Alaux, Anne-Françoise Adam-Blondon, Caroline Pont, Hadi Quesneville, et al.

### ► To cite this version:

Raphael Flores, Cécile Huneau, Laura Burlot, Mathilde Lainé, Erik Kimmel, et al.. SyntenyViewer: a comparative genomics-driven translational research tool. Database - The journal of Biological Databases and Curation, 2023, 2023, 10.1093/database/baad027 . hal-04099923

**HAL Id: hal-04099923**

**<https://hal.inrae.fr/hal-04099923>**

Submitted on 17 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# SyntenyViewer: a comparative genomics-driven translational research tool

Raphael Flores<sup>1,2,†</sup>, Cécile Huneau<sup>3,†</sup>, Laura Burlot<sup>1</sup>, Mathilde Lainé<sup>1</sup>, Erik Kimmel<sup>1,2</sup>, Cyril Pommier<sup>1,2</sup>, Michael Alaux<sup>1,2</sup>, Anne-Françoise Adam-Blondon<sup>1,2</sup>, Caroline Pont<sup>3</sup>, Hadi Quesneville<sup>1,2</sup> and Jerome Salse<sup>1,2,\*</sup>

<sup>1</sup>Université Paris-Saclay, INRAE, URGI, Versailles 78026, France

<sup>2</sup>Plant Bioinformatics Facility, Université Paris-Saclay, INRAE, Bioinformatics, Versailles 78026, France

<sup>3</sup>UCA, INRAE, GDEC 1095, Clermont-Ferrand 63000, France

\*Corresponding author: Tel: +00 33 4 43 76 15 29; Email: [jerome.salse@inrae.fr](mailto:jerome.salse@inrae.fr)

†These authors contributed equally to this work.

Citation details: Flores, R., Huneau, C., Burlot, L. *et al.* SyntenyViewer: a comparative genomics-driven translational research tool. *Database* (2023) Vol. 2023: article ID baad027; DOI: <https://doi.org/10.1093/database/baad027>

## Abstract

**SyntenyViewer** is a public web-based tool relying on a relational database available at <https://urgi.versailles.inrae.fr/synteny> delivering comparative genomics data and associated reservoir of conserved genes between angiosperm species for both fundamental (evolutionary studies) and applied (translational research) applications. SyntenyViewer is made available for (i) providing comparative genomics data for seven major botanical families of flowering plants, (ii) delivering a robust catalog of 103 465 conserved genes between 44 species and inferred ancestral genomes, (iii) allowing us to investigate the evolutionary fate of ancestral genes and genomic regions in modern species through duplications, inversions, deletions, fusions, fissions and translocations, (iv) use as a tool to conduct translational research of key trait-related genes from model species to crops and (v) offering to host any comparative genomics data following simplified procedures and formats

Database URL: <https://urgi.versailles.inrae.fr/synteny>

### Key points

- (i) SyntenyViewer is a web resource to perform comparative genomics in plants;
- (ii) SyntenyViewer allows access to expertised data and to download novel analyses;
- (iii) SyntenyViewer provides methods, scripts, documents and procedures to generate comparative genomics data.

## Introduction

Flowering plants, or angiosperms, emerged some 120–250 million years ago, depending on the dating approach (1–3), to rapidly diversify into 350 000 species alive today (4–7). These species are divided into two main groups, the monocots and eudicots, which, respectively, account for 20% and 75% of the plant diversity characterized to date (6). Cost reduction and technical improvements in sequencing technology make increasingly available public high-quality plant genome sequences offering the opportunity to conduct in-depth comparative genomics (8). Knowledge on gene

functions in relation to traits and processes as well as genome evolutionary dynamics is gained from accurate comparative genomics investigation. In that regard, several public tools are available to query comparative genomics data between plant genomes such as PLAZA (9), Gramene (10), Ensembl (11), CoGe (12) and Genomicus (13). However, methodologies can differ in defining conserved genes between species, making it particularly difficult to take into account recurrent whole-genome duplication (WGD) events in plant paleohistory, leading to artefactual identification of conserved genes (14). All extant species are either ancient (paleo-) or modern (neo-) polyploids derived from either the doubling of a single parental genome (autopolyploidy, AA deriving AAAA) or the hybridization of two parental genomes (allopolyploidy, AA × BB deriving AABB) (15). Consequently, all extant genomes may contain more than one copy of each ancestral gene. However, the accepted subgenome fractionation mechanism following polyploidization and consisting in the bias erosion of the ancestral gene content between the two parental genomes in the newly formed polyploid species, and then leading to least fractionated (LF) and most fractionated (MF) genomic fractions, leads to the progressive deletion of duplicated genes over time (16). Then,

Received 23 October 2022; Revised 18 March 2023; Accepted 14 April 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

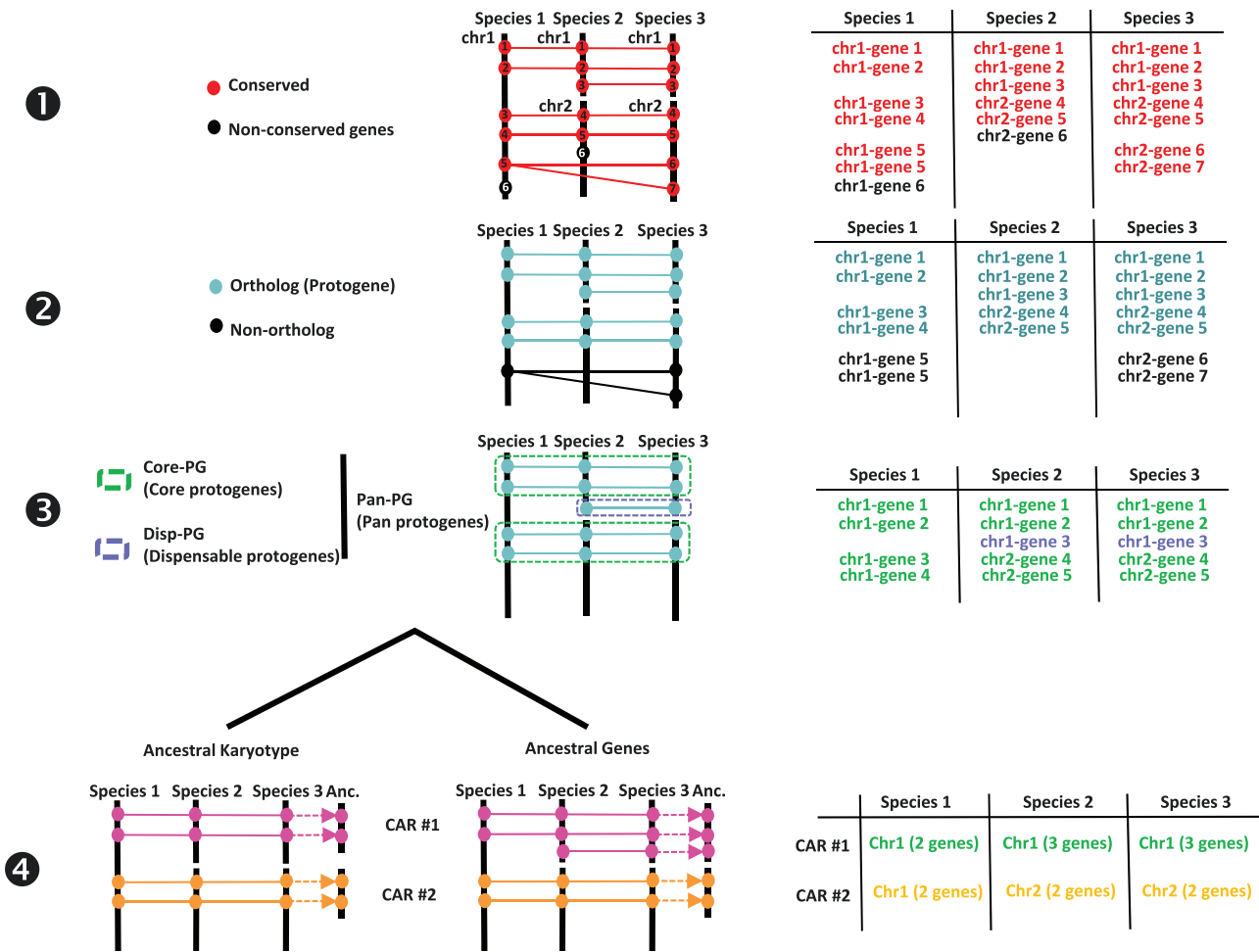
recurrent polyploidization–fractionation cycles in the course of plant evolution make the precise identification of conserved (orthologs) and duplicated (paralogs) genes in plant comparative genomics studies difficult. This article presents SyntenyViewer, a web-based tool hosting expertised synteny relationships between angiosperm genomes through the reconstruction of ancestral genomes (17), and discusses potential uses of the delivered catalog of conserved genes for evolutionary studies as well translational research investigation.

Materials and methods

Synteny inference through ancestral genome reconstruction

From an ancestral (possibly extinct) genome that evolved into different extant species through speciation and distinct chromosome shuffling events (fusions, fissions, inversions and translations), each of the ancestral chromosomes will derive a subset of extant chromosomal regions sharing synteny. Following this evolutionary evidence when reconstructing ancestral karyotypes *in silico*, comparative genomics of modern genomes should produce genomic fragments showing

independent (non-shared) syntenic blocks, referred to as conserved ancestral regions (CARs), which are considered as ancestral chromosomes in the inferred ancestral karyotype. We proposed (14, 17–20) a four-step method to infer ancestral genomes from BLAST-based comparison of modern genomes (Figure 1). The genes (protein sequences) from the investigated precise are compared using BLASTP with thresholds for cumulative identity percentage (CIP) ≥ 50% and cumulative alignment length percentage blast parameter (CALP) ≥ 50%, which deliver conserved genes between the investigated species using the following formulas:  $CIP = \sum nb ID by (\frac{HSP}{AL}) \times 100$ , where CIP corresponds to the cumulative percent of sequence identity observed for all the high-scoring pairs (HSPs) divided by the cumulative aligned length (AL), which corresponds to the sum of all HSP lengths; and  $CALP = \frac{AL}{Query length}$ , where CALP is the sum of the HSP lengths (AL) for all HSPs divided by the length of the query sequence. With these parameters, BLAST produces the highest cumulative percentage identity over the longest cumulative length, thereby increasing stringency in defining conserved genes between two genome sequences. From the previous BLAST comparison, the first step consists in



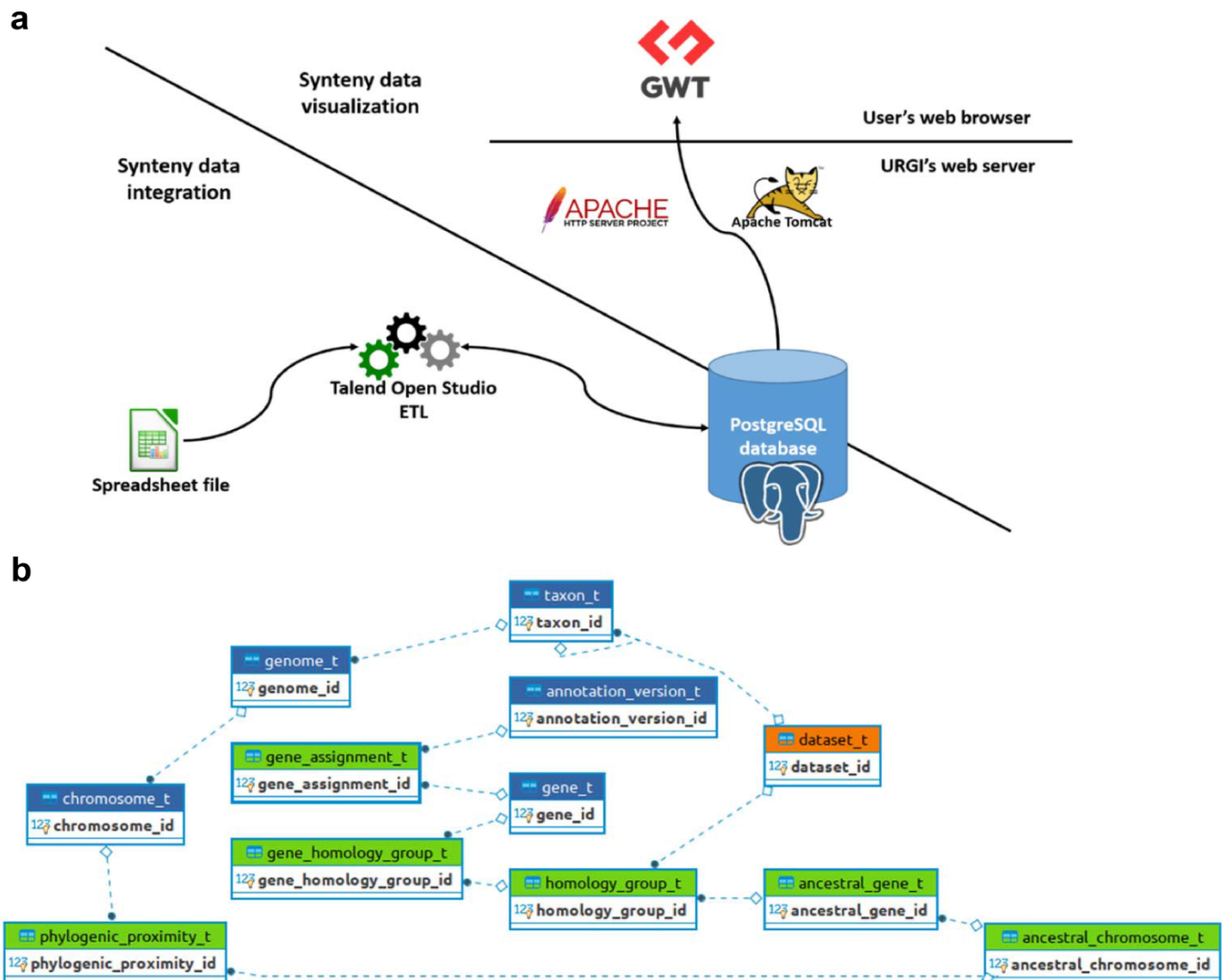
**Figure 1.** Procedure for reconstructing ancestral karyotypes. Ancestral genomes are inferred from (see the Materials and methods section) conserved genes (Step 1), orthologous relationships (Step 2), SBs (Step 3) and CARs (Step 4), to provide the best scenario explaining the transition between ancestral and modern genomes. Types of tabular files derived from each step are illustrated at the right to help readers to properly follow the procedure (described in and adapted from (20)).

retaining conserved genes. The second step consists in retaining single-copy orthologs and removing species-specific and tandem duplicates. This step consists in extracting one-to-one gene relationships (or 1- $n$  relationships for  $n$  WGD events) between species from the Step 1 output file. The third step consists in clustering or chaining groups of conserved genes into synteny blocks (SBs), which reveal core protogenes (Core-PGs) conserved in all the investigated species or dispensable PGs (Disp-PGs) between conserved genes in a subset (at least two) of the investigated species. This step consists of extracting all combinations of chromosome-to-chromosome relationships (for SBs sharing more than five orthologous genes) from the Step 2 output file. In the fourth step, SBs from the previous output file are then merged into ancestral protochromosomes (also referred to as CARs). This step

consists of defining independent groups of SBs sharing synteny between the modern species investigated. When the ancestral karyotype has been defined in its chromosome structure, conserved genes beyond one-to-one gene relationships between species (from Step 1) can be included in each protochromosome.

### SyntenViewer database interface

SyntenViewer is a tool relying on a relational database (DB), aiming at displaying and making publicly available the previously described comparative genomics data at <https://urgi.versailles.inrae.fr/synteny>. The Java web application uses the Google Web Toolkit (GWT) framework for graphical dynamic web content processing. On the back end, the web server uses



**Figure 2.** SyntenViewer data processing and database description. (a) Illustration of the SyntenViewer architecture including the data integration step into a PostgreSQL instance, and the data visualization based on GWT powered by Apache Tomcat and Apache HTTP server. (b) Illustration of the SyntenViewer database model with for each box a table named with its primary key term (referenced below). Tables colored in green store data in an append only fashion when a new synteny dataset is submitted, blue tables contain new data as well as data shared between different datasets, orange table stores new data as well as updated data from a previously inserted dataset (i.e. 'dataset\_t' that handles several versions of a dataset: a new tuple is inserted for Version 2 of Dataset A, while the tuple with Version 1 is marked obsolete). Some technical relationships between tables have been hidden for clarity. The database is structured below.

'Table dataset\_t': this contains information about each dataset (name, version, DOI). 'Table gene\_assignment\_t': this contains genomic information of a gene (position, strand, phase...). 'Table homology\_group\_t': this represents a group of conserved genes. 'Table gene\_homology\_group\_t': this makes the link between a gene and its orthologous groups. 'Table ancestral\_gene\_t': this contains the information of PGs. 'Table ancestral\_chromosome\_t': this stores the reconstructed protochromosomes. 'Table phylogenetic\_proximity\_t': this allows us to store the evolutionary distance between a reconstructed protochromosome and its descendant (modern) chromosomes.

Apache HTTP and Apache Tomcat, while the DB management system relies on a PostgreSQL 9.6 instance to store the data ensuring referential integrity (Figure 2).

Results

Genome synteny between angiosperm species and within major botanical families

Genome synteny has been obtained following a four-step method consisting in the identification of conserved genes (Step 1), orthologous relationships (Step 2), SBs (Step 3) and CARs (Step 4), Figure 1. Following this methodology, SyntenyViewer delivers published comparative genomics data (listed in Table 1) obtained for the two major angiosperm families with the grasses within the monocots [ancestral grass karyotype (AGK) with 12 protochromosomes and 16 560 PGs (21, 22)] and the eudicots [ancestral eudicot karyotype (AEK) with 21 chromosomes and 10 286 PGs (23)]. SyntenyViewer also provides published comparative genomics data for angiosperm lineages of agronomical interest such as *Rosaceae* [ancestral *Rosaceae* karyotype (ARK) with nine protochromosomes and 8861 PGs (24, 25)], *Brassicaceae* [ancestral *Brassicaceae* karyotype (ABK) with eight protochromosomes and 20 037 PGs (26, 27)], *Cucurbitaceae* [ancestral *Cucurbitaceae* karyotype (ACuK) with 22 protochromosomes and 17 969 PGs (28, 29)], legumes [ancestral legume karyotype (ALK) with 16 protochromosomes and 13 181 PGs (30–32)] and *Solanaceae* [ancestral *Solanaceae* karyotype (ASK) with 17 protochromosomes and 17 879 PGs (33)]. All datasets are licensed under the Open Licence Version 2.0 (CC-BY-compatible) as described in the ‘Terms’ tab of each dataset from the French national scientific data repository Recherche Data Gouv (RDG) <https://recherche.data.gouv.fr/en>. For example, see the Terms tab

of ‘PlantSyntenyViewer Solanaceae submission file’ where the license is also prompted when a user attempts to download the associated file (Table 1). Genome version and information are available at the RDG portal (see references).

Data integration and query in SyntenyViewer

Previous synteny data are integrated into the SyntenyViewer tool relying on a DB with a Java web application for graphical dynamic web content processing, a web server (Apache HTTP and Apache Tomcat) and a PostgreSQL 9.6 instance to store the data (Figure 2). A spreadsheet-based data exchange format allows synteny data submission to SyntenyViewer, available at <https://urgi.versailles.inrae.fr/Data/Synteny/Data-submission>. It consists of a four-sheet file (in addition to a README that explains how to properly complete the whole file) regarding (i) the person to contact and authors of the data, (ii) genomic features (genes, position on chromosomes, annotation and genome versions) mainly from the Phytozome database (34), (iii) phylogenic relatedness between chromosomes of extant species and chromosomes of their inferred ancestors and (iv) homology groups that are used to store relationships between genes of several species, each gene being declared in the genome description sheet (see Point ii). Excel format is provided as an example for users to be completed, but text format is also possible for simplicity in the data submission process. An Extract-Transform-Load (ETL) toolbox, using open-source Talend Open Studio, is dedicated to validate the dataset consistency and completeness as well as its database insertion. This ETL is able to manage data updates on previously integrated datasets, by only inserting the changes, masking the previous versions and ultimately validating consistency and unicity of the data provided by users. Part of this validation step to avoid

Table 1. Ancestral plant genomes

Family	Dating	Ancestor	Chromosome number	Gene number	Species	Data accessed
Eudicots	87–109	AEK (post-γ)	21	10286	Papaya, Arabidopsis thaliana, cacao, soybean, lotus, apple, strawberry, poplar, grape	Murat <i>et al.</i> 2015 (23)
Grasses	65–81	AGK (post-ρ)	12	16560	Rice, <i>Brachypodium</i> , barley, wheat, setaria, sorghum, maize	Murat <i>et al.</i> 2010 (21), 2014 (22)
<i>Brassicaceae</i>	27–40	ABK (post-α/β)	8	20037	<i>Arabidopsis thaliana</i> , <i>Arabidopsis lyrata</i> , <i>Capsella rubella</i> , <i>Brassica rapa</i> , <i>Thellungiella parvula</i>	<a href="https://doi.org/10.15454/DKXVAC">https://doi.org/10.15454/DKXVAC</a>
<i>Rosaceae</i>	70–90	ARK (post-WGD)	9	8861	Strawberry, rose, peach, apricot, apple, pear	<a href="https://doi.org/10.15454/GUJBZB">https://doi.org/10.15454/GUJBZB</a>
<i>Cucurbitaceae</i>	25–50	ACuK (post-WGD)	22	17969	Melon, cucumber, gourd, watermelon, squash	<a href="https://doi.org/10.15454/A96TW6">https://doi.org/10.15454/A96TW6</a>
Legumes	56–59	ALK (post-WGD)	16	13181	Peanut, lotus, chickpea, garden pea, barrel medic, pigeon pea, soybean, common bean, mung bean, adzuki bean, lupin	<a href="https://doi.org/10.15454/J9RN5S">https://doi.org/10.15454/J9RN5S</a>
<i>Solanaceae</i>	20–25	ASK (post-WGD 49 mya)	17	17879	Tomato, pepper, tobacco, sesame	<a href="https://doi.org/10.15454/TRBVMD">https://doi.org/10.15454/TRBVMD</a>

Summary of reconstructed ancestral angiosperm genomes listing the targeted botanical family, dating (in million years), the ancestral genome name (with WGD defining the delivered post-polyploidization ancestors in parentheses), number of protochromosomes, number of PGs, associated extant species involved and the link to the raw data information (README: description of the data provided in the table; ‘CONTACT’: person to contact for information on the data provided; ‘GENOME’: all versions of genomes used; ‘PHYLOGENY’: synteny information between chromosomes and derived ancestral chromosomes; ‘HOMOLOGY\_GROUP’: number of conserved genes and corresponding conserved chromosomes)



conflicts is handled into the ETL tool directly before integration into the database. The database itself relies heavily on unicity constraints over identifiers, some composite keys and some concatenations of entity (gene, dataset and chromosome) name and version. It contains several constraints used to ensure complete consistency over time between several integrations of updated versions. The version of the application sticks to the GnpIS information system (35). Dataset's versions are displayed in the dataset form and detailed in the associated downloadable dataset from the RDG repository. Synteny data is then made publicly available when the format described on our website (<https://urgi.versailles.inrae.fr/Data/Synteny/Data-submission>), and the aims of SyntenyViewer are met. The usage of an all-in-one Excel file simplifies the data exchange between both parties. It also eases its manipulation by scientists for filling and submitting a unique file, which includes some static data extracted from the database (i.e. taxon scientific names), and hence guides the submitter with correct data at the beginning of the process and reduces the need for interactions between both parties. For data upload, the file can be provided through a 'minimal web form', allowing us to track the submission versions. Also, exchanges with the application maintainers (The Plant Bioinformatics Facility from URGI) are still possible by e-mail using the [urgi-support@inrae.fr](mailto:urgi-support@inrae.fr) address.

### SyntenyViewer functionalities

There are two main entry points to visualize SyntenyViewer data (*cf* Supplementary Video). The first allows for selecting a dataset that provides gene conservation among a given botanic family or at a larger scale to whole monocot (grasses) or eudicot phylum, which then shows a query form described later. The other entry point allows for searching for a gene name (*via* its entire form or a prefix) across all datasets available. A search displays a popup with a short description of matching genes referenced in the database, and clicking on a selected gene loads the associated orthologous genes in the selected botanic family of interest, with an additional form for querying data in several manners as well as customizing display parameters. This query form offers users to enter the database through a gene ID, an extant or ancestral chromosome number and species of interest. The customized display parameters offer users to display windows with a specific number of (extant or ancestral) genes, to produce a compact view when having numerous genes visible, to swap gene order on chromosome or to hide chromosomes. Orthologous genes are given the same color code and are linked in a top-down manner between chromosomes, facilitating the identification of orthologous groups between genomes and species. Left clicking on a gene updates the synteny display centered on the selected gene. Right clicking a gene provides the associated gene information with its ID, its coordinates as well as links for redirecting toward numerous international databases (36, 37) following the Findable, Accessible, Interoperable, Reusable principles (38). Dedicated buttons make it possible to browse along the chromosomes. There is also a specific mode, which makes searches wider when no hit can be found for a specific gene on syntenic chromosomes: in such case, flanking genes are serially searched for orthologs on syntenic chromosomes until a first match is found, highlighting all relationships observed between all displayed genes of a given genomic region. At any stage, the SyntenyViewer's

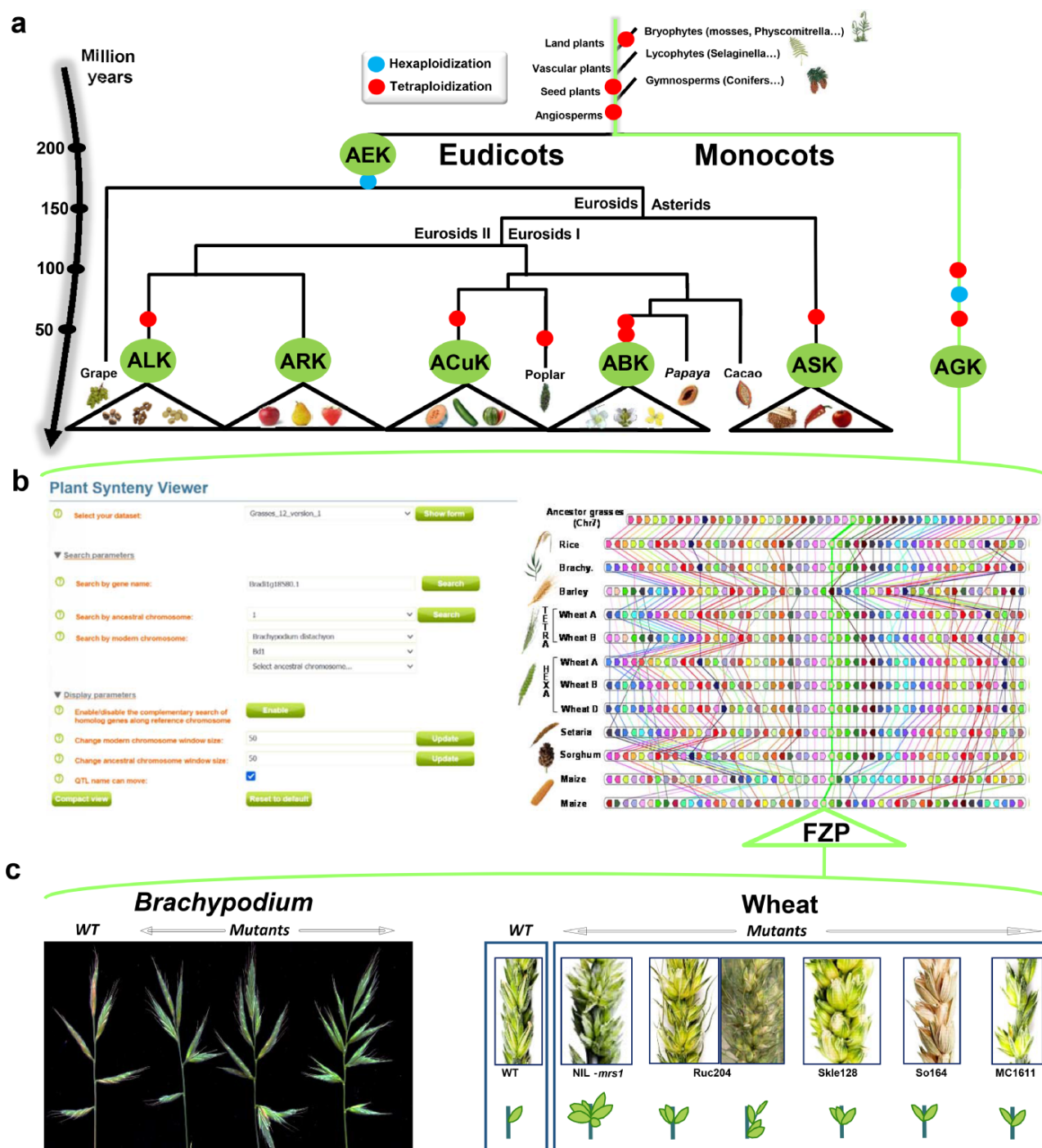
Uniform Resource Locator is dynamically updated, and it can be bookmarked in the browser for sharing the visualization link, which makes users able to go back to previous work and pursue exploration. Finally, a download button, available after dataset selection, allows users to access its publicly available submission file along with relevant metadata (authors, description and DOI) through the RDG repository or through any link provided by the submitter. Data are downloadable by users in a tabular format from which additional visualizations can be performed such as dotplots using classical R packages available.

### Discussion

The delivered SyntenyViewer tool gives public access to validated and reviewed comparative genomics data either between angiosperm species or within major botanical families that can be used as a backbone to investigate evolutionary trends of genes, perform translational research of traits and conduct evolutionary developmental biology (for Evo-Devo) investigation of traits.

With the use of reconstructed ancestral genomes, structural (intron and exon structure) and functional Gene Ontology annotations of genes can be improved by comparing orthologous gene sets that may share similar (ancestral) genomic features. The reconstructed ancestral karyotypes can also be used to infer a parsimonious evolutionary model that assumes minimal numbers of genomic rearrangements (including duplications, inversions, deletions, fusions, fissions and translocations). SyntenyViewer allows deep investigation of evolutionary fates of ancestral genes/genomes, through precise identification of the changes involved (gains and losses of genes and associated gene ontologies) and their assignment to specific species or botanical families (Figure 3a). Among major evolutionary events, WGD can be investigated in detail as well as post-polyploidization partitioning between paralogous blocks forming 'MF' (also known as S for sensitive) and 'LF' (also known as D for dominant) chromosomal compartments (39, 40).

SyntenyViewer can also be used as a useful tool for translational research on genes driving key agronomical traits, particularly from model species (such as *Arabidopsis thaliana*) to crops (41). Such translational-based dissection of traits has been performed successfully in several botanical families, including legumes [for example, between *Medicago truncatula* and pea (42)] or grasses (Figure 3b) with *Brachypodium* used as a pivotal genome to dissect wheat traits (43) related to yield [i.e. NUE for nitrogen use efficiency (44)] or bread-making quality [GFC for grain fiber content (45), as well as carotenoid content (46)], among other cases. As a case example, Figure 3c illustrates the translational-based cloning of FRIZZY PANICLE (FZP) genes in bread wheat (47). Bread wheat inflorescences, or spikes, are characteristically unbranched and normally bear one spikelet per rachis node. From the gene conservation information delivered from SyntenyViewer, further validation needs to be conducted to establish the conservation of the phenotype or trait between the investigated species. In the case of FZP, based on wheat mutants with supernumerary spikelets (SS) and comparative genomics data between *Brachypodium* and wheat, it has been shown that the orthologous FZP gene, encoding a member of the APETALA2/ethylene response factor (AP2/ERF) transcription factor family, drives the SS trait in *Brachypodium*, bread



**Figure 3.** SyntenyViewer, a comparative genomics-driven translational research tool. (a) ‘Plant genome evolution from reconstructed ancestors’. The present-day angiosperm species (bottom) are represented along the evolutionary tree of the Angiosperms from founder ancestors (AGK, AEL, ACuK, ASK, ARK, ABK and ALK) of major botanical families with the time scale shown on the left (in million years). The polyploidization events that have shaped the structure of modern plant genomes during their evolution from inferred ancestors are indicated by red dots (duplication) and blue dots (triplication). (b) ‘SyntenyViewer screen capture’. SyntenyViewer tool with the setting parameters (search by gene name and ancestral or modern chromosomes) illustrated at the left and the derived comparative genomics data visualization, as detailed in the text, at the right (here for cereals). Genes are illustrated as colored boxes for each species (in lines), so that conserved genes are linked with colored lines between species. (c) ‘Synteny-based translational research of FZP gene in grasses’. FZP gene characterization in grasses with orthologs from SyntenyViewer (Panel b) and functional validation in wheat and Brachypodium (in mutants compared to wild type) in deriving similar SS phenotypes (adapted from (47)).

wheat and rice (47, 48). Structural and functional characterization of the three wheat FZP homologous genes (WFZP-A-B-D of the allohexaploid bread wheat) revealed that coding mutations of WFZP-D cause the SS phenotype with the most

severe effect when WFZP-D lesions are combined with a frameshift mutation in WFZP-A (47–49).

Beside translational research of genes, SyntenyViewer allows us to conduct ‘Evo-Devo dissection of traits’.

SyntenViewer allows us to compare a group of angiosperm species that acquired new phenotypes (or traits in the broad sense) in the course of evolution, compared to a group of species that did not acquire this trait. Following this strategy, comparing woody to herbaceous angiosperms allowed us to link the life history of trees to the amplification in tandem of genes involved in immunity, which has thus been proposed as a key process underpinning longevity of such a long lifespan species (50). This comparative Evo-Devo framework can be used to provide a better understanding of the molecular bases of major agronomical interest, such as seasonality (comparing annual *vs.* perennial species), photosynthesis (comparing C3 *vs.* C4 species) as well as grain and fruit developmental and quality traits in crops.

Overall, SyntenViewer is a web-based tool delivering comparative genomics data either between angiosperm species or within major botanical families (including the *Rosaceae*, *Brassicaceae*, *Cucurbitaceae*, legume and *Solanaceae*) for evolutionary and translational research purposes.

## Supplementary material

Supplementary material is available at Database online.

## Data availability

All required links or identifiers are provided in the current article.

## Funding

The ‘Région Auvergne-Rhône-Alpes’ and ‘Fonds Européen de Développement Régional’ (#23000816 project ‘SRESRI-2015’), the Institut Carnot Plant2Pro (project ‘SyntenViewer’), the Initiative-Science-Innovation-Territoires-Economie (ISITE) CAP2025 (#00002146 SRESRI 2015, Pack Ambition Recherche Project ‘TransBlé’), the Agence Nationale de la Recherche (ANR) projects ‘BREEDWHEAT’ (ANR-10-BTBR-03-06), the Biomass For the Future project (ANR-11-BTBR-0006), the ANR project, for Plant and Animal Genome Evolution (PAGE) (ANR-11-BSV6-0008) and the Plant Bioinformatics Facility (<https://doi.org/10.15454/1.5572414581735654E12>).

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

The authors would like to thank the following staff members for their involvement at various stages of the project: Jérémy Destin and Julie Bogoin from INRAE-URGI and Florent Murat and Sébastien Guizard from INRAE-GDEC.

## References

1. Bell, C.D., Soltis, D.E. and Soltis, P.S. (2005) The age of the angiosperms: a molecular timescale without a clock. *Evolution*, **59**, 1245–1258.
2. Magallon, S. (2010) Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Syst. Biol.*, **59**, 384–399.
3. Barba-Montoya, J., Dos Reis, M., Schneider, H. *et al.* (2018) Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol.*, **218**, 819–834.
4. Friis, E.M., Pedersen, K.R. and Crane, P.R. (2006) Cretaceous angiosperm flowers: innovation and evolution in plant reproduction. *Palaeogeogr Palaeoclimatol Palaeoecol.*, **232**, 251–293.
5. Friis, E.M., Pedersen, K.R. and Crane, P.R. (2010) Diversity in obscurity: fossil flowers and the early history of angiosperms. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **365**, 369–382.
6. Soltis, D.E., Bell, C.D., Kim, S. *et al.* (2008) Origin and early evolution of angiosperms. *Ann. N. Y. Acad. Sci.*, **1133**, 3–25.
7. Doyle, J.A. (2012) Molecular and fossil evidence on the origin of angiosperms. *Annu. Rev. Earth Planet Sci.*, **40**, 301–326.
8. Marks, R.A., Hotaling, S., Frandsen, P.B. *et al.* (2021) Representation and participation across 20 years of plant genome sequencing. *Nat. Plants*, **7**, 1571–1578.
9. Van Bel, M., Diels, T., Vancaester, E. *et al.* (2018) PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.*, **46**, D1190–D1196.
10. Tello-Ruiz, M.K., Naithani, S., Stein, J.C. *et al.* (2018) Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Res.*, **46**, D1181–D1189.
11. Herrero, J., Muffato, M., Beal, K. *et al.* (2016) Ensembl comparative genomics resources. *Database (Oxford)*, **2016**, bav096.
12. Haug-Baltzell, A., Stephens, S.A., Davey, S. *et al.* (2017) SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*, **33**, 2197–2198.
13. Nguyen, N.T.T., Vincens, P., Dufayard, J.F. *et al.* (2022) Genomicus in 2022: comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Res.*, **50**, D1025–D1031.
14. Murat, F., Armero, A., Pont, C. *et al.* (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.*, **49**, 490–496.
15. Clark, J.W. and Donoghue, P.C.J. (2018) Whole-genome duplication and plant macroevolution. *Trends Plant Sci.*, **23**, 933–945.
16. Cheng, F., Wu, J., Cai, X. *et al.* (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants*, **4**, 258–268.
17. Salse, J. (2016) Ancestors of modern plant crops. *Curr. Opin. Plant Biol.*, **30**, 134–142.
18. Salse, J., Abrouk, M., Murat, F. *et al.* (2009) Improved criteria and comparative genomics tool provide new insights into grass paleogenomics. *Brief. Bioinformatics*, **10**, 619–630.
19. Pont, C., Wagner, S., Kremer, A. *et al.* (2019) Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol.*, **20**, 29.
20. Shi, T., Huneau, C., Zhang, Y. *et al.* (2022) The slow-evolving *Acorus tatarinowii* genome sheds light on ancestral monocot evolution. *Nat. Plants*, **8**, 764–777.
21. Murat, F., Xu, J.H., Tannier, E. *et al.* (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.*, **20**, 1545–1557.
22. Murat, F., Zhang, R., Guizard, S. *et al.* (2014) Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biol. Evol.*, **6**, 12–33.
23. Murat, F., Zhang, R., Guizard, S. *et al.* (2015) Karyotype and gene order evolution from reconstructed extinct ancestors highlight contrasts in genome plasticity of modern rosoid crops. *Genome Biol. Evol.*, **7**, 735–749.
24. Raymond, O., Gouzy, J., Just, J. *et al.* (2018) The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.*, **50**, 772–777.
25. Huneau, C., Flores, R.-G. and Salse, J. (2021) [dataset] PlantSyntenViewer *Rosaceae* submission file. Portail Data INRAE, V1.



26. Murat,F., Louis,A., Maumus,F. *et al.* (2015) Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol.*, **16**, 262.
27. Huneau,C., Flores,R.-G. and Salse,J. (2021) [dataset] PlantSyn-tenyViewer Rosaceae submission file. Portail Data INRAE, V1.
28. Wu,S., Shamimuzzaman,M., Sun,H. *et al.* (2017) The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a papaya ring-spot virus resistance locus. *Plant J.*, **92**, 963–975.
29. Huneau,C., Flores,R.-G. and Salse,J. (2021) [dataset] PlantSyn-tenyViewer Cucurbitaceae submission file. Portail Data INRAE, V1.
30. Kreplak,J., Madoui,M.A., Cápál,P. *et al.* (2019) A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.*, **51**, 1411–1422.
31. Hufnagel,B., Marques,A., Soriano,A. *et al.* (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.*, **11**, 492.
32. Huneau,C., Flores,R.-G. and Salse,J. (2021) [dataset] PlantSyn-tenyViewer Legumes submission file. Portail Data INRAE, V1.
33. Huneau,C., Flores,R.-G. and Salse,J. (2021) [dataset] PlantSyn-tenyViewer Solanaceae submission file. Portail Data INRAE, V1.
34. Goodstein,D.M., Shu,S., Howson,R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.
35. Adam-Blondon,A.F., Alaux,M., Durand,S. *et al.* (2017) Mining plant genomic and genetic data using the GnpIS information system. *Methods Mol. Biol.*, **1533**, 103–117.
36. Alaux,M., Rogers,J., Letellier,T., International Wheat Genome Sequencing Consortium. (2018) Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.*, **19**, 111.
37. Sen,T.Z., Caccamo,M., Edwards,D. *et al.* (2020) Building a successful international research community through data sharing: the case of the Wheat Information System (WheatIS). *F1000Research*, **9**, 536.
38. Wilkinson,M., Dumontier,M., Aalbersberg,I. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.
39. Soltis,P.S., Marchant,D.B., Van de Peer,Y. *et al.* (2015) Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.*, **35**, 119–125.
40. Salse,J. (2016) Deciphering the evolutionary interplay between subgenomes following polyploidy: a paleogenomics approach in grasses. *Am. J. Bot.*, **103**, 1167–1174.
41. Valluru,R., Reynolds,M.P. and Salse,J. (2014) Genetic and molecular bases of yield-associated traits: a translational biology approach between rice and wheat. *Theor. Appl. Genet.*, **127**, 1463–1489.
42. Bordat,A., Savoie,V., Nicolas,M. *et al.* (2011) Translational genomics in legumes allowed placing in silico 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3 (Bethesda)*, **1**, 93–103.
43. Quraishi,U.M., Pont,C., Ain,Q.U. *et al.* (2017) Combined genomic and genetic data integration of major agronomical traits in bread wheat (*Triticum aestivum* L.). *Front. Plant Sci.*, **8**, 1843.
44. Quraishi,U.M., Abrouk,M., Murat,F. *et al.* (2011) Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution. *Plant J.*, **65**, 745–756.
45. Quraishi,U.M., Murat,F., Abrouk,M. *et al.* (2011) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics*, **11**, 71–83.
46. Dibari,B., Murat,F., Chosson,A. *et al.* (2012) Deciphering the genomic structure, function and evolution of carotenogenesis related phytoene synthases in grasses. *BMC Genom.*, **13**, 221.
47. Dobrovolskaya,O., Pont,C., Sibout,R. *et al.* (2015) FRIZZY PANICLE drives supernumerary spikelets in bread wheat. *Plant Physiol.*, **167**, 189–199.
48. Wang,Y., Du,F., Wang,J. *et al.* (2022) Improving bread wheat yield through modulating an unselected AP2/ERF gene. *Nat. Plants*, **8**, 930–939.
49. Bai,X., Huang,Y., Mao,D. *et al.* (2016) Regulatory role of FZP in the determination of panicle branching and spikelet formation in rice. *Sci. Rep.*, **6**, 19022.
50. Plomion,C., Aury,J.M., Amselem,J. *et al.* (2018) Oak genome reveals facets of long lifespan. *Nat. Plants*, **4**, 440–452.