Exchange your knowledge on plant gene families

Guignon V[1], Cenci A[1], Rouard M[1]

[1] Bioversity International - Commodity Systems & Genetic Resources Programme, Parc Scientifique Agropolis II, 34397Montpellier - Cedex 5, France

ABSTRACT: Every year, hundreds of **gene families** are characterized through peer-review publications but their structure and classification is hardly captured by dedicated databases such as Phytozome, Plaza, Ensembl or GreenPhylDB. To address this situation, we developed a user-friendly interface allowing either to customize pre-computed protein sequence clusters or to create new ones based on prior knowledge of a given gene family. Information can be then **shared with collaborators and/or reviewers with a unique URL**.

1. Rationale

Many studies on plant gene families...

...but is there any repository?

Get started with **GreenPhyl** to add value to your gene families!

- easy gene family creation
- feature management
- nice display

GreenPhyl v4

What is GreenPhyl?
A published public web resource designed for **comparative** and **functional genomics** in plants:

- 4th release
- 37 complete plant **genomes**
- 1 300 000 gene sequences
- 8 300 sequence clusters
- 2 900 annotated clusters

2. Searching

Family directory

Family search

3. Display

a. Species composition

b. Phylogeny analysis display

External tools: Galaxy

c. (Sub)family relationship

d. Sequence export and mapping

4. Creation

a. Building gene family **from GreenPhyl** data using "*My List*" feature

"*My List*" is a clipboard-like space storing a set of sequences. This set can be modified using another set of selected sequences or gene families by applying some basic operations:

Selection — My List

operations

Merge — Replace — Intersect — Substract

b. Building gene family **from FASTA** sequences

step 1: Multiple-FASTA import

step 2: Species validation

supports all plant UniProt codes (>4000 species)

5. Management

Content
a. Sequence management

b. Relationship management

c. Phylogeny analysis

Properties
dynamic edition

User account

Free, required for gene family creation, **OpenID** supported: use your Google, Yahoo!, etc. account to log in!

5 access levels

public - read only
private
restricted by password - read only
restricted by password - editable
public - read only
public - editable

Access control

Designed for **peer-review** process

private
for work in progress

restricted (read or read/edit)
for publication review process, collaboration, ...

"private" URL or password

Public access (read or read/edit)
for visibility or wiki-style gene family
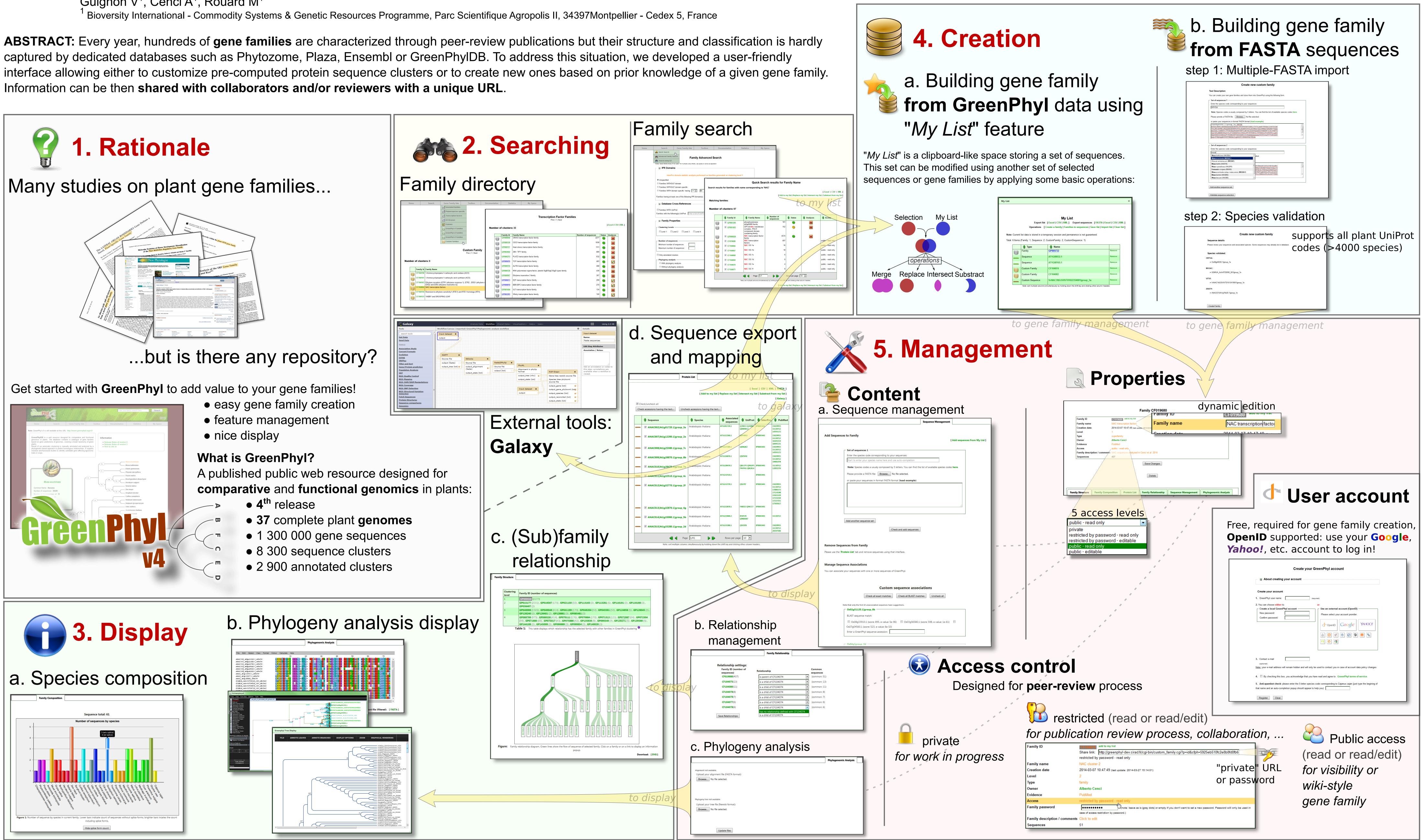
References:
1. Cenci A, Guignon V, Roux N, Rouard M. **Genomic analysis of NAC transcription factors in banana (Musa acuminata) and definition of NAC orthologous groups for monocots and dicots.** Plant Mol Biol. 2014 Feb 26.
2. Jourda, C., Cardi, C., Mbéguié-A-Mbéguié, D., Bocs, S., Garsmeur, O., D'Hont, A. and Yahiaoui, N, (2014), **Expansion of banana (Musa acuminata) gene families involved in ethylene biosynthesis and signalling after lineage-specific whole-genome duplications.** New Phytologist. doi: 10.1111/nph.12710.
3. Rouard,M., Guignon,V., Aluome,C., Laporte,M.-A., Droc,G., Walde,C., Zmasek,C.M., Perin,C. et Conte,M.G. (2011) **GreenPhylDB v2.0: comparative and functional genomics in plants.** Nucleic Acids Research, 39, D1095-D1102, 10.1093/nar/gkq811.
4. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009), **Jalview Version 2 - a multiple sequence alignment editor and analysis workbench.** Bioinformatics25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033
5. Zmasek C.M., Eddy S.R. **ATV: display and manipulation of annotated phylogenetic trees.** Bioinformatics. 2001, Apr;17(4):383-4.