Toward the improvement of the gene centered information in grapevine genomics

Jérôme Grimplet, Anne-Francoise Adam-Blondon, Pierre-François Bert, Olivier Bitz, Dario Cantu, Grant Cramer, Christopher Davies, Serge Delrot, Mario Pezzotti, Stephane Rombauts

To cite this version:

Jérôme Grimplet, Anne-Francoise Adam-Blondon, Pierre-François Bert, Olivier Bitz, Dario Cantu, et al.. Toward the improvement of the gene centered information in grapevine genomics. Plant and Animal Genome Conference, Jan 2014, San Diego, CA, United States. hal-02798091

HAL Id: hal-02798091
https://hal.inrae.fr/hal-02798091v1

Submitted on 5 Jun 2020

Toward the Improvement of the Gene Centered Information in Grapevine Genomics

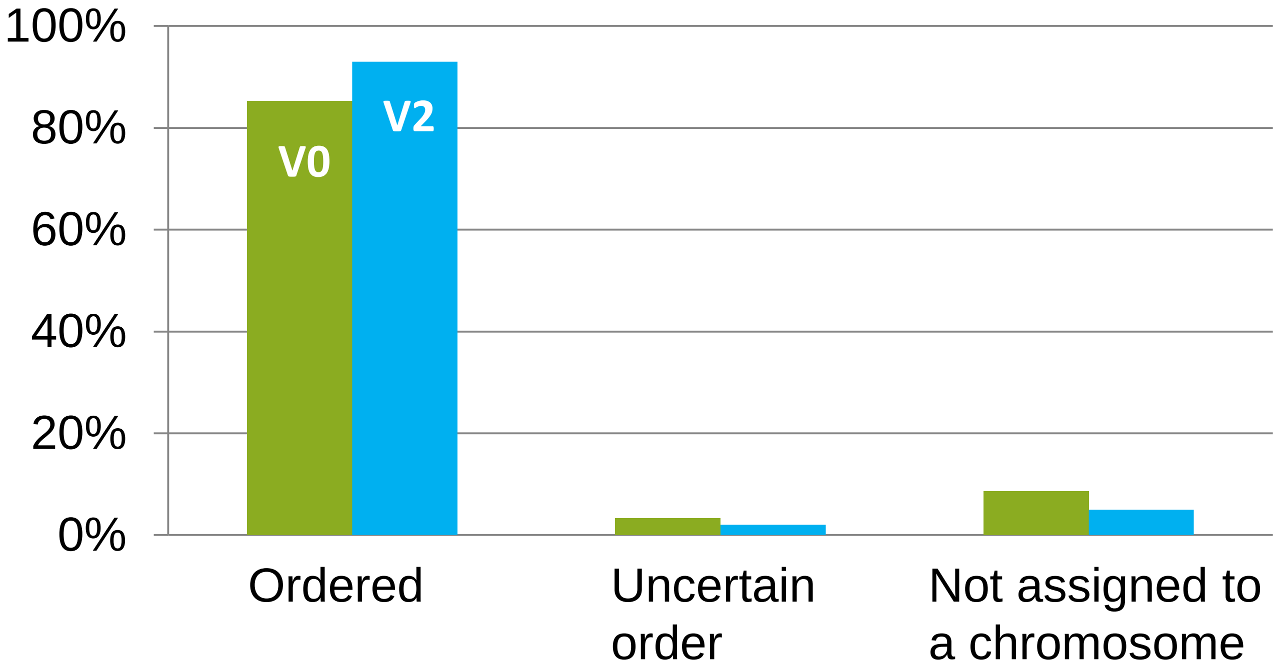
Super Nomenclature Committee

A-F Adam-Blondon
Unit of Research in Genomics Informatics

A little bit of context…

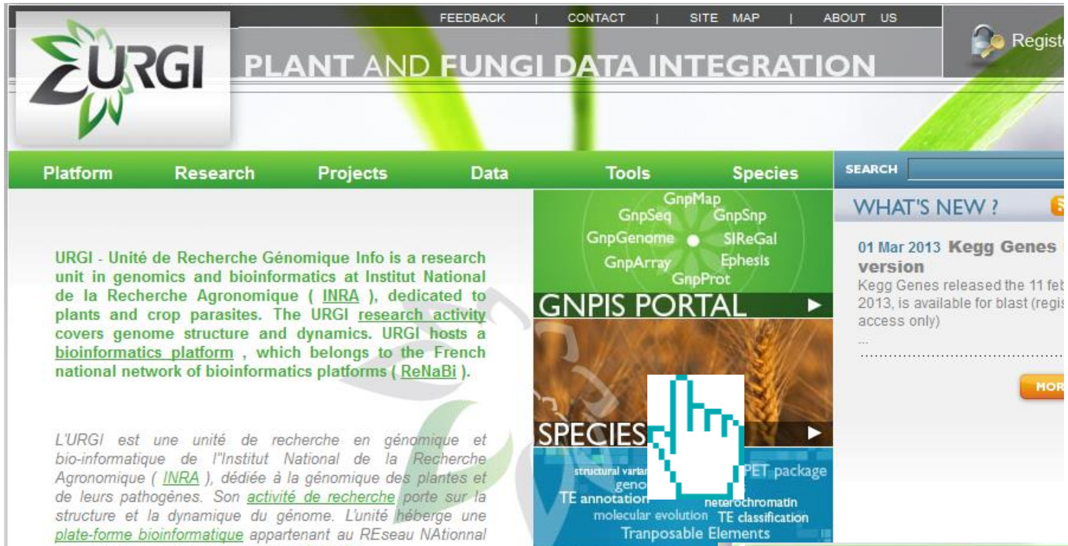
The grapevine reference genome sequence is evolving

- Improvement of the chromosome assembly (see P417)=>12X.2 version of the 19 chromosomes + 1 random



- Improvement of its annotation (talk of S. Rombauts) => V2 version of the annotation

A little bit of context…



URGI - Unité de Recherche Génomique Info is a research unit in genomics and bioinformatics at Institut National de la Recherche Agronomique (INRA), dedicated to plants and crop parasites. The URGI research activity covers genome structure and dynamics. URGI hosts a bioinformatics platform , which belongs to the French national network of bioinformatics platforms (ReNaBi).

L'URGI est une unité de recherche en génomique et bio-informatique de l''Institut National de la Recherche Agronomique (INRA), dédiée à la génomique des plantes et de leurs pathogènes. Son activité de recherche porte sur la structure et la dynamique du génome. L'unité héberge une plate-forme bioinformatique appartenant au REseau NAtionnal des plateformes BIo-informatiques (ReNaBi).

01 Mar 2013 **Kegg Genes version**

Kegg Genes released the 11 feb 2013, is available for blast (regis access only)

...

EVENT & PUBLICATIONS

28 Feb 2013 **Genomes to germplasm**

15 Feb 2013 **Efficient compa**

pecies

You are here : Home / Home URGI / Species

Species

You find here, the main species stored in the URGI repository

Vitis

Why develop grape genomics?
The International Grape Genome Program (IGGP, launched in 2001) a international efforts for the development of genomic resources for the years, more than 380 000 complete or partial sequences ...

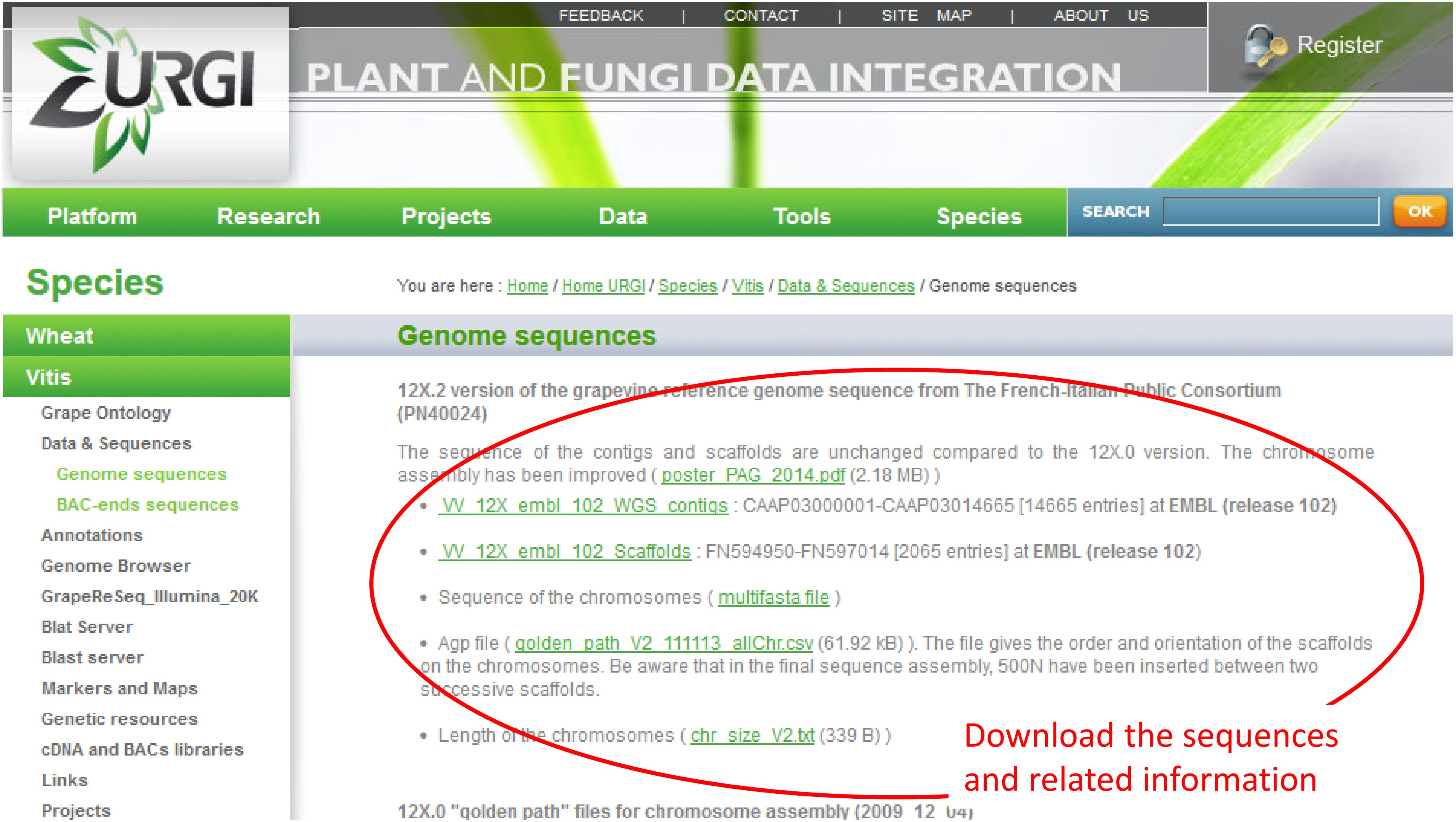
Wheat

Wheat URGI website

The new chromosome assembly is available at

https://urgi.versailles.inra.fr/

https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences



You are here : Home / Home URGI / Species / Vitis / Data & Sequences / Genome sequences

Genome sequences

12X.2 version of the grapevine reference genome sequence from The French-Italian Public Consortium (PN40024)

The sequence of the contigs and scaffolds are unchanged compared to the 12X.0 version. The chromosome assembly has been improved (poster_PAG_2014.pdf (2.18 MB))

- VV_12X_embl_102_WGS_contigs : CAAP03000001-CAAP03014665 [14665 entries] at EMBL **(release 102)**
- VV_12X_embl_102_Scaffolds : FN594950-FN597014 [2065 entries] at **EMBL (release 102)**
- Sequence of the chromosomes (multifasta file)
- Agp file (golden_path_V2_111113_allChr.csv (61.92 kB)). The file gives the order and orientation of the scaffolds on the chromosomes. Be aware that in the final sequence assembly, 500N have been inserted between two successive scaffolds.
- Length of the chromosomes (chr_size_V2.txt (339 B))

Download the sequences and related information

12X.0 "golden path" files for chromosome assembly (2009_12_04)

A little bit of context…

A lot of knowledge is continuously beeing gained on grapevine gene and their functions

- Expert-based structural annotation of gene families

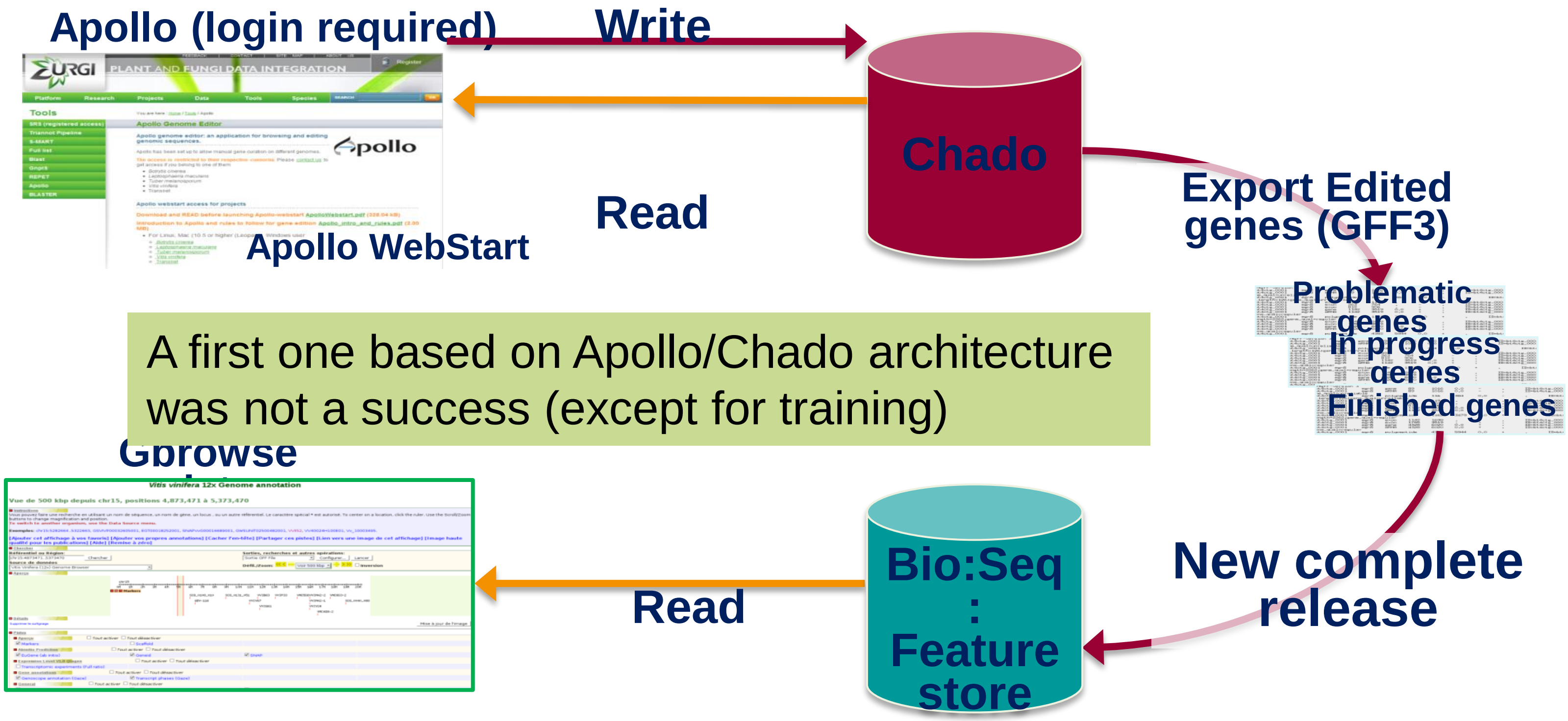
	Missed genes/pseudogenes			Bad models for coding genes		
	TPS	**STS&CHS**	**NBS-LRR**	**TPS**	**STS&CHS**	**NBS LRR**
V0	35%	58%	44%	77%	77%	65%
V1	na	na	34%	na	na	68%
Eugene	8%	2%	6%	19%	36%	53%

- Gene functional characterisation: transcriptomics, mutants, etc…

→ Highly decentralized information, many inconsistent gene naming, poorly implemented in databases

A little bit of context…

Centralization of gene related data requires a central information systems, user friendly interface and query systems and time dedicated to tailoring



You are here : Home / Tools / Apollo

Apollo Genome Editor

Apollo genome editor: an application for browsing and editing genomic sequences.

Apollo has been set up to allow manual gene curation on different genomes.

The access is restricted to their respective consortia. Please contact us to get access if you belong to one of them:

- Botrytis cinerea
- Leptosphaeria maculans
- Tuber melanosporum
- Vitis vinifera
- Transmet

Apollo webstart access for projects

Download and READ before launching Apollo-webstart ApolloWebstart.pdf (328.04 kB)

Introduction to Apollo and rules to follow for gene edition Apollo_intro_and_rules.pdf (2.00 MB)

- For Linux, Mac (10.5 or higher (Leopard), Windows user
  - Botrytis cinerea
  - Leptosphaeria maculans
  - Tuber melanosporum
  - Vitis vinifera
  - Transmet

Vitis vinifera 12x Genome annotation

Vue de 500 kbp depuis chr15, positions 4,873,471 à 5,373,470

A first one based on Apollo/Chado architecture was not a success (except for training)

New start with the ORCHAE information system (S. Rombauts)

A little bit of context…

There is a need for common rules for gene naming in order to improve:

- consistency between studies
- integration of decentralized data, recent and historical

- Formation of an International Super Nomenclature Committee coordinated by J. Grimplet under the auspices of the IGGP consortium
- Major output of the GRCN NSF program (G. Cramer coord) and will follow up through the EU COST FA 1106 (M. Pezzotti co-coord)

Super Nomenclature Committee

- **Coordinator: Jérôme Grimplet** (CSIC-Universidad de La Rioja-Gobierno de La Rioja, Sp)
- Anne-Francoise Adam-Blondon (INRA Versailles, Fr)
- Pierre-François BERT (Université de Bordeaux, Fr)
- Oliver Bitz (Geisenheim University, D)
- Dario Cantu (University of California Davis, USA)
- Grant R. Cramer (University of Nevada, Reno, USA)
- Christopher Davis (CSIRO Plant Industry, Aus)
- Serge Delrot (Université de Bordeaux, Fr)
- Mario Pezzotti (University of Verona, It)
- Stephane Rombauts (VIB-Ghent University, B)

Conventions

Elements	Locus ID	Full Name	Symbol	Synonyms
Example	VIT01g00000-1.1	(V. vinifera) Alcohol dehydrogenase 1	(vvi)ADH1	GV-ADH1 **aldehyde reductase** **ethanol dehydrogenase** **GSVIVT01010024001** **VIT_18s0001g15410**
Description	Genome localization	-Descriptive function -Includes the level of quality of annotation	- Size limited - Descriptive of function if possible	Any known synonyms

Physical address

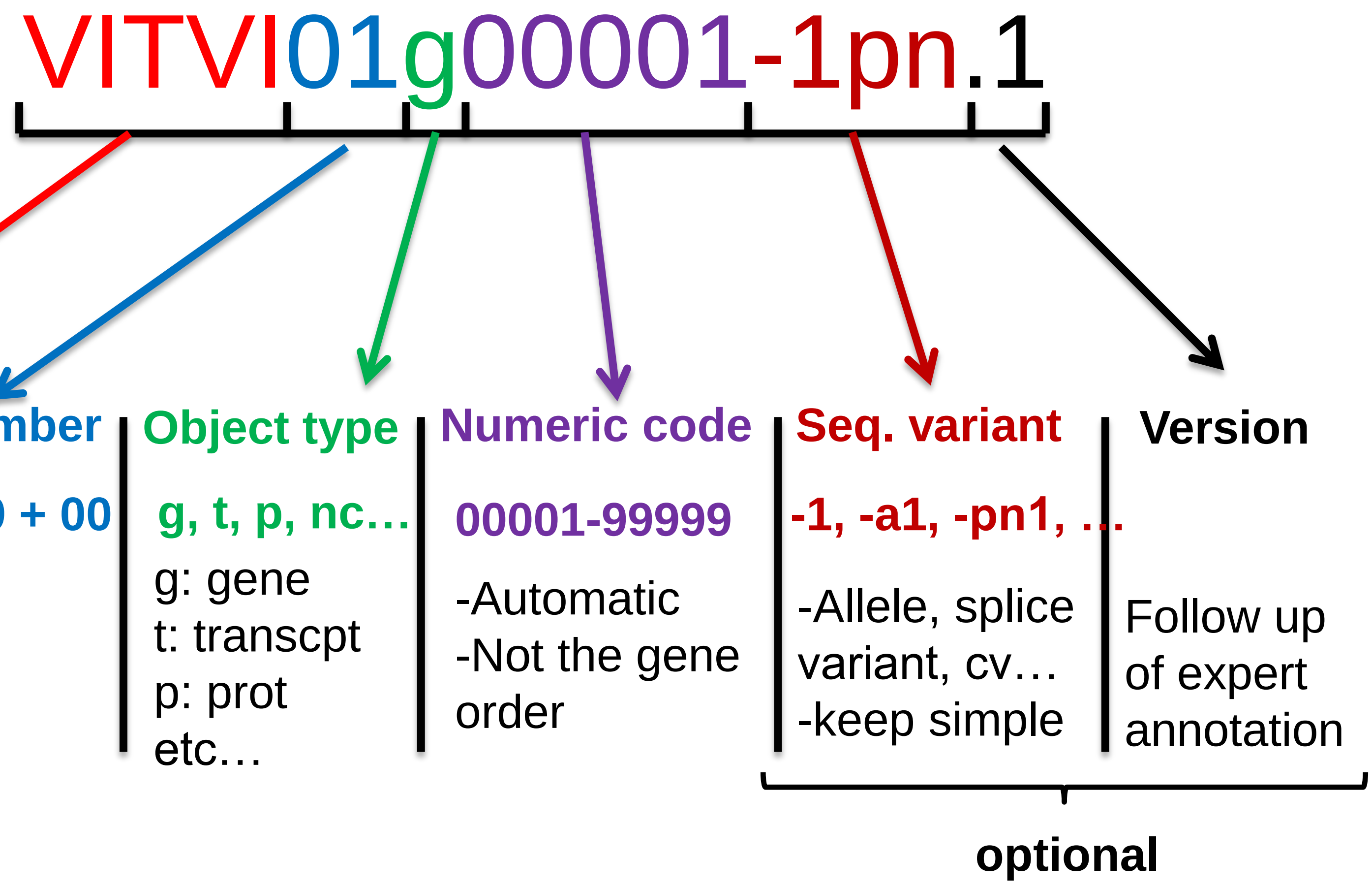
Function

History
Redundancy

Must be unique

Locus ID = unique genome address

Standard system for locus Identifier in whole genome annotation

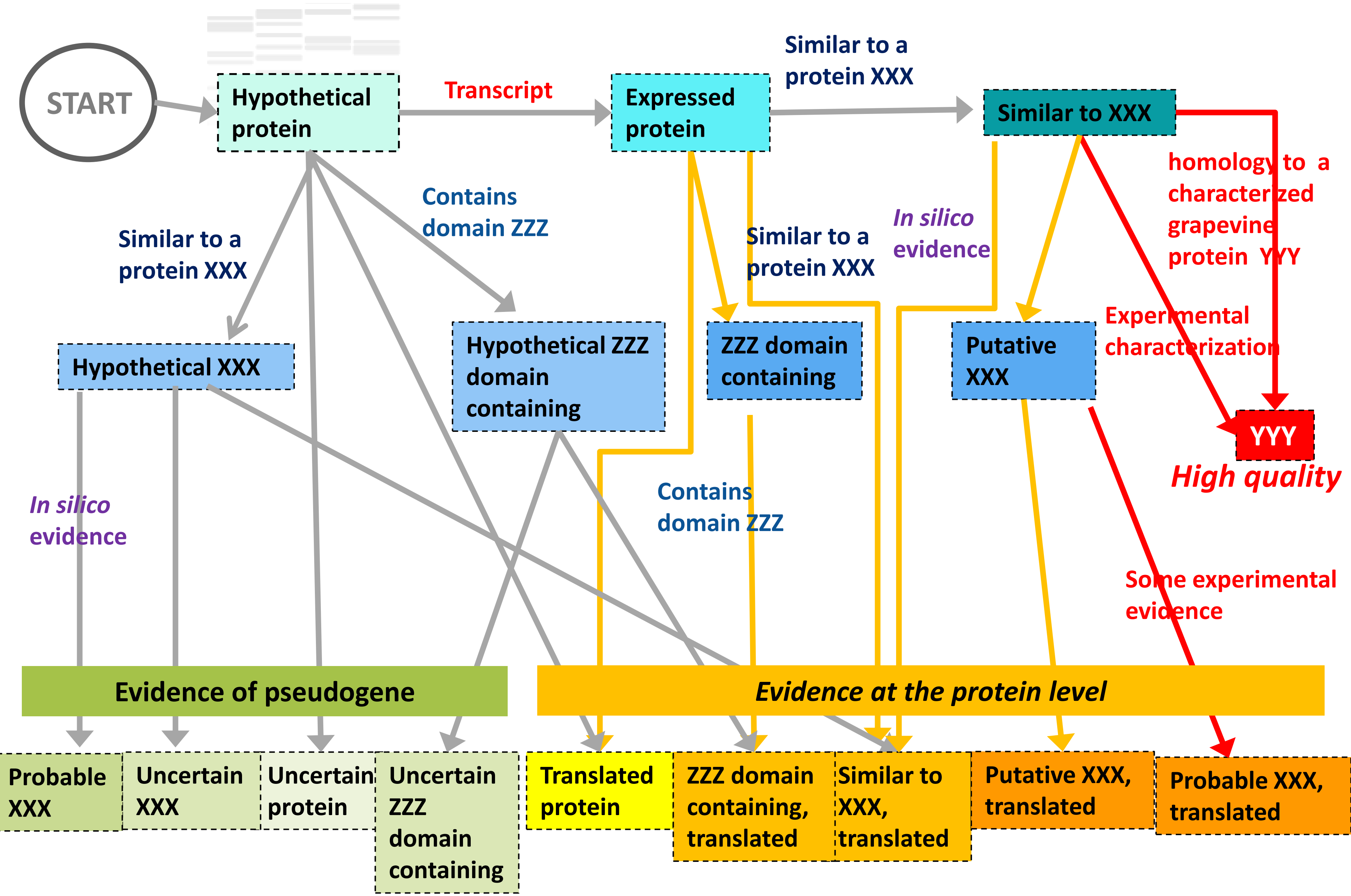


VITVI01g00001-1pn.1

Taxonomy ID	**Chr number**	**Object type**	**Numeric code**	**Seq. variant**	**Version**
UniProt std	**01 to 19 + 00**	**g, t, p, nc…**	**00001-99999**	**-1, -a1, -pn1, …**	
VITVI: vinifera VITRI: riparia etc..		g: gene t: transcpt p: prot etc…	-Automatic -Not the gene order	-Allele, splice variant, cv… -keep simple	Follow up of expert annotation

optional

Full name and quality level of the annotation



START

Hypothetical protein

Transcript

Expressed protein

Similar to a protein XXX

Similar to XXX

homology to a characterized grapevine protein YYY

Experimental characterization

YYY

High quality

Similar to a protein XXX

Contains domain ZZZ

Similar to a protein XXX

In silico evidence

Hypothetical XXX

Hypothetical ZZZ domain containing

ZZZ domain containing

Putative XXX

In silico evidence

Contains domain ZZZ

Some experimental evidence

Evidence of pseudogene

Evidence at the protein level

Probable XXX

Uncertain XXX

Uncertain protein

Uncertain ZZZ domain containing

Translated protein

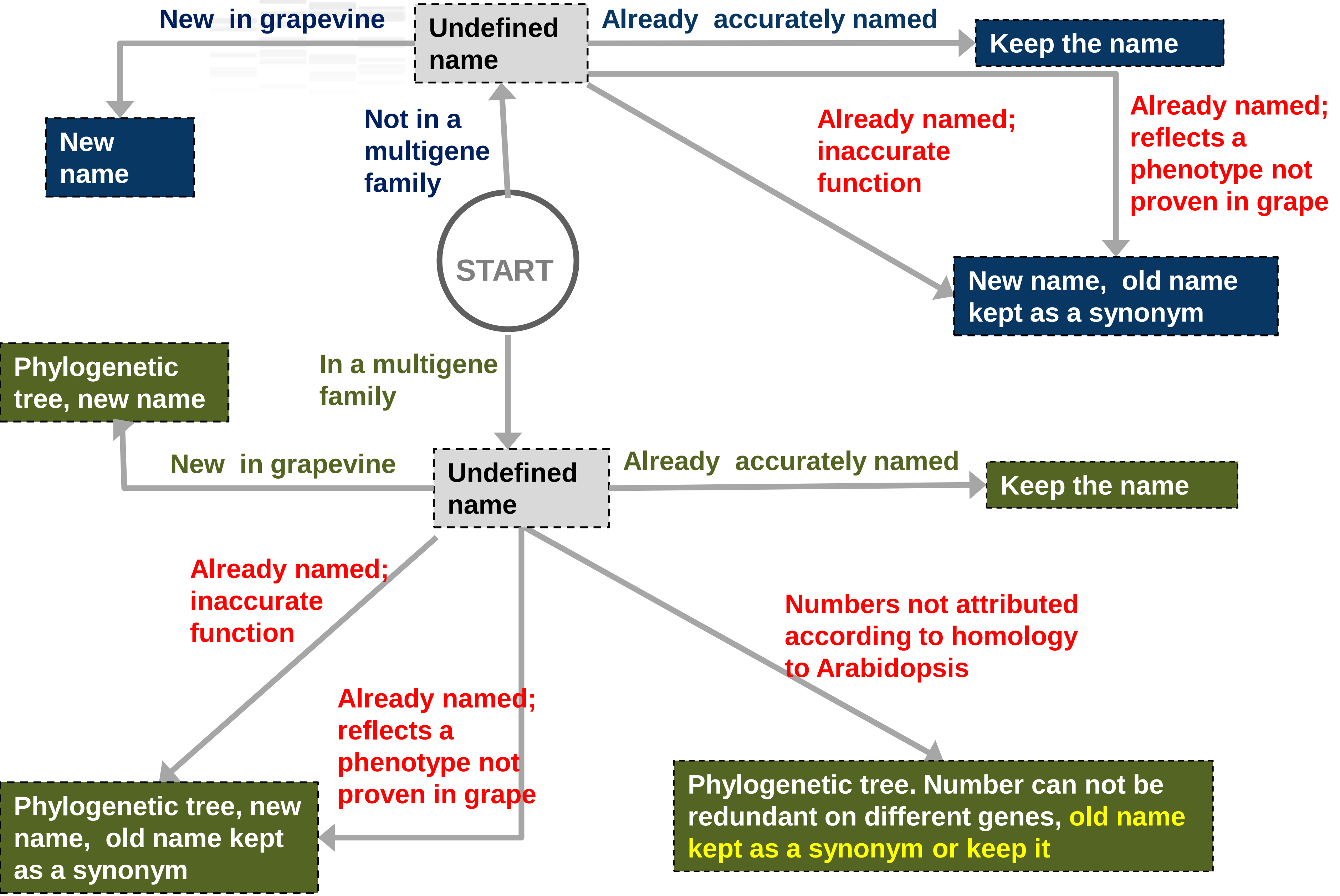
ZZZ domain containing, translated

Similar to XXX, translated

Putative XXX, translated

Probable XXX, translated

Symbol : multigene families and historical data



START

Not in a multigene family

Undefined name

New in grapevine

New name

Already accurately named

Keep the name

Already named; inaccurate function

Already named; reflects a phenotype not proven in grape

New name, old name kept as a synonym

In a multigene family

Undefined name

New in grapevine

Phylogenetic tree, new name

Already accurately named

Keep the name

Already named; inaccurate function

Already named; reflects a phenotype not proven in grape

Phylogenetic tree, new name, old name kept as a synonym

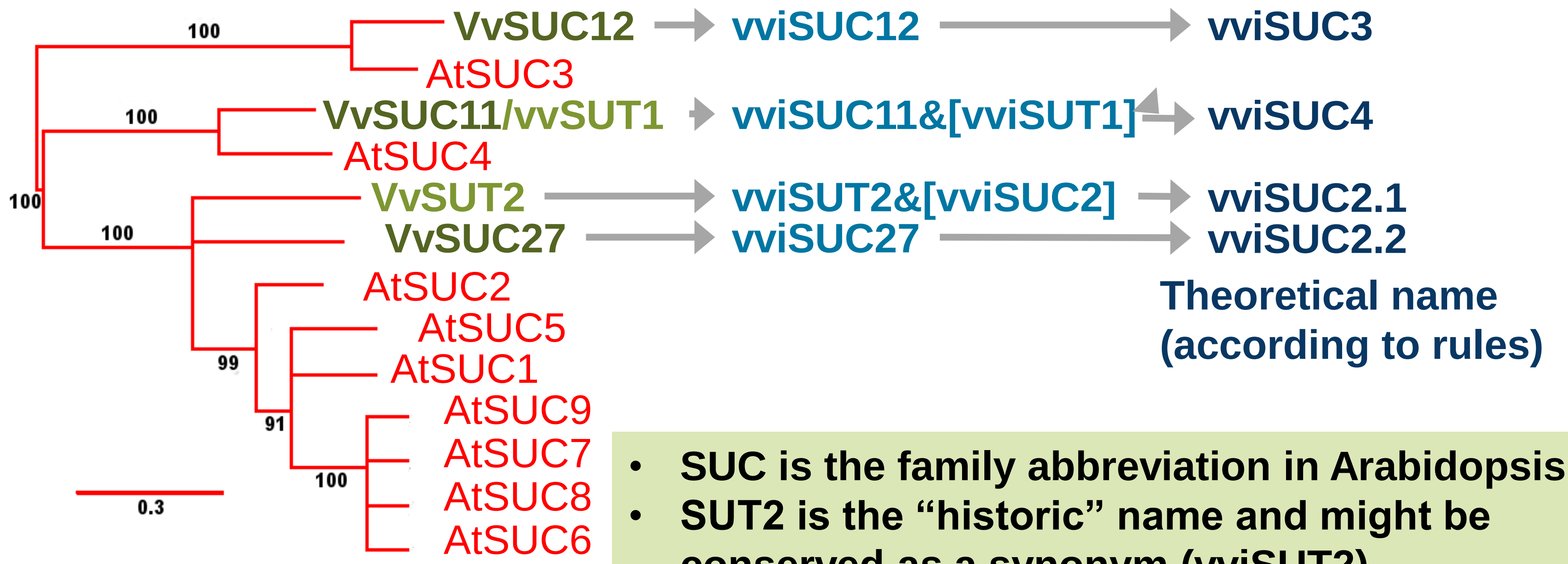
Numbers not attributed according to homology to Arabidopsis

Phylogenetic tree. Number can not be redundant on different genes, old name kept as a synonym or keep it

Symbol choice : example of the Sucrose transporters

Issue : genes have multiple names in the litterature

- *Davies C, Wolf T, Robinson SP*. Three putative sucrose transporters are differentially expressed in grapevine tissues. Plant Sci 1999, 147: 93-100.
- Ageorges A, Issaly N, Picaud S, Delrot S, Romieu C. Identification and functional expression in yeast of a grape berry sucrose carrier. Plant Physiol Biochem. 2000;38(3):177–185



VvSUC12 → vviSUC12 → vviSUC3

VvSUC11/vvSUT1 → vviSUC11&[vviSUT1] → vviSUC4

VvSUT2 → vviSUT2&[vviSUC2] → vviSUC2.1

VvSUC27 → vviSUC27 → vviSUC2.2

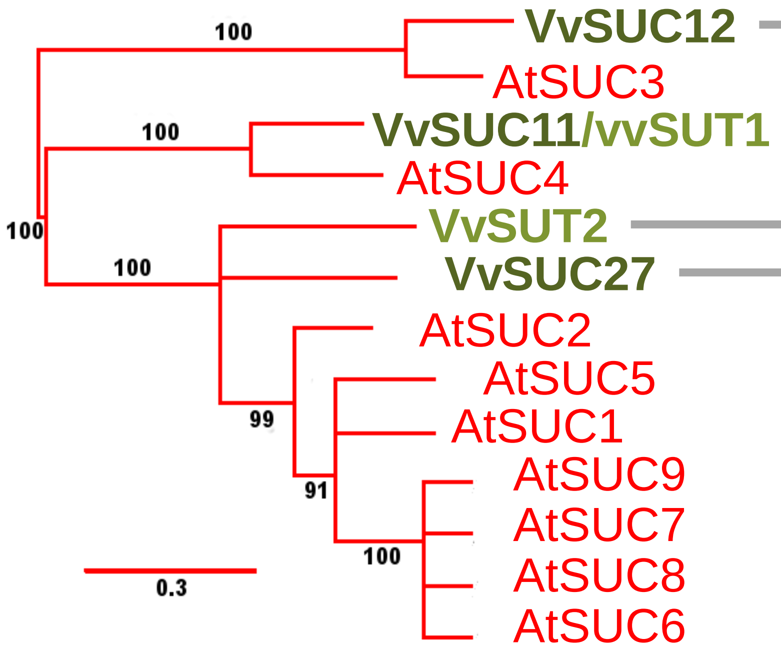
Theoretical name (according to rules)

- SUC is the family abbreviation in Arabidopsis
- SUT2 is the "historic" name and might be conserved as a synonym (vviSUT2)

Symbol choice : example of the Sucrose transporters

Issue : genes have multiple names in the litterature

- ***Davies C, Wolf T, Robinson SP*. Three putative sucrose transporters are differentially expressed in grapevine tissues. Plant Sci 1999, 147: 93-100.**
- **Ageorges A, Issaly N, Picaud S, Delrot S, Romieu C. Identification and functional expression in yeast of a grape berry sucrose carrier. Plant Physiol Biochem. 2000;38(3):177–185**



VvSUC12 → **vviSUC12** → **vviSUC3**

VvSUC11/vvSUT1 → **vviSUC11&[vviSUT1]** → **vviSUC4**

VvSUT2 → **vviSUT2&[vviSUC2]** → **vviSUC2.1**

VvSUC27 → **vviSUC27** → **vviSUC2.2**

Theoretical name (according to rules)

Expert based Annotation only!

Perspectives

- Manuscript to be soon submitted
- Rules will be applied in the V2 version of the annotation of the 12X.2 version of the genome (S. Rombauts)

Update on the IGGP

- General Assembly made at the Grapevine Physiology and Biotechnology conference, La Serena, Chile, April 2013

- New chair : A-F Adam-Blondon
- New deputy chair : Mario Pezzotti and Laurent Deluc
- Modification of the reste of the steering Committee

www.vitaceae.org