Grape gene catalogue guidelines

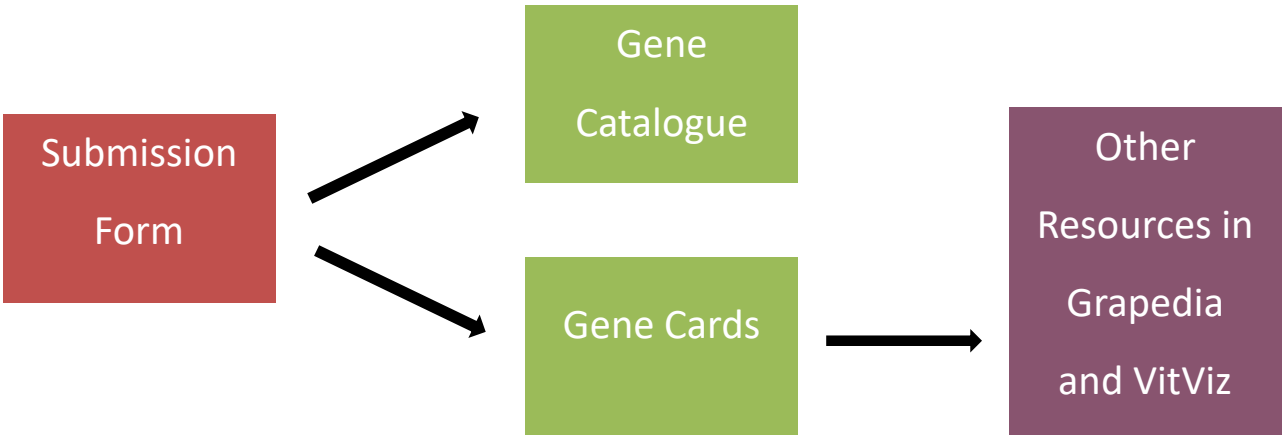
(v2.1 guidelines, click here to get most up to date catalogue, guidelines and submission form)

The current Grapevine gene catalogue has around 2,000 entries consisting of unique gene id and gene symbol pairs centred on the PN40024 assembly (Navarro-Payá *et al.* 2023). Each entry has a description of function with associations to the literature and a given confidence level for the evidence available. *Arabidopsis thaliana* has a similar protein-coding gene count to *Vitis vinifera* and over 10,000 gene symbols are currently reported in TAIR, which allow referring to a gene by its symbol rather than a genomic id (that will depend on the cultivar and also change over time).

So far, the catalogue has focused on *Vitis vinifera ssp. vinifera* genes (more specifically on the reference genome assembly and annotation of cv. PN40024). The catalogue will still have a focus on the reference genome, as most gene-centred studies refer to this assembly in the literature due to its quality. However, with the recent emergence of pangenome strategies, and the fact that some genes may be cultivar or even species-specific this catalogue is also prepared to accept gene submissions with gene ids from other assemblies as long as no ortholog has been found in the PN40024 genome. At the moment we expect submissions the catalogue to reflect *Vitis vinifera* as a species and perhaps some cultivar-specific genes but the forms are now prepared for submissions from other species as well. This will pave the way for a pangenome catalogue in the future that takes into account genes within the *Vitis* genus in a broader sense.

This guide should help you with the submission process of your genes of interest, that will then be incorporated into the different genomic resources available (gene catalogue, Gene Cards, and other resources in Grapedia and VitViz). There are two main aims regarding catalogue gene submissions:

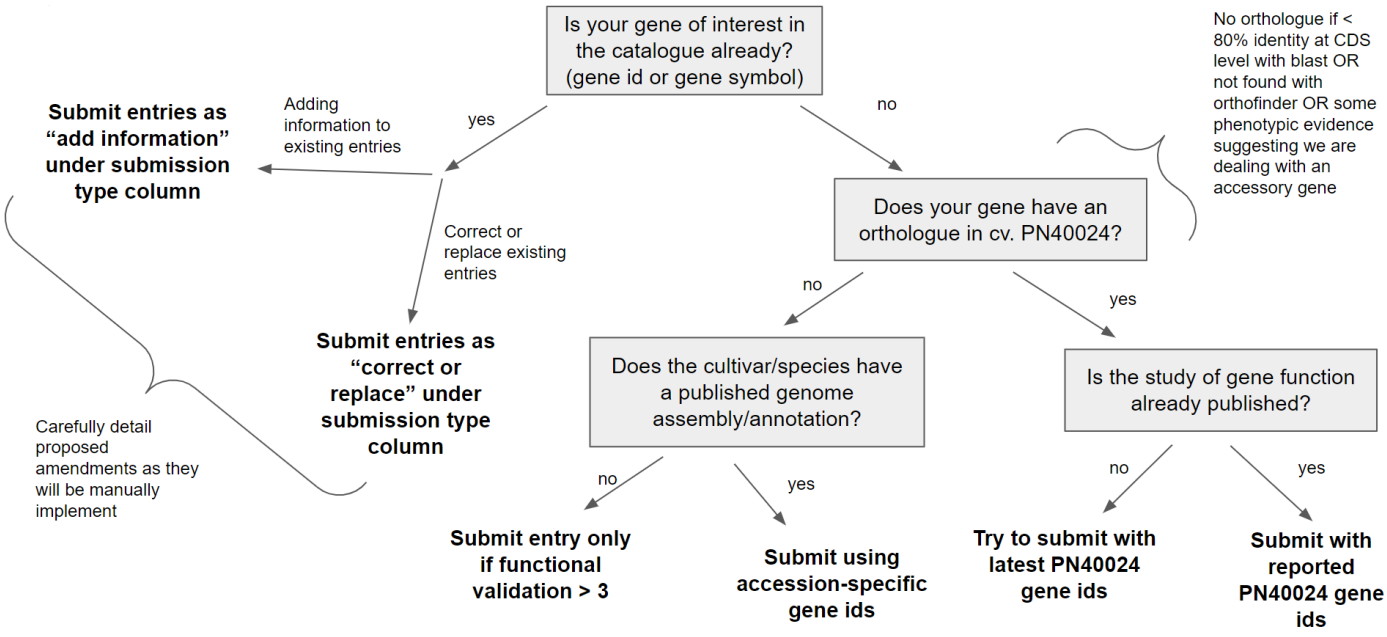
- Assign unique gene symbols to gene ids (preferably from the reference genome unless it is known to be missing in it)
- Describe the function of the submitted genes (link this to the literature and provide a confidence level of that association)



We recommend to follow the grapevine gene nomenclature system (Grimplet et al. 2014) to assign gene symbols and names.

Gene ids can be provided in any of the PN40024 gene annotation versions, in fact, gene ids as reported in a publication are preferred for traceability. The gene ids will be then updated to the current annotation version once the data is incorporated into the catalogue. It is now using the v3 annotation and the 12Xv2 assembly, however, it will be updated in the next round to the T2T assembly and the new annotation in preparation (proposed gene id format: Vitvi05_01chr07g00000 for the reference and Vitvi05_01chr07g00000_CS for other cultivars).

If your gene of interest is absent in the reference assembly you can use its species or cultivar specific gene id to give it a unique gene symbol. Bear in mind that the catalogue table in its final version will only include PN40024 gene ids, however, the submitted information regarding ids from other accessions will be shown in Gene Cards (temporarily in VitViz before moving to the Grapedia portal). It will include gene id equivalences to different species and cultivars as well as older PN40024 gene ids, as recent pangenome analyses suggest that there are many unique genes in each Vitis spp. The tables below describe the different fields in the submission forms available which can be filled-in and how the data will be incorporated into the catalogue table and Gene Cards app. Submissions can be both to add new genes as well as replacing or amending existing entries. Any conflicts arising from entry clashes (the aim is to have both unique gene ids and gene symbols) will be dealt with manually so the comments column in the submission becomes important in these cases.



Grapedia
The Grapevine Genomics Encyclopedia

Submission Form

gene symbol		Gene name. Include a single, unique name per row.
gene id		Based on PN40024 unless the gene of interest is absent in the cultivar's assembly but it is in fact present in the assembly/annotation of another species/cultivar. Include a single, unique gene ID per row.
species		Species corresponding to gene id
cultivar		Cultivar corresponding to gene id
assembly version		Go to the genomes section in Grapedia for more info on the different PN40024 genome annotations and assemblies. If your gene of interest is absent from the PN40024 assembly and you have chosen a gene id from a different cultivar/species please indicate both the genome and annotation version if appropriate.
annotation version		
group		Gene group
subgroup		Gene subgroup (if any)
pathway or role		Metabolic pathway(s) or general role(s)
protein family		Protein family or families (based on domains etc.)
full name(s)		Unique gene name in full (i.e. Laccase 1 not Laccase)
gene name synonym(s)		Gene name/symbol synonyms in grapevine
name(s) in other species		Gene name/symbol synonyms of orthologs in species (e.g., At) other than *Vitis* spp.
pseudogene		pseudogene or gene to allow easy catalogue filtering
cell localisation		Subcellular localisation when there is experimental evidence
type		Type of protein
EC number		Enzyme Commission number
functional validation level		See different levels described below
description of function		Description of the functional association
functional evidence	**species/cultivar(s)**	In which species/cultivar(s) was it demonstrated i.e., overexpressed in Syrah and silenced in Cabernet Sauvignon
	citation(s)	Reference(s) for functional association(s)
	doi(s)	Reference doi(s) for functional association(s)
***in silico* evidence**	**phylogenetic method**	Phylogenetic method used (if any) for analysis
	species/cultivar(s)	Which species/cultivar sequences were analysed
	citation(s)	Reference(s) for gene family characterisation(s)
	doi(s)	Reference doi(s) for gene family characterisation(s)
submission type		Choose "New Gene" for new entries, "Add Information" for extra data to an existing entry and "Correct or Replace" to modify it.
comments		Any other detail worth noting. Particularly necessary when aiming to replace existing genes in the catalogue, please explain the rationale so any conflict with existing entries can be resolved.

Grapedia
The Grapevine Genomics Encyclopedia

Functional Validation Levels

1	**Hypothetical:** only based on similarity to other proteins (e.g. BLAST, Hidden Markov Models/PFAM search).
2	**Putative:** complete family identification (using phylogenetic trees and including other species) and/or expression data validation (e.g., RNA-seq, qPCR).
3	**Proposed:** correlation experiments such as gene co-expression networks, gene expression correlation to metabolite profiles or phenotypes.
4	**Candidate:** gene associated to a QTL with some level of support (expression behaviour, homology-based inference of function) suggesting it is a candidate gene behind the trait (not any gene present in a QTL that may have hundreds or thousands of genes).
5	**Validated in other sp.:** overexpression (transient or stable) of gene or dominant-negative form in **another species**.
6	**Validated:** knock-out/loss-of-function, silencing, overexpression (transient or stable) of gene or dominant-negative in the **same species**. Enzyme-encoding genes, *in vitro* or cell culture (e.g., *E. coli*/yeast) experiments or even overexpression of enzyme (transient, stable) in another species will suffice.

Assigned gene functions will have different levels of supporting evidence, ranging from a simple BLAST or phylogenetic tree to gene knock-outs and knock-ins. The scale presented takes into account the fact that *Vitis* sp. are not model organisms and knock-out or knock-in experiments are not as feasible as in other species. Nevertheless, note that the top score is reserved for experimental evidence carried out in grapevine plants except for enzymatic characterisation, which is given a pass since substrate/product analyses are generally reliable even if carried out in a different species. The function of transcription factors or other genes, on the other hand, could be more dependent on the genetic background of the tested species and hence experiments in grapevine plants are required to reach the top score.

Grapedia
The Grapevine Genomics Encyclopedia

Key submission columns processing into catalogue

gene symbol	The **name will be checked to be unique in the catalogue** unless it is an amendment to an existing gene in which case the previous name will be moved to the synonym(s) column.
gene id	Submitted gene id **will be updated to the latest PN40024 IDs** and genes missing in the reference assembly will be given an NA. Information submitted regarding other species or cultivar IDs will now be held in Gene Cards. Checks will be in place to ensure single, unique IDs per row.
group	These three categories will be used to order the rows in the catalogue table. Enzymatic pathways will try to be ordered sequentially if possible.
subgroup	
pathway or role	
protein family	
gene name synonym(s)	In case of amendments the previous name will be moved here.
pseudogene	Will also be used to order rows.
type	Will also be used to order rows.

Grapedia
The Grapevine Genomics Encyclopedia

Catalogue Table Examples

		Example 1	Example 2	Example 3
gene symbol		MYB14	LAC50	CHS1
gene id		Vitvi07g00598	Vitvi18g03024	Vitvi14g01448
group		R2R3-MYBs	LAC	CHS
subgroup		S2	8	CHS
pathway or role		Phenylpropanoid Pathway (Stilbenoid Branch)	Phenylpropanoid Pathway (Stilbenoid Branch)	Phenylpropanoid Pathway (Early Flavonoid Branch)
protein family		MYBs	LAC	PKS
full name(s)		R2R3-type myeloblastosis transcription factor 14	Laccase 50	Chalcone synthase 1
gene name synonym(s)			Laccase6	
name(s) in other species				
pseudogene		gene	gene	gene
cell localisation		Nucleus		
type		TF	Enzyme	Enzyme
EC number				2.3.1.74
functional validation level		6	3	3
description of function		Stilbenoid, early pp and shikimate pathway regulator	Putative involvement in resveratrol modifications	CHS by homology and coexpression with berry phenotype
functional evidence	**species/ cultivar(s)**	Overexpression in cv. Chardonnay hairy roots, expression monitored in cvs. Shiraz, Pinot Noir; DAP-Seq in cv. Pinot Noir, overexpression in cv. Shiraz leaves.	Gene co-expression networks connect LAC50 specifically to STS gene expression rather than lignin pathway enzymes (multiple cvs.).	mRNA of CHS1 accumulated in the berry skin and young leaves of cv. Chardonnay and in the berry skin of cv. Merlot
	citation(s)	Holl et al. 2013; Orduña et al. 2022	Pilati et al. 2021	Goto-Yamamoto et al. 2002
	doi(s)	10.1105/tpc.113.117127; 10.1111/tpj.15686	10.3390/biom11121744	doi.org/10.1016/S0168-9452(02)00042-0
***in silico* evidence**	**phylogenetic method**	Maximum Likelihood	Bayesian Inference	
	species/ cultivar(s)		cv. P40024 for phylogeny	
	citation(s)	Wong et al. 2016	Pilati et al. 2021	
	doi(s)	10.1093/dnares/dsw028	10.3390/biom11121744	
comments				