

## Cosmoss locus and gene identifiers (CGI)

On cosmoss each gene model has an additional unique gene (locus) ID (CGI). The CGI provides a unique address to a gene (model). Using a clustering procedure, all overlapping gene models were grouped into a unique locus. For each locus a unique number was assigned, which is specific for a given assembly. All CGIs include the number of the scaffold and the number of the locus they belong as well as information on the version of the assembly and annotation or the gene predictor they are derived from.

### CGI Syntax

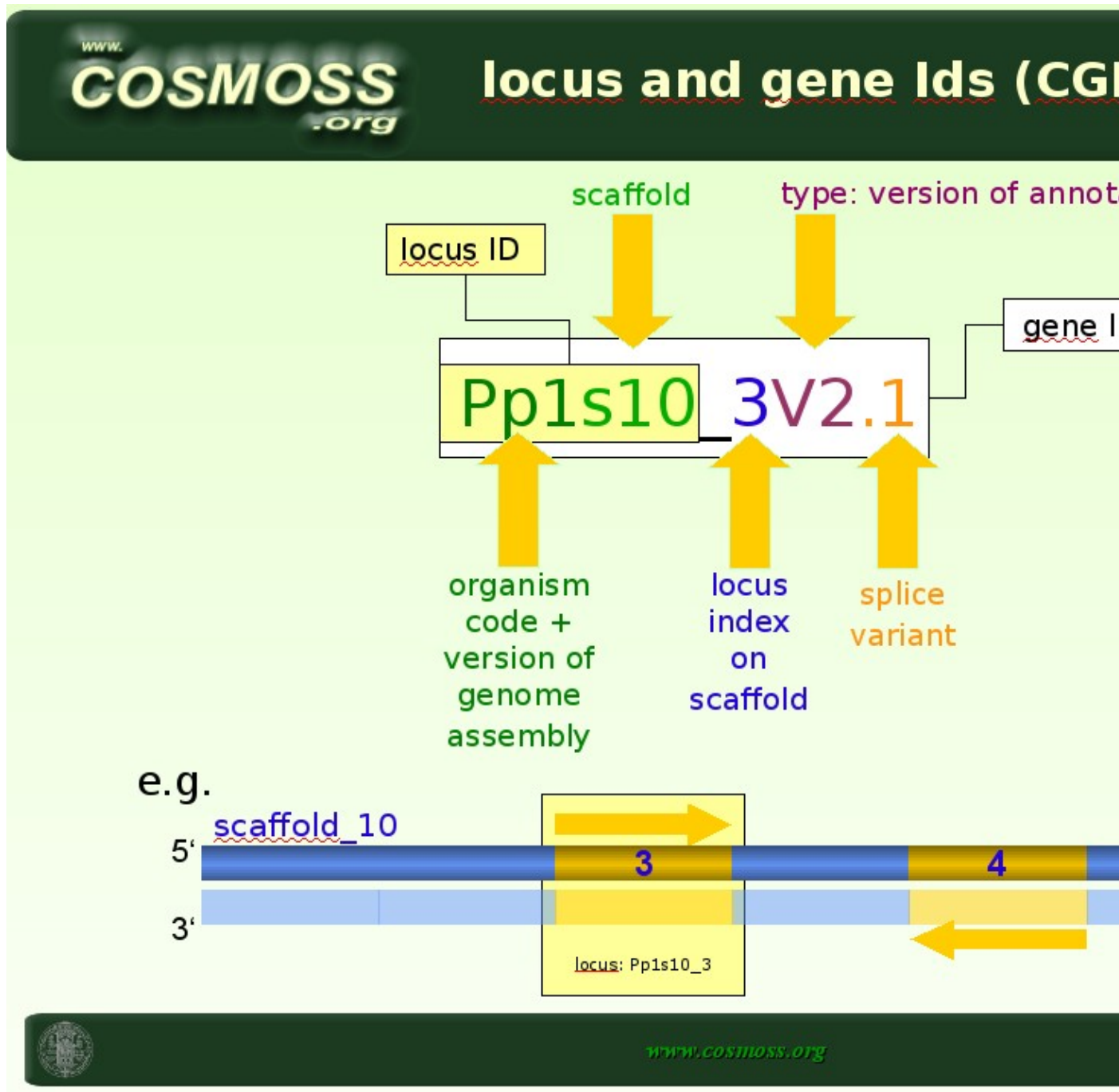
They include information on:

- the assembly version
- the number of the scaffold
- the number of the locus
- the class: the version of the annotation release or the gene predictor that predicted the model
- the splice variant

OrganismCode+AssemblyVersion+ScaffoldNumber+\_+LocusNumber+Type+. +  
SpliceVariant

e.g

Pp1s275\_3V2.1 (= Phypa\_196781)



## Type field

The type field indicates either the version of a released model (e.g. V1.2 or V1.5) or the predictor of a gene model in the all\_models catalog.

Source	Type	Description
JGL_FM1	V0	V1.0
JGL_FM3	V1	V1.1

## Cosmoss\_Gene\_IDs

JGI_cosmoss	V2	V1.2
JGI_e_gw1	G3	
JGI_estExt_fgenesh1_kg	F2	
JGI_estExt_fgenesh1_pg	F4	
JGI_estExt_fgenesh1_pm	F5	
JGI_estExt_fgenesh2_kg	F7	
JGI_estExt_fgenesh2_pg	F9	
JGI_estExt_fgenesh2_pm	F10	
JGI_estExt_Genewise1	G2	
JGI_estExt_gwp_gw1	G4	
JGI_fgenesh1_kg	F1	
JGI_fgenesh1_pg	F3	
JGI_fgenesh2_kg	F6	
JGI_fgenesh2_pg	F8	
JGI_gw1	G1	
JGI_user	U1	User models
EuGene_newIMM	E1	
EuGene_newIMM3	E2	
EuGene_newIMM3_trim_UTR	E3	
EVM_gth	E4	