



CRISPRCasdb UML class diagram :

- **Entity**: the parent class of most other classes, it stores the name of the child classes, comments etc.. and methods common to other classes,
- **Strain**: Assembles metadata (RefSeq and GenBank accessions (ex:GCA_000258245.1), release level ("Major, Minor"), assembly status ("latest"),
- **Sequence**: Metadata for each sequence of a strain : Length of each sequence, count of N in sequence. The complete nucleotide sequences are not stored in the database, but in fasta files (for Cas gene sequence viewing),
Values : None = 0, Circular = 1, Linear = 2, Chromosome = 4, Plasmid = 8
- **CrisprLocus (+Coordinates)**: position of each detected CRISPR locus, evidence level (1 to 4), direct repeat consensus sequence (in **Region** class), direct repeats conservation level, spacers conservation level, orientation of CRISPR,
- **Region**: Ordered list of sequences
Values : Unknown = 0, DirectRepeat = 1, Leader = 2, Spacer = 3, Tracr = 4, LeftFLANK = 5, RightFLANK = 6,
- **ClusterCas (+Coordinates)**: position of each detected Cas cluster, Type (class) of Cas cluster (ex : TypeIE), and composition of genes (cas 1, 2, 3...),
- **Taxon**: Complete taxonomy for each strain: taxonomic ranks (genus, species...), and name associated (Escherichia, Escherichia coli).