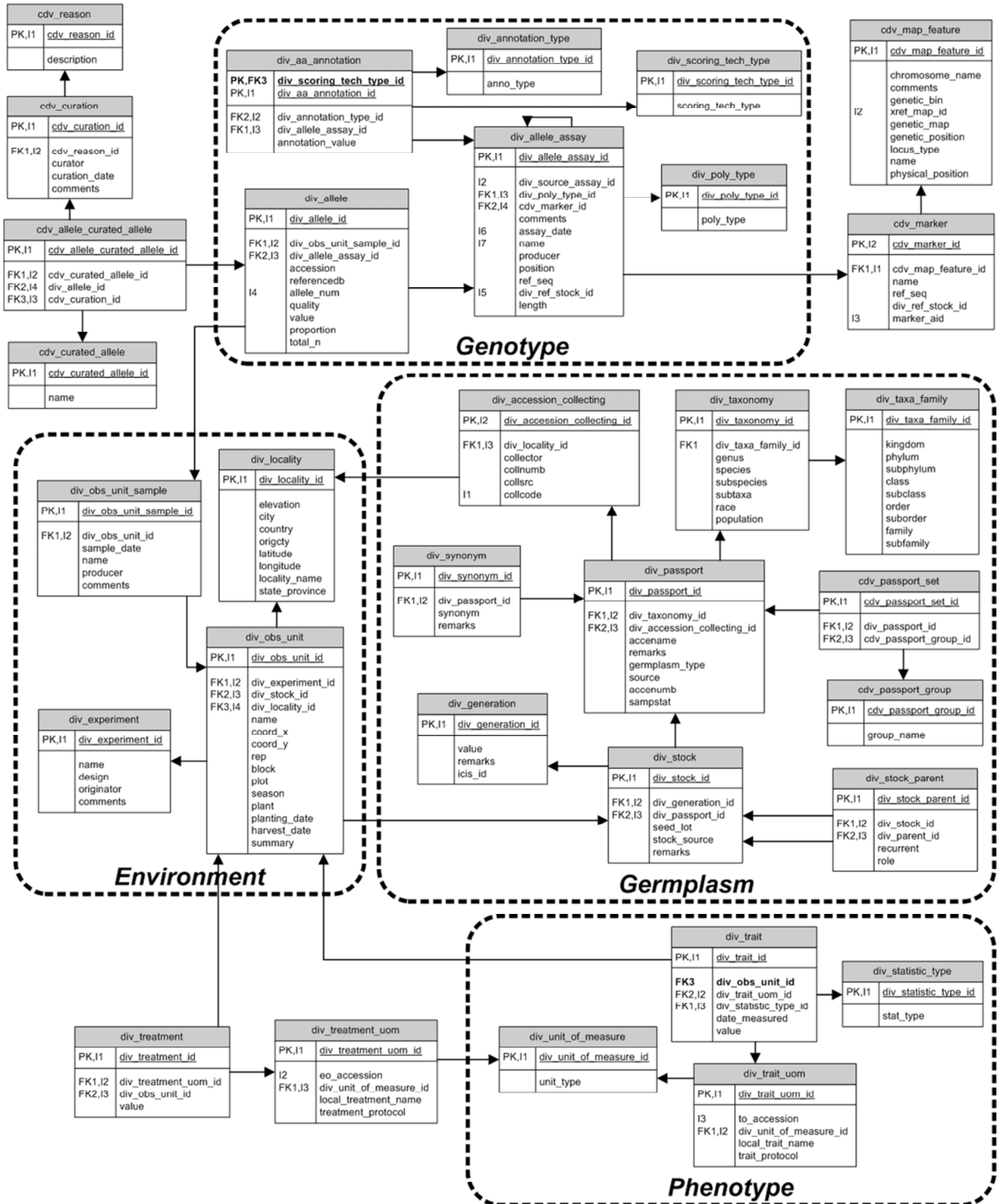


The Genomic Diversity and Phenotype Data Model (GDPDM) - April 25, 2005



**Table/Field descriptions for
The Genomic Diversity and Phenotype Data Model (GDPDM)
April 25, 2004**

| Table/Field | Type | Req'd | Description |
|---------------------------------|-------------|--------------|---|
| div_aa_annotation | table | | this table is used to annotate an allele assay. Any given allele assay can have multiple annotations. |
| div_aa_annotation_id | int | yes | primary key |
| div_annotation_type_id | int | yes | foreign key referencing the annotation type. |
| div_allele_assay_id | int | yes | foreign key referencing the allele assay this is annotating. |
| div_scoring_tech_type_id | int | no | foreign key referencing its scoring technology type. |
| annotation_value | text | yes | value for given type of this annotation. |
| div_accession_collecting | table | | This table stores information about where and by whom an accession was collected. This information is expected only for germplasm from collections maintained by germplasm centers. |
| div_accession_collecting_id | int | yes | primary key |
| div_locality_id | int | yes | foreign key to locality |
| collector | varchar | no | Collector of the accession |
| collnumb | varchar | yes | Collecting number: Original number assigned by the collector(s) of the sample, normally composed of the name or initials of the collector(s) followed by a number. This number is essential for identifying duplicates held in different collections. |
| collsrc | varchar | no | Collecting/acquisition source: The coding scheme proposed can be used at 2 different levels of detail: either by using the general codes (in boldface) such as 10, 20, 30, 40 or by using the more specific codes such as 11, 12 etc. 10) Wild habitat 11) Forest/woodland 12) Shrubland 13) Grassland 14) Desert/tundra 15) Aquatic habitat 20) Farm or cultivated habitat 21) Field 22) Orchard 23) Backyard, kitchen or home garden (urban, peri-urban or rural) 24) Fallow land 25) Pasture 26) Farm store 27) Threshing floor 28) Park 30) Market or shop 40) Institute, Experimental station, Research organization, Genebank 50) Seed company 60) Weedy, disturbed or ruderal habitat 61) Roadside 62) Field margin 99) Other (Elaborate in REMARKS field) |
| collcode | varchar | no | Collecting institute code: Code of the Institute collecting the sample. If the holding institute has collected the material, the collecting institute code (COLLCODE) should be the same as the holding institute code (INSTCODE). Follows INSTCODE standard. |
| div_allele | table | | this table is used to record allele values. |
| div_allele_id | int | yes | primary key |
| div_obs_unit_sample_id | int | yes | foreign key referencing observational unit sample. |
| div_allele_assay_id | int | yes | foreign key referencing allele assay. |
| allele_num | int | yes | number used to uniquely identify this allele from other alleles with the same observational unit sample and allele assay |
| quality | text | no | percentage indicating the accuracy of this allele value. |
| value | text | yes | div_allele.value - Value of the allele. Examples include: For a sequence-type polymorphism, the value would be the actual sequence. For an SSR-type polymorphism, the value would be the actual length in bp of this allele. |
| proportion | int | no | number of alleles with given value divided by total_n |

| Table/Field | Type | Req'd | Description |
|----------------------------|---------|-------|--|
| total_n | int | no | total number of homologous chromosomes surveyed (for diploids it is the number of gametes) |
| accession | varchar | no | |
| referencedb | varchar | no | |
| | | | |
| div_allele_assay | table | | table used to record allele assays. |
| div_allele_assay_id | int | yes | primary key |
| cdv_marker_id | int | no | foreign key referencing the marker associated with the reference sequence. |
| div_source_assay_id | int | no | foreign key referencing source assay within this same table. |
| div_poly_type_id | int | yes | foreign key referencing the polymorphism type of this assay |
| comments | text | no | comments about this allele assay. |
| assay_date | date | yes | date of this allele assay. |
| name | varchar | yes | name of this assay |
| producer | varchar | no | producer's name of this assay. |
| length | int | no | length of polymorphism being assayed relative to reference sequence. |
| position | varchar | no | position of polymorphism being assayed relative to reference sequence. |
| | | | Examples: 1) A SNP would have a position=1 and length=1; 2) A sequence assay might have a position and 500bp of length (length of the alignment); 3) A SSR would have a position and particular length in the reference.; 4) A Affy hybridization scoring reaction would have a position and length of 25. |
| ref_seq | text | yes | reference sequence used. |
| div_ref_stock_id | int | no | foreign key referencing stock associated with the reference sequence. |
| | | | |
| div_annotation_type | table | | table used to record different types of annotations. |
| div_annotation_type_id | int | yes | primary key |
| anno_type | varchar | yes | value of this annotation type |
| | | | |
| div_experiment | table | | table of experiments. An experiment is a collection of observational units with one originator that are part of a specific study. An experiment may be conducted at more than one location and in more than one season or year. |
| div_experiment_id | int | yes | primary key |
| name | varchar | yes | name of this experiment |
| design | varchar | no | design of this experiment |
| originator | varchar | yes | originator of this experiment |
| comments | text | no | comments |
| | | | |
| div_generation | table | | table of generation values. A generation value describes the number and type of generations that separate a taxon from an original source cross or population as defined in the div_stock_parent table. For example, F2, F7, or BC2F4. |
| div_generation_id | int | yes | primary key |
| value | varchar | yes | generation value |
| remarks | text | no | text description |
| icis_id | text | no | reference to ICIS generation entity |
| | | | |
| div_locality | table | | table used to store localities |
| div_locality_id | int | yes | primary key |

| Table/Field | Type | Req'd | Description |
|-----------------------------|---------|-------|--|
| elevation | int | no | Elevation of collecting site: [m asl] Elevation of collecting site expressed in metres above sea level. Negative values are allowed. |
| city | varchar | no | city |
| country | varchar | no | country in plain text |
| origcty | varchar | no | Country of origin: Code of the country in which the sample was originally collected. Use the 3-letter ISO 3166-1 extended country codes. |
| latitude | double | no | Latitude of collecting site: this should be recorded as a decimal number. |
| longitude | double | no | Longitude of collecting site: this should be recorded as a decimal number. |
| locality_name | varchar | no | name of this locality |
| state_province | varchar | no | state or province |
| div_obs_unit | table | | table of observational units. An observational unit is a plant or group of plants on which trait measurements are recorded or from which DNA samples are derived. Typically, it is a plot in a field but can also be a plant or pot in a greenhouse or growth chamber. |
| div_obs_unit_id | int | yes | primary key |
| div_experiment_id | int | no | foreign key referencing associated experiment |
| div_stock_id | int | yes | foreign key referencing stock planted |
| div_locality_id | int | no | foreign key referencing associated locality |
| name | varchar | yes | name of observational unit |
| coord_x | int | no | x coordinate |
| coord_y | int | no | y coordinate |
| rep | varchar | no | group of blocks |
| block | varchar | no | group of plots |
| plot | varchar | no | smallest measure observed. (i.e. row in field) |
| season | varchar | no | season |
| plant | varchar | no | Arbitrary plant number (i.e. 1, 2, 3, pool). |
| planting_date | date | no | date planted (example format: 4/19/2004) |
| harvest_date | date | no | date harvested (example format: 4/19/2004) |
| summary | text | no | summary |
| div_obs_unit_sample | table | | table used to define observation unit samples |
| div_obs_unit_sample_id | int | yes | primary key |
| div_obs_unit_id | int | yes | foreign key referencing observational unit this sample was collected from |
| sample_date | date | yes | date sample collected |
| name | varchar | yes | name of this sample |
| producer | varchar | yes | producer of this sample |
| comments | text | no | comments |
| div_passport | table | | table of taxa. Each record in this table represents a distinct genetic entity, which is expected to be represented by one or more physical stocks. NOTE: Many of the passport, taxonomy, and accession_collecting fields were derived from the FAO/IPGRI Multi-Crop Passport Descriptors (http://www.ipgri.cgiar.org/publications/pdf/124.pdf) |
| div_passport_id | int | yes | primary key |
| div_taxonomy_id | int | no | foreign key to taxonomy table |
| div_accession_collecting_id | int | no | foreign key to accession_collecting table |

| Table/Field | Type | Req'd | Description |
|------------------------------|---------|-------|---|
| accename | varchar | yes | Accession name: Either a registered or other formal designation given to the accession. First letter uppercase. Alternate names should be entered in the synonym table |
| remarks | text | no | Remarks: The remarks field is used to add notes or to elaborate on descriptors with value 99 or 999 (=Other). Prefix remarks with the field name they refer to and a colon (e.g. COLLSRC:roadside). Separate remarks referring to different fields are separated by semicolons without space. |
| germplasm_type | varchar | no | Deprecated: Use sampstat instead. Germplasm type of accession (i.e. Tripsacum, inbred) |
| source | varchar | no | Institute code or person's name. Immediate source of germplasm. |
| accenumb | varchar | no | Accession number: Actually an ASCII string, which serves as the unique identifier for accessions within a genebank collection, and is assigned when a sample is entered into the genebank collection. |
| sampstat | varchar | no | Biological status of accession: The coding scheme proposed can be used at 3 different levels of detail: either by using the general codes such as 100, 200, 300, 400 or by using the more specific codes such as 110, 120 etc. 100) Wild 110) Natural 120) Semi-natural/wild 200) Weedy 300) Traditional cultivar/landrace 400) Breeding/research material 410) Breeder's line 411) Synthetic population 412) Hybrid 413) Founder stock/base population 414) Inbred line (parent of hybrid cultivar) 415) Segregating population 420) Mutant/genetic stock 500) Advanced/improved cultivar 999) Other (Elaborate in REMARKS field). |
| div_poly_type | table | | table used to store/control possible polymorphism types |
| div_poly_type_id | int | yes | primary key |
| poly_type | varchar | yes | value of polymorphism (SNP, LENGTH, SEQALIGNMENT, SEQUENCE, RAWDATA, CATEGORICAL). |
| div_scoring_tech_type | table | | |
| div_scoring_tech_type_id | int | yes | primary key |
| scoring_tech_type | varchar | yes | |
| div_statistic_type | table | | table used to store/control possible statistic types |
| div_statistic_type_id | int | yes | primary key |
| stat_type | varchar | yes | value of statistic type (measure, mean, variance) |
| div_stock | table | | table of stocks. A stock is the specific source of a genetic entity from which an observational unit is taken. For most species, a stock will be a packet of seed in cold storage. An established line or cultivar is represented as a passport while a specific seedlot of that line is represented as a stock. |
| div_stock_id | int | yes | primary key |
| div_generation_id | int | no | foreign key referencing stock's generation |
| div_passport_id | int | yes | foreign key referencing stock's passport information |
| seed_lot | varchar | yes | an alphanumeric id that can be used by the source to access a specific stock and information about it. The combination of source and seedlot will be unique for a stock. |
| stock_source | varchar | no | the individual or institution that PROVIDED the stock directly used to create a specific obs_unit [the person may no longer maintain or have it, i.e., if the entire stock was used in the experiment] |
| remarks | text | no | remarks |

| Table/Field | Type | Req'd | Description |
|-------------------------|---------|-------|---|
| div_stock_parent | table | | table used to define the parents of the cross or population from which a stock is derived |
| div_stock_parent_id | int | yes | primary key |
| div_stock_id | int | yes | foreign key referencing child stock |
| div_parent_id | int | yes | foreign key referencing parent stock (the stock_id of the parent) |
| recurrent | tinyint | yes | Boolean value indicating if recurrent. |
| role | varchar | yes | Value indicating role of this parent (i.e. self, male, female). |
| div_synonym | table | | Synonyms or alternate names for passport accession names (passport.accename) |
| div_synonym_id | int | yes | primary key |
| div_passport_id | int | yes | Foreign key referencing div_passport table. |
| synonym | varchar | yes | The synonym. |
| remarks | text | no | Notes about the synonym. |
| div_taxa_family | table | | table used to further specify the taxonomy |
| div_taxa_family_id | int | yes | primary key |
| kingdom | varchar | no | top level grouping of organisms. |
| phylum | varchar | no | |
| subphylum | varchar | no | |
| class | varchar | no | |
| subclass | varchar | no | |
| order | varchar | no | |
| suborder | varchar | no | |
| family | varchar | no | |
| subfamily | varchar | no | |
| div_taxonomy | table | | table used to store taxonomy |
| div_taxonomy_id | int | yes | primary key |
| div_taxa_family_id | int | no | foreign key referencing the div_taxa_family table. |
| genus | varchar | no | Genus: Genus name for taxon. Initial uppercase letter required. |
| species | varchar | yes | Species: Specific epithet portion of the scientific name in lowercase letters. Following abbreviation is allowed: 'sp.' |
| subspecies | varchar | no | Subspecies |
| subtaxa | varchar | no | Subtaxa: Subtaxa can be used to store any additional taxonomic identifier. Following abbreviations are allowed: 'subsp.' (for subspecies); 'convar.' (for convariety); 'var.' (for variety); 'f.' (for form). |
| race | varchar | no | |
| population | varchar | no | Name of the population |
| div_trait | table | | table used to store trait values for an observational unit |
| div_trait_id | int | yes | primary key |
| div_trait_uom_id | int | yes | foreign key referencing trait being measured |
| div_statistic_type_id | int | yes | foreign key referencing statistic type |
| div_obs_unit_id | int | yes | foreign key referencing observational unit |
| date_measured | date | yes | date trait measured |
| value | varchar | yes | value of trait |
| div_trait_uom | table | | table used to control trait names, units of measure, and a link to the Gramene Trait Ontology. |
| div_trait_uom_id | int | yes | primary key |
| to_accession | int | no | foreign key referencing trait being measured (reference into Gramene database). |
| div_unit_of_measure_id | int | yes | foreign key referencing unit of measure |

| Table/Field | Type | Req'd | Description |
|----------------------------------|---------|-------|--|
| local_trait_name | varchar | yes | this defines the trait being measured and should correspond with the qtl_trait_ontology_id. This will always be defined even when a qtl_trait_ontology_id for the measured trait does not exist. |
| trait_protocol | text | no | Very important. A brief description of how the trait was collected, and if a rating, the minimum and maximum values |
| div_treatment | table | | table used to define treatment values or levels |
| div_treatment_id | int | yes | primary key |
| div_obs_unit_id | int | yes | foreign key referencing observational unit being treated |
| div_treatment_uom_id | int | yes | foreign key referencing type of treatment being applied |
| value | varchar | yes | value of this treatment |
| div_treatment_uom | table | | table used to control treatment units of measure and descriptors. Treatments can be experimental treatments or environment variables. |
| div_treatment_uom_id | int | yes | primary key |
| eo_accession | int | yes | foreign key referencing type of treatment being applied (reference into Gramene database). |
| div_unit_of_measure_id | int | yes | foreign key referencing unit of measure |
| local_treatment_name | varchar | yes | local treatment name |
| treatment_protocol | varchar | no | comments |
| div_unit_of_measure | table | | used to store/control possible units of measure |
| div_unit_of_measure_id | int | yes | primary key |
| unit_type | varchar | yes | defines unit of measure. (i.e. inches, lbs, etc) |
| cdv_allele_curated_allele | table | | Many-to-many between div_allele, cdv_curated_allele, and cdv_curation |
| cdv_allele_curated_allele_id | int | yes | primary key |
| cdv_curated_allele_id | int | yes | foreign key to cdv_curated_allele |
| div_allele_id | int | yes | foreign key to div_allele |
| cdv_curation_id | int | yes | foreign key to cdv_curation: Who assigned this observation to this curated allele? |
| cdv_curated_allele | table | | Curated allele name corresponding to observed div_alleles |
| cdv_curated_allele_id | int | yes | Primary key |
| name | varchar | yes | Curated name for this allele. |
| cdv_curation | table | | General curation table. |
| cdv_curation_id | int | yes | Primary key |
| cdv_reason_id | int | yes | Reason for an allele grouping, e.g., same SSR length. Foreign key to cdv_reason |
| curator | varchar | yes | Name of curator |
| curation_date | date | yes | Date curated |
| comments | text | no | Comments |
| cdv_map_feature | table | | table used to store map feature |
| cdv_map_feature_id | int | yes | Primary key |
| chromosome_name | text | yes | Chromosome name where this locus resides. |
| comments | text | no | Comments |
| genetic_bin | varchar | no | Position of the locus within a bin |
| genetic_map | varchar | no | Name of genetic map (map set) |
| xref_map_id | varchar | no | External database reference to a genetic map (e.g., "cu-dh-2001-6") |
| genetic_position | double | no | Position in cM along a chromosome |

| Table/Field | Type | Req'd | Description |
|---------------------------|-------------|--------------|---|
| locus_type | varchar | yes | Type of this locus (i.e. Gene, Cytological). |
| name | varchar | yes | Name of this locus. |
| physical_position | double | no | Physical position of this locus. |
| | | | |
| cdv_marker | table | | table used to store markers |
| cdv_marker_id | int | yes | Primary key |
| cdv_map_feature_id | int | yes | Foreign key to cdv_map_feature |
| div_ref_stock_id | int | no | Foreign key referencing associated stock |
| marker_aid | varchar | no | reference into Gramene database |
| name | varchar | yes | name |
| ref_seq | text | no | reference sequence |
| | | | |
| cdv_passport_group | table | | A user-selected working set of germplasm lines. |
| cdv_passport_group_id | int | yes | Primary key |
| name | varchar | no | User's name for this set of lines. |
| | | | |
| cdv_passport_set | table | | Many/many relation table for cdv_passport_group to div_passport |
| cdv_passport_set_id | int | yes | Primary key |
| cdv_passport_group_id | int | yes | Foreign key |
| div_passport_id | int | yes | Foreign key |
| | | | |
| cdv_reason | table | | Controlled vocabulary of reasons for creating a curation entry. |
| cdv_reason_id | int | yes | Primary key |
| description | text | yes | E.g., Same SSR length |