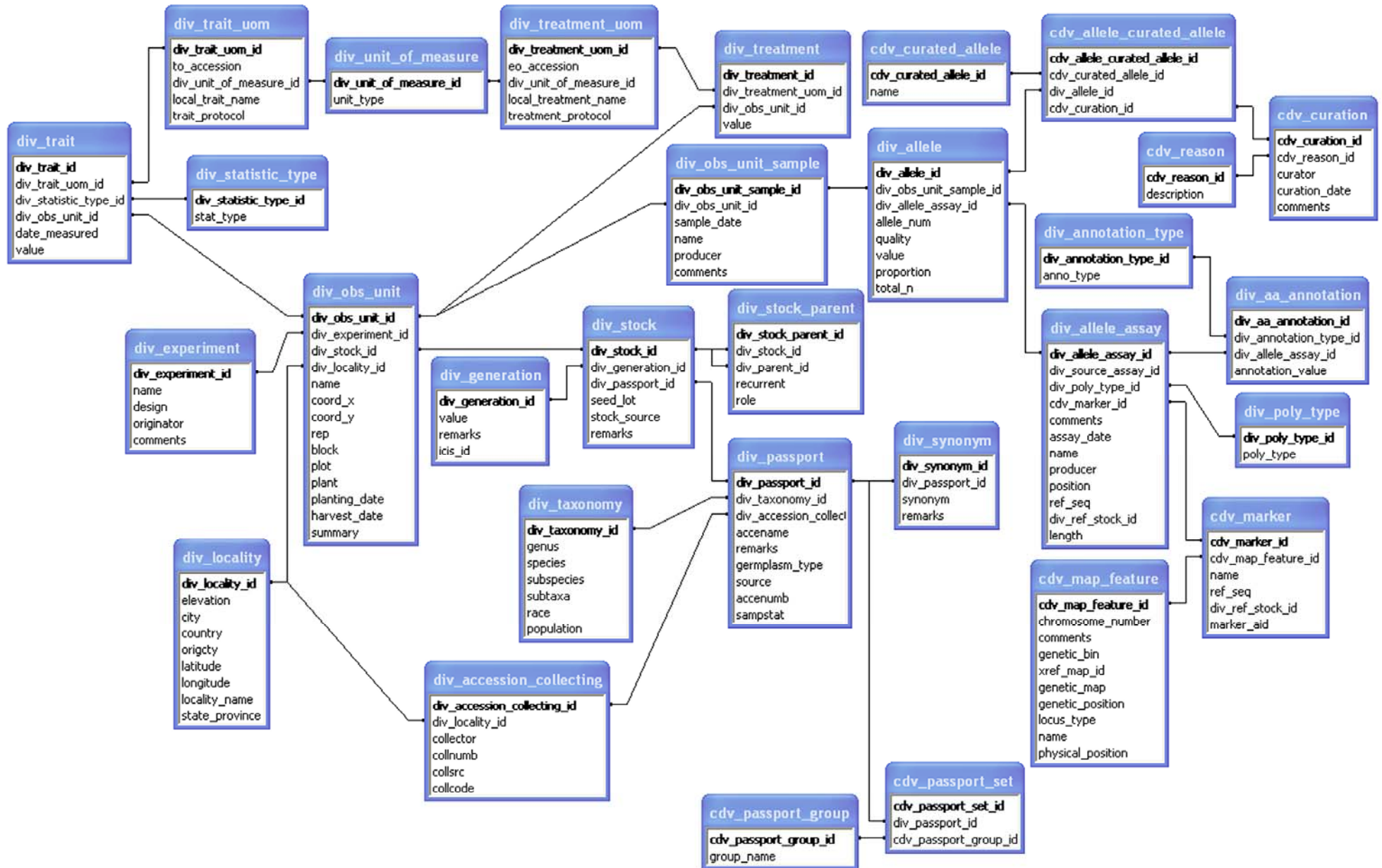


The Genomic Diversity and Phenotype Data Model (GDPDM) - August 24, 2004



**Table/Field descriptions for
The Genomic Diversity and Phenotype Data Model (GDPDM)
August 24, 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	number	yes	primary key
div_annotation_type_id	number	yes	foreign key pointing to the annotation type.
div_allele_assay_id	number	yes	foreign key pointing to the allele assay this is annotating.
annotation_value	text	yes	value for given type of this annotation.
div_accession_collecting	table		Optional, only for germplasm not in germplasm center. Original source of germplasm.
div_accession_collecting_id	number	yes	primary key
div_locality_id	number	yes	foreign key to locality
collector	text	no	Collector of the accession
collnumb	text	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	text	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	text	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	number	yes	primary key
div_obs_unit_sample_id	number	yes	foreign key pointing to observational unit sample.
div_allele_assay_id	number	yes	foreign key pointing to allele assay.
allele_num	number	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	memo	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	number	no	number of alleles with given value divided by total_n

Table/Field	Type	Req'd	Description
total_n	number	no	total number of homologous chromosomes surveyed (for diploids it is the number of gametes)
div_allele_assay	table		table used to record allele assays.
div_allele_assay_id	number	yes	primary key
div_marker_id	number	no	foreign key pointing to the marker associated with the reference sequence.
div_source_assay_id	number	no	foreign key pointing to source assay within this same table.
div_poly_type_id	number	yes	foreign key pointing to the polymorphism type of this assay
comments	text	no	comments about this allele assay.
assay_date	date	yes	date of this allele assay.
name	text	yes	name of this assay
producer	text	no	producer's name of this assay.
length	number	no	length of polymorphism being assayed relative to reference sequence.
position	text	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	number	no	foreign key pointing to stock associated with the reference sequence.
div_annotation_type	table		table used to record different types of annotations.
div_annotation_type_id	number	yes	primary key
anno_type	text	yes	value of this annotation type
div_experiment	table		table used to record high level experiments
div_experiment_id	number	yes	primary key
name	text	yes	name of this experiment
design	text	no	design of this experiment
originator	text	no	originator of this experiment
comments	text	no	comments
div_generation	table		table used to store/control all possible generation values
div_generation_id	number	yes	primary key
value	text	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	number	yes	primary key
elevation	number	no	Elevation of collecting site: [m asl] Elevation of collecting site expressed in metres above sea level. Negative values are allowed.
city	text	no	city
country	text	no	country in plain text

Table/Field	Type	Req'd	Description
origcty	text	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	number	no	Latitude of collecting site: this should be recorded as a decimal number.
longitude	number	no	Longitude of collecting site: this should be recorded as a decimal number.
locality_name	text	no	name of this locality
state_province	text	no	state or province
div_obs_unit	table		table used to store observational units
div_obs_unit_id	number	yes	primary key
div_experiment_id	number	no	foreign key pointing to associated experiment
div_stock_id	number	yes	foreign key pointing to stock planted
div_locality_id	number	no	foreign key pointing to associated locality
name	text	yes	name of observational unit
coord_x	number	no	x coordinate
coord_y	number	no	y coordinate
rep	text	no	group of blocks
block	text	no	group of plots
plot	text	no	smallest measure observed. (i.e. row in field)
plant	text	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	number	yes	primary key
div_obs_unit_id	number	yes	foreign key pointing to observational unit this sample was collected from
sample_date	date	yes	date sample collected
name	text	yes	name of this sample
producer	text	yes	producer of this sample
comments	text	no	comments
div_passport	table		It still needs to be determined how the codes used by these fields will be handled. May import them into separate tables? NOTE: Many of the passport fields were derived from the FAO/IPGRI Multi-Crop Passport Descriptors (http://www.ipgri.cgiar.org/publications/pdf/124.pdf)
div_passport_id	number	yes	primary key
div_taxonomy_id	number	no	foreign key to taxonomy table
div_accession_collecting_id	number	no	foreign key to accession_collecting table
accename	text	yes	Accession name: Either a registered or other formal designation given to the accession. First letter uppercase. Multiple names separated with semicolon without space. For example: Rheinische Vorgebirgstrauben;Emma;Avlon
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.

Table/Field	Type	Req'd	Description
germplasm_type	text	no	Deprecated: Use sampstat instead. Germplasm type of accession (i.e. Tripsacum, inbred)
source	text	no	Institute code or person's name. Immediate source of germplasm.
accenumb	text	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	number	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	number	yes	primary key
poly_type	text	yes	value of polymorphism (SNP, LENGTH, SEQALIGNMENT, SEQUENCE, RAWDATA, MOBILITY).
div_statistic_type	table		table used to store/control possible statistic types
div_statistic_type_id	number	yes	primary key
stat_type	text	yes	value of statistic type (measure, mean, variance)
div_stock	table		table used to store different stocks
div_stock_id	number	yes	primary key
div_generation_id	number	no	foreign key pointing to stock's generation
div_passport_id	number	yes	foreign key pointing to stock's passport information
seed_lot	text	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	text	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
div_stock_parent	table		table used to define any number of parents/ancestors of a given stock
div_stock_parent_id	number	yes	primary key
div_stock_id	number	yes	foreign key pointing to relevant stock
div_parent_id	number	yes	foreign key pointing to parent stock
recurrent	boolean	yes	Boolean value indicating if recurrent.
role	text	yes	Value indicating role of this parent (i.e. self, male, female).
div_synonym	table		Synonyms table.
div_synonym_id	number	yes	primary key
div_passport_id	number	yes	Foreign key pointing to div_passport table.
synonym	text	yes	The synonym.
remarks	text	no	Notes about the synonym.

Table/Field	Type	Req'd	Description
div_taxonomy	table		table used to store taxonomy
div_taxonomy_id	number	yes	primary key
genus	text	no	Genus: Genus name for taxon. Initial uppercase letter required.
species	text	yes	Species: Specific epithet portion of the scientific name in lowercase letters. Following abbreviation is allowed: 'sp.'
subspecies	text	no	Subspecies
subtaxa	text	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	text	no	
population	text	no	Name of the population
div_trait	table		table used to store trait values collection for a given phenotype and observational unit
div_trait_id	number	yes	primary key
div_trait_uom_id	number	yes	foreign key pointing to trait being measured
div_statistic_type_id	number	no	foreign key pointing to statistic type
div_obs_unit_id	number	yes	foreign key pointing to observational unit
date_measured	date	yes	date trait measured
value	text	yes	value of trait
div_trait_uom	table		table used to control unit of measure always used with given trait ontology.
div_trait_uom_id	number	yes	primary key
to_accession	number	no	foreign key pointing to trait being measured (pointer into Gramene database).
div_unit_of_measure_id	number	yes	foreign key pointing to unit of measure
local_trait_name	text	yes	this defines the trait being measured and should correspond with the qtl_trait_ontology_id. This also always a trait to be defined even when a qtl_trait_ontology_id for the measured trait does not exist.
trait_protocol	text	no	comments
div_treatment	table		table used to define a treatment to an observational unit. Any given observational unit could be treated in multiple ways
div_treatment_id	number	yes	primary key
div_obs_unit_id	number	yes	foreign key pointing to observational unit being treated
div_treatment_uom_id	number	yes	foreign key pointing to type of treatment being applied
value	text	yes	value of this treatment
div_treatment_uom	table		table used to control unit of measure always used with given treatment ontology
div_treatment_uom_id	number	yes	primary key
eo_accession	number	yes	foreign key pointing to type of treatment being applied (pointer into Gramene database).
div_unit_of_measure_id	number	yes	foreign key pointing to unit of measure
local_treatment_name	text	yes	local treatment name
treatment_protocol	text	no	comments
div_unit_of_measure	table		used to store/control possible units of measure
div_unit_of_measure_id	number	yes	primary key
unit_type	text	yes	defines unit of measure. (i.e. inches, lbs, etc)

Table/Field	Type	Req'd	Description
cdv_allele_curated_allele	table		Many-to-many between div_allele, cdv_curated_allele, and cdv_curation
cdv_allele_curated_allele_id	number	yes	primary key
cdv_curated_allele_id	number	yes	foreign key to cdv_curated_allele
div_allele_id	number	yes	foreign key to div_allele
cdv_curation_id	number	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	number	yes	Primary key
name	text	yes	Curated name for this allele.
cdv_curation	table		General curation table.
cdv_curation_id	number	yes	Primary key
cdv_reason_id	number	yes	Reason for an allele grouping, e.g., same SSR length. Foreign key to cdv_reason
curator	text	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	number	yes	Primary key
chromosome_number	number	yes	Chromosome number where this locus resides.
comments	text	no	Comments
genetic_bin	text	no	Position of the locus within a bin
genetic_map	text	no	Name of genetic map (map set)
xref_map_id	text	no	External database reference to a genetic map (e.g., "cu-dh-2001-6")
genetic_position	number	no	Position in cM along a chromosome
locus_type	text	yes	Type of this locus (i.e. Gene, Cytological).
name	text	yes	Name of this locus.
physical_position	number	no	Physical position of this locus.
cdv_marker	table		table used to store markers
cdv_marker_id	number	yes	Primary key
cdv_map_feature_id	number	yes	Foreign key to cdv_map_feature
div_ref_stock_id	number	no	Foreign key pointing to associated stock
marker_aid	text	no	Pointer into Gramene database
name	text	yes	name
ref_seq	text	no	reference sequence
cdv_passport_group	table		A user-selected working set of germplasm lines.
cdv_passport_group_id	number	yes	Primary key
name	text	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	number	yes	Primary key
cdv_passport_group_id	number	yes	Foreign key
div_passport_id	number	yes	Foreign key
cdv_reason	table		Controlled vocabulary of reasons for creating a curation entry.
cdv_reason_id	number	yes	Primary key
description	text	yes	E.g., Same SSR length