

**Table/Field descriptions for
The Genomic Diversity and Phenotype Data Model (GDPDM)
May 10, 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_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
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.
position	text	no	
ref_seq	text	yes	reference sequence used.
div_ref_stock_id	number	no	foreign key pointing to stock associated with the reference sequence.
length	number	no	
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

Table/Field	Type	Req'd	Description
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
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
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_locus	table		table used to store loci
div_locus_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	number	no	Position in cM along a chromosome
genetic_position	number	no	Genetic Position of this locus.
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.
div_marker	table		table used to store markers
div_marker_id	number	yes	primary key
div_ref_stock_id	number	no	foreign key pointing to associated stock
div_locus_id	number	no	foreign key pointing to associated locus
name	text	yes	name
ref_seq	text	no	reference sequence
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
field_coord_x	number	no	x coordinate

Table/Field	Type	Req'd	Description
field_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_locality_id	number	no	foreign key pointing to this passport's locality.
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
collnumb	text	no	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.
collector	text	no	Collector of the accession.
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.
genus	text	no	Genus: Genus name for taxon. Initial uppercase letter required.
germplasm_type	text	no	Germplasm type of accession (i.e. Tripsacum, inbred)
local_name	text	no	Local name of the accession.
population	text	no	Name of the population
race_name	text	no	Name of the race
reference	text	no	Reference for collection
secondary_source	text	no	Secondary source for the accession
source	text	no	Source of the accession.
species	text	yes	Species: Specific epithet portion of the scientific name in lowercase letters. Following abbreviation is allowed: 'sp.'
subspecies	text	no	Subspecies

Table/Field	Type	Req'd	Description
instcode	text	no	Institute code: Code of the institute where the accession is maintained. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus a number. The current set of Institute Codes is available from the FAO website (http://apps3.fao.org/wiews/).
accnumb	number	no	Accession number: This number serves as a unique identifier for accessions within a genebank collection, and is assigned when a sample is entered into the genebank collection.
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.
spauthor	text	no	Species authority: Provide the authority for the species name.
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).
subtauthor	text	no	Subtaxa authority: Provide the subtaxa authority at the most detailed taxonomic level.
cropname	text	no	Common crop name: Name of the crop in colloquial language, preferably English (i.e. 'malting barley', 'cauliflower', or 'white cabbage')
acqdate	date	no	Acquisition date: [YYYYMMDD] Date on which the accession entered the collection where YYYY is the year, MM is the month and DD is the day. Missing data (MM or DD) should be indicated with hyphens. Leading zeros are required.
colldate	date	no	Collecting date of sample: [YYYYMMDD] Collecting date of the sample where YYYY is the year, MM is the month and DD is the day. Missing data (MM or DD) should be indicated with hyphens. Leading zeros are required.
bredcode	text	no	Breeding institute code: Institute code of the institute that has bred the material. If the holding institute has bred the material, the breeding institute code (BREDCODE) should be the same as the holding institute code (INSTCODE). Follows INSTCODE standard.
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).

Table/Field	Type	Req'd	Description
collsrc	number	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)
donorcode	text	no	Donor institute code: Code for the donor institute. Follows INSTCODE standard.
donor numb		no	Donor accession number: Number assigned to an accession by the donor. Follows ACCENUMB standard.
other numb	number	no	Other identification: (numbers) associated with the accession (OTHERNUMB) Any other identification (numbers) known to exist in other collections for this accession. Use the following system: INSTCODE:ACCENUMB;INSTCODE:ACCENUMB;... INSTCODE and ACCENUMB follow the standard described above and are separated by a colon. Pairs of INSTCODE and ACCENUMB are separated by a semicolon without space. When the institute is not known, the number should be preceded by a colon.
duplsite	text	no	Location of safety duplicates: Code of the institute where a safety duplicate of the accession is maintained. Follows INSTCODE standard.
storage	number	no	Type of germplasm storage: If germplasm is maintained under different types of storage, multiple choices are allowed, separated by a semicolon (e.g. 20;30). (Refer to FAO/IPGRI Genebank Standards 1994 for details on storage type.) 10) Seed collection 11) Short term 12) Medium term 13) Long term 20) Field collection 30) In vitro collection (Slow growth) 40) Cryopreserved collection 99) 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, RAWDATA).
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	yes	foreign key pointing to stock's generation
div_passport_id	number	yes	foreign key pointing to stock's passport information
seed_lot	text	yes	Field Book Entry

Table/Field	Type	Req'd	Description
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_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
qtl_trait_ontology_id	number	no	foreign key pointing to trait being measured
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.
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
qtl_treatment_ontology_id	number	yes	foreign key pointing to type of treatment being applied
div_unit_of_measure_id	number	yes	foreign key pointing to unit of measure
local_treatment_name	text	yes	local treatment name
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)
qtl_trait_ontology	table		table used to store phenotypes
qtl_trait_ontology_id	number	yes	primary key
qtl_treatment_ontology	table		table used to store treatments
qtl_treatment_ontology_id	number	yes	primary key