VectorBase PopBio Map Data Dictionary

Data types

There are 5 specialized views in PopBio Map to reflect the major kinds of data contained within our database.

Each data type will have different associated fields, although some fields will be common to many or all data types.

Many of our fields utilize a controlled vocabulary - these terms may be hierarchical and links to the tree is provided.

Note that for many data types, not every field name will have a value for every record. For example, a pathogen assay that only determined a single “presence” vs “absence” in the Infection status field on the entire pool (e.g. result of a PCR assay for an arbovirus on a pool of mosquitoes) would not report a percentage infected mosquitoes, whereas an assay dissecting mosquito midguts for the presence of malaria would report a Phenotype value, Phenotype value unit, and Phenotype value type because the percent of infected mosquitoes would be reported.

Common fields

Information on fields common to all or many data types

Sample ID Unique ID for every pool (or individual, if non-pooled data) of organisms collected per trap per collection per species. See abundance fields for special notes.

Species The taxonomic identification of the organism(s) collected/assay - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=VBCV&t=VBcv%3A0000981

Specimens collected The number (“count”) of organisms collected (for example in a trap, in a given collection). See note under abundance collections.

Sample type

Pool – assay was performed or reported on multiple organisms (e.g. 30 mosquitoes are ground up and an assay performed on the lysate)

Individual – assay was performed on a single organism (e.g. the malaria status of a single mosquito was determined)

Sample size Number of organisms that were assayed to determine

Collection ID Unique ID for all organisms collected at a single time. For example, all the organisms collected in a single passive trap during one collection period would share the same Collection ID

Projects Unique VectorBase data ID for a given project. A project may reflect all data from a given publication. For large multi-year surveillance data sets, a project typically represents a year's worth of data from a given data provider.

Collection date range Date *or range of dates* that the organism could have been collected from. When known this could be the exact dates a trap was set. Alternately, this could be the most precise possible dates of collection known (e.g. year sample was collected).

Label Human readable description of the sample

Collection protocols Method by which the organisms were collected. Often, this a particular trap type - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A30000044

Latitudes GPS latitude for the location of the collection event (WGS84)

Longitudes GPS longitude for the location of the collection event (WGS84)

Locations Geographic place name. VectorBase uses a modified version of the GADM (https://gadm.org) database to assign place names from GPS points - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=GADM&t=VBGEO%3A0000003

Citations PubMed IDs and DOIs of the data

Sex The sex of the organism(s) collected.

Male
Female
Mixed – most often used when sex ratio not reported/determined

Developmental stage The developmental stage of the organism assyed (which may be different from the stage collected in the field, see collection protocols) - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=IDOMAL&t=IDOMAL%3A0000644

Attractants If known, the attractants used to attract organisms - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=IRO%3A0000034

Usage license If known, the creative commons re-use license. Please see https://creativecommons.org/share-your-work/licensing-types-examples/ for more information

CC BY – data can be reused, with attribution
CC BY-NC - data can be reused, with attribution, for non-commercial applications

Tag Link projects together, such as all data collected by a given mosquito abatement authority. Could also link data providers/curators curation work together such as all projects curated by a given partner - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=VBCV&t=VBcv%3A0001076

Abundance fields

Sample ID Unique ID for every pool of organisms collected per trap per collection per species. For abundance records, all zero values (confirmed absences) for all species at that collection share the same Sample ID

Collection duration days For passive traps, the number of days the trap was set. Often, for mosquitoes, this actually is "trap nights."

Specimens collected The number ("count") of organisms collected (for example in a trap, in a given collection). Note, most data is from fixed trapping sites. Meaning that the same trap was checked on a repeated basis (e.g. daily/weekly) Therefore, zeros can be interpreted as a true absence. VectorBase infers zeros when for a given data submitter (e.g. a mosquito abatement district) we know they are looking for a given species (because they reported that species during the year in at least one of their traps one time); we know a trap was set and successfully collected because they either (1) reported a collection of at least one species from that trap on a given day or (2) explicitly reported an empty trap. Traps where malfunctions / failures (e.g. flooded, infested with ants) are removed from our records (would not appear as zero counts).

Insecticide resistance phenotype fields

Sample ID Unique ID for every individual or pool collected per trap per collection per species.

Assay ID A unique VectorBase identifier for a given assay performed on a pool/individual mosquito. A given pool/individual mosquito (which will have a unique Sample ID) could have multiple assays performed on it (e.g. a pool of mosquito lysate could be assayed for activity against three different insecticides, and would thus have 3 Assay IDs).

Concentration Unitless concentration of insecticide compound assayed

Concentration unit Unit of concentration of insecticide (e.g. “percent”, “microgram per milliliter”

Duration The amount of time that the organisms were exposed to the insecticidal substance

Duration unit Unit of duration

Phenotype value Unitless result of insecticidal assay

Phenotype value unit Unit of experimentally determined phenotype value (e.g. “percent”, “milligram per liter”, or “microgram per milliliter”)

Phenotype value type Assay end point measure type (e.g. “LC50, LT50, mortality rate) - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=VBCV&t=VBcv%3A0000702

Insecticide The insecticide the organisms were exposed to - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A10000239

Protocols The insecticide resistance test that was performed on the organisms - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A20000001

Insecticide resistance genotype fields

Phenotypes

Assay ID A unique VectorBase identifier for each given assay performed on a pool/individual mosquito. A given pool/individual mosquito (which will have a unique Sample ID) could have multiple assays performed on it (e.g. a given mosquito sample could be assayed for three different genotypes at a given loci or even tested at multiple loci and would thus have 3 Assay IDs)

Genotype Type

Genotype Name The SNP name. Genomic location name is based on the originally reported name of mutation, often from Drosophila - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A10000220

Mutated Protein Value the percentage of a sample of mosquitoes with the given genotype.

Pathogen status fields

Assay ID A unique VectorBase identifier for each given assay performed on a pool/individual mosquito. A given pool/individual mosquito (which will have a unique Sample ID) could have multiple assays performed on it (e.g. a given mosquito sample could be assayed for three different arbovirus and would thus have 3 Assay IDs).

Protocols The method used to determine pathogen status - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A20000001

Pathogen Name of the pathogen that was assayed for - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=VSMO&t=VSMO%3A0000042

Infection status The result of an assay for if a given pathogen was present (or not) in the given mosquito sample. ("present" or "absent")

Phenotype value Unitless value for infection prevalence - if quantified

Phenotype value unit unit of Phenotype value

Phenotype value type Assay end point measure type - if quantified (e.g. "infection prevalence")

Bloodmeal host fields

Assay ID A unique VectorBase identifier for a given assay performed on a pool/individual organism. A given pool/individual mosquito (which will have a unique Sample ID) could have multiple assays performed on it (e.g. a mosquito could be assayed by PCRs for 3 different hosts, and would thus have 3 Assay IDs).

Blood meal host The host species(s) of organism assayed as the source of the blood - see our hierarchical ontology for this field at https://www.vectorbase.org/ontology-browser?cv=VBCV&t=VBcv%3A0001004

Protocols The method that was used to determine blood meal host (e.g. "PCR" , enzyme-linked immunosorbent assay , "sequencing") - Most of the terms for this field can be found on our hierarchical ontology for at https://www.vectorbase.org/ontology-browser?cv=MIRO&t=MIRO%3A20000004

Phenotype value Unitless value of host bloodmeal source

Phenotype value unit Unit of phenotype value (e.g. percent)

Phenotype value type Assay end point measure type (e.g. “host blood index”)