

Description of PostgreSQL amphibians database

Roland Knapp, 07 April 2020

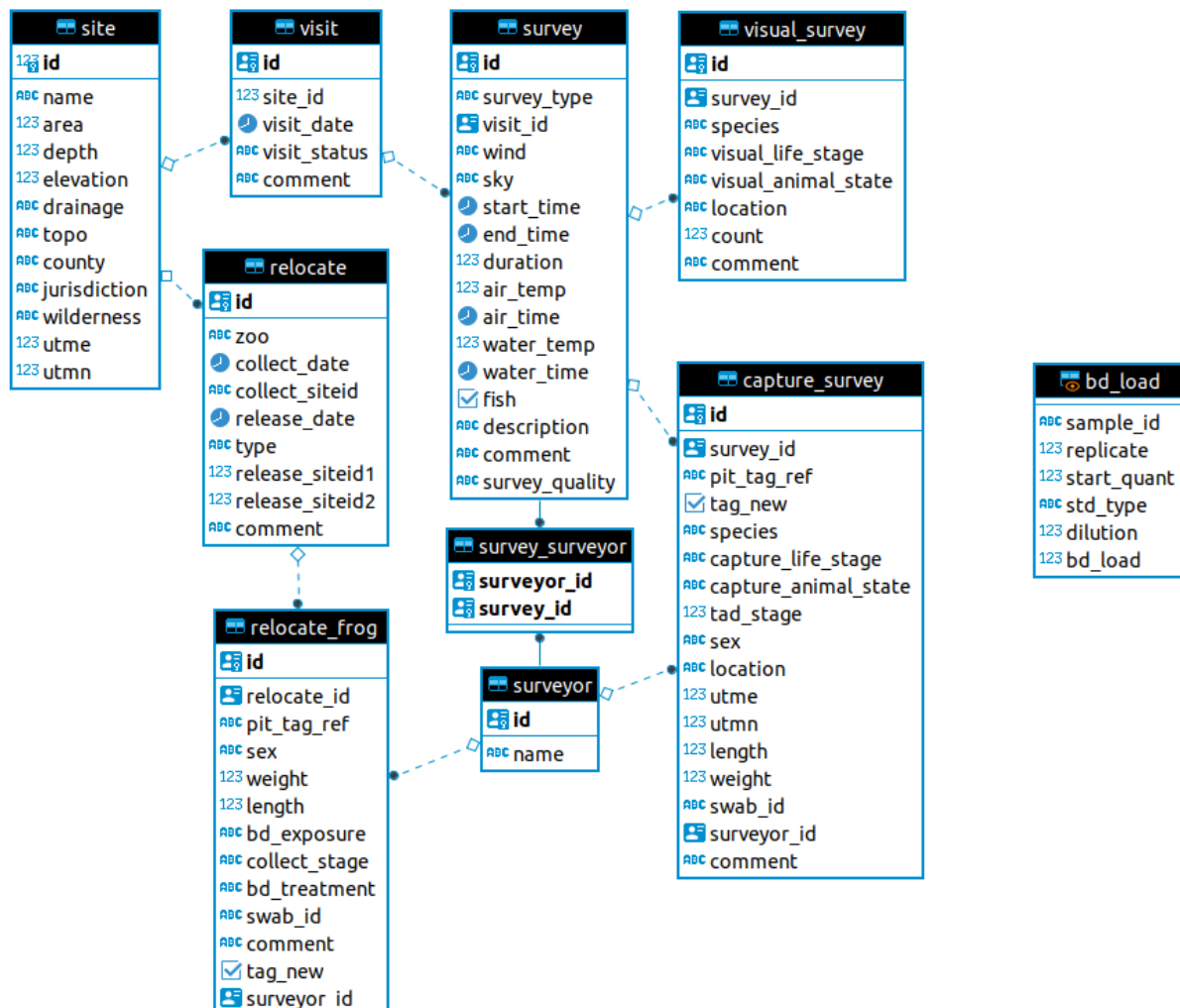
General comments:

In the following database description, the format xxxxx.yyyyy is used to refer to a specific table.column combination (e.g., “site.id” refers to the “id” column in the “site” table). When column names in database tables include two or more words, those words are separated with an underline (“_”) both in the database columns and in the descriptive text below. These conventions are particularly important to understand because of their relevance for understanding table relationships. Specifically, the foreign key in the database tables typically reflects the name of the column in the parent table that it references. For example, the “visit_id” column in the survey (child) table (survey.visit_id) references the “id” column in the visit (parent) table (visit.id).

Regarding Null values and empty strings, in general Null values are those that are not known/not given/missing, and empty strings are those that should be empty or are an optional entry. Column-specific information provides additional details.

See “Additional notes” at end of document for additional information about the database.

Database structure (from DBeaver, as of 18-January-2021)



Lines between tables show foreign key constraints, with dashed lines representing foreign key constraints set to a non-primary key column in the child table, and solid lines representing those set to a primary key column in the child table. For each line, white squares indicate the parent table and black circles indicate the child table.

Site: Characteristics of each site

- Id (integer, primary key): 5-digit SLIP site id. In general, 1XXXX = Kings Canyon NP, 2XXXX = Sequoia National Park, 4XXXX = CDFW Region 4/Sierra and Sequoia National Forests, 5XXXX = CDFW Region 6/Inyo National Forest, 7XXXX = Yosemite NP, 8XXXX = areas north of Yosemite including Humboldt-Toiyabe NF, Desolation Wilderness.
- Name (text): Official name for site from 7.5' USGS topographic map. Unofficial name sometimes included, in quotes. Use empty string for unnamed sites
- Area (integer): Site area in square meters from GIS. $10,000 \text{ m}^2 = 1 \text{ hectare}$. Missing values indicated by Null.
- Depth (real): Maximum water depth of site. Missing values indicated by Null.
- Elevation (integer): Altitude in meters, from USGS 7.5' topographic map. Missing values indicated by Null.
- Drainage (text): String of named waterways into which the site drains, starting with the highest stream order (e.g., sn_jqn_sf_goddard_ck_disappearing_ck). Missing values indicated by Null.
- Topo (text): Name of 7.5' topographic map containing the site. Missing values indicated by Null.
- County (text): County in which the site is located. Missing values indicated by Null.
- Jurisdiction (text): Name of land management agency with jurisdiction over site. Missing values indicated by Null.
- Wilderness (text): Name of designated wilderness in which site is located. Missing values indicated by Null.
- Utme, Utmn (integer): UTM coordinates of site centroid, from GIS. UTM zone is 11 for all sites except those in Desolation Wilderness (zone 10). Projection is NAD83. Missing values indicated by Null.

Visit: Site status during visit, with one visit.id per site_id/visit_date

- Id (uuid, primary key): uuid generated with gen_random_uuid().
- Site_id (integer; foreign key, references site.id): See site.id above. Null values not allowed.
- Visit_date (date): Date on which site was visited (as yyyy-mm-dd). Null values not allowed.
- Visit_status (enum): Status of the site at time of visit. Possible values are "suitable", "dry", "frozen", or "other". Null values not allowed.
- Comment (text): Description of the site and/or site conditions during visit, especially those that could affect survey results. Comments directly related to any survey conducted during the visit (ves, cmr, swab) should not be included here. No comment indicated by empty string.

Survey: Description of the surveys conducted at a site on a particular date, including ves, cmr, and swab.

- Id (uuid, primary key): uuid generated with gen_random_uuid(). Note that only one survey type (see next bullet) is allowed per survey.id.
- Survey_type (enum): Possible values are “ves”, “cmr”, or “swab” Null values not allowed.
- Visit_id (uuid; foreign key, references visit.id): See visit.id above.
- Wind (enum): Wind conditions during majority of survey. Possible values are “calm”, “breezy”, “windy”, or Null.
 - Calm: wind is <5 mph. No or minimal ripples on lake surface, no movement of tree branches.
 - Breezy: wind is 5-15 mph. Significant ripples on lake surface, tree branches show some movement.
 - Windy: Wind is >15 mph. Waves (and whitecaps with stronger wind) on lake surface, tree branches show marked swaying.
- Sky (enum): Sky conditions during majority of survey. Possible values are “clear”, “p_cloudy”, “m_cloudy”, “precip”, or Null.
 - Clear: <5% of sky with clouds.
 - P_cloudy (partly cloudy): 5-50% of sky with clouds.
 - M_cloudy (mostly cloudy): >50% of sky with clouds.
 - Precip (precipitating): Rain or snowfall occurring.
- Start_time (time, 24-hr format): Time when survey started. Missing values indicated by Null.
- End_time (time, 24-hr format): Time when survey ended. Missing values indicated by Null.
- Duration (integer): Total time spent surveying, in minutes. For ves and swab, survey duration of each surveyor should be summed. For cmr, survey duration is not summed across each surveyor. Missing values indicated by Null.
- Air_temp (integer): Air temperature, typically measured at the beginning or end of the survey. Measurement is made in the shade and at a height of 1 m. Missing values indicated by Null.
- Air_time (time, 24-hr format): Time at which air temperature was measured. Missing values indicated by Null.
- Water_temp (integer): Water temperature, typically measured at the beginning or end of the survey. Measured approximately 0.5 m out from shore and 10 cm below the water surface. Missing values indicated by Null.
- Water_time (time): Time at which water temperature was measured. Missing values indicated by Null.
- Fish (boolean): Whether or not fish were seen during survey. Possible values are “true”, “false”, or Null.

- Description (text): Description of how the survey was conducted, in particular what habitats were visited. Missing values indicated by Null.
- Comment (text): Information about the survey, including conditions experienced that could have affected survey results in a favorable or unfavorable manner (e.g., weather, water clarity). No comment indicated by empty string.

Visual_survey: Results of visual encounter surveys

- Id (uuid; primary key): uuid generated with gen_random_uuid().
- Survey_id (uuid; foreign key, references survey.id): See survey.id above.
- Species (enum): Abbreviation of scientific name for amphibian/reptile species. Possible values are “amma” (*Ambystoma macrodactylum*), “bubo” (*Bufo* [= *Anaxyrus*] *boreas*), “buca” (*Bufo* [= *Anaxyrus*] *canorus*), “clma” (*Clemmys marmorata*), “hyre” (*Hyla* [= *Hyllola*] *regilla*), “raca” (*Rana cascadae*), “ramu” (*Rana muscosa/sierrae*), “tato” (*Taricha torosa/sierrae*), “thco” (*Thamnophis couchii*), “thel” (*Thamnophis elegans*), and “thsi” (*Thamnophis sirtalis*). Null values not allowed.
- Visual_life_stage (enum): Life stage observed. Possible values are “adult”, “subadult”, “tadpole”, or “eggmass”. Null values not allowed.
- Visual_animal_state (enum): State of observed animal. Possible values are “alive” or “dead”. Null values not allowed.
- Location (enum): Location where animal was observed. Possible values are “lake”, “inlet”, “outlet”, or “fringing”. Null values not allowed.
- Count (integer): Count made during ves. Null values not allowed.
- Comment (text): Additional information regarding the species/life stage/animal state/location record. No comment indicated by empty string.

Capture_survey: Results of cmr and swab surveys.

- Id (uuid; primary key): uuid generated with gen_random_uuid().
- Survey_id (uuid; foreign key, references survey.id): See survey.id above.
- Pit_tag_ref (text): 15-digit pit_tag_id. Missing values indicated by Null.
- Tag_new (boolean): Indicates whether a new PIT tag was inserted into the captured frog. Possible values are “true”, “false”, or Null. Column was added prior to the 2008 field season, but for 2008-2018 the possible values were “true” or Null. This was somewhat error-prone because an entry was not required. To correct this, prior to the 2019 field season the field was changed to require an entry (“true” or “false”). To increase consistency across years, for the 2008-2018 data Null values were subsequently changed to “false”, and for the 2005-2007 data were updated to include “true”, “false”, or Null values. In 2008, ten subadult frogs at 70439 were tagged with alpha tags. For these records, pit_tag_ref is Null and tag_new is “true”.
- Species (enum): See visual_survey.species above. Null values not allowed.
- Capture_life_stage (enum): See visual_survey.capture_life_stage above. Null values not allowed.
- Capture_animal_state (enum): State of captured animal. Possible values are “healthy”, “sick”, “dead”, or Null.
- Tad_stage (integer): Gosner stage of captured tadpole.
- Sex (char(1)): Sex of captured adult. Possible values are “f”, “m”, Null.
- Location (enum): Location where animal was captured. Possible values are “lake”, “inlet”, “outlet”, “fringing”, or Null.

- **Utme, utmn (integer):** UTM coordinates of location where animal was captured, from handheld GPS. UTM zone is 11 for all sites except those in Desolation Wilderness (zone 10). Projection is NAD83. Missing values indicated by Null.
- **Length (integer):** Snout-vent length of captured animal, in millimeters. Missing values indicated by Null.
- **Weight (integer):** Weight of captured animal, in grams (total weight – bag weight). Missing values indicated by Null.
- **Swab_id (text; foreign key, references swab.id):** Identifier of skin swab sample. When no sample was collected, value is Null.
- **Surveyor_id (uuid; foreign key, references surveyor.id):** uuid generated with `gen_random_uuid()`. Null values not allowed. Identifies surveyor who conducted swabbing.
- **Comment (text):** Additional information about the captured animal. No comment indicated by empty string.

View: *bd_load*: Collected skin swabs described in `capture_survey` table are analyzed using a qPCR Bd assay. Results are stored in the `qpcr` database (along with other assay metadata). These data are made available to the amphibians database via a View (i.e., saved query) and foreign data wrapper of the `qpcr.public` schema (both stored in the amphibians database). The View can be added to any query using a join statement in the same way that a table is added. Opening the View in the amphibians database displays all of the swabs collected to date starting in 2004 by our research group, partners (San Francisco and Oakland Zoos), and clients (Yosemite National Park, Plumas National Forest, Stanislaus National Forest, etc.) and analyzed at UCSB or SNARL.

- **Sample_id (text):** This corresponds to `swab_id` in the `capture_survey` table.
- **Replicate (integer):** Replicate number, range = 1-5. For samples run singly, replicate = 1.
- **Start_quant (real):** Starting quantity of Bd DNA (copies of ITS1 region) in subsample of swab DNA extract.
- **Std_type (text):** Standard type used in the qPCR Bd assay (genomic, gblock, or plasmid).
- **Dilution (integer):** Factor by which the sample was diluted to reduce or remove the effect of qPCR inhibitors (caused by contaminants such as tannins). For samples that amplify normally, dilution = 1. For samples showing inhibition using dilution = 1, sample is diluted 10-fold and re-analyzed (dilution = 10).
- **Bd_load (real):** Number of ITS1 copies per swab. For qPCR assays using a plasmid or gblock standard, $bd_load = start_quant * dilution * 80$. Multiplying by 80 accounts for the 80-fold dilution that occurs during DNA extraction and qPCR. For qPCR assays using genomic standard, $bd_load = start_quant * dilution * 80 * 60$ (multiply by 60 because there are approximately 60 ITS copies per zoospore in Sierra Nevada Bd strains).

Relocate: Description of all translocation and reintroduction events.

- **Id (uuid; primary key):** uuid generated with `gen_random_uuid()`.
- **Zoo:** Zoo where collected animals were raised to adulthood. Current values are “sfz” (San Francisco Zoo) and “oak” (Oakland Zoo). **Change to enum data type?**
- **Collect_date:** Date on which collection was made (yyyy-mm-dd).

- Collect_siteid: Site id(s) from which collection was made. Entered data is non-standard in that more than one site id can be entered, separated by commas (no spaces).
- Release date: Date on which collected animals were released (yyyy-mm-dd).
- Type: Type of relocation ("translocation", "reintroduction"). **Change to enum data type?**
- Release_siteid1: Site id at which relocation was conducted.
- Release_siteid2: Site id of second relocation site, if relocation was conducted at a second site. This is a non-standard table structure, and the second site should probably best be entered as a second row (with its own uuid).
- Comment about relocation event.

Relocate_frog: Description of translocated or reintroduced frogs.

- Id (uuid; primary key): uuid generated with gen_random_uuid().
- Relocat_id (uuid; foreign key, references relocate.id): See relocate.id above.
- Pit_tag_ref (text): 15-digit pit_tag_id. Missing values indicated by Null.
- Sex (char(1)): Sex of adult at time of release. Possible values are "f", "m", Null.
- Weight (integer): Weight of adult at time of release, in grams. Missing values indicated by Null.
- Length (integer): Snout-vent length of adult at time of release, in millimeters. Missing values indicated by Null.
- Bd_exposure (text): Whether or not animal was exposed to Bd prior to release in an effort to initiate an immune response ("exposed", "unexposed").
- Collect_stage (text): Life stage of animal at time of collection. Exact life stage is sometimes unknown, for example if tadpoles and metamorphs were collected and these animals were not kept separately at zoo. When more than one life stage was collected, the possibilities are listed, separated by commas (non-standard format). c
- Bd_treatment (text): When animals were treated with itraconazole prior to relocation, "treated" and "untreated" (control) animals are categorized as such.
- Swab_id (text): Identifier of skin swab sample. When no sample was collected, value is empty string. Note that this is different from how swab_id is treated in the capture_survey table, because non-adult life stages are generally not swabbed at the time of collection. In this case, swab_id should be empty and is not missing.
- Comment (text): Comment about adult at time of release.
- Tag_new (boolean): For translocated adults, whether the animal was tagged as a part of relocation event (TRUE), or was either not tagged or had been tagged prior to the relocation event (FALSE). Missing values are indicated by Null.
- Surveyor_id (uuid; foreign key, references surveyor.id): uuid generated with gen_random_uuid(). **Null values not allowed**. Identifies surveyor who conducted swabbing.

Swab_result_deprecated: Note: This table is in the amphibians database, but has been superceded by the bd_load View. To remove any links between this table and all tables in the amphibians database, all foreign keys have been deleted.

Results from qPCR Bd assay applied to skin swab samples. For specific information regarding qPCR plates and sample results (e.g., type of standard used in the analysis of the plate/sample), see the "qpcr" database and accompanying database description.

- Id (uuid, primary key): uuid generated with gen_random_uuid().

- Swab_id (text; foreign key, references swab.id). Swab_id of analyzed sample. Null values not allowed.
- Replicate (integer): Replicate number. For samples run singly, replicate = 1.
- Start_quant (real): Starting quantity of ITS DNA in subsample of extract from swab sample.
- Bd_load (real): Number of ITS copies per swab. For qPCR assays using a plasmid standard, $\text{bd_load} = \text{start_quant} * 80$ (to account for 80-fold dilution during DNA extraction and qPCR). For qPCR assays using genomic standard, $\text{bd_load} = \text{start_quant} * 80 * 60$ (multiply by 60 because there are approximately 60 ITS copies per zoospore in Sierra Nevada Bd strains).

Additional notes:

- All comment fields (in the visit, survey, visual_survey, and capture_survey tables) and description field (in the survey table) are cleaned up using the comment_cleaner.sql script. Script replaces null values with empty strings (or empty strings with null values, depending on the column in question), comments containing only spaces or periods with empty strings, carriage returns and line feed symbols with spaces, and multiple consecutive spaces with single spaces. It also formats all text to lower case.