# Targeted development of registries of biological parts – online supplement

**Jean Peccoud\*, Megan F. Blauvelt, Yizhi Cai, Kristal L. Cooper, Oswald Crasta, Emily DeLalla, Clive Evans, Otto Folkerts, Blair Lyons, Rebecca Shelton, Shrinivasrao P. Mane, Matthew A. Sweede, Sally A. Waldon**

**\*Corresponding author.**

## Database infrastructure

A custom MySQL database was developed to organize the data generated by this project. The database can be accessed at [www.genocad.org/phpMyAdmin](www.genocad.org/phpMyAdmin) using the parameters below:

login name: biobrick_viewer
password: parts

Below is the description of the tables in this database along with a short description of the fields when needed.

- **dna_repository**: Our database is organized around the notion of clones which are the plasmids distributed to the iGEM teams. This table is therefore the spine of our database.

    - **CloneID**: a unique identifier of the clones.

    - **Bba_number**: this is the key used by the MIT Registry to identify its entries. This key points to the biobrick_parts table.

    - **Plasmid**: This key points to the plasmid table used to record the plasmid backbone in which the parts are inserted.

    - **physical_sequence**: this is the sequence assembled from multiple sequencing runs.

    - **Selected_for_sequencing**: Indicates if the clone was sequenced

    - **Rec_date**: date of last modification of the record.

- **Biobrick_parts:** This table records information relative to biological parts, i.e. functional DNA fragment.

    - **ID:** a unique identifier for the biobricks.

    - **Bba_number:** The identifier used by the Registry of Standard Biological Parts

    - **Dna_sequence:** the part theoretical or published sequence. This sequence information was imported from the Registry.

    - **Part_type:** An integer used to indicate the category of the parts. Categories are listed in the part_type table. They correspond to the

functional categories used by the Registry such as Ribosome Binding Site, Regulatory, Coding, etc.

- o **description**: A short description of the part.
- o **Rec_date**: date of last modification of the record.

- **Plate_key**: The DNA Repository came in 4 384-well plates that we remapped to a set of 16 96-well plates. For each clone in the repository, this table indicates in which well it can be found in the 384-well and 96-well sets.

- **Peaks**: This table is used to record the electrophoresis data after the PCR step. Each clone is associated through the cloneID field to 0 or more DNA fragments or electrophoresis peaks in this table. The fields describing the peaks are documented in the user manual of the Agilent 2100 Expert software.

- **Nanodrop**: This table is used to record the spectrophotometric data generated by the software controlling the Nanodrop instrument.

- **Parts_contained_sequences**: This table is used to record the sequence inclusion relationships used to draw Figure 1. The master_sequence_id is the identifier of a BioBrick whose sequence contains the contained_sequence_id. Master_part_type and contained_sequence_type points to the category of each of the corresponding parts. The fields corresponding to start and end positions of the alignment are self-explanatory. Relationship_flagged is used to indicated the reduced set of relationships. Relations flagged with 1 constitute the reduced set used to generate Figure 1.

# Downloadable files

The following supporting files are available in the online supplement:

- `Registry.cys` is the files used to generate figure 1. By importing this file into Cytoscape, it is possible to interactively explore the Registry connectivity diagram.

- `MyVBIonBBa_tophits.xls` is an Excel file summarizing the results of the Blast analysis that can be used to analyze the quality of the match between published and assembled sequence on a clone by clone basis.

In addition, trace files can be downloaded from [www.genocad.org/biobricks](www.genocad.org/biobricks) `.Trace_files.zip` is an archive containing the trace files generated by the sequencers. They can be used to reassemble the sequences using different parameters. This archive also contains sequence data for clones that could not be assembled.

## parts_contained_sequences

| 🔑 ID | int unsigned |
|---|---|
| **master_sequence_id** | varchar (25) |
| **master_part_type** | int |
| **contained_sequence_id** | varchar (25) |
| **contained_part_type** | int |
| **contained_sequence_length** | int |
| **master_seq_start_position** | int |
| **master_seq_end_position** | int |
| relationship_flagged | char (1) |

FK1parts_contained_sequences

FK2parts_contained_sequences

## part_type

| 🔑 **ID** | int |
|---|---|
| **name** | varchar (35) |

FK1biobrick_parts

## biobrick_parts

| 🔑 **ID** | int unsigned |
|---|---|
| **bba_number** | varchar (15) |
| **dna_sequence** | varchar (48000) |
| **part_type** | int |
| description | varchar (255) |
| **rec_date** | date |

FK2biobrick_parts

## nanodrop

| 🔑 **cloneID** | int |
|---|---|
| concentration | decimal (8,2) |
| A260 | decimal (8,3) |
| A280 | decimal (8,3) |
| A260_A280 | decimal (8,2) |
| A260_A230 | decimal (8,2) |
| constant | decimal (2) |
| cursor_pos | decimal (3) |
| cursor_abs | decimal (5,3) |
| 340_raw | decimal (4,3) |
| **rec_date** | date |

Fk1nanodrop

## dna_repository

| 🔑 **cloneID** | int |
|---|---|
| bba_number | varchar (15) |
| plasmid | int |
| physical_sequence | varchar (48000) |
| **selected_for_sequencing** | enum (2) |
| **rec_date** | date |

FK1dna_repository

FK2dna_repository

FK1peaks

## plate_key

| 🔑 **cloneID** | int |
|---|---|
| **384_plate_name** | varchar (10) |
| **384_plate_well** | varchar (10) |
| **96_plate_name** | varchar (10) |
| **96_plate_well** | varchar (10) |

## plasmid

| 🔑 **ID** | int |
|---|---|
| **name** | varchar (35) |

## peaks

| 🔑 **ID** | int unsigned |
|---|---|
| **cloneID** | int |
| **size** | int |
| conc | double (22) |
| molarity | double (22) |
| observations | double (22) |
| area | double (22) |
| aligned_migration_time | double (22) |
| peak_height | double (22) |
| peak_width | double (22) |
| percent_total | double (22) |
| time_corrected_area | double (22) |
| **rec_date** | date |