**Model Organism Taxonomy and Evolution Database Project**

Hood College, IT/ BIFX 530-02, Fall 2022

Nicholus Vaughan, Emily Whisenhunt,

Michael Cook, J. Jedediah Smith

**Table of Contents**

[**Description** 3](#_Toc120776005)

[**Scope** 3](#_Toc120776006)

[**Requirements Analysis** 4](#_Toc120776007)

[**ER Diagram** 5](#_Toc120776008)

[**Assumptions** 5](#_Toc120776009)

[**Relational Schema** 6](#_Toc120776010)

[**SQL Create Tables** 7](#_Toc120776011)

[**SQL Load Data** 9](#_Toc120776012)

[**Website Diagram** 11](#_Toc120776013)

[**Page Functionality** 11](#_Toc120776014)

[**Website Link** 11](#_Toc120776015)

# **Description**

Welcome to our TED Talk. Biology 201 (BIOL 201) is Hood College’s Evolution and Ecology course for undergraduates. This course introduces students to the fundamentals of evolution and naturally, the branching (or cladistics) of species. The purpose of this project is to create an interactive relational database to educate students on taxonomy (the study of organism classification) and phylogeny (the study of evolutionary relationships between organisms). Each entity will be described by its taxonomic classification and have biologically relevant features as attributes. The hope is that students in this course will find a database of searchable animal species helpful in understanding the development of organ systems over the course of evolutionary time and how this relates to interspecies relatedness.

# **Scope**

This iteration of the Taxonomic Database will initially include 12 academically relevant animal species (i.e., common model organisms for biological/genetic studies) such as Human, Xenopus, Mouse, Rat, C. elegans (a nematode), and Drosophila (fruit fly). The database will detail attributes of features, organ systems, and groups closely related organisms based on a theoretical most recent common ancestor, or MRCA. Students will be able to use this database to search for species and their features for the purpose of cladistic and phylogenetic comparisons. Eventually, the idea of this database is to include more animal species and possibly genomic and transcriptomic data for these species in the future.

The user interface  will allow simple searches that return species based on taxonomic classifications (i.e., Phylum, Class, Order, etc.) and advanced searches  based on attributes of interest (e.g., the number of heart chambers, the mode of oxygen intake). Additionally, each species can be compared to one another to find how long ago (in millions of years) their MRCA likely lived. The goal is to present students with tangible evidence of the interrelation of seemingly unrelated distinct animal species, as well as learn about the basics of species classification and the branching nature of evolution and biodiversity.

# **Requirements Analysis**

Each account has a unique student ID and general information about the user. They must be associated with an institution, as this relationship is what will grant access to the database.

Each institution has a unique institute ID to distinguish places with similar names. They must purchase a set number of licenses for a set period to grant their students access to the database. Should the license expire, their student accounts will be deleted. Should their licenses be used up, new student accounts cannot be associated with that institution.

Saved searches are searches an account has made that they wish to revisit in the future. This just saves the search ID in an easy to access place. Account may have numerous saved searches.

Each search has a unique search ID and a date so it can be easily referenced. The key words are what will be used to find various species for the search results display.

Each species has a unique scientific name (e.g., Mus musculus). The other attributes for species include ploidy and reproductive rate for the model organism. These are important attributes that help establish their value as model organisms.

Each MRCA (most recent common ancestor) must be associated with one species and vice versa. The MRCA entity contains the name of the ancestor and the estimated time (in MYA, millions of years ago) that the ancestor lived.

# **ER Diagram**

Diagram

Description automatically generated

# **Assumptions**

1. Each Account must be associated with one Institution. Institutions may have multiple accounts.
2. Each Search is made by one account. One account may make zero, one, or many searches.
3. Each Search may find multiple Species. Each Species may be found by multiple searches.
4. Saved\_Searches must be associated with one Account. One Account may be associated with multiple saved searches.
5. Each Species has one MRCA. Each MRCA is associated with one Species.

# **Relational Schema**

Species(Scientific\_Name: VARCHAR (50), Common\_Name: VARCHAR (30), Vertebrate: BIT, Region: VARCHAR (25), Reproduction: VARCHAR(30), Ploidy: INTEGER, MRCA\_Name: VARCHAR(30), MYA: INTEGER)

The Species table stores biological data about each of the model organisms in the database. This includes the scientific name, common name, where it can be found, mode of reproduction, and ploidy (the number of chromosomes). The Vertebrate/Invertebrate attribute is coded as a bit field, with 0 = Invertebrate and 1 = Vertebrate. The Scientific Name serves as the primary key for Species. The MRCA (Most Recent Common Ancestor) table stores information about a Species’ MRCA with humans (Homo sapiens). The primary key for MRCA is MRCA\_name. The MRCA table also contains how long ago this ancestor lived in millions of years as MYA.

Institution(Institute\_ID: VARCHAR(9), Name: VARCHAR (50), Expiration: DATE, Licence\_Amt: INTEGER)

The Institution table stores data for the universities that have purchased database licenses. Each Institution has a unique Institute\_ID as a primary key, the name of the university, the license expiration, and number of accounts available under that license.

Account(Student\_ID: VARCHAR(9), Name: VARCHAR(30), Major: VARCHAR(20), Class: VARCHAR(20), Institute\_ID: VARCHAR(9))

The Account table stores information pertaining to its student user. This includes Student ID, their major, class, and Institute ID.

Search(Search\_ID: VARCHAR (9), Search\_Date: DATE, Key\_Words: VARCHAR (75), Student\_ID: VARCHAR (9))

The Search table stores all the searches made by all accounts. This includes IDs to connect searches to accounts, the date of the search, and any keywords. They removed at regular intervals, much like NCBI database searches. However, if one wishes to preserve a specific search and return to it later, they can save it to their account.

Find(Search\_ID: VARCHAR(9), Scientific\_Name: VARCHAR (50))

The Find table contains the unique Search\_ID associated with the generated relevant Scientific\_Name both as primary keys.

Saved(Student\_ID: VARCHAR(9), Search\_ID: VARCHAR(9))

The Saved table stores searches that an account wishes to preserve from deletion and return to later. This just includes IDs to connect accounts and searches.

# **SQL Create Tables**

USE xxx##db;

DROP TABLE IF EXISTS saved;

DROP TABLE IF EXISTS find;

DROP TABLE IF EXISTS search;

DROP TABLE IF EXISTS account;

DROP TABLE IF EXISTS institution;

DROP TABLE IF EXISTS species;

CREATE TABLE species (

Scientific\_Name VARCHAR (50),

Common\_Name VARCHAR (30),

Vertebrate INTEGER,

Region VARCHAR (25),

Reproduction VARCHAR(30),

Ploidy INTEGER,

MRCA\_Name VARCHAR (30),

MYA INTEGER,

PRIMARY KEY (Scientific\_Name)

);

CREATE TABLE institution (

Institute\_ID VARCHAR(9) NOT NULL,

Name VARCHAR(30),

Expiration DATE,

License\_Amt INTEGER,

PRIMARY KEY (Institute\_ID)

);

/\* When an institution ceases to pay for its license or is deleted, we do not want user data to be deleted. Accounts should be able to update though if their intuition changes. Hence ON CASCADE is only for update, not delete. \*/

CREATE TABLE account(

Student\_ID VARCHAR(9),

Name VARCHAR(30),

Major VARCHAR(20),

Class VARCHAR(20),

Institute\_ID VARCHAR(9) NOT NULL,

PRIMARY KEY (Student\_ID),

FOREIGN KEY (Institute\_ID) REFERENCES institution(Institute\_ID)

ON UPDATE CASCADE

);

/\* We set Student\_ID to NOT NULL because every search must be made by an account, which in turn will have a student ID.\*/

CREATE TABLE search(

Search\_ID VARCHAR (9),

Search\_Date DATE,

Key\_Words VARCHAR (75),

Student\_ID VARCHAR (9) NOT NULL,

PRIMARY KEY (Search\_ID),

FOREIGN KEY (Student\_ID) REFERENCES account(Student\_ID)

ON DELETE CASCADE

)ENGINE=InnoDB;

/\* Scientific\_Name is NOT NULL because all Species will have the MRCA information. We also want entries to match the behavior of its associated species. There is no point in preserving this information if the species is deleted and it should be kept up to date, hence ON CASCADE. \*/

CREATE TABLE mrca (

MRCA\_Name VARCHAR (30),

MYA INTEGER,

Scientific\_Name VARCHAR(50) NOT NULL,

PRIMARY KEY (MRCA\_Name),

FOREIGN KEY (Scientific\_Name) REFERENCES species(Scientific\_Name)

ON DELETE CASCADE

ON UPDATE CASCADE

)ENGINE=InnoDB;

/\* When a search or species is deleted or updated, whatever results were produced should be affected the same way. Results will be misleading without the updates and will never be displayed if not associated with a search. Hence the ON CASCADE constraint. \*/

CREATE TABLE find(

Search\_ID VARCHAR(9),

Scientific\_Name VARCHAR (50),

PRIMARY KEY(Scientific\_Name, Search\_ID),

FOREIGN KEY (Scientific\_Name) REFERENCES species(Scientific\_Name)

ON DELETE CASCADE

ON UPDATE CASCADE,

FOREIGN KEY (Search\_ID) REFERENCES search(Search\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)ENGINE=InnoDB;

/\* When an account or search is deleted or updated, we want the saved searches to be affected the same way. No use in saved searches if they are no longer associated with an account or their account ID is not kept up to date. Hence the ON CASCADE constraint. \*/

CREATE TABLE saved(

Student\_ID VARCHAR(9) NOT NULL,

Search\_ID VARCHAR(9),

PRIMARY KEY (Student\_ID, Search\_ID),

FOREIGN KEY (Student\_ID) REFERENCES account(Student\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE,

)ENGINE=InnoDB;

# **SQL Load Data**

INSERT INTO species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy, MRCA\_NAME, MYA) VALUES

('Xenopus tropicalis', 'African Clawed Frog', 1, 'Africa', 'Ovoviviparous', 20, 'Tetrapoda', 390),

('Danio rerio', 'Zebrafish', 1, 'Asia', 'Oviparous', 50, 'Gnathostomata', 419),

('Drosophila melanogaster', 'Fruit Fly', 0, 'Asia', 'Oviparous', 4, 'Urbilateria', 580),

('Hydra vulgaris', 'Freshwater Polyp', 0, 'S. America', 'Budding', 32, 'Eumetazoa', 630),

('Caenorhabditis elegans', 'C. elegans', 0, 'N. America', 'Autogamy', 12, 'Nephrozoa', 558),

('Macaca mulatta', 'Rhesus Macacque', 1, 'Asia', 'Viviparous', 42, 'Catarrhini', 25),

('Mus musculus', 'Common Mouse', 1, 'Europe', 'Viviparous', 40, 'Euarchontoglires', 90),

('Schmidtea mediterranea', 'Planaria Worm', 0, 'Europe', 'Binary Fission', 8, 'Xenacoelomorpha', 526),

('Arbacia punctulata', 'Purple-Spined Sea Urchin', 0, 'Pacific', 'Allogamy', 44, 'Ambulacraria', 533),

('Euprymna scolopes', 'Hawaiian Bobtail Squid', 0, 'Pacific', 'Iteroparous', 46, 'Lophotrochozoa', 600);

INSERT INTO institution(Institute\_ID, Name, Expiration, License\_Amt) VALUES

('000000001', 'Hood College', '2026-01-01', 200),

('000000002', 'Shepherd University', '2024-05-22', 500),

('000000003', 'Montgomery College', '2025-03-08', 300),

('000000004', 'Shippensburg University', '2023-11-07', 1000),

('000000005', 'Loma Linda University', '2029-08-01', 100),

('000000006', 'University of Maryland', '2023-03-08', 700),

('000000007', 'University of Virginia', '2028-01-09', 8000),

('000000008', 'University of California', '2029-09-09', 9999),

('000000009', 'Andrews University', '2024-05-02', 802),

('000000010', 'Blueridge Community College', '2099-11-17', 1999);

INSERT INTO account(Student\_ID, Name, Major, Class, Institute\_ID) VALUES

('123456789','John Doe','Biology','BIOL-156','000000001'),

('987654321','Jane Snow','Chemistry','CHEM-262','000000002'),

('789456123','Richard Roe','Biochemistry','CHEM-442','000000003'),

('789123456','Jenny Crow','Biochemistry','BIOL-445','000000004'),

('456123789','Reginald Loe','Ecology','BIOL-243','000000005'),

('123123123','Bob Ross','Art','ARTL-123','000000006'),

('789789789','Rowan Atkinson','Biology','BIOL-333','000000007'),

('456456456','David Attenborough','Ecology','BIOL-212','000000008'),

('295739234','Billy Boe','Chemistry','CHEM-121','000000009'),

('129549849','Major Major','Biochemistry','CHEM-455','000000010');

INSERT INTO search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) VALUES

('abcde1234', '2022-01-01', 'Mouse', '123456789'),

('qwert0009', '2022-02-02', 'Urchin', '789123456'),

('poikl9876', '2022-03-03', 'Fish', '789123456'),

('fooba5555', '2022-04-04', 'Elegans', '456123789'),

('ytryx7539', '2022-05-05', 'Worm', '987654321'),

('tmrwl6682', '2022-06-06', 'Fly', '123456789'),

('rotfl9999', '2022-07/07', 'Squid', '789123456'),

('fghij5678', '2022-08-08', 'Polyp', '789123456'),

('sqqab9876', '2022-09-09', 'Frog', '456123789'),

('edcab4123', '2022-10-10', 'Macaca', '987654321');

INSERT INTO find(Search\_ID, Scientific\_Name) VALUES

('abcde1234', 'Mus musculus'),

('qwert0009', 'Arbacia punctulata'),

('poikl9876', 'Danio rerio'),

('fooba5555', 'Caenorhabditis elegans'),

('ytryx7539', 'Schmidtea mediterranea'),

('tmrwl6682', 'Drosophila melanogaster'),

('rotfl9999', 'Euprymna scolopes'),

('fghij5678', 'Hydra vulgaris'),

('sqqab9876', 'Xenopus tropicalis'),

('edcab4123', 'Macaca mulatta');

INSERT INTO saved(Student\_ID, Search\_ID) VALUES

('123456789', 'abcde1234'),

('789123456', 'qwert0009'),

('789123456', 'poikl9876'),

('456123789', 'fooba5555'),

('987654321', 'ytryx7539'),

('789456123', 'tmrwl6682'),

('987654321', 'rotfl9999'),

('789456123', 'fghij5678'),

('123456789', 'sqqab9876'),

('456123789', 'edcab4123');

# **Website Diagram**

Diagram

Description automatically generated 

# **Page Functionality**

Home: Main landing page that has links to all the other features of our admin view. Each of the functions involves interactions with the user accounts.

Browse Accounts: This is a basic display page that shows all the different user accounts registered with our model species database.

Insert Account: This interface allows the insertion of new accounts to the account table. Afterwards, it displays an updated list of accounts.

Remove Account: This interface allows the removal of accounts from the accounts table. Afterwards, it displays an updated list of accounts.

Update Account: This interface allows the modification of accounts in the account table. Afterwards, it displays an updated list of accounts.

# **Website Link**

http://pluto.hood.edu/~jjs14/Group\_Project/main.html