**Model Organism Taxonomy and Evolution Database Project**

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

Nicholus Vaughan, Emily Whisenhunt,

Michael Cook, J. Jedediah Smith

Updated 11/16/2022

**Data 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.

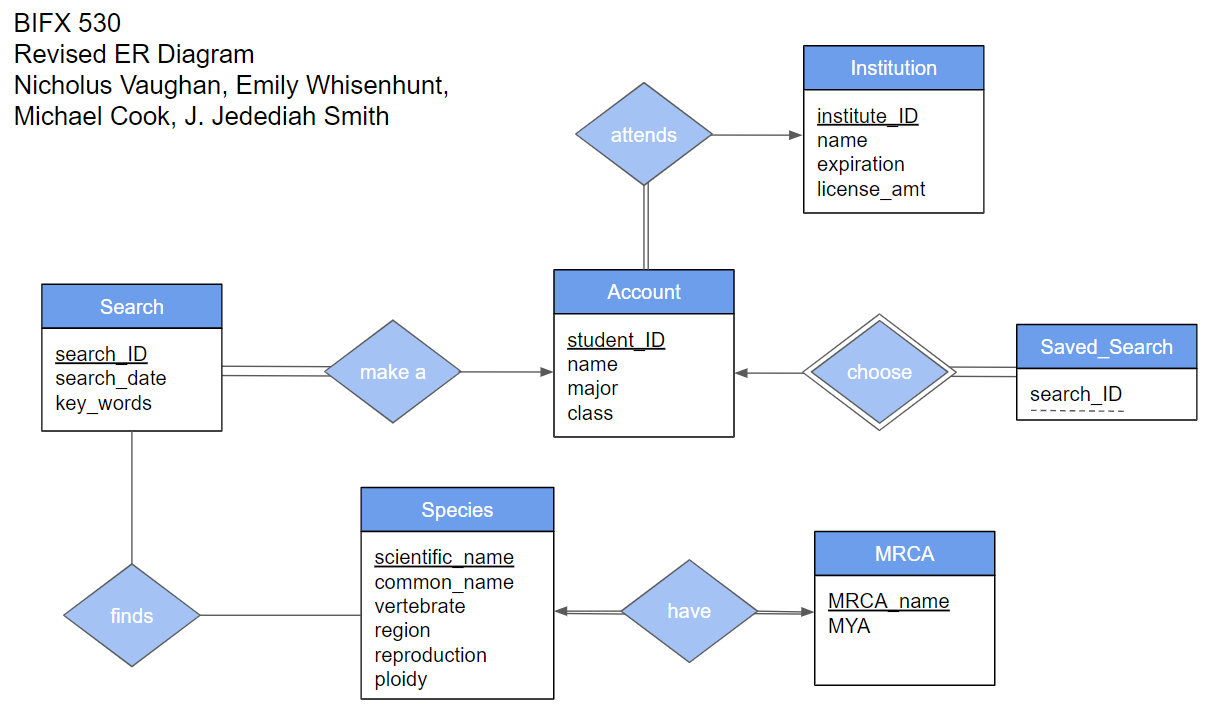
Saved searches are searches an account has made that they wish to revisit in the future. This just saves the search ID in an easy to access place. Account may have numerous saved searches.

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.

**Revised ER Diagram**



**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)

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.

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.

MRCA(MRCA\_Name: VARCHAR (30), MYA: INTEGER, Scientific\_Name: VARCHAR (50))

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.

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 Data Tables**

use XXX;

DROP TABLE IF EXISTS search;

DROP TABLE IF EXISTS species;

DROP TABLE IF EXISTS account;

DROP TABLE IF EXISTS institution;

DROP TABLE IF EXISTS mrca;

CREATE TABLE species (

Scientific\_Name VARCHAR (50),

Common\_Name VARCHAR (30),

Vertebrate BIT,

Region VARCHAR (25),

Reproduction VARCHAR(30),

Ploidy INTEGER,

PRIMARY KEY (Scientific\_Name)

)ENGINE=InnoDB;

CREATE TABLE institution (

Institute\_ID VARCHAR(9),

Name VARCHAR(30),

Expiration DATE,

License\_Amt INTEGER,

PRIMARY KEY (Institute\_ID)

)ENGINE=InnoDB;

/\* 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

)ENGINE=InnoDB;

/\* 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),

Search\_ID VARCHAR(9),

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

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

ON DELETE CASCADE

ON UPDATE CASCADE,

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

ON DELETE CASCADE

ON UPDATE CASCADE

)ENGINE=InnoDB;

**SQL Load Data**

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000001’, 'Hood College’, ‘01/01/2026’, 200);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000002’, ‘Shepherd University’, ‘05/22/2024’, 500);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000003’, ‘Montgomery College’, ‘03/08/2025’, 300);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000004’, ‘Shippensburg University’, ‘11/07/2023’, 1000);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000005’, ‘Loma Linda University’, ‘08/01/2029’, 100);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000006’, ‘University of Maryland’, ‘03/08/2023’, 700);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000007’, ‘University of Virginia’, ‘01/09/2028’, 8000);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000008’, ‘University of California’, ‘09/09/2029’, 9999);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000009’, ‘Andrew’s University’, ‘05/02/2024’, 802);

insert into institution(Institute\_ID, Name, Expiration, License\_Amt) values (‘000000010’, ‘Blueridge Community College’, ‘11/17/2099’, 1999);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Xenopus tropicalis’, ‘African Clawed Frog’, 1, ‘Africa’, ‘Ovoviviparous’, 20);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Danio rerio’, ‘Zebrafish’, 1, ‘Asia’, ‘Oviparous’, 50);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Drosophila melanogaster’, ‘Fruit Fly’, 0, ‘Asia’, ‘Oviparous’, 4);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Hydra vulgaris’, ‘Freshwater Polyp’, 0, ‘S. America’, ‘Budding’, 32);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Caenorhabditis elegans’, ‘C. elegans’, 0, ‘N. America’, ‘Autogamy’, 12);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Macaca mulatta’, ‘Rhesus Macacque’, 1, ‘Asia’, ‘Viviparous’, 42);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Mus musculus’, ‘Common Mouse’, 1, ‘Europe’, ‘Viviparous’, 40);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Schmidtea mediterranea’, ‘Planaria Worm’, 0, ‘Europe’, ‘Binary Fission’, 8);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Arbacia punctulata’, ‘Purple-Spined Sea Urchin’, 0, ‘Pacific’, ‘Allogamy’, 44);

insert into species(Scientific\_Name, Common\_Name, Vertebrate, Region, Reproduction, Ploidy) values (‘Euprymna scolopes’, ‘Hawaiian Bobtail Squid’, 0, ‘Pacific’, ‘Iteroparous’, 46);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values ('Eumetazoa', 630, ‘Hydra vulgaris’);  
insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Gnathostomata’, 419, ‘Danio Rerio’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Urbilateria’, 580, ‘Drosophila melanogaster’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Tetrapoda’, 390, ‘Xenopus tropicalis’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Nephrozoa’, 558, ‘Caenorhabditis elegans’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Catarrhini’, 25, ‘Macaca mulatta’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Euarchontoglires’, 90, ‘Mus musculus’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Xenacoelomorpha’, 526, ‘Schmidtea mediterranea’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Ambulacraria’, 533, ‘Arbacia punctulata’);

insert into mrca(MRCA\_Name, MYA, Scientific\_Name) values (‘Lophotrochozoa’, 600, ‘Euprymna scolopes’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘abcde1234’, ‘01/01/2022’, ‘Mouse’, ‘123456789’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘qwert0009’, ‘02/02/2022’, ‘Urchin’, ‘789123456’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘poikl9876’, ‘03/03/2022’, ‘Fish’, ‘789123456’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘fooba5555’, ‘04/04/2022’, ‘Elegans’, ‘456123789’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘ytryx7539’, ‘05/05/2022’, ‘Worm’, ‘987654321’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘tmrwl6682’, ‘06/06/2022’, ‘Fly’, ‘123456789’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘rotfl9999’, ‘07/07/2022’, ‘Squid’, ‘789123456’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘fghij5678’, ‘08/08/2022’, ‘Polyp’, ‘789123456’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘sqqab9876’, ‘09/09/2022’, ‘Frog’, ‘456123789’);

insert into search(Search\_ID, Search\_Date, Key\_Words, Student\_ID) values (‘edcab4123’, ‘10/10/2022’, ‘Macaca, ‘987654321’);

insert into saved(Student\_ID, Search\_ID) values (‘123456789’, ‘abcde1234’);

insert into saved(Student\_ID, Search\_ID) values (‘789123456’, ‘qwert0009’);

insert into saved(Student\_ID, Search\_ID) values (‘789123456’, ‘poikl9876’);

insert into saved(Student\_ID, Search\_ID) values (‘456123789’, ‘fooba5555’);

insert into saved(Student\_ID, Search\_ID) values ('987654321’, ‘ytryx7539’);

insert into saved(Student\_ID, Search\_ID) values (‘789456123’, 'tmrwl6682’);

insert into saved(Student\_ID, Search\_ID) values (‘987654321’, ‘rotfl9999’);

insert into saved(Student\_ID, Search\_ID) values (‘789456123’, ‘fghij5678’);

insert into saved(Student\_ID, Search\_ID) values (‘123456789’, ‘sqqab9876’);

insert into saved(Student\_ID, Search\_ID) values (‘456123789’, ‘edcab4123’);

insert into find(Search\_ID, Scientific\_Name) values (‘abcde1234’, ‘Mus musculus’);

insert into find(Search\_ID, Scientific\_Name) values (‘qwert0009’, ‘Arbacia punctulata’);

insert into find(Search\_ID, Scientific\_Name) values (‘poikl9876’, ‘Danio rerio’);

insert into find(Search\_ID, Scientific\_Name) values (‘fooba5555’, ‘Caenorhabditis elegans’);

insert into find(Search\_ID, Scientific\_Name) values (‘ytryx7539’, ‘Schmidtea mediterranea’);

insert into find(Search\_ID, Scientific\_Name) values (‘tmrwl6682, ‘Drosophila melanogaster’);

insert into find(Search\_ID, Scientific\_Name) values (‘rotfl9999’, ‘Euprymna scolopes’);

insert into find(Search\_ID, Scientific\_Name) values (‘fghij5678’, ‘Hydra vulgaris’);

insert into find(Search\_ID, Scientific\_Name) values (‘sqqab9876’, ‘Xenopus tropicalis’);

insert into find(Search\_ID, Scientific\_Name) values (‘edcab4123’, ‘Macaca mulatta’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘123456789’,’John Doe’,’ Biology’,’BIOL-156’,’000000001’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘987654321’,’Jane Snow’,’Chemistry’,’CHEM-262’,’000000002’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘789465123’,’Richard Roe’,’Biochemistry’,’CHEM-442’,’000000003’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘789123456’,’Jenny Crow’,’Biochemistry’,’BIOL-445’,’000000004’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘456123789’,’Reginald Loe’,’Ecology’,’BIOL-243’,’000000005’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘123123123’,’Bob Ross’,'Art’,’ARTL-123’,’000000006’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘789789789’,’Rowan Atkinson’,’Biology’,’BIOL-333’,’000000007’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘456456456’,’David Attenborough’,’Ecology’,’BIO-212’,’000000008’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘295739234’,’Billy Boe’,’Chemistry’,’CHEM-121’,’000000009’);

insert into account(Student\_ID, Name, Major, Class, Institute\_ID) values (‘129549849’,’Major Major’,’Biochemistry’,’CHEM-455’,’000000010’);