PROJECT INGA DATABASE

IS-6420

Student Names List
Date

Table of Contents

Business Description

The **Coley-Kursar Lab** is a research lab in the Department of Biology at the University of Utah in Salt Lake City, Utah.

The Coley-Kursar lab studies the evolution of anti-herbivore defenses in a Neotropical tree genus called Inga. The lab conducts visits to tropical forests to gather information about individual Inga trees. They record a variety of observations and collect a variety of insect and plant matter for later chemical and DNA analysis, which is then synthesized to draw various evolutionary, botanical, and ecological conclusions.

The lab is looking to move away from storing this data in Microsoft Excel workbooks and wishes to implement a relational database management system that will simplify how data is recorded and retrieved. This change will eliminate duplicate data entry and minimize the chance for error. The ability to quickly and accurately retrieve information from the database is critically important to the analysis the lab conducts and for the scholarly papers the lab ultimately publishes.

Glossary

Chemistry Collection: Plant matter (young leaf, mature leaf, flower, or flower bud) collected from a plant in the field for the purpose of later chemical extraction.

Extraction: A chemical extraction performed on the plant matter contained in one or many chemistry collections. Each extraction has a method number which identifies the array of chemical procedures to be performed on the extraction.

Field Event: An event marking each time an individual Inga plant is visited in the field and observations are recorded about that plant. Things recorded as observations related to a field event can change over time; for example, an individual Inga plant's height.

LC/MS Image: Stands for "Liquid Chromatography Mass Spectrometry Image," which is a reference to an image of the output from a Liquid Chromatography Mass Spectrometry procedure (LCMS Image is an attribute of the Result entity type).

Percent Expansion: The rate at which a young leaf is growing (Percent Expansion is an attribute of the Chemistry Collection entity type).

Plant DBH: Stands for "Plant Diameter at Breast Height," a common measure of plant trunk thickness (Plant_DBH is an attribute of the Field Event entity type).

Plant Number: An identifier used to designate individual Inga plants at a site, unique across that site, but not unique across all sites. For example, the site called "Tiputini" has plant numbers ranging from 1 - n, and the site called "Los Amigos" also has plant numbers ranging from 1 - n (where n is the total number of plants observed at a site).

Result: A result is an outcome of a chemical procedure. It consists of the name of the chemical class that was isolated by the chemical procedure as well as the weight of the chemical class obtained.

Species Code: A code used as a tentative species designation for Inga plants observed at a site, of the format “C-#” (where “C” is a one-letter character abbreviation denoting the site where the species was observed and “#” is a positive integer). Because many of the observed Inga plants at a site may not have been formally identified yet, each individual Inga plant is given a species code that serves as a tentative designation that may be confirmed or changed through later DNA analysis. An example species code is: “T67”; the “T” indicates that the species was observed at a specific site (in this case, “Tiputini”) and “67” denotes that it was the 67th unique observed species at that site.

Data Requirements

The lab visits (and may re-visit) a number of sites in order to observe and collect physical samples from individual plants of the genus Inga. Each site has the following attributes: a site name (which is unique); the country in which the site is located (e.g., Peru, Brazil, etc.); latitude and longitude (in degrees and minutes) and altitude; average temperature; annual rainfall and rainfall seasonality (and references); soil type (and references); and notes. Sites may or may not have individual Inga plants associated with them, depending upon whether or not that site has been visited yet.

During a visit to a site, the research team records information about individual Inga plants (each plant must belong to a site, and a site has multiple plants). Each plant is given a number unique to that site (called the "plant number") and the trail address where that plant is located is recorded. General notes about the plant are also recorded.

Because many of the Inga plants observed on these trips may not have been formally identified, each individual Inga plant is given a tentative species code that serves as a tentative species designation that may be confirmed or changed through later DNA analysis. An example species code is: "T67"; the "T" indicates that that the species was observed at a specific site (in this case, "Tiputini") and "67" denotes that it was the 67th unique observed species at that site. Each time a new species is observed (i.e., one that exhibits a unique set of traits not previously observed at that site), the numeric portion of the species code is increased by one. So the very next Inga plant that exhibits a new suite of traits not previously observed would receive a species code of "T68."

The collectors may take notes about each species code. Additionally, if DNA analysis confirms that a given species code has been formally named, that name is then applied to all individual plants that were designated with that specific code. For instance, if DNA analysis reveals that species code T67 is the formally identified species "edulus," then all individual plants with species code T67 should relate to the name "edulus." Finally, if there is an authoritative reference on a particular species code, a link to that authoritative paper is recorded. Each plant has only one species; however, each species designation can be shared by many individual plants.

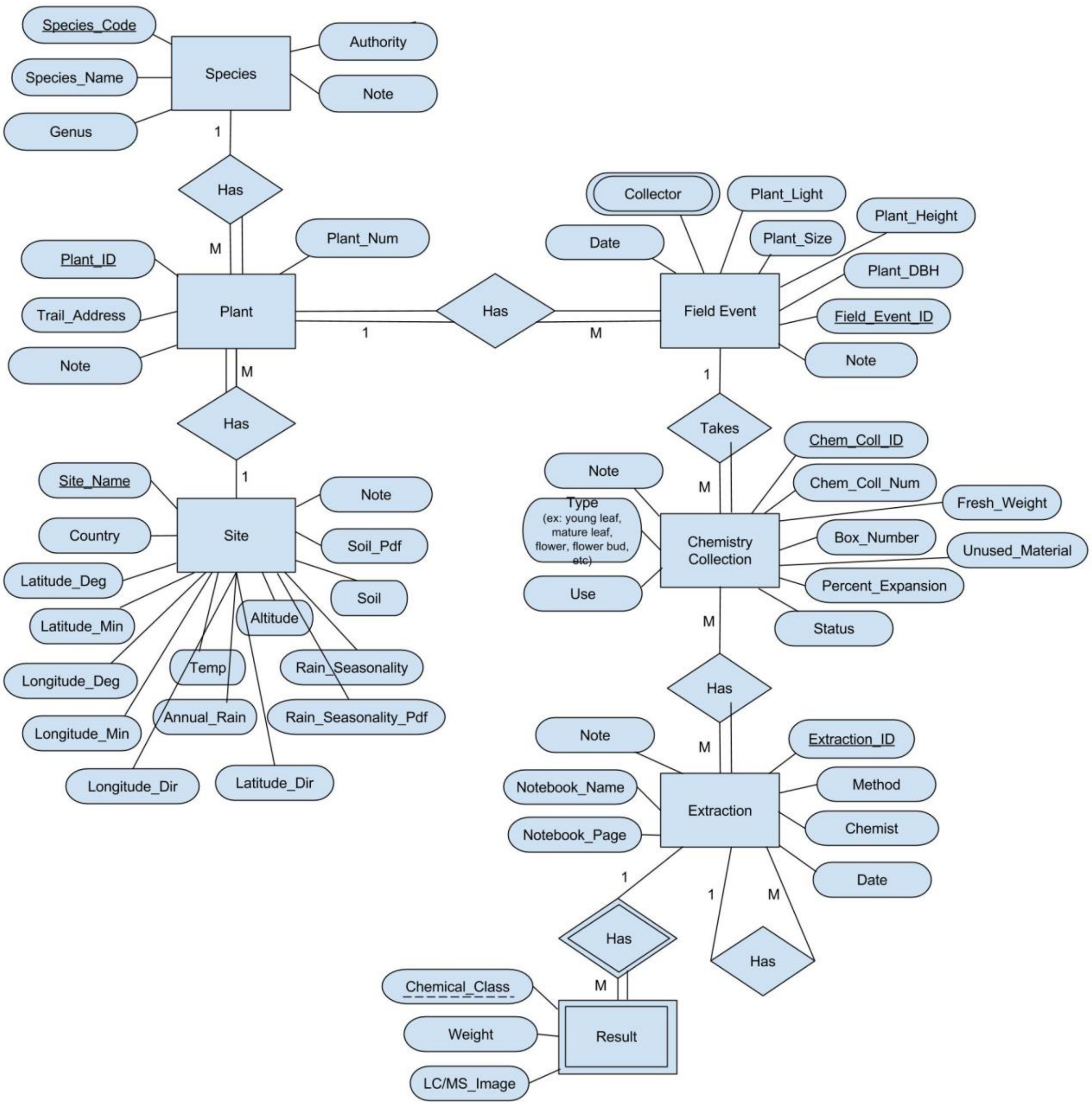
Each time an individual tree is visited, information about that field event is recorded. The field event is a snapshot in time; the information recorded about each Inga tree can change between visits. This field event information includes: the date of the event; the names of one or more collectors conducting the field event; how much light the plant receives (e.g., "sun," "shade," or "intermittent"); the size of the plant (e.g., "sapling," "sprout," or "adult"); the tree height; the tree diameter at breast-height (a common measurement in the field of botany); and any notes about the field event. Each field event relates to only one plant; but a single plant may have multiple field events over time.

During a field event, plant matter from the tree is collected for later chemical analysis. A field event can give rise to many such “chemistry collections” – for instance, a visit to a plant might involve collecting a young leaf, a mature leaf, a flower, and a flower bud (four separate chemistry collections). Each one of these collections is assigned a unique chemistry collection number, and the type of collection (“young leaf,” “mature leaf,” etc.) is recorded. In addition, each chemistry collection has the following attributes: % expansion (the rate that a young leaf is growing, if it is a young leaf), a use (what the collection will be used for), a fresh weight, an unused weight (updated as the material is used for extractions), and a status (e.g., “ground,” or “extracted”) that is updated as the collection is processed.

Chemistry collections are eventually ground for chemical extraction. Ground material from chemistry collections (one or many) is then chemically extracted according to a method number (which identifies the series of chemical procedures to be performed on the extract). For each extraction, the following information is also collected: a unique extraction number, the name of the chemist performing the extraction, the date of the extraction, the notebook name in which the extraction is recorded, and the notebook page on which the extraction is recorded. A chemistry collection may give rise to multiple extractions. Additionally, an extraction can itself be extracted to create a new “child” extraction on which additional experiments are performed.

Finally, the chemical procedures performed on each extract give rise to one or many “results” – expressed as weights of specific chemical classes. Additionally, an image of the liquid chromatography mass spectrometry (LC/MS) output can also be stored with the result. Each result arises from only one extraction.

ERD


Species_Code
Species_Name
Genus
Species
Authority
Note
1
Has
M
Plant_ID
Trail_Address
Note
Plant
Plant_Num
Has
1
M
M
Has
1
Site_Name
Country
Latitude_Deg
Latitude_Min
Longitude_Deg
Longitude_Min
Longitude_Dir
Site
Note
Soil_Pdf
Soil
Altitude
Temp
Annual_Rain
Rain_Seasonality
Rain_Seasonality_Pdf
Latitude_Dir
Collector
Plant_Light
Plant_Height
Date
Plant_Size
Plant_DBH
Field Event
Field_Event_ID
Note
1
Takes
M
Note
Type
(ex: young leaf, mature leaf, flower, flower bud, etc)
Use
Chemistry Collection
Chem_Coll_ID
Chem_Coll_Num
Fresh_Weight
Box_Number
Unused_Material
Percent_Expansion
Status
M
Has
M
Note
Notebook_Name
Notebook_Page
Extraction
Extraction_ID
Method
Chemist
Date
1
1
M
Has
Has
M
Chemical_Class
Weight
LC/MS_Image
Result

Relational Model

Plant (Plant_ID, Plant_Num, Species_Code, Site_Name, Trail_Address, Note)
Foreign Key (Species_Code) References Species (Species_Code)
Foreign Key (Site_Name) References Site (Site_Name)

Species (Species_Code, Species_Name, Genus, Authority, Note)

Site (Site_Name, Country, Latitude_Deg, Latitude_Dir, Latitude_Min, Longitude_Deg, Longitude_Dil, Longitude_Min, Temp, Altitude, Annual_Rain, Rain_Seasonality, Rain_Seasonality_Pdf, Soil, Soil_Pdf, Note)

Field_Event (Field_Event_ID, Plant_ID, Date, Plant_Light, Plant_Size, Plant_Height, Plant_DBH, Note)
Foreign Key (Plant_ID) References Plant (Plant_ID)

Collector (Field_Event_ID, Collector_Name)
Foreign Key (Field_Event_ID) References Field_Event (Field_Event_ID)

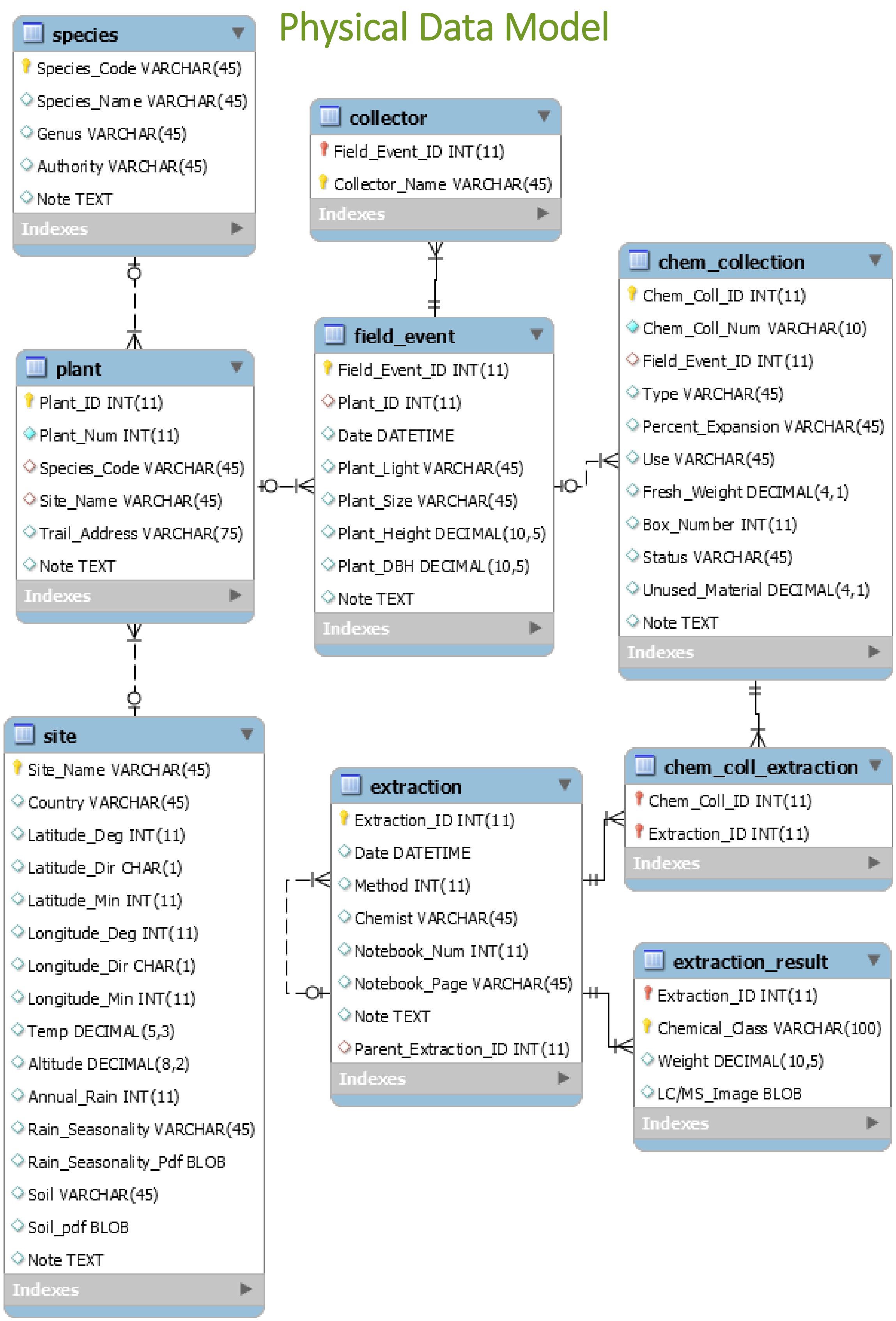
Chem_Collection (Chem_Coll_ID, Chem_Coll_Num, Field_Event_ID, Type, Percent_Expansion, Use, Fresh_Weight, Box_Number, Status, Unused_Material, Note)
Foreign Key (Field_Event_ID) References Field_Event (Field_Event_ID)

Extraction (Extraction_ID, Date, Method, Chemist, Notebook_Num, Notebook_Page, Note, Parent_Extraction_ID)
Foreign Key (Parent_Extraction_ID) References Extraction (Extraction_ID)

Extraction_Result (Extraction_ID, Chemical_Class, Weight, LC/MS_Image)
Foreign Key (Extraction_ID) References Extraction (Extraction_ID)

Chem_Coll_Extraction (Chem_Coll_ID, Extraction_ID)
Foreign Key (Chem_Coll_ID) References Chem_Collection (Chem_Coll_ID)
Foreign Key (Extraction_ID) References Extraction (Extraction_ID)

Notes:
Chem_Collection.Type = Young Leaf, Mature Leaf, Flower, Flower Bud, etc.
Chem_Collection.Use = Chem, Chem Bulk, Nitrogen, etc.
Chem_Collection.Status = Leaf, Ground, Extracted, Missing
Also see Inga Glossary for more details



Physical Data Model
species
Species_Code VARCHAR(45)
Species_Name VARCHAR(45)
Genus VARCHAR(45)
Authority VARCHAR(45)
Note TEXT
Indexes
collector
Field_Event_ID INT(11)
Collector_Name VARCHAR(45)
Indexes
chem_collection
Chem_Coll_ID INT(11)
Chem_Coll_Num VARCHAR(10)
Field_Event_ID INT(11)
Type VARCHAR(45)
Percent_Expansion VARCHAR(45)
Use VARCHAR(45)
Fresh_Weight DECIMAL(4,1)
Box_Number INT(11)
Status VARCHAR(45)
Unused_Material DECIMAL(4,1)
Note TEXT
Indexes
field_event
Field_Event_ID INT(11)
Plant_ID INT(11)
Date DATETIME
Plant_Light VARCHAR(45)
Plant_Size VARCHAR(45)
Plant_Height DECIMAL(10,5)
Plant_DBH DECIMAL(10,5)
Note TEXT
Indexes
plant
Plant_ID INT(11)
Plant_Num INT(11)
Species_Code VARCHAR(45)
Site_Name VARCHAR(45)
Trail_Address VARCHAR(75)
Note TEXT
Indexes
site
Site_Name VARCHAR(45)
Country VARCHAR(45)
Latitude_Deg INT(11)
Latitude_Dir CHAR(1)
Latitude_Min INT(11)
Longitude_Deg INT(11)
Longitude_Dir CHAR(1)
Longitude_Min INT(11)
Temp DECIMAL(5,3)
Altitude DECIMAL(8,2)
Annual_Rain INT(11)
Rain_Seasonality VARCHAR(45)
Rain_Seasonality_Pdf BLOB
Soil VARCHAR(45)
Soil_pdf BLOB
Note TEXT
Indexes
extraction
Extraction_ID INT(11)
Date DATETIME
Method INT(11)
Chemist VARCHAR(45)
Notebook_Num INT(11)
Notebook_Page VARCHAR(45)
Note TEXT
Parent_Extraction_ID INT(11)
Indexes
chem_coll_extraction
Chem_Coll_ID INT(11)
Extraction_ID INT(11)
Indexes
extraction_result
Extraction_ID INT(11)
Chemical_Class VARCHAR(100)
Weight DECIMAL(10,5)
LC/MS_Image BLOB
Indexes

MySQL Syntax

```
CREATE TABLE `ingadb`.`site` (
 `Site_Name` VARCHAR(45) NOT NULL,
 `Country` VARCHAR(45) NULL,
 `Latitude_Deg` INT NULL,
 `Latitude_Dir` CHAR(1) NULL,
 `Latitude_Min` INT NULL,
 `Longitude_Deg` INT NULL,
 `Longitude_Dir` CHAR(1) NULL,
 `Longitude_Min` INT NULL,
 `Temp` DECIMAL(5,3) NULL,
 `Altitude` DECIMAL(8,2) NULL,
 `Annual_Rain` INT NULL,
 `Rain_Seasonality` VARCHAR(45) NULL,
 `Rain_Seasonality_Pdf` BLOB NULL,
 `Soil` VARCHAR(45) NULL,
 `Soil_pdf` BLOB NULL,
 `Note` TEXT NULL,
 CONSTRAINT pk_site PRIMARY KEY (`Site_Name`));

CREATE TABLE `ingadb`.`species` (
 `Species_Code` VARCHAR(45) NOT NULL,
 `Species_Name` VARCHAR(45) NULL,
 `Genus` VARCHAR(45) NULL,
 `Authority` VARCHAR(45) NULL,
 `Note` TEXT NULL,
 CONSTRAINT pk_species PRIMARY KEY (`Species_Code`));

CREATE TABLE `ingadb`.`plant` (
 `Plant_ID` INT NOT NULL,
 `Plant_Num` INT NOT NULL,
 `Species_Code` VARCHAR(45) NULL,
 `Site_Name` VARCHAR(45) NULL,
 `Trail_Address` VARCHAR(75) NULL,
 `Note` TEXT NULL,
 CONSTRAINT pk_plant PRIMARY KEY (`Plant_ID`),
 INDEX `Species_Code_idx` (`Species_Code` ASC),
```

```
 INDEX `Site_Name_idx` (`Site_Name` ASC),
 CONSTRAINT fk1_plant FOREIGN KEY (`Species_Code`)
	REFERENCES `ingadb`.`species` (`Species_Code`)
	ON DELETE NO ACTION
	ON UPDATE NO ACTION,
 CONSTRAINT fk2_plant FOREIGN KEY (`Site_Name`)
	REFERENCES `ingadb`.`site` (`Site_Name`)
	ON DELETE NO ACTION
	ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`field_event` (
 `Field_Event_ID` INT NOT NULL,
 `Plant_ID` INT NULL,
 `Date` DATETIME NULL,
 `Plant_Light` VARCHAR(45) NULL,
 `Plant_Size` VARCHAR(45) NULL,
 `Plant_Height` DECIMAL(10,5) NULL,
 `Plant_DBH` DECIMAL(10,5) NULL,
 `Note` TEXT NULL,
 CONSTRAINT pk_field_event PRIMARY KEY (`Field_Event_ID`),
 INDEX `Plant_ID_idx` (`Plant_ID` ASC),
 CONSTRAINT fk_field_event FOREIGN KEY (`Plant_ID`)
	REFERENCES `ingadb`.`plant` (`Plant_ID`)
	ON DELETE NO ACTION
	ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`collector` (
 `Field_Event_ID` INT(11) NOT NULL,
 `Collector_Name` VARCHAR(45) NOT NULL,
 CONSTRAINT pk_collector PRIMARY KEY (`Field_Event_ID`, `Collector_Name`),
 CONSTRAINT fk_collector FOREIGN KEY (`Field_Event_ID`)
	REFERENCES `ingadb`.`field_event` (`Field_Event_ID`)
	ON DELETE NO ACTION
	ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`chem_collection` (
 `Chem_Coll_ID` INT NOT NULL,
 `Chem_Coll_Num` VARCHAR(10) NOT NULL,
```

```
 `Field_Event_ID` INT(11) NULL,
 `Type` VARCHAR(45) NULL,
 `Percent_Expansion` VARCHAR(45) NULL,
 `Use` VARCHAR(45) NULL,
 `Fresh_Weight` DECIMAL(4,1) NULL,
 `Box_Number` INT NULL,
 `Status` VARCHAR(45) NULL,
 `Unused_Material` DECIMAL(4,1) NULL,
 `Note`TEXT NULL,
 CONSTRAINT pk_chem_collection PRIMARY KEY (`Chem_Coll_ID`),
 INDEX `Field_Event_ID_idx` (`Field_Event_ID` ASC),
 CONSTRAINT fk_chem_collection     FOREIGN KEY (`Field_Event_ID`)
         REFERENCES `ingadb`.`field_event` (`Field_Event_ID`)
         ON DELETE NO ACTION
         ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`extraction` (
 `Extraction_ID` INT NOT NULL,
 `Date` DATETIME NULL,
 `Method` INT NULL,
 `Chemist` VARCHAR(45) NULL,
 `Notebook_Num` INT(11) NULL,
 `Notebook_Page` VARCHAR(45) NULL,
 `Note` TEXT NULL,
 `Parent_Extraction_ID` INT NULL,
 CONSTRAINT pk_extraction PRIMARY KEY (`Extraction_ID`),
 INDEX `Parent_Extraction_ID_idx` (`Parent_Extraction_ID` ASC),
 CONSTRAINT fk_extraction FOREIGN KEY (`Parent_Extraction_ID`)
         REFERENCES `ingadb`.`extraction` (`Extraction_ID`)
         ON DELETE NO ACTION
         ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`extraction_result` (
 `Extraction_ID` INT NOT NULL,
 `Chemical_Class` VARCHAR(100) NOT NULL,
 `Weight` DECIMAL(10,5) NULL,
 `LC/MS_Image` BLOB NULL,
 CONSTRAINT pk_extraction_result PRIMARY KEY (`Extraction_ID`, `Chemical_Class`),
```

```
 CONSTRAINT fk_extraction_result FOREIGN KEY (`Extraction_ID`)
        REFERENCES `ingadb`.`extraction` (`Extraction_ID`)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION);

CREATE TABLE `ingadb`.`chem_coll_extraction` (
 `Chem_Coll_ID` INT NOT NULL,
 `Extraction_ID` INT NOT NULL,
 CONSTRAINT pk_chem_coll_extraction PRIMARY KEY (`Chem_Coll_ID`, `Extraction_ID`),
 INDEX `Extraction_ID_idx` (`Extraction_ID` ASC),
 CONSTRAINT fk1_chem_coll_extraction FOREIGN KEY (`Chem_Coll_ID`)
        REFERENCES `ingadb`.`chem_collection` (`Chem_Coll_ID`)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
 CONSTRAINT fk2_chem_coll_extraction FOREIGN KEY (`Extraction_ID`)
        REFERENCES `ingadb`.`extraction` (`Extraction_ID`)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION);
```

Database View and Top Five Queries

Extraction Results View

create or replace view extraction_results_view as select site_name as ‘Site name’, plant.plant_num as ‘Plant #’, species.species_name as ‘Plant species’, chemical_class as ‘Chemical class’, weight as ‘Result weight’

from extraction_result, extraction, chem_coll_extraction, chem_collection, field_event, plant, species

where extraction_result.extraction_ID = extraction.extraction_ID and

extraction.extraction_ID = chem_coll_extraction.extraction_ID and

chem_coll_extraction.chem_coll_ID = chem_collection.chem_coll_ID and

chem_collection.field_event_id = field_event.field_event_id and

field_event.plant_id = plant.plant_id and

plant.species_code = species.species_code

order by site_name asc, plant.plant_id asc;

Average Plant Height at each Site by Species Code

select plant.Site_Name as 'Site name', plant.Species_Code as 'Species code', avg(field_event.plant_height) as Avg_Plant_Height

from field_event, plant

where field_event.plant_id = plant.plant_id

group by species_code

order by avg_plant_height;

Number of Species of each Chemical Collection Status type

Select cc.status as 'Chem Collection Status', Count(distinct p.Species_Code) as ‘# of distinct species’

from chem_collection as cc, plant as p, field_event as f

where cc.Field_Event_Id=f.Field_Event_Id and f.Plant_ID=p.Plant_ID

group by cc.status;

List all Species found at each Site

select plant.Site_Name as 'Site name', plant.Species_Code as 'Species Code', species.Species_Name as ‘Species name’

from plant, species

where plant.species_code = species.species_code

group by plant.species_code

order by plant.site_name;

Average Weight by Chemical Class for different Types of Collection (young leaf, mature leaf, etc)

select chem_collection.type as ‘Type of plant collection’, extraction_result.Chemical_Class as ‘Chemical class result’, avg(extraction_result.weight) as ‘Avg result weight’ from extraction_result, chem_collection, extraction, chem_coll_extraction

where extraction_result.extraction_ID = extraction.extraction_ID and

extraction.extraction_ID = chem_coll_extraction.extraction_ID and

chem_coll_extraction.chem_coll_ID = chem_collection.chem_coll_ID

group by chem_collection.type, extraction_result.Chemical_Class;

Average Weight by Chemical Class for Plants in different kinds of Light (shade vs. sun vs. int)

select field_event.plant_light as ‘Light Condition of Plant’, extraction_result.Chemical_Class as ‘Chemical Class’, avg(extraction_result.weight) as ‘Avg Result Weight’ from extraction_result, chem_collection, extraction, chem_coll_extraction, field_event

where extraction_result.extraction_ID = extraction.extraction_ID and

extraction.extraction_ID = chem_coll_extraction.extraction_ID and

chem_coll_extraction.chem_coll_ID = chem_collection.chem_coll_ID and

chem_collection.field_event_ID = field_event.field_event_ID

group by field_event.plant_light, extraction_result.Chemical_Class;