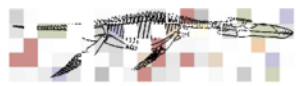


SQL databases

An introduction

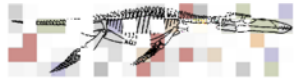


<input type="checkbox"/>			3	11798	Carnivora	Felidae	Felis	sp.	S. Kuemmell	C. Linnaeus	1758	10883	Lufeng	China	Yunnan
<input type="checkbox"/>			4	11798	Carnivora	Felidae	Metailurus	sp.	H. O'Regan	Zdansky	1924	10883	Lufeng	China	Yunnan
<input type="checkbox"/>			5	11803	Carnivora	Felidae	Panthera	pardus	W. Clyde	C. Linnaeus	1758	10604	Tabun Cave Level C & D	Israel	Mount Carmel
<input type="checkbox"/>			6	13066	Carnivora	Felidae	Panthera	onca	M. Uhen	C. Linnaeus	1758	10272	Ladds Quarry	United States	Georgia
<input type="checkbox"/>			7	13066	Carnivora	Felidae	Lynx	rufus	M. Uhen	Schreber	1777	10272	Ladds Quarry	United States	Georgia
<input type="checkbox"/>			8	13066	Carnivora	Felidae	Miracinonyx	inexpectatus	M. Uhen	E. D. Cope	1895	10272	Ladds Quarry	United States	Georgia
<input type="checkbox"/>			9	13066	Carnivora	Felidae	Puma	concolor	J. Alroy	Linnaeus	1771	10272	Ladds Quarry	United States	Georgia
<input type="checkbox"/>			10	13066	Carnivora	Felidae	Felis	sp.	M. Uhen	C. Linnaeus	1758	10272	Ladds Quarry	United States	Georgia



Today's Session

1. Selecting data from Paleobiology Database
2. Downloading and reformatting PaleoDB data
3. Setting up AMP servers and understanding them
4. Understanding a database server
5. Using phpMyAdmin client to operate MySQL
6. Importing data into MySQL
7. Break
8. Querying data using the SQL language
9. Assignment



The Paleobiology Database

- * Community effort to record fossil occurrence data from published literature for use by the community
- * Started in 1998 by project funded through the National Center for Ecological Analysis and Synthesis
- * John Alroy of Macquarie University coordinates the PaleoDB
- * Data are peer-reviewed and entered by scientists working on volunteer basis
- * Data are extensive
- * Data are incomplete: not all time, localities, or taxa are represented and there may be many omissions and errors
- * Other, more specialized occurrence databases include:
 - NOW (Neogene Old World Database, coordinated by Mikael Fortelius)
 - FaunaMap (Database of mammal occurrences in the Pleistocene of North America, coordinated by the Illinois State Museum)
 - MioMap (Database of mammal occurrences in the Miocene of North America, coordinated by Tony Barnosky)
- * The above are similar kinds of community efforts that focus on particular times and taxa, with similar strengths and weaknesses
- * Always acknowledge or cite the database, the data enterers, and the authors of the original literature from which the data was derived

Totals right now: 42,731 references • 246,069 taxa • 132,831 fossil collections • 1,051,885 taxonomic occurrences contributed by 308 scientists from 124 institutions in 22 countries





PaleoDB Download: 1. Basic Options Page

Paleobiology Database

Quick search Full search Download Analyze About Log in

Basic options Included collections Collection fields Occurrence fields

Download request form

This form allows you to download taxonomic occurrences, genera or species, or fossil collections. The [full form](#) includes extra options and fields. To compute a diversity curve, you need to check off either "period," "epoch," "Cenozoic subepoch," or "10 m.y. bin" in the "Collection fields" part of the form.

Your name (optional): David Polly

Output data: occurrence list

Output delimiter: comma-delimited text

Taxon or taxa to include: Genus level or above Felidae

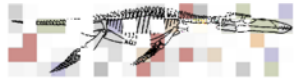
Oldest and youngest intervals (by name or by age in Ma):

Research group or project includes:

Create data set

Move between tabs here

Taxonomic name here



PaleoDB Download: 2. Included Collections Page

Paleobiology Database

Quick search Full search Download Analyze About Log in

Basic options **Included collections** Collection fields Occurrence fields

Download request form

Lithologies: carbonate mixed siliciclastic unknown

Environments: carbonate siliciclastic terrestrial unknown

Environmental zones: lacustrine fluvial karst other terrestrial
marginal marine reef shallow subtidal
deep subtidal offshore slope/basin

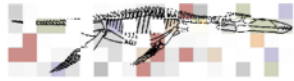
Continents or paleocontinents:

- Modern
- Europe Asia North America South America
- Africa Indian Ocean Australia Oceania Antarctica
- Western Laurasia
 - Barentsia Eastern USA Laurentia
 - North Britain Stikinia Western Avalonia
 - Wrangellia
- Eastern Laurasia
 - Baltica Eastern Avalonia Japan Kazakhstan
 - Siberia
- Western Gondwana
 - Caribbean Costa Rica-Panama
 - Precordillera South America Yucatan
- Eastern Paleotethys
 - North China Shan-Thai South China
- Eastern Gondwana
 - Africa Antarctica Arabia Australia Cimmeria
 - India New Zealand Southern Europe

Collection number(s):

Create data set

Limit search by continent here if you want



PaleoDB Download: 3. Collection Fields Page

Useful fields to include

The screenshot shows the 'Download request form' for 'Collection fields' in the Paleobiology Database. The form is organized into several sections, each with a title and a list of fields with checkboxes. The 'Collection fields' section is highlighted in red. The fields are:

- Basic fields:** collection name, collection aka, reference number, publication year, authorizer, enterer, modifier, research group.
- Geography fields:** country, state, county, geographic scale, tectonic plate id, latitude/longitude in decimal, paleolatitude/longitude in decimal.
- Time fields:** period, epoch, Cenozoic subepoch, stage, 10 m.y. bin, Fossil Record 2 bin, maximum interval, minimum interval, interval base (Ma), interval top (Ma), interval midpoint (Ma), interpolated maximum (Ma), interpolated minimum (Ma), interpolated midpoint (Ma). A note below states: "Interpolation can be used when time interval boundaries are not directly dated but fall in between boundaries that are."
- Stratigraphy fields:** formation, member, stratigraphic scale, stratigraphic comments.
- Lithology fields:** lithification 1, minor lithology 1, lithology 1, lithification 2, minor lithology 2, lithology 2, lithological description, environment, geology comments.

A 'Create data set' button is located at the bottom of the form.

Screen capture from Paleobiology Database (<http://pbdb.org/>)



PaleoDB Download: 4. Occurrence Fields Page

Paleobiology Database

Quick search Full search Download Analyze About Log in

Basic options Included collections Collection fields **Occurrence fields**

Download request form

Basic fields

<input type="checkbox"/> Check all	<input type="checkbox"/> authorizer	<input checked="" type="checkbox"/> enterer	<input type="checkbox"/> modifier
<input type="checkbox"/> reference number	<input checked="" type="checkbox"/> order name	<input checked="" type="checkbox"/> family name	
<input checked="" type="checkbox"/> class name	<input checked="" type="checkbox"/> species name	<input checked="" type="checkbox"/> abundance	<input type="checkbox"/> comments
<input type="checkbox"/> subgenus name	<input type="checkbox"/> plant organ 2	<input type="checkbox"/> preservation category	
<input type="checkbox"/> plant organ			

Taxonomic fields

<input type="checkbox"/> Check all	<input type="checkbox"/> second author	<input type="checkbox"/> other authors	<input checked="" type="checkbox"/> year named
<input checked="" type="checkbox"/> first author	<input type="checkbox"/> body part of type	<input type="checkbox"/> extant	<input checked="" type="checkbox"/> common name
<input type="checkbox"/> type specimen			

Note that the collection ID and genus name fields are always output.

Create data set

More useful fields

Click to search when ready

Save all three CSV files to your computer



Remember to give credit...

- * to the Paleobiology Database
- * to the contributors of the data
- * to the original authors of the data

Your download is large. If you intend to use the data in a publication, we ask you to:

- Notify some of the major contributors about your research.
- Acknowledge the major contributors by name in your paper.
- Cite the relevant [Online Systematics Archive](#).
- If none are relevant, give a generic citation to the Paleobiology Database including the date of access.
- [Join the PaleoDB](#) and contribute more data.
- At a bare minimum, [request a publication number](#) once your paper is accepted.

The major contributors to this data set were:

- John Alroy (44.7%)
- Kay Behrensmeyer (23.0%)
- Alan Turner (19.2%)

Each percentage is based on the estimated amount of time spent entering data by the authorizer and associated data enterers. The grand total is 748 hours (equivalent to 4.5 work months). The estimate is based on (1) date stamps for keystroked collections and (2) counts of published references used to document uploaded collections.

Additional contributors include Chris Bell, Matt Carrano, Will Clyde, Darin Croft, Emmanuel Fara, Jason Head, Denne Reed, Mark Uhen, Xiaoming Wang, and Lars van den Hoek Ostende.

If you agree to these terms, [click here](#) to access the data.

Here are the publications that yielded the most records in this data set, in order of importance. Please cite some or all of them in any publication based on it. If possible, please also include the [DavidPolly-refs.csv](#) file as online supplementary information.

- M. D. Leakey and J. M. Harris. 1987. Laetoli: a Pliocene Site in Northern Tanzania. *Clarendon Press, Oxford, Great Britain* [17 collections, 45 occurrences]
- M. Bon, G. Piccoli, and B. Sala. 1991. I giacimenti Quaternari di vertebrati fossili nell'Italia Nord-Orientale. *Memorie degli Istituti di Geologia e Mineralogia dell'Universita di Padova* **47**:185-231 [11 collections, 27 occurrences]
- E. I. Belyaeva. 1948. *Catalogue of Tertiary Fossil Sites of the Land Mammals in the U.S.S.R.* [10 collections, 26 occurrences]
- W. von Koenigswald and W.-D. Heinrich. 1999. Mittelpleistozane Saugetierfaunen aus Mitteleuropa - der Versuch einer biostratigraphischen Zuordnung. *Kaupia* **9**:53-112 [7 collections, 29 occurrences]
- E. Heintz, C. Guerin, R. Martin and F. Prat. 1974. Principaux gisements villafranchiens de France: Listes fauniques et biostratigraphie. *Memoires du Bureau de Recherches Geologiques et Minières* **78(1)**:169-182 [7 collections, 25 occurrences]
- G. D. Koufos. 2001. The Villafranchian mammalian faunas and biochronology of Greece. *Bollettino della Societa Paleontologica Italiana* **40(2)**:217-223 [10 collections, 15 occurrences]
- D. Janossy. 1986. Pleistocene vertebrate faunas of Hungary. *Developments in Palaeontology and Stratigraphy*, **8**. Elsevier, Amsterdam 1-208 [9 collections, 16 occurrences]
- M. L. Cassiliano. 1999. Biostratigraphy of Blancan and Irvingtonian mammals in the Fish Creek-Vallecito Creek section, southern California, and a review of the Blancan-Irvingtonian boundary. *Journal of Vertebrate Paleontology* **19(1)**:169-186 [12 collections, 10 occurrences]



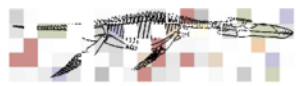
AMP: Apache, mySQL, PHP

This installation installs the Apache webserver, the PHP scripting language, and the mySQL database on your computer:

- * **Apache:** runs in the background on your computer and waits for requests from web browsers, sending back the requested files (<http://localhost/> to see the site now running on your laptop)
- * **mySQL:** a relational database server that runs in the background of your computer and waits for requests from SQL clients, returning to them the requested data
- * **PHP:** a computer language that was specially designed for use on internet servers, making it easy to extract customized data from databases and to reform it into web pages



MAMP: One-click-solution for setting up your personal webserver



Client-server model

A method of distributing access to a program

Server is the main program that handles information, calculations, data analysis, etc.

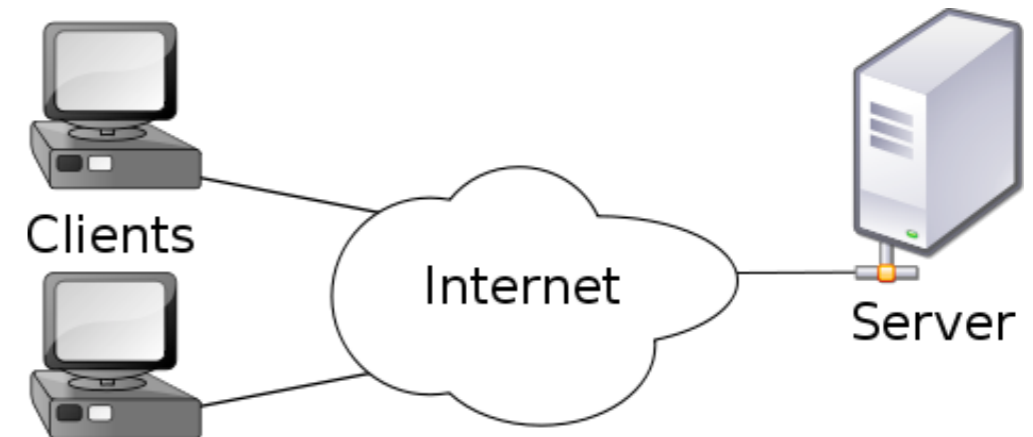
Client is the computer or program that asks for and receives information from the server

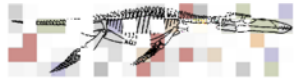
Web: client is your web browser, server is a web server application like Apache.

Databases: client is the program you use to query database, server is MySQL installed on a machine somewhere

iPhone Apps: client is the app installed on iPhone, servers is (usually) a web server like Apache

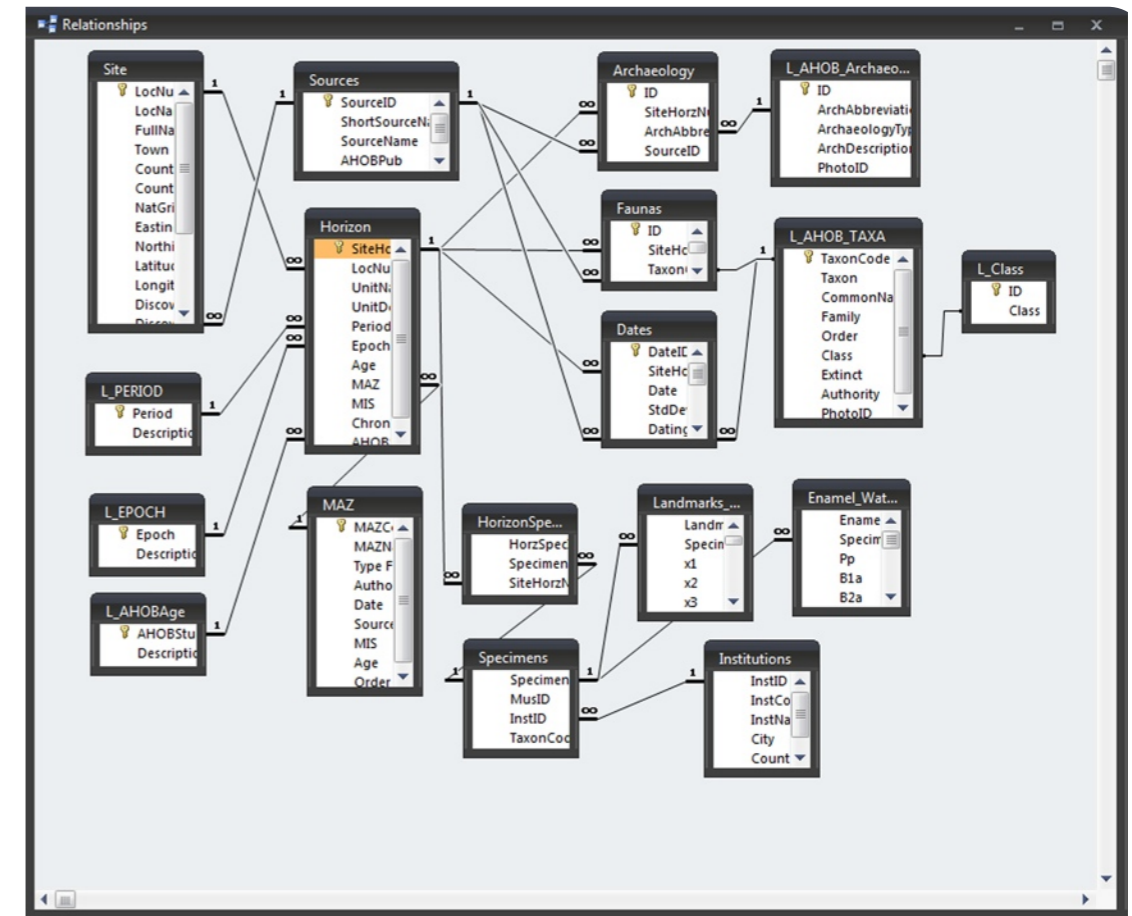
Advantage of client server model for databases: you can access data from MANY different programs without reformatting your data.





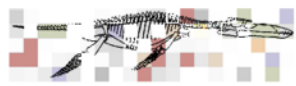
What is a relational SQL database?

- * Software that stores and processes data
- * Relational aspect refers to the ability to keep data in many tables that are linked by particular variables (e.g., “specimen” table may list of data about specimens in a museum, “taxon” table may contain information about species, the link allows users of the specimen table to find out about the species).
- * SQL (Structured Query Language) is a simple language or syntax for issuing instructions to the database software
- * Databases are useful for managing large amounts of data and for reformatting or collating them



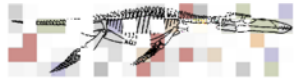
from Polly & Stringer, 2011

Graphic summary of relationships in a complex database with eighteen tables



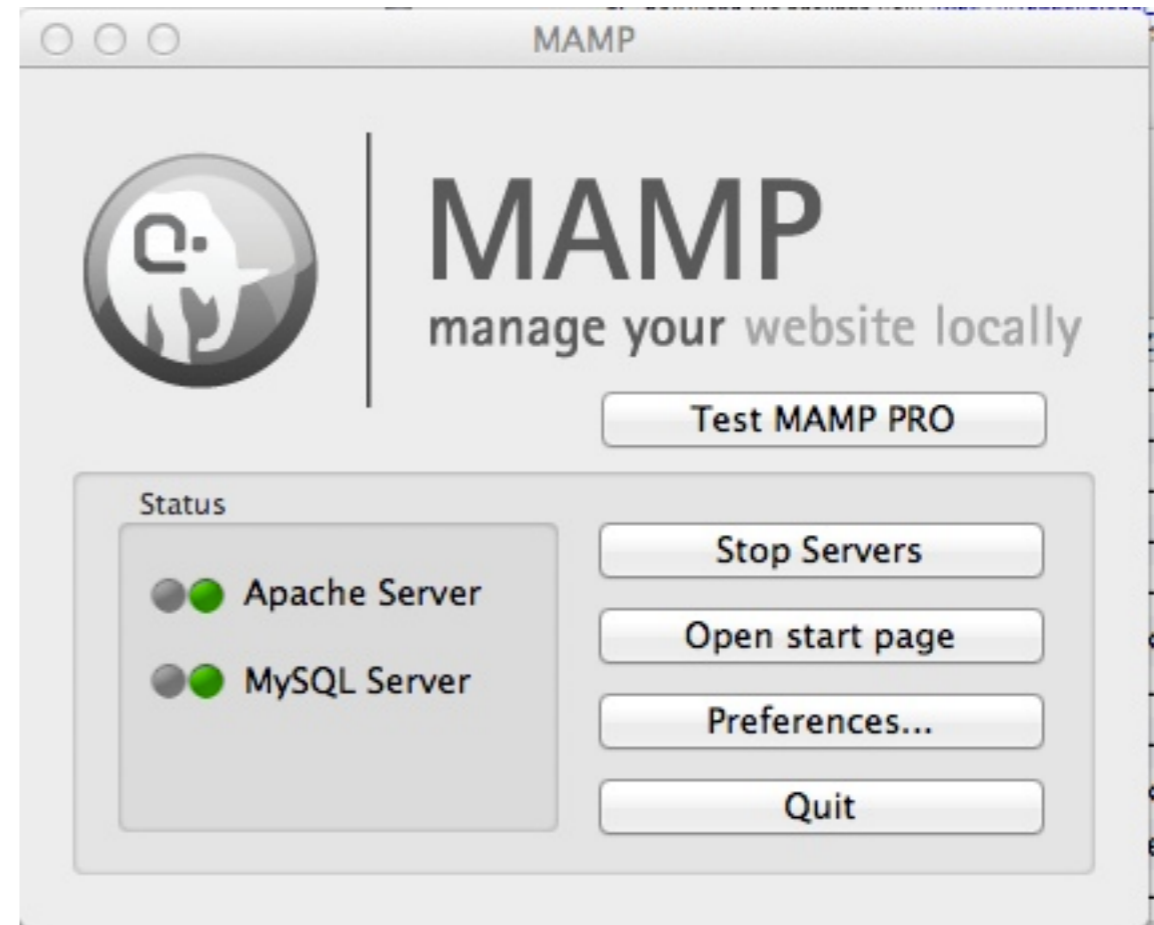
mySQL database

- * A free, open-source relational database (now owned by Oracle)
- * SQL compliant
- * Built on a client-server model
- * Fast, can handle huge data sets
- * Cross-platform
- * Specially designed to interface with web applications using scripts (PHP, PERL, Python, Java, etc.)

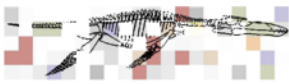


Start the MySQL Database Server and phpMyAdmin

- * “Start Servers” starts both MySQL database and Apache web server
- * Always stop servers when you are not using them
- * Start page opens the PHPMyAdmin control interface



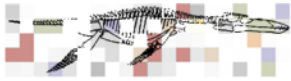
Mac OS X version. WAMP control looks different.



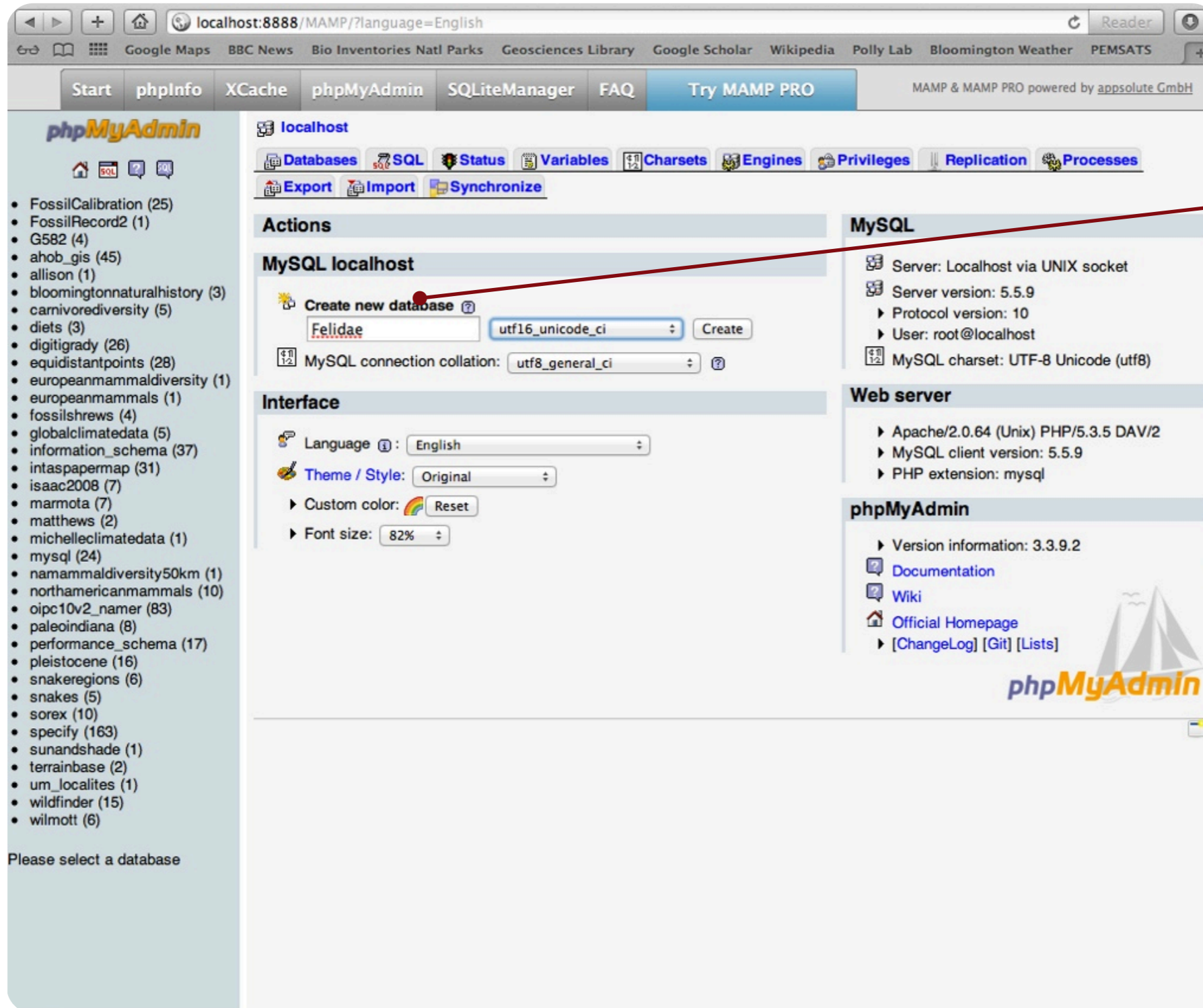
phpMyAdmin layout

The screenshot shows the phpMyAdmin web interface with several key areas highlighted by red circles and lines:

- Choose databases:** Points to the left sidebar containing a list of databases such as 'FossilCalibration (24)', 'FossilRecord2 (1)', and 'ahob_gis (45)'.
- Create queries:** Points to the 'SQL' tab in the top navigation bar.
- Create databases:** Points to the 'Create new database' form in the 'Actions' section, which includes a text input field, a 'Collation' dropdown, and a 'Create' button.
- Manage users:** Points to the 'Privileges' tab in the top navigation bar.
- Import / export data:** Points to the 'Export' and 'Import' tabs in the top navigation bar.
- Server status:** Points to the 'MySQL localhost' status panel on the right, which displays server details like 'Server: Localhost via UNIX socket', 'Server version: 5.5.9', and 'MySQL charset: UTF-8 Unicode (utf8)'.



Creating a database in phpMyAdmin



Create new database form (on home page)

Name your database and choose utf16_unicode_ci from the Collation dropdown menu

Press Create when ready



Reformat PaleoDB data for import into mySQL

Using Excel (or other spreadsheet):

Delete columns that do not seem well-populated or useful.

Rename columns with a short and intelligible label. For example:

“order_name” -> “Order”

“occurrence.genus_name” -> “Genus”

“latdec” -> “Latitude”

“longdec” -> “Longitude”

“interval_base” -> “MaxAge”

“interval_base” -> “MinAge”

“interval_midpoint” -> “MedAge”

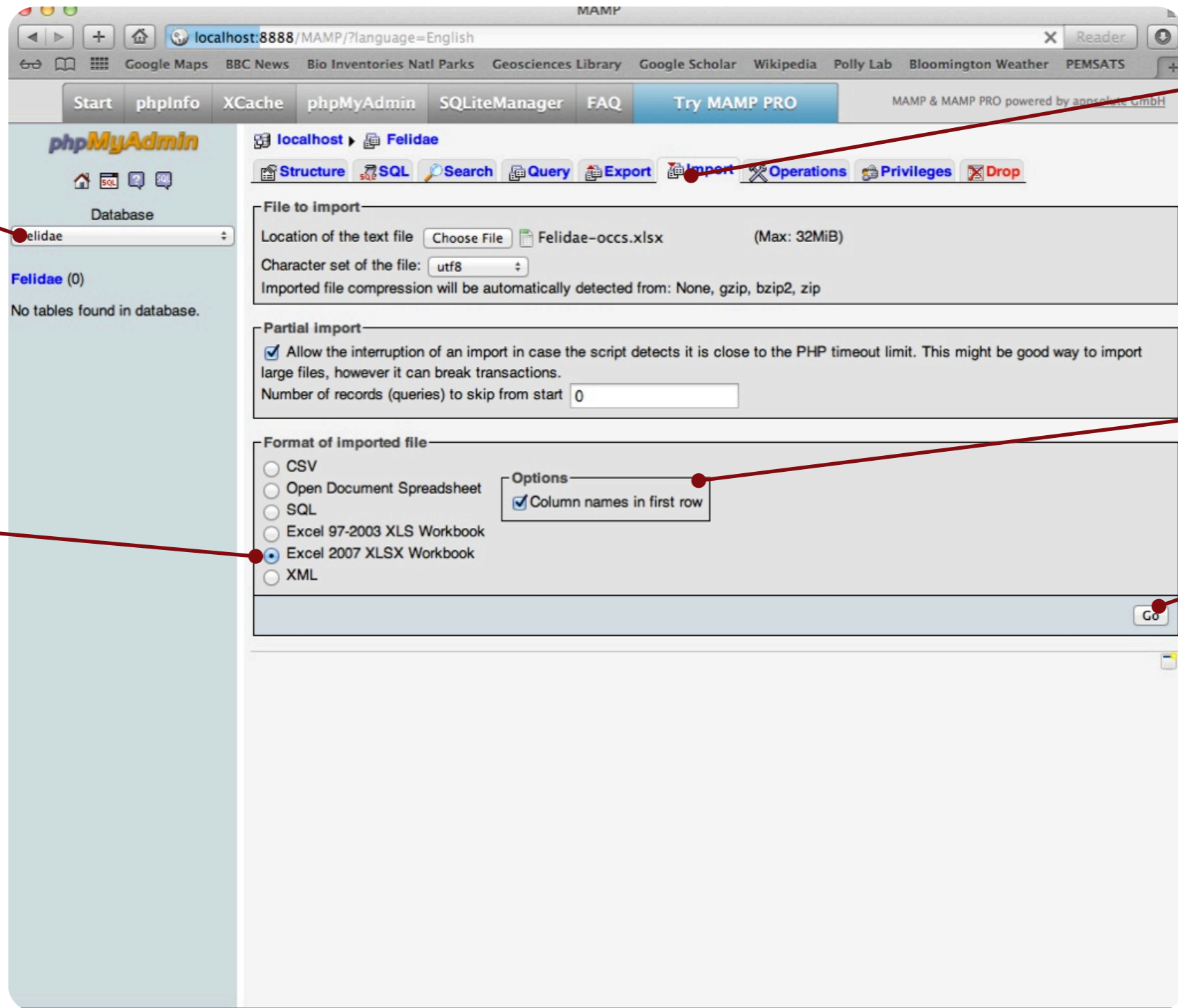
Do not use the same column name twice.

Save as an Excel workbook instead of comma-delimited file.

collection_no	class_name	order_name	family_name	occurrence.g	occurrence.g	occurrence.s	occurrence.s	original.genus	original.genus	original.spec	original.spec	occurrence.e	abund_value	abu
1	11797	Carnivora	Felidae	Vishnufelis								H. O'Regan		
2	11798	Carnivora	Felidae	Machairodus	fires			Epimachairodus		fires		H. O'Regan		
3	11798	Carnivora	Felidae	Felis								S. Kuemmell		
4	11798	Carnivora	Felidae	Metailurus				Pseudaelurus		sp.		H. O'Regan		
5	11803	Carnivora	Felidae	Panthera	pardus							W. Clyde		
6	13066	Carnivora	Felidae	Panthera	onca							M. Uhen	4 spe	
7	13066	Carnivora	Felidae	Lynx	rufus			Felis		rufus		M. Uhen	10 spe	
8	13066	Carnivora	Felidae	Miracinonyx	cf.	inexpectatus		Felis	cf.	inexpectata		M. Uhen	2 spe	
9	13066	Carnivora	Felidae	Puma	concolor			Felis		concolor		J. Alroy		
10	13066	Carnivora	Felidae	Felis								M. Uhen	1 spe	
11	13293	Carnivora	Felidae	Lynx								R. Whatley		
12	13293	Carnivora	Felidae	Panthera	pardus							R. Whatley		
13	13293	Carnivora	Felidae	Panthera	tigris							R. Whatley		
14	13456	Carnivora	Felidae	Panthera	cf.	leo						J. Finarelli		
15	13473	Carnivora	Felidae	Megantereon	praecox							J. Finarelli		
16	13473	Carnivora	Felidae	Vinayakia	nocturna							J. Finarelli		
17	13738	Carnivora	Felidae	Panthera	tigris							J. Finarelli	1 spe	
18	13907	Carnivora	Felidae	Leopardus	n. sp.	vorohuensis		Felis	n. sp.	vorohuensis		J. Wertheim	1 indi	
19	17988	Carnivora	Felidae	Pseudaelurus		intrepidus						J. Alroy		
20	18033	Carnivora	Felidae	Machairodus		coloradensis						J. Alroy		
21	18034	Carnivora	Felidae	Nimravides	cf.	thinobates						J. Alroy		
22	18036	Carnivora	Felidae	Machairodus		sp.						J. Alroy		
23	18046	Carnivora	Felidae	Nimravides		thinobates		Nimravides		catacopis		J. Alroy		
24	18060	Carnivora	Felidae	Pseudaelurus		sp.						J. Alroy		
25	18063	Carnivora	Felidae	Pseudaelurus		sp.						J. Alroy		
26	18072	Carnivora	Felidae	Machairodus		sp.						J. Alroy		
27	18072	Carnivora	Felidae	Nimravides	cf.	thinobates						J. Alroy		
28	18073	Carnivora	Felidae	Machairodus	cf.	coloradensis						J. Alroy		
29	18076	Carnivora	Felidae	Felis		rexroadensis		Felis		rexroadensis		J. Alroy		
30	18086	Carnivora	Felidae	Nimravides		sp.						J. Alroy		
31	18097	Carnivora	Felidae	Machairodus	cf.	coloradensis						J. Alroy		
32	18097	Carnivora	Felidae	Nimravides	n. sp.	hibbardi		Pseudaeluru: n. sp.		hibbardi		J. Alroy		
33	18097	Carnivora	Felidae	Lynx		proterolyncis		Felis		proterolyncis		J. Alroy		
34	18126	Carnivora	Felidae	Machairodus		sp.						J. Alroy		
35	18140	Carnivora	Felidae	Machairodus		coloradensis		Machairodus cf.		catocopis		J. Alroy		
36	18140	Carnivora	Felidae	Adelphailuru	n. sp.	kansensis						J. Alroy		
37	18164	Carnivora	Felidae	Pseudaelurus		intrepidus						J. Alroy		
38	18165	Carnivora	Felidae	Pseudaelurus		sp.						J. Alroy		
39	18166	Carnivora	Felidae	Adelphailurus		kansensis						J. Alroy		
40	18180	Carnivora	Felidae	Nimravides		galiani						J. Alroy		
41	18181	Carnivora	Felidae	Nimravides		thinobates		Nimravides		catacopis		J. Alroy		
42	18181	Carnivora	Felidae	Machairodus		sp.						J. Alroy		
43	18198	Carnivora	Felidae	Nimravides	cf.	thinobates						J. Alroy		
44	18208	Carnivora	Felidae	Pseudaelurus		sp.						J. Alroy		
45	18209	Carnivora	Felidae	Pseudaelurus		marshi						J. Alroy		
46	18219	Carnivora	Felidae	Nimravides		sp.						J. Alroy		
47	18220	Carnivora	Felidae	Pseudaelurus		stouti		Lynx		stouti		J. Alroy		
48	18225	Carnivora	Felidae	cf.	Nimravides	sp.						J. Alroy		



Importing data into your new database



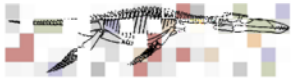
Make sure the database you want is selected

Select Excel

Import tab

Tick column headers

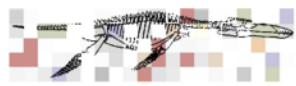
Start importing



If successful....

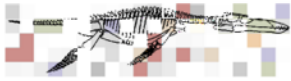
New table is listed here

The screenshot shows the phpMyAdmin interface for a MAMP environment. The browser address bar indicates the URL is localhost:8888/MAMP/?language=English. The interface includes a navigation menu with options like Start, phpInfo, XCache, phpMyAdmin, SQLiteManager, and FAQ. The main content area shows the 'Felidae' database selected, with a list of tables including 'Felidae-occs.csv'. A green message box indicates that the import was successful, stating 'Import has been successfully finished, 2 queries executed.' Below this, a list of structures is shown, including 'Felidae (Options)' and 'Felidae-occs.csv (Structure) (Options)'. The interface also features sections for 'File to import', 'Partial import', and 'Format of imported file'.



To import data from comma-delimited text file

1. Create new database to be home to your imported tables
2. Select database and click on Import tab
3. Choose file
4. Note that “import” is context dependent. If you have selected database, it will import as new table, but if you have selected a table it will import into that table rather than a new one
5. Once imported, select table from list at left
6. Change table name using the “Operations” tab
7. Add a “key” to your table by selecting your table then selecting the “structure” tab. At bottom add one variable at the beginning of the table. Name that (e.g., “ID”), give it an Integer type (INT), make it the primary index, and tick autoincrement (A_I). This key assigns a unique number to each line in the table, which is important for some operations.
8. Browse data with browse tab. Note that you can delete or edit lines of data using the “X” and the pencil icons.



Browsing and editing data

Browse tab

Statistics about data

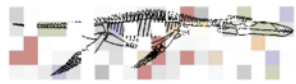
The screenshot shows the phpMyAdmin interface for a database named 'Felidae'. The 'Browse' tab is selected, displaying a table of occurrences. The table has columns for ID, CollNo, Order, Family, Genus, Species, Enterer, Author, Year, CollRefNo, CollName, and Country. The first few rows are:

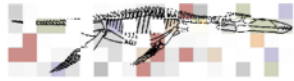
ID	CollNo	Order	Family	Genus	Species	Enterer	Author	Year	CollRefNo	CollName	Country
1	11797	Carnivora	Felidae	Vishnufelis	sp.	H. O'Regan	Pilgrim	1932	4195	Ramanagar	India
2	11798	Carnivora	Felidae	Machairodus	fires	H. O'Regan	NULL	NULL	10883	Lufeng	China
3	11798	Carnivora	Felidae	Felis	sp.	S. Kuemmell	C. Linnaeus	1758	10883	Lufeng	China
4	11798	Carnivora	Felidae	Metailurus	sp.	H. O'Regan	Zdansky	1924	10883	Lufeng	China
5	11803	Carnivora	Felidae	Panthera	pardus	W. Clyde	C. Linnaeus	1758	10604	Tabun Cave Level C & D	Israel
6	13066	Carnivora	Felidae	Panthera	onca	M. Uhen	C. Linnaeus	1758	10272	Ladds Quarry	United States
7	13066	Carnivora	Felidae	Lynx	rufus	M. Uhen	Schreber	1777	10272	Ladds Quarry	United States
8	13066	Carnivora	Felidae	Miracinonyx	inexpectatus	M. Uhen	E. D. Cope	1895	10272	Ladds Quarry	United States
9	13066	Carnivora	Felidae	Puma	concolor	J. Alroy	Linnaeus	1771	10272	Ladds Quarry	United States
10	13066	Carnivora	Felidae	Felis	sp.	M. Uhen	C. Linnaeus	1758	10272	Ladds Quarry	United States
11	13293	Carnivora	Felidae	Lynx	sp.	R. Whatley	Kerr	1792	4412	Xiashan Cave lower part (Guangdong Province)	China
12	13293	Carnivora	Felidae	Panthera	pardus	R. Whatley	C. Linnaeus	1758	4412	Xiashan Cave lower part (Guangdong Province)	China

Click pencil to edit a record

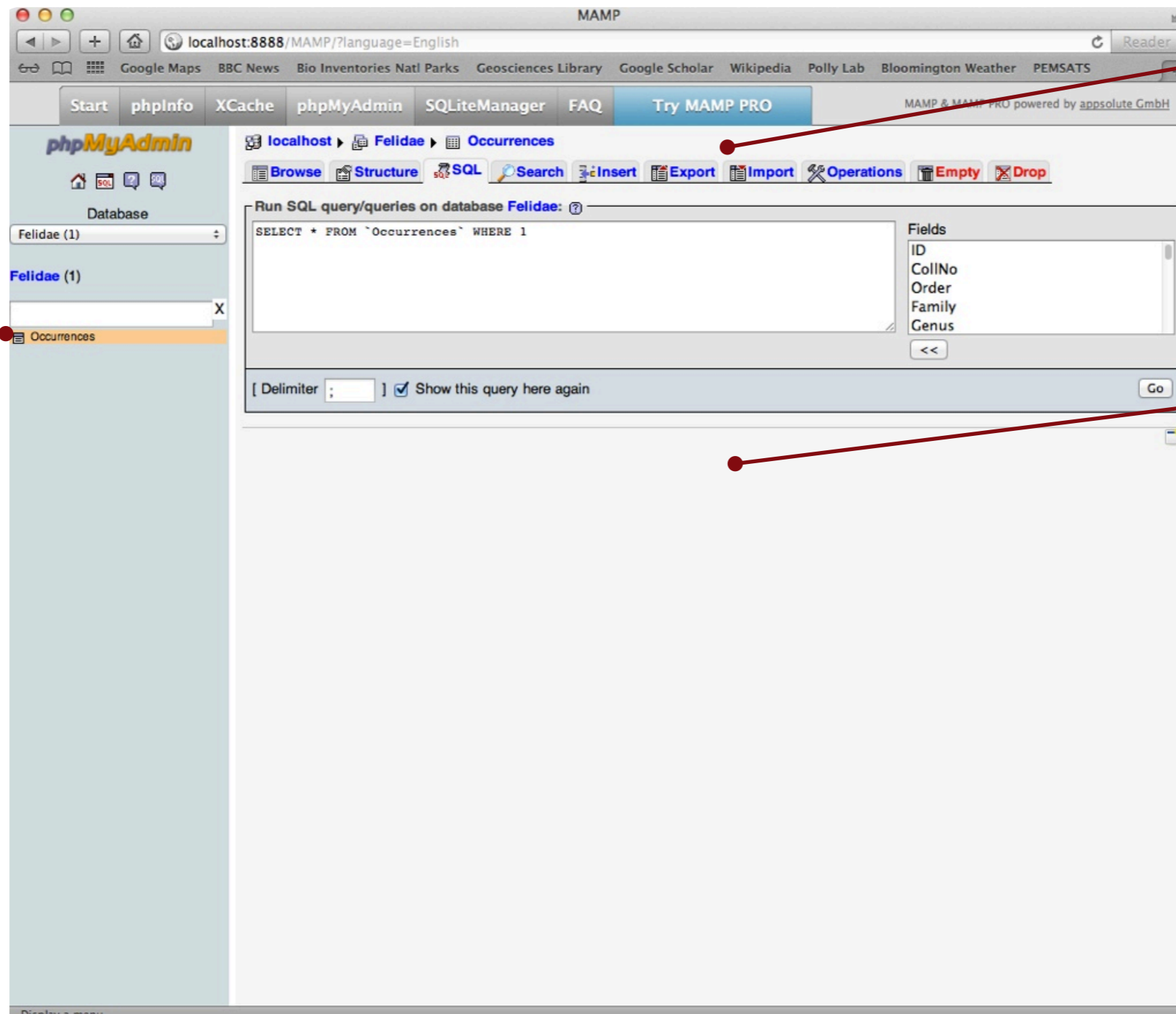
Click x to delete a record

Tick box to select a record for editing or deleting multiple lines with control at bottom of the page





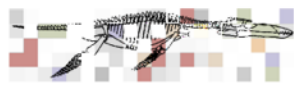
Querying your database with SQL



Make sure the database and table you want is selected

Import tab

Tick column headers



Structured Query Language (SQL)

SQL is a standardized way to format instructions to a database server. It resembles English in its grammar and syntax.

Verb (required), **Object** (required), **Object Modifiers** (optional),
From Clause (required), **Modifying Clause** (optional)

```
SELECT * FROM pbdb WHERE Genus='Hesperocyon';
```

SELECT	*	FROM	pbdb	WHERE	Genus='Hesperocyon';
Select Clause		From Clause		Modifying Clause	

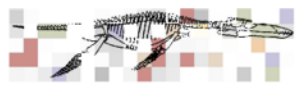
Verb gives the command (e.g., “Select”)

Object indicates what fields to select

Clause indicates what table the fields should be selected from

Modifying clause tells what criteria should be used to make the selection

* is the wildcard character, meaning “all fields”



SQL queries

Sentence form:

SELECT *whichever columns you want* **FROM** *whichever table you want* **WHERE** *your criteria are true*

For example, the following query....

SELECT * FROM occurrences

...generates the same browse page you already saw because it selects all columns (*) from the table *taxa* where no criteria are specified (i.e., it selects all rows).

And the following query....

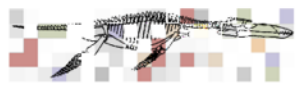
SELECT * FROM occurrences WHERE Country='United States'

selects all the columns from only those rows where the Country column contains the word “United States”.

Finally, the following query....

SELECT Genus FROM taxa WHERE Country='United States'

selects only the taxon name from the rows where the TaxonLevel column contains the word “Species”.



Some powerful modifiers for SQL queries

Select a list of all the unique Countries the database

```
SELECT DISTINCT Country FROM occurrences
```

Count the number of entries for each Country

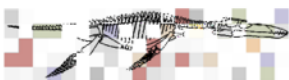
```
SELECT Country, Count(Country) FROM occurrences GROUP BY Country
```

note that “group by” is used to define the grouping variable and then the Count() function counts all the entries of Country in each group

Count the number of genera in the database from the USA

```
SELECT Genus, Count(Genus) FROM occurrences WHERE Country='United States' GROUP BY Genus
```

Useful functions that can be used in the SELECT statement in conjunction with a GROUP BY statement include: Count(), Avg(), Max(), Min(), Std(), Sum(), Variance(). You can also use multiplication, division, addition, etc. (e.g., “Sum(TaxonID)+Count(TaxonLevel)”).



Quickly tabulate data with SQL

localhost ▶ Felidae ▶ occurrences

Browse Structure SQL Search Insert Export Import Operations Empty Drop

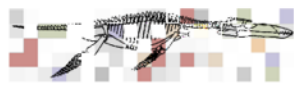
Run SQL query/queries on database Felidae: ?

```
SELECT Genus, Count(Genus) FROM occurrences WHERE Country='United States' GROUP BY Genus ORDER BY Count(Genus)
```

Fields
ID
CollNo
Order
Family
Genus

[Delimiter ;] Show this query here again Go

Genus	Count(Genus) ▲
Metailurus	1
Pratifelis	1
Xenosmilus	2
Dinofelis	4
Adelphailurus	5
Megantereon	7
Leopardus	7
Felis	14
Nimravides	19
Miracinonyx	26
Machairodus	26
Pseudaelurus	30
Homotherium	33
Puma	35
Panthera	48
Smilodon	53
Lynx	68



Assignment for next week

Summarize your data

Use SQL and phpMyAdmin to summarize your data. Compile the following information and summarize it neatly, perhaps putting some of the information in a nicely formatted table:

1. Name of your group
2. Total number of occurrences in your database table
3. Number of distinct species
4. Number of distinct genera
5. Number of distinct families
6. Continents represented in your database
7. Number of distinct localities (i.e., distinct latitude and longitude combinations)
8. Using one age variable (e.g., median age), the maximum and minimum age of the occurrences in your data set
9. Number of distinct ages

Your data should have at least a few hundred occurrences, minimum of 5-8 distinct taxa at species or genus level, minimum of 20 distinct localities, and minimum of 10 distinct ages.

Reading (available in OnCourse)

Sepkoski Jr, J. 1993 Ten years in the library: new data confirm paleontological patterns. *Paleobiology* 19, 43-51.

Lloyd, G., Davis, K., Pisani, D., Tarver, J., Ruta, M., Sakamoto, M., Hone, D., Jennings, R. & Benton, M. 2008 Dinosaurs and the Cretaceous terrestrial revolution. *Proceedings of the Royal Society B: Biological Sciences* 275, 2483-2490.