

---

# **VoSeq Documentation**

***Release 1.5.0***

**Carlos Peña, Tobias Malm**

April 06, 2013

# CONTENTS

<b>1</b>	<b>News</b>	<b>1</b>
<b>2</b>	<b>Hello there</b>	<b>2</b>
2.1	How to cite VoSeq . . . . .	3
2.2	Help and contact information . . . . .	3
<b>3</b>	<b>Getting Started</b>	<b>4</b>
3.1	Install in Linux . . . . .	4
3.2	Install in Mac OS X . . . . .	5
3.3	Manual Install in Windows 7 / Vista / XP . . . . .	6
3.4	Install in Windows with XAMPP . . . . .	8
<b>4</b>	<b>Quick Guide</b>	<b>10</b>
4.1	Adding vouchers . . . . .	10
4.2	Adding genes . . . . .	11
4.3	Adding sequences . . . . .	13
4.4	Create taxonset . . . . .	13
4.5	Create datasets . . . . .	15
4.6	Search . . . . .	16
4.7	Upload voucher photos . . . . .	16
4.8	Create Excel table . . . . .	17
4.9	Update voucher . . . . .	18
<b>5</b>	<b>Update VoSeq</b>	<b>20</b>
<b>6</b>	<b>Plugins</b>	<b>21</b>
6.1	Yahoo! Maps plugin . . . . .	21
6.2	Flickr plugin . . . . .	21
6.3	BLAST plugin . . . . .	22
6.4	Integration with EOL . . . . .	23
6.5	Sharing Photos with EOL . . . . .	23
6.6	Integration with GBIF . . . . .	24
<b>7</b>	<b>FAQ</b>	<b>27</b>
7.1	Installation FAQ: . . . . .	27
7.2	POST request error . . . . .	27
7.3	Execution time error . . . . .	27
7.4	Too many variables problem . . . . .	27
7.5	Mac permission problem . . . . .	28
7.6	How to find php.ini and see your PHP settings . . . . .	28

# NEWS

---

**Note:** Here we will describe new information regarding VoSeq, as well as list changes in new versions of the software.

---

As of version 1.5.0 we have implemented more information to be stored and used for genes/alignments!

- For example introns (or other non-wanted regions), and genetic codes and reading frames for protein coding genes are now set for each gene.
- Now can also gene tables be made with characteristics for each gene/alignment (for chosen taxonset) such as length, gene type, parsimony informative sites, nucleotide content etc.

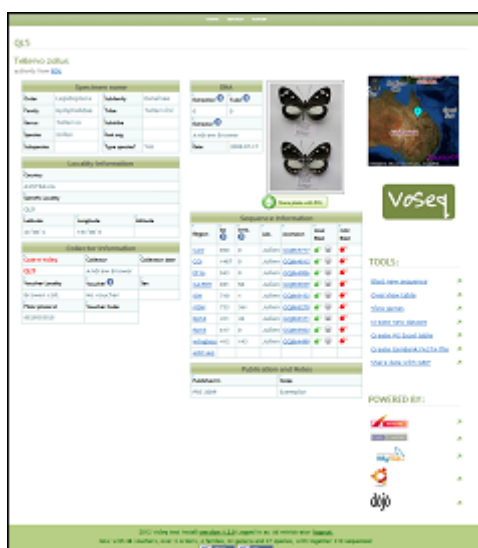
We have tried out a perhaps easier way of installation for Apache, MySQL and PHP with the XAMPP package (<http://www.apachefriends.org/en/xampp.html>). See installation instructions for windows systems. Hope it makes things easier!

# HELLO THERE

Welcome to **VoSeq**, a database to store voucher and DNA sequence data for phylogenetic analysis. It's a unique database that enables to digitize biological data of museum specimens and molecular data such as DNA sequences, primers and genes.

VoSeq has tools that facilitate the batch upload of lots of voucher data and DNA sequences with a few clicks. It also has **BLAST** (<http://en.wikipedia.org/wiki/BLAST/>) capabilities, meaning that you can find out whether one particular DNA sequence is most similar to other sequence in **NCBI GenBank** (<http://www.ncbi.nlm.nih.gov/genbank/>). You can also BLAST your sequence against all others in your VoSeq database (see *BLAST plugin* (page 22) section for details).

VoSeq is written mainly in **PHP** (<http://www.php.net/>). It uses **MySQL** (<http://www.mysql.com>) as back-end and it is designed to run in a local server (for example by installing **Apache** (<http://httpd.apache.org/>) on your computer) or to run on any commercial server service.



Dataset information

Dataset name

Dataset creator

Description

Show seqs

☐ ArgPin
 ☐ CAP
 ☐ DCC
 ☐ EFTs
 ☐ GAPDH
 ☐ IGH
 ☐ MDH
 ☐ RpS2
 ☐ RpS5
 ☐ template\_COI
 ☐ wingless

Filter names

Filter

Enter filter string here:

Sort by

☐ Order
 ☐ Family
 ☐ Subfamily
 ☐ Genus
 ☐ Species
 ☐ Host org
 ☒ Code
 ☐ X

Mark all

Unmark all

Filter/Sort

Add dataset

#	Code	Order	Family	Subfamily	Genus	Species	Host org
<input type="checkbox"/>	CP-B01	Lepidoptera	Nymphalidae	Satyrinae	Belenesia	circa	
<input type="checkbox"/>	CP-B02	Lepidoptera	Nymphalidae	Satyrinae	Penthesia	dorica	
<input type="checkbox"/>	CP01.33	Lepidoptera	Nymphalidae	Satyrinae	Egyptichia	typ2	
<input type="checkbox"/>	CP03-63	Lepidoptera	Nymphalidae	Satyrinae	Marcebia	cyclopina	
<input type="checkbox"/>	CP03-70	Lepidoptera	Nymphalidae	Satyrinae	Promophila	thelobis	
<input type="checkbox"/>	CP03-82	Lepidoptera	Nymphalidae	Satyrinae	Morpho	varona	

## 2.1 How to cite VoSeq

If you think VoSeq is useful and you happen to use it during your work, it would be great if you cite us as a source:

- Peña, C. & Malm, T. **2012**. VoSeq: a Voucher and DNA Sequence Web Application. *PLOS ONE*, 7(6): e39071. doi (<http://dx.doi.org/10.1371/journal.pone.0039071>)

## 2.2 Help and contact information

If you need help regarding installation or usage of the application, please contact [Carlos Peña](mailto:mycale-sis@gmail.com) (mycale-sis@gmail.com) or [Tobias Malm](mailto:tobemalm@gmail.com) (tobemalm@gmail.com).

You can also subscribe to VoSeq's discussion list on [Google Groups](https://groups.google.com/d/forum/voseq-discussion-list) (<https://groups.google.com/d/forum/voseq-discussion-list>).

# GETTING STARTED

Once you have successfully downloaded VoSeq, you can find out how to:

- *Install in Linux* (page 4)
- *Install in Mac OS X* (page 5)
- *Manual Install in Windows 7 / Vista / XP* (page 6)
- *Quick Guide* (page 10) to get started with VoSeq.

## 3.1 Install in Linux

Before installing VoSeq, you need to install in your computer a web server (such as [Apache](http://httpd.apache.org/) (<http://httpd.apache.org/>)) and the relational database [MySQL](http://www.mysql.com) (<http://www.mysql.com>).

### 3.1.1 Required software

- Web server with PHP 5.0 or higher (<http://www.php.net/manual/install.unix.php>). **Compile it with the library CURL**, which is needed to do BLASTs against GenBank.
  - Apache HTTP Server
  - PHP
- A MySQL server 5.0 or higher (see <http://www.mysql.com>)
- GD library

---

**Note:** These instructions assume that you are using Linux and Apache, and have installed [LAMP](http://en.wikipedia.org/wiki/LAMP) (<http://en.wikipedia.org/wiki/LAMP>) (Linux, Apache, MySQL and PHP on your computer).

---

1. Compile PHP with support for the graphics library GD. More info [here](http://www.php.net/manual/en/image.installation.php) (<http://www.php.net/manual/en/image.installation.php>).
2. Download VoSeq: [Download from github](https://github.com/carlosp420/VoSeq/tags) (<https://github.com/carlosp420/VoSeq/tags>).
3. Unzip the source files in some directory: `unzip VoSeq_X.Y.Z.zip`
4. If you are not a Linux Guru and you have [WinRAR](http://www.rarlab.com/) (<http://www.rarlab.com/>) (like WinZip but works with gzipped files) on your Windows system you can cheat a little bit here. You can download the file to your Windows machine, use WinRAR to unzip the gzipped file into a directory in Windows and then use an FTP program like [WinSCP](http://winscp.net/eng/index.php) (<http://winscp.net/eng/index.php>) to transfer the entire VoSeq directory for you to a commercial server for example.
5. Move the directory into your web directory: e.g. `mv VoSeq /usr/local/apache2/htdocs/myVoSeq` or `mv VoSeq public_html/myVoSeq` or use your FTP software to do this for you.

6. To run the installation script, you'll need to temporarily make your myVoSeq directory writable by the web server. The simplest way to do this on a Unix/Linux system is to make it world-writable by typing: `chmod -R 777 myVoSeq`. To do this into a commercial server you will need a telnet client like PuTTY (<http://www.chiark.greenend.org.uk/~sgtatham/putty/>) on your system.
7. At this point you should have Apache and MySQL running (this varies between distributions and setups, see their documentations for details).
8. Go to your web browser and surf into the VoSeq installation directory (under `htdocs` or `public_html` folders of Apache). It will direct you to the config script (if it doesn't, just load up the `http://localhost/myVoSeq/index.php` file. Fill out the forms.
9. If all goes well, the installer will create a configuration file named `conf.php` in your myVoSeq installation directory. This file will contain all the important variables and information needed to run VoSeq in your system.

## 3.2 Install in Mac OS X

We have successfully installed VoSeq in a MacBook OS X Lion. It appears that the Mac operative systems **come already with Apache and PHP installed**. However you will need to enable Apache to read and run PHP files.

### 3.2.1 To connect Apache and PHP so that they work together:

1. Edit Apache's configuration text file:
  - `sudo nano /etc/apache2/httpd.conf`
2. Make sure that the line: `LoadModule php5_module libexec/apache2/libphp5.so` is in the file and it is not commented (there is no `#` symbol at the beginning of the line).
3. Find the section `<IfModule mime_module>` and write the following line `AddType application/x-httpd-php .php` so that Apache will run any file with the extension `.php` as a script and will not show it as plain text.

### 3.2.2 Install MySQL

Unfortunately Mac OS X systems don't come with MySQL installed. You can download it from here:

1. Download MySQL from here: <http://dev.mysql.com/downloads/mysql/5.1.html>
  - Download the `.dmg` package according to your systems specifications (32 bits or 64 bits).
2. You might also want to install MySQL GUI Tools <http://dev.mysql.com/downloads/gui-tools/5.0.html>
3. The following is a quick guide to installing MySQL on your computer. **It is not comprehensive and you will find much more info in the documentation for installing MySQL here:** <http://dev.mysql.com/doc/mysql-macosx-excerpt/5.5/en/index.html>
4. Unpack and install both pieces of software. Make sure you install the package, in my case, `mysql-5.1.60-osx10.6-x86_64.pkg` and `MySQLStartupitem.pkg`
5. Start the MySQL server by typing in the terminal: `sudo /Library/StartupItems/MySQLCOM/MySQLCOM.start`
6. Create a password for the user **root** by typing: `/usr/local/mysql/bin/mysqladmin -uroot password 'myownpassword'`

### 3.2.3 Install VoSeq

1. To start Apache, go to System Preferences>Sharing> and tick Web Sharing to start your web server. Your assigned folder to host your webpages and VoSeq installation is the folder Sites in your Home directory: /Users/YourName/Sites. You will need to place there the source files of VoSeq\_X.Y.Z.zip
2. You need to click the button “create personal share folder” to create the folder “Sites”.
3. Open a Terminal: go to Applications>Utilities>Terminal. In the Terminal window, type `cd ~/Sites` to go to the folder where the file VoSeq\_X.Y.Z.zip should be.
4. Unpack the contents by typing `unzip VoSeq_X.Y.Z.zip`
5. Start the MySQL server: `sudo /Library/StartupItems/MySQLCOM/MySQLCOM start`
6. Go to your web browser and point it to the VoSeq installation directory: `http://localhost/~YourName/VoSeq`. It will direct you to the config script. Fill out the forms.
7. If all goes well, the installer will create a configuration file named `conf.php` in your VoSeq installation directory. This file will contain all the important variables and information needed to run VoSeq in your system.
8. If during installation, VoSeq cannot connect to MySQL server, you might need to modify your /usr/local/mysql/support-files/my-large.cnf file parameters:
  - Modify the lines `/var/mysqld/mysqld.sock` to this `/tmp/mysql.sock`
  - Save the file as `/etc/my.cnf`

## 3.3 Manual Install in Windows 7 / Vista / XP

Follow these instructions to install Apache, PHP and MySQL and lastly VoSeq on Windows 7 / Vista / XP systems - its not as hard as it looks!

### 3.3.1 Apache

1. **Download and install “Apache2.x”** (tested on 2.2.x) <http://httpd.apache.org/> as recommended, preferably use “localhost” as Network Domain and Server Name. Start the service and try it out by opening <http://localhost> in your web browser - the output should be **It works!**.

---

**Note:** Notice that Apache will want to use port 0.0.0.0:80, which may be used by other programs, if Apache doesnt start (may say something about port occupied), write `netstat -nab` in Terminal and check if some other process is using that adress - then close that process if appropriate.

---

### 3.3.2 PHP

1. **Download “PHP 5.x.zip”** (tested on version 5.2.17) <http://windows.php.net/download/>. We recommend that you download the VC6 Thread Safe version if using Apache. Unpack the PHP5.x.zip file to a folder named PHP (ex. C:\PHP or C:\Program Files\PHP). Then copy the `php.ini-???` to C:\WINDOWS and rename it `php.ini`. (??? can be dist, production or development).
2. Open the apache configuration file `httpd.conf` in a text editor (found in the C:\Program Files\Apache Software Foundation\Apache2.2\conf folder after standard install).
3. Add the following 4 lines at the end of the LoadModule section (now assuming php installed to C:\PHP otherwise change this to correct installation folder):



```
LoadModule php5_module "c:/PHP/php5apache2_2.dll"
AddHandler application/x-httpd-php .php
# configure the path to php.ini
PHPIniDir "c:/windows"
```

4. Add a file called `info.php` containing `<?php phpinfo();?>` to the `C:\Program Files\Apache Software Foundation\Apache2.2\htdocs` folder.
5. Restart your Apache Server to confirm changes: “Start > All Programs > Apache HTTP Server 4.2.4 > Control Apache Server > Restart”.
6. Open up your web browser and type in: <http://localhost/info.php>. If you get a page with blue tables containing PHP and Apache info, then **installation is successful!**
7. Finish installing PHP by modifying your PHP Configuration File (`C:\WINDOWS\php.ini`) in a text editor:
  - Find the line containing: (Delete the “;” at the beginning of the lines)
    - `;extension_dir = "./"` and change it to
    - `extension_dir = "C:\php\ext"`
  - and the line containing:
    - `;session.save_path = "/tmp"` and change it to
    - `session.save_path = "C:\WINDOWS\temp"`

### 3.3.3 Enable the curl protocol

Curl is needed to get the Flickr plugin to work and enable VoSeq to interact with other databases.

1. Copy the file `php_curl.dll` from the folder `C:\PHP\ext` into the folder `C:\WINDOWS\system32`
2. Remove the semicolon `;` from the line `;extension=php_curl.dll` in your file `C:\WINDOWS\php.ini`
3. Restart the apache server.

### 3.3.4 MySQL

1. **Download and install MySQL** (tested on 5.5) from <http://dev.mysql.com/downloads/mysql/> with typical install - check the “skip Sign-Up” and “Configure the MySQL server now” boxes when they arrive. Finish installation.
2. The MySQL Server Instance Configuration Wizard should appear.
  - Click “next” ->
  - Select “Detailed Configuraton” and click “next” ->
  - Select “Developer Machine” and click “next” ->
  - Select “Multifunctional Database” and click “next” -> click “next” ->
  - Select “Decision support (DSS)/OLAP” and click “next” ->
  - Check “Enable TCP/IP Networking”
  - “Port Number” should be set to “3306” and
  - Check “Enable strict mode”, click “next” ->
  - Select “Standard Character Set” and click “next” ->
  - Check “Install As Windows Service, set the name to “MySQL” and check “Launch the MySQL Server automatically

- Make sure that the “Include Bin Directory in Windows Path” is **NOT** checked.
  - Click “Next”. -> Check the box that says “Modify Security Settings”.
  - Enter a password for the default “root” account, and confirm the password in the box below.
  - **Do NOT check the boxes** “Enable root access from remote machines” or “Create An Anonymous Account”.
  - Click “Next” -> Click “Execute” and let it finish.
  - Click “Finish”. Now MySQL should be installed.
3. **To enable PHP to use the MySQL databases**, open the `php.ini` (`C:/WINDOWS/php.ini`) file in your text editor and find the line `;extension=php_mysql.dll`. Delete the “`;`” at the beginning of the line and save the file.
  4. Add the PHP directory to Windows PATH - To do this, click:
    - Start > My Computer > Properties > Advanced > Environment Variables.
    - Under the second list (System Variables), there will be a variable called “Path”.
    - Select it and click “Edit”. Add “`;` `C:\php`” (or your own path to PHP if installed as other) to the very end of the string and click “OK”.
  5. Restart your computer and try out the database.
  6. (Optional) In order to easily makes changes or additions in your database download and try out the [MySQL Workbench](http://dev.mysql.com/downloads/workbench/5.2.html) (<http://dev.mysql.com/downloads/workbench/5.2.html>)

### 3.3.5 Install VoSeq

1. Download and unzip the file `Voseq_VersionNumber.zip` in the Apache folder (rename the new folder if necessary):
  - `C:\Program Files\Apache Software Foundation\Apache2.2\htdocs`
2. Point your web browser to the address (that is - localhost + the name of your VoSeq folder): `http://localhost/Voseq_VersionNumber` and follow the instructions for installing the software.
3. If all goes well, the installer will create a configuration file named `conf.php` in your VoSeq installation directory. This file will contain all the important variables and information needed to run VoSeq in your system.

## 3.4 Install in Windows with XAMPP

If you dont want to install Apache, MySQL and PHP manually you may want to try using a 3-rd party installer. XAMPP installs all three as well as the extra protocols needed for PHP.

### 3.4.1 Install XAMPP

You can download XAMPP installer at <http://www.apachefriends.org/en/xampp.html> Then install it with the installer (tested with version 1.8.1). With XAMPP MySQL is installed without password, **for security you need to create a new password for MySQL** (as well as for the XAMPP web directory which by default is accessible for everyone that know your IP adress, though you may still be somewhat protected behind a router ([read here](http://www.apachefriends.org/en/xampp-windows.html#1221) (<http://www.apachefriends.org/en/xampp-windows.html#1221>)).

- Goto `localhost/security` and check your security level and set passwords!

### 3.4.2 Install VoSeq

1. Download and unzip the file `Voseq_VersionNumber.zip` in the `XAMPP/htdocs` directory (rename the new folder if necessary):
  - `C:\XAMPP\htdocs`
2. Point your web browser to the address (that is - localhost + the name of your VoSeq folder): `http://localhost/VoSeq_VersionNumber` and follow the instructions for installing the software.
3. If all goes well, the installer will create a configuration file named `conf.php` in your myVoSeq installation directory. This file will contain all the important variables and information needed to run VoSeq in your system.

Configuration files after XAMPP install can be seen [XAMPP config](#) (page 9).

---

#### Note:

- This has only been tested quickly, and may not work for all computer systems!
    - We welcome all feedback for this type of installation!
  - If you already have MySQL install XAMPP SHOULD not overwrite your existing databases, but precaution is a virtue (or something...) and we advice making backups of stored data before installation. (We can not be held responsible for any loss of data)
  - More information regarding XAMPP for windows is found here: <http://www.apachefriends.org/en/xampp-windows.html>
- 

### 3.4.3 XAMPP config

File (Directory)	Usage
<code>\xampp\apache\conf\httpd.conf</code>	The main configuration file for Apache. It's including other files from the subdirectory "extra".
<code>\xampp\FileZillaFTP\FileZilla Server.xml</code>	The configuration file for the FileZilla FTP server. It's configured with the application " <i>FileZilla Server Interface.exe</i> " (button "Admin" in CP).
<code>\xampp\MercuryMail\mercury.ini</code>	The configuration file for the Mercury SMTP server. It's configured with the application " <i>mercury.exe</i> " (button "Admin" in CP).
<code>\mysql\bin\my.ini</code>	The configuration file for the MySQL Server.
<code>\xampp\php\php.ini</code>	The configuration file for PHP.
<code>\xampp\phpMyAdmin\config.inc.php</code>	The configuration file for phpMyAdmin.
<code>\xampp\sendmail\sendmail.ini</code>	The configuration file for msmtplib (sendmail compatible interface).
<code>\xampp\webalizer\webalizer.conf</code>	The configuration file for the Webalizer.

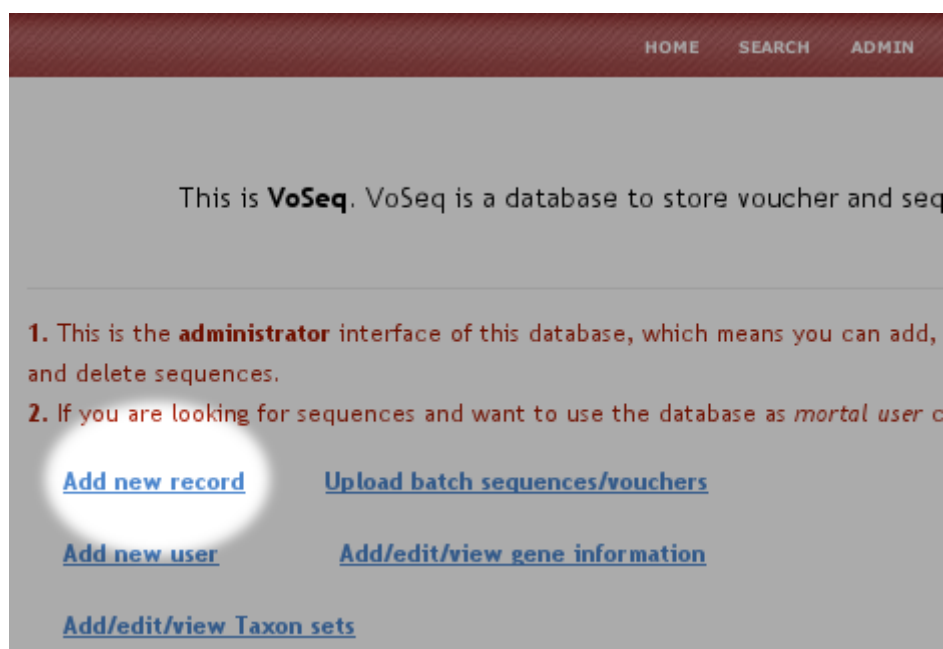
## QUICK GUIDE

Here is a quick overview on how to use VoSeq with your data.

1. *Adding vouchers* (page 10)
2. *Adding genes* (page 11)
3. *Adding sequences* (page 13)
4. *Create taxonset* (page 13)
5. *Create datasets* (page 15)
6. *Search* (page 16)
7. *Upload voucher photos* (page 16)
8. *Create Excel table* (page 17)
9. *Update voucher* (page 18)

### 4.1 Adding vouchers

After successful installation, the first thing to do is to add records (vouchers, or specimens). You can add a single record by going to the **Administrator interface** and clicking on the link **Add new record**.



The most important information to enter is the **code** of voucher, which has to be unique. VoSeq will refuse to accept duplicate codes and will issue error message if this happens. Another necessary field is the **Genus** entry, while all other fields are optional.

You can also upload a batch of records using the tool **Upload batch sequences/vouchers**. You will be shown a page to batch-upload sequences. By clicking the button **Upload vouchers instead** you will see instructions on how to upload specimen data. You can quickly import voucher data from a table in MS Excel by copying and pasting into the text area, provided that you use the right field headers.

### SEQUENCE UPLOAD

Upload vouchers instead

- You can upload batches of sequence data.
- Copy tab delimited data, with one row per sequence (including code), into the field below - see pre-printed example
- The first line should have the field names (code, gene, primers etc.). Gene, code and sequence must be listed for completion.
- This is only for adding data - not updating! If data is already present - the new data will be discarded (but listed as error)

Please edit your field headers names accordingly (not necessarily all or in that order):

Code Genecode Sequences Laborator Accession Primer1 Primer2 Primer3 Primer4 Primer5 Primer6 Creation\_date

Process data

Your input data: copy your tab-delimited data here

Something like this should be fine:

Code	genecode	Laborator	Primer1	Primer2	Sequences
10	COI	Tobias Malm	fly-Ci-J-1514	A2590	ATGATGATGATGATGATGATGATG
11	CAD	Tobias Malm	743nFi	1028Ri	GTGTGAGTGGTAGTGGTAGTGGTA

## 4.2 Adding genes

The second thing after adding vouchers that should be done is to create **new genes** or “alignments” for your database. This must be done in order to be able to **add sequences** to the database. You can add a single record by going to the **Administrator interface** and clicking on the link **Add/edit/view gene information**, followed by **Add gene**.

Delete me

Gene information			
Gene code	<input type="text" value="EF1a"/>	Aligned or not?	<input checked="" type="radio"/> yes <input type="radio"/> no
	Length (if aligned): <input type="text" value="1240"/>		
Description	<input type="text" value="elongation factor alpha"/>		
Notes:	<input type="text"/>		
Gene type:	<input type="radio"/> mitochondrial <input checked="" type="radio"/> nuclear <input type="radio"/> ribosomal <input type="radio"/> plastid <input type="radio"/> other:		
	<input type="text"/>		
Protein coding genes			
This section will only be applied if 'Protein coding' is set to 'yes'			
Protein coding	<input type="radio"/> yes <input checked="" type="radio"/> no	Reading frame	<input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Translation table	<input type="text" value="Standard"/>		
Introns			
Introns	Do you have introns in your alignment? Enter number and press "Update/Add introns" Then enter starting and ending nucelotide number for each intron region. To remove introns just select fewer and press button again.		
	<input type="text" value="2"/> <input type="button" value="Update/Add introns"/>		
Intron 1:	<input type="text" value="100"/> - <input type="text" value="200"/>	Intron 2:	<input type="text" value="300"/> - <input type="text" value="500"/>
<input type="button" value="Update gene"/>			

The most important and the only field that is obligatory is the **gene code** field, this will be the name of your gene when using the database. This could be a simple short version (e.g. **COI** for **Cytochrome oxidase I**) or any other name (no spaces allowed, but \_ are ok). You can for example create a gene code for aligned data, say the barcode version of COI of 658 bp could be named COI\_658 or similar. Other genecodes could be made for unaligned sequences (e.g. COI\_raw).

First the gene/alignment have to be **specified if aligned or not**. For example may raw sequences be set to **no** (and these may be retrieved as FASTA-files, whereas if you want to build other datasets (Nexus, PHYLIP, TNT) they need to be set to aligned. If set to **no**, then other information regarding reading frames and such will be ignored!

Then (if aligned) you should include the **length** of an aligned gene - this will be used for dataset creation and will there warn for sequences longer than the specified length.

You may also add a **description** for the gene - these should be the **full name of the gene** (e.g. **Cytochrome oxidase I**) - as this field is used for example in creating tables to submit to **GenBank**.

Aligned genes may be set as **protein-coding** for **additional prot-coding gene functionality** when retrieveing datasets for example (e.g. position choices, translation).

For aligned protein coding (=yes) genes you may choose to specify the **reading frame** as well as the genetic code for translation - this will be used for dataset creation and is a must f you want to partition your genes according to **codon positions** or **amino acids**.

**Introns** may be added - enter number of introns in your alignment and click 'update introns', that will give new fields for entering starting and finsihing positions for your introns. (Remember that positions in an alignment here is counted from 1 and upwards).

## 4.3 Adding sequences

In the **Administrator interface**, the tool **Upload batch sequences/vouchers** allows you to upload DNA sequences into VoSeq. Along with the DNA sequences, you have to upload the required fields **gene code** and **voucher code**, optionally the primer names, laborator and creation date. Each sequence and its related data goes into one line, with fields separated by tabs. If you have your data in a spreadsheet such as MS Excel, you can copy and paste the data into the text area.

- It is important to use the same headers provided in the text area.
- It is also necessary that the **code** of each sequence matches the **code** of voucher specimens that had been uploaded into VoSeq. This is the unique identifier that is used to connect the voucher data and their sequences.

(Aligned sequences should for best use of the database have missing data coded as questionmark (?) and gaps as a dash (-))

### SEQUENCE UPLOAD

**Upload vouchers instead**

- ☐ You can upload batches of sequence data.
- ☐ Copy tab delimited data, with one row per sequence (including code), into the field below - see pre-printed example
- ☐ The first line should have the field names (code, gene, primers etc.). Gene, code and sequence must be listed for completion.
- ☐ This is only for adding data - not updating! If data is already present - the new data will be discarded (but listed as error)

Please edit your field headers names accordingly (not necessarily all or in that order):

Code GeneCode Sequences Laborator Accession Primer1 Primer2 Primer3 Primer4 Primer5 Primer6 Creation\_date

**Process data**

Your input data: copy your tab-delimited data here

Something like this should be fine:

Code	genecode	Laborator	Primer1	Primer2	Sequences
10	COI	Tobias Malm	fly-Ci-J-1514	A2590	ATGATGATGATGATGATGATGATG
11	CAD	Tobias Malm	743nFi	1028Ri	GTGTGAGTGGTAGTGGTAGTGGA

## 4.4 Create taxonset

Taxonsets is a way to make a list of taxa that are being used for a specific project or analysis. A Taxonset is just a list of voucher codes. By having Taxonsets, you can quickly create datasets and tables for them.

If you have not set Taxonsets you will need to type specimen codes everytime you create a dataset. Instead, if you have a Taxonset for a particular project, you could select it when creating Tables for manuscripts.

Go to the **Administrator interface** and click on the link **Add/edit/view Taxon sets**.

A taxonset must have a name in order to be saved and usable later!

You can create a **Taxonset** by entering a list of specimen codes, each separated by a return:

Dataset information	
Taxonset name	<input type="text"/>
Taxonset creator	<input type="text"/>
Description	<input type="text"/>
<input type="button" value="Add dataset"/>	
Add a list of codes: tA1 tA2 and so on...	
<div style="border: 1px solid black; height: 100px; width: 100%;"></div>	

Or by browsing the data in VoSeq and choosing the specimens you are interested in by marking them in the X field:

Dataset information							
Taxonset name	<input type="text"/>						
Taxonset creator	<input type="text"/>						
Description	<input type="text"/>						
Show seqs:	<input type="checkbox"/> ArgKin <input type="checkbox"/> CAD <input type="checkbox"/> DDC <input type="checkbox"/> EF1a <input type="checkbox"/> GAPDH <input type="checkbox"/> IDH <input type="checkbox"/> MDH <input type="checkbox"/> Rp52 <input type="checkbox"/> Rp55 <input type="checkbox"/> template_COI <input type="checkbox"/> wingless						
Filter names:	<input type="text" value="Filter"/> Enter filter string here: <input type="text"/>						
Sort by:	<input type="checkbox"/> Order <input type="checkbox"/> Family <input type="checkbox"/> Subfamily <input type="checkbox"/> Genus <input type="checkbox"/> Species <input type="checkbox"/> Host org. <input checked="" type="checkbox"/> Code <input type="checkbox"/> X						
<input type="button" value="Mark all"/> <input type="button" value="Unmark all"/> <input type="button" value="Filter/Sort"/> <input type="button" value="Add dataset"/>							
X	Code	Order	Family	Subfamily	Genus	Species	Host org.
<input type="checkbox"/>	<a href="#">CP-B01</a>	Lepidoptera	Nymphalidae	Satyrinae	Brintesia	circe	
<input type="checkbox"/>	<a href="#">CP-B02</a>	Lepidoptera	Nymphalidae	Satyrinae	Penthema	darlisa	
<input type="checkbox"/>	<a href="#">CP01-33</a>	Lepidoptera	Nymphalidae	Satyrinae	Euptychia	spn2	
<input type="checkbox"/>	<a href="#">CP03-63</a>	Lepidoptera	Nymphalidae	Satyrinae	Manerebia	cyclopina	
<input type="checkbox"/>	<a href="#">CP03-70</a>	Lepidoptera	Nymphalidae	Satyrinae	Pronophila	thelebe	
<input type="checkbox"/>	<a href="#">CP03-82</a>	Lepidoptera	Nymphalidae	Satyrinae	Morpho	aurora	

Here you can sort the table according to chosen information (taxonomic level, code, X-marked or not), as well as choose genes to display information of. If you have chosen one or several genes, you can sort the table according to sequence availability for selected genes. You can also press **mark all** or **unmark all** to add or remove X's to or from each taxa that are displayed (works well with filtering). In order to perform a filtering or after selection of a new genecode you must press **Sort/Filter** to proceed. Your already marked taxa will be remembered.

After completing your selection of taxa and adding name and descriptions - press the **Add dataset** button to save it. If you are updating an already existing taxonset - press **Update taxon set**.



## 4.5 Create datasets

We believe that one of VoSeq's important features is the **capability to create dataset files of molecular sequences that are ready-to-run in phylogenetic software** such as MrBayes, TNT, PAUP, RAxML, etc.

Now that you have voucher and sequence data in your installation of VoSeq, you might want to create datasets for analysis of sequences in phylogenetic software.

In the **user interface**, you will find under the **Tools** section the link **Create new dataset**. You will be shown a page to select the sequences you want by entering the **voucher codes** and **gene codes**. You can select your data to be in several formats (FASTA, NEXUS, etc), choose between codon positions, as well as choosing what information your taxon names should include.

This will create a **ready-to-use** data set for analyses!

Enter the required info to make yourself a ready-to-run dataset

Choose file format:

☐ TNT format  
☐ NEXUS format  
☐ PHYLIP format  
☒ Unaligned FASTA format

Outgroup (code, for NEXUS and TNT):

Only for Protein coding:  
(others will be output as standard)  
Choose codon positions to use

☒ all ☐ 1st ☐ 2nd ☐ 3rd ☐ Special ☐ Amino acids  
 Partition by (positions): ☒ as one ☐ each ☐ 1st-2nd, 3rd  
 ⚠ Warning! your dataset will not necessarily be properly aligned!

(Override priority:  
Amino Acids->Special->All->1st-2nd,3rd):

What info do you want in the taxon names?

☒ Code ☐ Order ☐ Family ☐ Subfamily ☐ Tribe ☐ Subtribe  
☒ Genus ☒ Species ☐ Subspecies ☐ Auctor ☐ Host org. ☐ Gene code

For single gene datasets,  
exclude taxa missing this gene?

☒ yes ☐ no

For multigene datasets,  
exclude taxa with less than X genes?

Minimum number of genes:

Ignore introns?  
(yes' will not use them in the data set)

☒ yes ☐ no

Choose taxonset

Choose taxonset

...and/or a  
list of codes:

Enter one code per line  
 in the box below.  
 (with a -- sign before taxon  
 names to disable them from  
 the taxon set)  
 For example:  
 tA 1  
 S077  
 --S078  
 and so on...

Check to select  
your alignment/gene:

☐ ArgKin ☐ CAD ☐ COI ☐ DDC  
☐ EF1a ☐ GAPDH ☐ IDH ☐ MDH  
☐ RpS2 ☐ RpS5 ☐ wingless

Create dataset

The **Outgroup** field, if needed, should include the voucher code for the chosen outgroup taxa.

**Codon positions** Marking **1st**, **2nd** or **3rd** and unmarking **all** positions will create a dataset with only the chosen position(s) for all genes. **Special** will take you to a new page where you will be able to choose which codon positions to include for each gene. Marking **amino acids** will translate **protein-coding** genes with a set **genetic code**, the others will be treated as normal dna, i.e. making "mixed" datatype in Nexus for MrBayes, and setting partitions correctly in Nexus and PHYLIP format. Note that codon position choices as well as translation to aminoacids are only able to function if the user have specified a **reading frame** for the chosen gene(s) (see [Adding genes](#) (page 11)).

**Partition by (position)** Here you can choose how to do partitioning for each gene. **as one** will create one partition per gene, regardless of which codon positions you include. **each** will create a partition per codon position, whereas **1st-2nd, 3rd** will create one combined partition for the 1st and 2nd positions and one separate for the 3rd codon positions. Note that **each** and **all** are only possible to process with a per gene specified **reading frame** (see [Adding genes](#) (page 11)).

You can also chose to **omit taxa from a taxonset that contains less than a specific number sequences**. Say you have a 10 gene data set and want to remove all taxa with 5 or less! Easy! Just enter a minimum number of genes!

If you have introns in your alignment you can choose to include or remove them from the output data set. If included they will be treated as separate data blocks and partitions for the Nexus and PHYLIP outputs!

The voucher codes can be entered one by one (separated by return) in the text area or you could create a [Create taxonset](#) (page 13) (a list of voucher codes for a specific project).

**Note:** As of version 1.5.0, protein-coding ability, aligned or not, introns and genetic code will be set for each gene/alignment in the admin gene section!

## 4.6 Search

You can search for records by queries using single fields or any combination of them. The autocomplete dropboxes will help you query existing data easily. This can be done in both the **user interface** and the **administrator interface** - where the latter have more options to search (e.g. record history).

The screenshot displays the 'Search for records' interface. On the left, a grid of taxonomic filters includes 'Order', 'Family', 'Genus' (with an autocomplete dropdown showing 'Caerois', 'Calisto', 'Celastrina', 'Curetis', 'Cyllogenes', and 'Cyrestis'), 'Subspecies', 'Subfamily', 'Tribe', 'Subtribe', 'Host org.', and 'Type species' (with 'Yes' and 'No' radio buttons). Below these are 'Country', 'Specific Locality', 'Latitude', 'Longitude', and 'Altitude'. On the right, the 'DNA' section contains 'Extraction', 'Tube', 'Extractor', and 'Date' fields. The 'Sequence Information' section includes 'Gene Code', 'In Genbank?' (with 'Yes' and 'No' radio buttons), and 'Accession'. At the bottom, the 'Publication and Notes' section is partially visible with a 'Published in' field.

## 4.7 Upload voucher photos

In the **Administrator interface** you will see that some records have the link **Picture missing**. By clicking on this link, you will be able to upload a photo for that voucher.



If you want to replace an existing picture with another, you will need to click the **Change picture** icon.

After you upload your photo, VoSeq will automatically **post the picture in Flickr** and save the necessary URL addresses in the MySQL database. Thus, you will see your photo in the corresponding voucher page.

If you have not enabled the *Flickr plugin* (page 21), VoSeq will instruct you how to do this. However, it is also possible to host your photos locally on your own server or computer. For this you will need to edit a line in your `conf.php` file:

- Change the line:
  - `$photos_repository = 'flickr';` to this one:
  - `$photos_repository = 'local';`

## 4.8 Create Excel table

You can create a MS Excel table with specimen codes, genus and species names, genes used in analysis along with their accession numbers.

Go to the **User interface** and under the **Toolbox** click on the link **Create MS Excel table**

Instead of typing your specimen codes in the text area below, you could select a Taxonset (provided that it has been set before (*Create taxonset* (page 13))). This table will be ready to attach to a manuscript for publication.

You can also change the way sequence information is displayed in the table by choosing between **number of bases** (displays number of bases - does not count questionmarks `?`), **accession numbers** (displays stored accession numbers instead of sequence length) or **X/-** (displays **X** if sequence is present and **-** if sequence is missing).

**Display missing sequence beginnings/ends with star(\*)?:** will show search for questionmarks (`?`) in the beginning or end of the sequences (when displayed by number of bases) and show if the sequence misses bases in those positions with an asterisk (`*`). Easy then to see during laboratory phase then where sequence information might be missing for your taxa.

You may also change between comma (,) and tab-delimited table mode.

**CREATE TABLE:**

Enter the required info to make yourself a table in MS Excel format

<b>Voucher info:</b>	<input checked="" type="checkbox"/> Code <input type="checkbox"/> Order <input type="checkbox"/> Family <input type="checkbox"/> Subfamily <input type="checkbox"/> Tribe <input type="checkbox"/> Subtribe <input checked="" type="checkbox"/> Genus <input checked="" type="checkbox"/> Species <input type="checkbox"/> Subspecies <input type="checkbox"/> Host org.
<b>Locality and collector info:</b>	<input type="checkbox"/> Country <input type="checkbox"/> Locality <input type="checkbox"/> Collector <input type="checkbox"/> Coll. date <input type="checkbox"/> Altitude <input type="checkbox"/> Latitude <input type="checkbox"/> Longitude
<b>Choose what gene info to display:</b>	<input checked="" type="radio"/> Number of bases <input type="radio"/> Accession number <input type="radio"/> X/- (exist/empty)
<b>Choose your field delimiter:</b>	<input type="radio"/> comma <input checked="" type="radio"/> tab Display missing sequence beginnings/ends with star(*)?: <input type="radio"/>

<b>Choose ready-made taxonset</b>	Choose taxonset ▼
-----------------------------------	-------------------

...and/or a list of codes, one code per line:	For example: tA1 S077 and so on...	<b>Check to select your Gene codes:</b>	<input type="checkbox"/> ArgKin <input type="checkbox"/> CAD <input type="checkbox"/> DDC <input type="checkbox"/> EF1a <input type="checkbox"/> GAPDH <input type="checkbox"/> IDH <input type="checkbox"/> MDH <input type="checkbox"/> RpS2 <input type="checkbox"/> RpS5 <input type="checkbox"/> template_COI <input type="checkbox"/> wingless
<div style="border: 1px solid black; height: 40px; width: 100%;"></div>			

## 4.9 Update voucher

When you **click on the code** of an already existing voucher in the **administrator interface** you will be transferred to its **voucher information page**.

Here you may make changes to all the fields - and these will be updated after pressing **Update record**.

A changed **voucher code** will automatically change the code in the connected fields for sequences and primer informations, so as to keep them connected.

There is also a **record history** displayed for administrators that list what changes have been made to the voucher information previously, with time and the user responsible for the changes.

## CP07-04

Specimen name		
Order	<input type="text" value="Lepidoptera"/>	Subfamily
Family	<input type="text" value="Nymphalidae"/>	Tribe
Genus	<input type="text" value="Morpho"/>	Subtribe
Species	<input type="text" value="sulkowskyi"/>	Host org.
Subspecies	<input type="text"/>	Type species? <input type="radio"/> Yes <input checked="" type="radio"/> No


Locality Information		
Country		
<input type="text" value="PERU"/>		
Specific Locality		
<input type="text" value="Oxapampa"/>		
Latitude ?	Longitude ?	Altitude
<input type="text" value="-10.666600"/>	<input type="text" value="-75.500000"/>	<input type="text"/>

Collector Information		
Code in VoSeq	Collector	Date (yyyy-mm-dd)
<input type="text" value="CP07-04"/>	<input type="text"/>	<input type="text"/>
Voucher Locality	Voucher ?	Voucher Code
<input type="text" value="MUSM, Lima"/>	<input type="text" value="Voucher photo"/>	<input type="text"/>
Sex	<input type="radio"/> Larva <input type="radio"/> Male <input type="radio"/> Female	

DNA	
Extraction ?	Tube ?
<input type="text" value="1007"/>	<input type="text" value="4"/>
Extractor ?	
<input type="text" value="Niklas Wahlberg"/>	
Date (yyyy-mm-dd)	<input type="text" value="2007-04-2"/>

Publication and Notes	
Published in	Notes
<input type="text"/>	<input type="text" value="coll fresh, Exemplar"/>

Publication and Notes	
Record history	
Latitude, longitude edited by Ad ministrator on 2012-03-08 14:53:00 IDH sequence added by Carlos Pena on 2012-03-07 16:10:34 RpS5 sequence added by Carlos Pena on 2012-	

Update record

## UPDATE VOSEQ

The easiest way to update VoSeq (that does not require new install of software or database):

- Download the new files from [github](https://github.com/carlosp420/VoSeq/tags) (<https://github.com/carlosp420/VoSeq/tags>).
  - **Unpack** the new files to your **webserver directory** (htdocs, webserver, etc).
  - **Rename your old “in use” VoSeq folder** something like, ‘VoSeq\_old’ or similar (e.g. “VoSeq-1.4.4” -> “VoSeq\_1.4.4\_old”).
  - **Give the newly downloaded VoSeq folder the same name as the old one had** (e.g. “VoSeq-1.4.4”).
  - **Copy the file “conf.php”** (in main folder) from the old version (e.g. “VoSeq\_1.4.4\_old”) to the new version (e.g. “VoSeq-1.4.4”).

Also:

- If you have used and installed blast files, make sure to copy the files **Blastdb\_aliastool**, **Blast**, **Makeblastdb** and **Makemindex** (.exe for all in windows), from the old version (blast/bin folder) to the new version (same folder). Then **set permissions** to read, write, and execute on the folder “blast/bin” and its content, as well as the folder “include/blast” (e.g. `chmod 777 -R path_to/~VoSeq_folder/blast/bin`).
  - For Mac users it may work better to use the `sudo chown -R _www VoSeq_folder` command instead, since files belong to user instead of root!
- **If you have voucher photos stored**, transfer them from the old one to the new one also (in pictures folder).

# PLUGINS

## 6.1 Yahoo! Maps plugin

**VoSeq** is able to interact with Yahoo! Maps to create on-the-fly maps for vouchers when geographic coordinates are present in voucher pages. After installing VoSeq, you can enable this capability by getting a **Yahoo! Maps API key** from them and writing them in your `conf.php` file:

1. Get an API key from <http://developer.yahoo.com/maps/simple/>
2. After filling in the required information you will be given a **Consumer Key** consisting of a long string of seemingly random characters that end with two dashes:

- MwRGV2Jm1zbWNHbmM9Y2Q9WVdrOVVHdj0yzlNQS0tJ9uc3VtZXJzZWNYZXQmeD1iMw--

3. Remove the two dashes from the end and copy your key into the `conf.php` file as a value for the variable `$yahoo_key`. Like the example below, including quotations and semicolon:

- `$yahoo_key = "MwRGV2Jm1zbWNHbmM9Y2Q9WVdrOVVHdj0yzlNQS0tJ9uc3VtZXJzZWNYZXQmeD1iMw--";`

4. Save the file and exit.

After doing this, VoSeq will be able to pull maps from Yahoo! whenever there is geographic information in your database. Note that you need to enter the geographic coordinates into VoSeq converted into decimal format, using the sign minus for the Southern and Western hemispheres.

## 6.2 Flickr plugin

**VoSeq** hosts all the specimen photos in [Flickr](http://www.flickr.com/) (<http://www.flickr.com/>). If you have a free account you can host up to 200 photos. The Pro account allows you hosting unlimited number of photos for a yearly fee (25 USD).

1. You need to get an API key from Flickr.
2. Create an account in [Flickr](http://www.flickr.com/) (<http://www.flickr.com/>) (if you don't own one already)
3. Go to <http://nymphalidae.utu.fi/cpena/VoSeq/>
4. Follow the instructions to get an **API key**, **Secret key** and **Token key**.
5. After submitting you will get your **Key**, **Secret** and **Token**. Write down those keys.
6. From a text editor software, edit the file `conf.php` by copying your keys in it.
7. For example [these are not real keys and will not work if you use them]:

- `$flickr_api_key = "2d7f59f9aaa2d5c0a2782d7f5d9083a6";`
- `$flickr_api_secret = "ef0def0f3d5f3f15f1";`
- `$flickr_api_token = "61607157718372495-f5524ead33b43129";`

8. Save and exit.

Thus, every picture that you upload into your VoSeq installation will be uploaded into your Flickr account.

**Note:** You can share your voucher photos with the Encyclopedia of Life. *Sharing Photos with EOL* (page 23)

## 6.3 BLAST plugin

VoSeq has **BLAST capabilities** (<http://en.wikipedia.org/wiki/BLAST>).

You can search for homologous sequences of your markers in GenBank. If you have a VoSeq installation in your work computer (or your server provider allows you to run the BLAST executable files), you do local BLASTs. For example, BLAST any of your sequences against all sequences of the same gene, or against all your sequences (full BLAST). Click on the “BLAST” icons in your voucher’s pages:

The screenshot shows the VoSeq web interface for specimen CP-B01, *Brintesia circe*. The page is divided into several sections:

- Specimen name:** A table with taxonomic details:
 

Order	Lepidoptera	Subfamily	Satyrinae
Family	Nymphalidae	Tribe	Satyrini
Genus	Brintesia	Subtribe	Satyrina
Species	circe	Host org.	
Subspecies		Type species?	Yes
- DNA:** A table with DNA data:
 

Extraction	Tube
1999	230
Extractor	
Carlos Peña	
Date	2004-07-19
- Locality Information:** A table with locality details:
 

Country	FRANCE	
Specific Locality	Aude, Villegly	
Latitude	Longitude	Altitude
43° 17' N	2° 27' E	
- Collector Information:** A table with collector details:
 

Code in VoSeq	Collector	Collection date
CP-B01	Niklas Wahlberg	2004-06-26
Voucher Locality	Voucher	Sex
SU	spread	
- Sequence Information:** A table with sequence data and BLAST results:
 

Region	bp	Amb	Lab	Accession	Local Blast	ncbi Blast
ArgKin	596	0	Niklas	EU141251		
CAD	850	0	Niklas	EU141291		
COI	1487	51	Carlos	DQ338865		
DDC	373	0	Niklas	EU141458		

On the right side, there is a map of the location, a "Share photo with EOL" button, and a "VoSeq" logo. At the bottom right, there is a "TOOLS:" section with a link to "Blast new sequence".

You can also copy and paste any new sequence into VoSeq’s **Blast new sequence** tool and see whether there are any similar sequence in your data (this tool is located on the sidebar on the right).

Remember that you need to download from NCBI the stand alone BLAST executable files and copy/install them in one of VoSeq’s folders:

- In Mac OS X: when you install from the .DMG package, the executable files will be written in the folder: `/usr/local/ncbi/blast/bin`. You just need to copy them to the right folder in VoSeq:
  - `mkdir ~/Sites/VoSeq/blast/bin`
  - `cp /usr/local/ncbi/blast/bin/* ~/Sites/VoSeq/blast/bin/.`
- In Linux: `/path/to/your/VoSeq/blast/bin/`
- In Windows: `C:\Program Files\Apache Software Foundation\Apache2.2\htdocs\VoSeq\blast\bin\`
- It is important that the executable files are placed inside the folder **bin**.



## 6.4 Integration with EOL

1. VoSeq makes it easy to share your voucher photos with EOL. More information here [Sharing Photos with EOL](#) (page 23).
2. VoSeq makes automated calls to EOL's web services for pulling information on authors and date of description for species. VoSeq sends genus and species names and waits for a response. If EOL response is positive, the full species name will be included in voucher pages:



## 6.5 Sharing Photos with EOL

VoSeq makes it easy to share your voucher photos with EOL. You can submit your best photos to EOL from VoSeq with just one click.

If you haven't done it already, you need to create an account in Flickr. Then log in to Flickr with your account and join the EOL group:

1. Go to [http://www.flickr.com/groups/encyclopedia\\_of\\_life](http://www.flickr.com/groups/encyclopedia_of_life)
2. Click **“Join This Group”**

Be aware that EOL requires that your photo is under any of the following licenses:

- Creative Commons Attribution (**CC-BY** (<http://www.flickr.com/creativecommons/>))
- Creative Commons Non-Commercial (**CC-BY-NC** (<http://www.flickr.com/creativecommons/>))
- Creative Commons Share-Alike (**CC-BY-SA** (<http://www.flickr.com/creativecommons/>))
- Creative Commons Non-Commercial Share Alike (**CC-BY-NC-SA** (<http://www.flickr.com/creativecommons/>))

In your VoSeq installation, you will see a button:



under your voucher photos. If you click this button, VoSeq will add a “machine tag” to the corresponding page in Flickr so that it can be harvested by EOL.

Thus you will be able to see your photo in EOL's pool of photos in their Flickr account [http://www.flickr.com/groups/encyclopedia\\_of\\_life/pool/with/4096153224/](http://www.flickr.com/groups/encyclopedia_of_life/pool/with/4096153224/)

EOL harvests the photos quite frequently, so in one day or two, you will be able to find your photo in the respective page in EOL.

## 6.6 Integration with GBIF

### 6.6.1 You can share your information hosted in VoSeq with GBIF

GBIF prefers data owners to use their [Integrated Publishing Toolkit \(IPT\)](http://www.gbif.org/informatics/infrastructure/publishing/#c889) (<http://www.gbif.org/informatics/infrastructure/publishing/#c889>). This means that you can install their IPT software to produce a resource in Darwin Core format that can be harvested by GBIF. In addition to the actual data in your VoSeq installation, IPT allows you to include a rich variety of metadata for GBIF.

VoSeq is able to produce a dump file containing all the data you own. Then you can import this file into a IPT installation and choose which types of data you want to publish via GBIF. Once you include all the metadata required by GBIF you have two choices in order to expose your data taken from [GBIF website](http://www.gbif.org/informatics/standards-and-tools/publishing-data/) (<http://www.gbif.org/informatics/standards-and-tools/publishing-data/>):

- By setting up a dynamic server software:
  1. Acquire hardware with a permanent Internet connection (a regular PC is sufficient).
  2. Install data publishing software. GBIF recommends the Integrated Publishing Toolkit (IPT). You will need a web server such as Apache.
  3. Configure the software for your local data structure; this is the ‘mapping’ process. Please follow the documentation of your chosen publishing software for this process.
  4. Register your service with GBIF and sign the GBIF Data Sharing Agreement.
  5. Create an archive for your entire dataset:
    - This scenario doesn’t require a permanent Internet connection. You simply need to create a Darwin Core Archive, upload it to a repository (for example an IPT repository installed by your GBIF Participant Node, an institutional FTP or web server, or a service like Dropbox or the Internet Archive). You then just need to register the public URL for the storage location of your archive with GBIF.

### 6.6.2 Create a dump file and use in an IPT installation

1. You can create a dump file with all the data in your VoSeq installation for submitting to GBIF. In VoSeq, on the sidebar on the right, click on **Share data with GBIF**. Save this file and open an installation of IPT.
2. In IPT, click on **Manage Resources** in the top menu, enter a name for your resource and click **Create**. Note: do not upload your file in this page (it will fail to recognize your tab delimited dump file).

GBIF free and open access to biodiversity data GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

Logged in as

Home Manage Resources Administration About

## Resources you have rights to manage

None, please create a resource or ask existing managers to add you to their resource(s).

### Create New Resource

You can create a new blank resource or upload an existing resource saved as a zipped Darwin Core archive. A short name is required within this IPT installation. Please use only alphanumeric characters, hyphens, or underscores.

Shortname

myTestResource

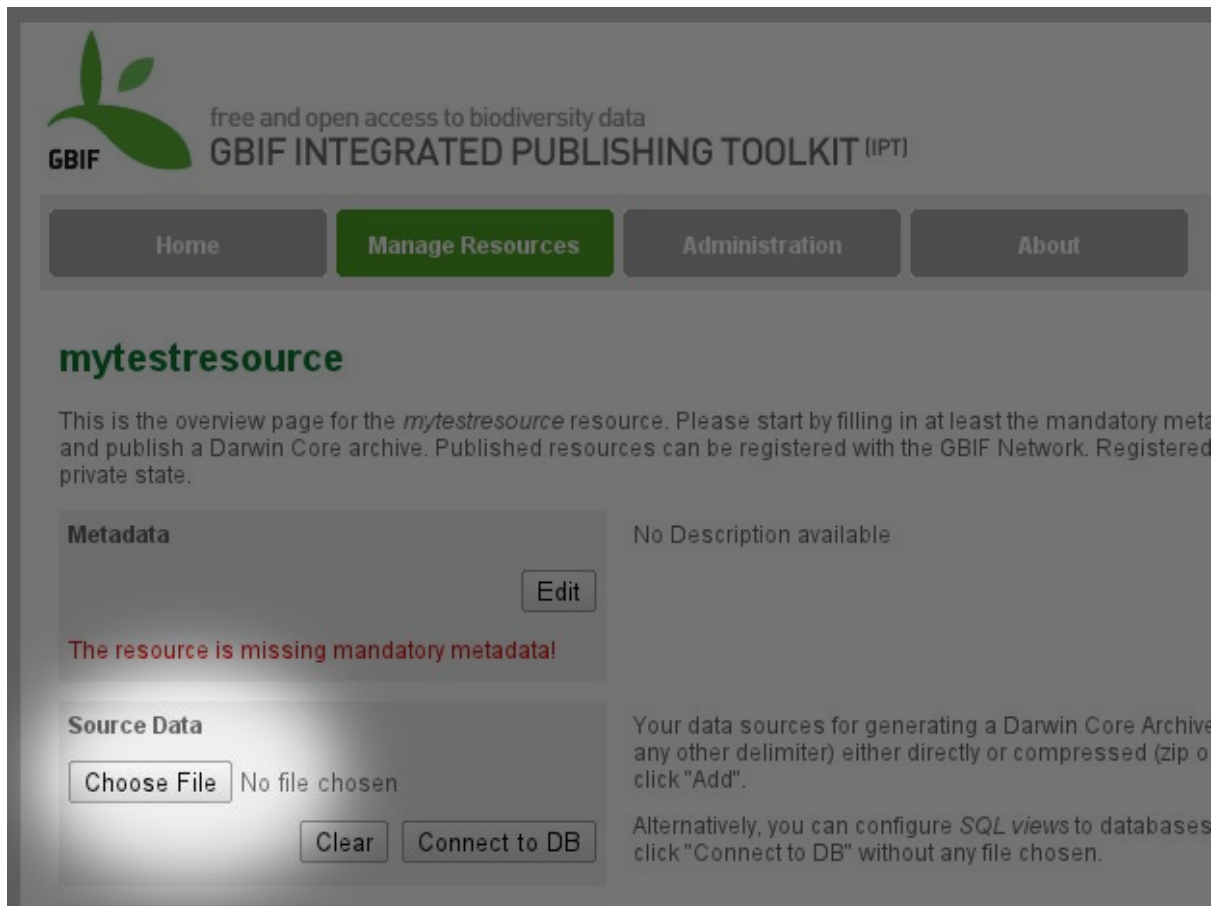
Optional archived resource to load

Choose File No file chosen

Create

Version 2.0.3-r3672 [About the IPT Project](#) [User Manual](#) [Report a bug](#) [Request new feature](#)

1. You will be directed to your test resource page. This is when you upload the dump file generated in VoSeq. And then you are ready to add a rich variety of metadata to your resource and become a provider of information to GBIF.



The screenshot shows the GBIF Integrated Publishing Toolkit (IPT) interface. At the top, the GBIF logo is on the left, followed by the text "free and open access to biodiversity data" and "GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)". Below this is a navigation bar with four buttons: "Home", "Manage Resources" (highlighted in green), "Administration", and "About".

The main content area is titled "mytestresource" in green. Below the title, a paragraph states: "This is the overview page for the *mytestresource* resource. Please start by filling in at least the mandatory metadata and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be set to public or private state."

There are two main sections on the left:

- Metadata:** Contains an "Edit" button. Below it, a red error message reads: "The resource is missing mandatory metadata!".
- Source Data:** Contains a "Choose File" button, the text "No file chosen", a "Clear" button, and a "Connect to DB" button.

On the right side of the page, there are two text blocks:

- "No Description available"
- A paragraph explaining data sources: "Your data sources for generating a Darwin Core Archive (or any other delimiter) either directly or compressed (zip or tar) can be added by clicking 'Add'." Below this, another paragraph states: "Alternatively, you can configure *SQL views* to databases by clicking 'Connect to DB' without any file chosen."

## FAQ

### 7.1 Installation FAQ:

**Q.- During installation I get the error 2002: “Can’t connect to local MySQL server through socket ....bla bla bla...”**

- VoSeq is trying to connect to MySQL using a file called **socket**. This error occurs when PHP tells VoSeq to find the socket in a folder where it is not placed. This can be fixed by telling MySQL to put the socket as the file `/tmp/mysql.sock` and by telling PHP to find it there and not to look for it in any other folder.
- From the installation folder of **PHP**, save the file `php.ini-development` in the folder `/usr/local/lib/` and name it `php.ini`
- Edit your file `php.ini` and look for the command `mysql.default_socket` and make sure it says:
  - `mysql.default_socket = /tmp/mysql.sock`
- Edit your MySQL installation file `/usr/local/mysql/support-files/my-large.cnf`:
  - File parameters: modify the lines `socket = /var/mysqld/mysqld.sock` to `socket = /tmp/mysql.sock`
  - Save the file as `/etc/my.cnf` and `/etc/mysql/my.cnf`
- Restart the server and resume the installation of VoSeq.

### 7.2 POST request error

**Q.- ...my computer complains that “The requested resource /VoSeq\_XXX/somefile.php does not allow request data with POST requests, or the amount of data provided in the request exceeds the capacity limit.”?**

**A.-** Open the PHP config file (see below “How to find PHP.ini”) and increase the value for `POST_max_size`, save file and restart webserver.

### 7.3 Execution time error

**Q.- ...my computer stops a VoSeq page from running due to execution timeout?**

**A.-** Open the PHP config file (see below “How to find PHP.ini”) and increase the value for `max_execution_time`, save file and restart webserver.

### 7.4 Too many variables problem

**Q.- ...my huge taxonsets or other lists doesnt include all the values I had marked and added for them?**

A.- PHP may have set a too low value to `max_input_vars`. Open the PHP config file (see below “How to find PHP.ini”) and increase the value for `max_input_vars`, save file and restart webserver.

## 7.5 Mac permission problem

Q.- ...if for example BLAST, storing pictures etc dont work on Mac!

A.- It happens specially when upgrading VoSeq. When you download a fresh copy of VoSeq from Github and copy the contents on your installation of VoSeq, it happens that all the files and folders have you as **owner**. So, VoSeq (and the Apache server) cannot write into the folders. To fix this it is necessary to set the Apache server as the **owner** of files and folder. In my MacBook the id for the Apache “user” is `_www`. So we need to do the following to transfer ownership of files and folders of VoSeq to the server: Open a terminal or console and use the command: `sudo chown -R _www VoSeq_folder`. This should give the permissions to VoSeq (actually the Apache server) to do this things!

## 7.6 How to find php.ini and see your PHP settings

Place a file named **info.php** containing `<?php phpinfo(); ?>` in your web server folder where you have your VoSeq folder. Open your browser and go to that file/page (ie. **`http://localhost/info.php`** for win/linux or **`http://127.0.0.1/~yourprivatefoldername/info.php`** for mac). This should get you the PHP config output, where you can find “Configuration File(`php.ini`) Path” and “Loaded Configuration File”. These fields should tell you where your config file (`php.ini`) is located. If these says “(none)” see below.

- Windows - In windows the PHP configuration file (`php.ini`) should be found under `C:Windows/`. If it's not there then copy the `php.ini-???` to `C:\WINDOWS` and rename it `php.ini`. (??? can be dist, production or development).
- Mac - on mac the the PHP configuration file (`php.ini`) should be found under `/private/etc/`. If no `php.ini` is found there but a `php.ini.default` is, run `sudo cp /private/etc/php.ini.default /private/etc/php.ini` in terminal create a `php.ini` file. Then restart server.
- Linux - Open a terminal or console and type `locate php.ini`. In my server I got this location: `/usr/local/lib/php.ini`