Python

First worked in «*Introduction to relational databases*» online course (2016).
Project summary:

"Writing a Python module that uses the PostgreSQL database to keep track of players and matches in a game tournament. The game tournament will use the Swiss system for pairing up players in each round: players are not eliminated, and each player should be paired with another player with the same number of wins, or as close as possible. This project has two parts: defining the database schema (SQL table definitions) and writing the code that will use it."

In 2017, I also took some private classes from Diana Kramer (Senior Security Engineer at King) while studying my degree, to learn basics of the language, object-oriented programming, basic python packages, etc.
Contact: Diana Kramer (**dia.kramer@gmail.com**)

Worked in 2018-2019 during a three months group project.
Project available at: https://github.com/olfatk96/PYSBI_Constructor
The project consisted on building a 3D protein modeller in python. The input where pdb files with at least two chains from a protein, and the program had to analyse and index them, then add chains in other of the files in the best position (considering molecular coordinates, chemical properties and RMSD), until all the pdbs from input were considered.

** *Also worked in Python Notebook during the MSc*

R

First used in 2015, during a degree project.
Project summary:
"Ecology subject. Analysis of the use of resources of two species of salamanders"
Used since then for all statistical analyses.

Used in 2019 for a group project in the MSc.
Project available at: https://github.com/olfatk96/IEO_project_UCEC
The project consisted on analysing a dataset of RNA-Seq results from a cancer cohort. The subjects were female patients of Uterine Corpus Endometrial Carcinoma (UCEC), and a quality assessment, differential expression analysis and functional analysis was performed on the aligned reads.
Contact: Robert Castelo (associate professor of Bioinformatics and Biostatistics at the Universitat Pompeu Fabra, Barcelona (Spain) and head of the Functional Genomics group of the IMIM-UPF joint Research Programme on Biomedical Informatics (GRIB).)
robert.castelo@upf.edu

Currently using R in the MSc thesis (2019-2020).
I'm working as a bioinformatician in the forensics department of the Medicine Faculty of KU Leuven. Due to confidential agreement, I cannot share information about the data or the objective of the project, but some technical specifications of the work done since August 2019:

- Reproduce results from previous year master student on a data set, improving an ANOVA test with Dunnett post hoc test, run on more than 10000 features and 14 individuals, reducing its running time from 3 days to 5h.
- Quality assessment of the RNA-seq results of the main dataset I am going to work with (filtering low expressed genes, normalizing values, etc.)
- Set out basic structure of a model I want to develop as a master thesis (until Forrester diagram until the moment)

Octave

3 months online course at Coursera, certified by Standford University.
Credential number: 753NYGVGT6XA
Project available at: https://github.com/laurag-antiga/Machine_Learning_Course
The course was an advanced Machine Learning overview. Different function implementation and optimization in Octave (like linear regression, logistic regression and Neural Networks models), and their cost functions and methods to minimize it.
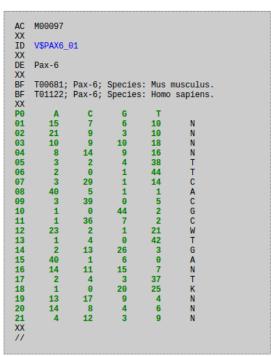Also saw advanced mathematics, like vectorization of matrix operations.

Perl	First worked in 2016 in the Genetics degree. Worked with MySQL in relational databases. Main objective on regular expression application on how to search information in databases and in files obtained from them.

Also worked in a 2 months projects of the MSc (2019). Evaluation depended on the difficulty of the project chosen and I chose the most difficult (the maximum grade worth).
Project:

Getting the Topmost Scoring Sequences from Position Weight Matrices

Position Weight Matrices are a simple way to model signals appearing on DNA and protein sequences. They summaryze the frequencies of every letter of the nucleotide or amino acid alphabets at a given position of the signal, for instance a splice site or a transcription factor binding site. This means that we can obtain a score, either from the absolute frequencies or by transforming them into log-likelihoods, for a given sequence to determine if it contains the signal pattern or not. Yet, they can also be used as sequence generators. The idea is to write a Perl script to produce all the possible sequences a PWM can match and report them ranked by score, for the shake of simplicity, we will work with nucleotides. On some cases, it can be unfeasible to produce all the possible sequences; the worst scenario is a PWM for a random sequence, where the score of any of the four nucleotides is 0.25, as the number of posible sequences will be 4^n (being n the length of the matrix). For instance a matrix for a nucleotide signal made of 10 nucleotides, $n=10$, can generate $4^{10} = 1048576$ different sequences. As we do not want to fill the disk with too many sequences, we can set a cut-off, say here M=10000 sequences. However, we still must produce only the M sequences that have the highest scores. The simplest way to achieve that is to use a secondary PWM where the positions are reordered by the score of the most frequent nucleotide per position, as well as an array with the original positions, taken as scalar values, recording the order of the new PWM. The best approach can be a recursive function that start generating new sequences iterating over the nucleotides per position sorted from higher frequencies to lower ones. The output records should contain two fields, the generated sequence and the corresponding score, both can be calculated simultaneously. Finally, the script should read a text file containing the PWM in TRANSFAC format, having a section with five fields to store in memory (the ones marked in green, for the position and the relative frequencies for A, C, G and T, in that order). Below you can find two matrices that you have to run separately through your program. Discuss the results obtained from them on the report.

```
AC  M00034
XX
ID  V$P53_01
XX
DE  tumor suppressor p53
XX
BF  T00671; p53; Species: Homo sapiens.
BF  T01806; p53; Species: Mus musculus.
XX
P0      A     C     G     T
01      4     0    13     0      G
02      5     0    12     0      G
03     15     0     2     0      A
04      0    17     0     0      C
05     17     0     0     0      A
06      0     0     0    17      T
07      0     0    17     0      G
08      0    13     0     4      C
09      0    17     0     0      C
10      0    17     0     0      C
11      0     0    17     0      G
12      0     0    17     0      G
13      2     0    15     0      G
14      0    17     0     0      C
15     17     0     0     0      A
16      0     0     0    17      T
17      0     0    17     0      G
18      0     2     0    15      T
19      0    13     0     4      C
20      0     7     2     7      Y
XX
//
```

```
AC  M00097
XX
ID  V$PAX6_01
XX
DE  Pax-6
XX
BF  T00681; Pax-6; Species: Mus musculus.
BF  T01122; Pax-6; Species: Homo sapiens.
XX
P0      A     C     G     T
01     15     7     6    10      N
02     21     9     3    10      N
03     10     9    10    18      N
04      8    14     9    16      N
05      3     2     4    38      T
06      2     0     1    44      T
07      3    29     1    14      C
08     40     5     1     1      A
09      3    39     0     5      C
10      1     0    44     2      G
11      1    36     7     2      C
12     23     2     1    21      W
13      1     4     0    42      T
14      2    13    26     3      G
15     40     1     6     0      A
16     14    11    15     7      N
17      2     4     3    37      T
18      1     0    20    25      K
19     13    17     9     4      N
20     14     8     4     6      N
21      4    12     3     9      N
XX
//
```

GRADES	Basic implementation (*working bug-free script*)	8.0 pt
	Using functions	+ 1.0 pt
	Report describing implementation and results	+ 1.0 pt

Project available at: https://github.com/laurag-antiga/Perl

MySQL	As mentioned, worked in an online course in 2016 (with python) and during Perl and Bioinformatics subjects during the degree (with perl).

Also used in an internship in 2017, while staying at the anthropological forensics department of the University Autonomous of Barcelona, where my task consisted on creating a database to store and manipulate the information of the skeletons they store in the department.

Web servers	php, html and CSS used to develop a web server in 2018 MSc project. Group project. Tasks: develop a database to store antibodies information and a web server to search, add and delete information from that database. The webserver connects to the database, and not only stores users' information, but also was linked to other webpages related to antibodies, such as PubMed or UniProt. The webserver was divided in two hierarchical levels: admins and users. Being admins the only users with deleting options. Thought to serve as a department database to store each laboratory's antibodies' information (which have been used, in which experiment, if it was successful, plus all the biological information about the antibody). Separate entities don't share antibodies information. **Project available at**: https://github.com/laurag-antiga/DasLab (documentation at the moment)
Galaxy	Used eventually during the degree. Main project: Locate and characterize CpG islands of a given sequence (2015)
Copasi	Used both in the degree and MSc (Systems Biology Subject, in both cases), to analyse chemical pathways, it's components, and the differential equations behind. Also worked the mathematics behind the behaviours in excel, to understand them step by step (Michelis-Menten equation and derivation, kernel optimization, etc.) In the same subject, for the MSc, a Network model was developed in python and R, in pajek format. **Project available at**: https://github.com/laurag-antiga/Sistems-Biology
Excel	Main program used while working at Boxwedl, in 2016-2019 summers. Mostly database management duties. The database is in excel format, storing clients' names (companies), webpages, telephone numbers, mail addresses and location. **Contact:** Eduard Esquerra (info-bcn@boxwedl.com) In 2016-2018, while I was working as a teacher in Ítaca, it also was the main software used there. Used to keep track of students' improvement, homework and skills with the other professors. **Contact:** Miguel Molina (Head of Ítaca academy) (miquelcdv@gmail.com)
Git and Github	Online course at Udacity: "How to use Git and Github" (2019) - In process