Luc Blassel

PhD Student, interested in Machine learning in Biology

luc.blassel@gmail.com lucblassel.com

Paris, France

Passionate about machine learning applied to biological problems, I did a PhD at the Institut Pasteur in Paris. I am currently a post-doctoral reseacher using deep attention networks to study evolution.

Experience

Post-Doctoral Researcher, CNRS, Paris, France 2023–

During this postdoc I will work on the phyhloformer project with Laurent Jacob and members of the LBBE in Lyon. I am working on improving performance of the model as well as it's generalisation. Extensions to the model are planned, such as taking unaligned sequences as input and including a differentiable tree-building step in an end-to-end pipeline.

PhD Student, Institut Pasteur, Paris, France 2020–2023

PhD in computational biology, focused on improving and learning from sequence alignments. With the ever growing quantity of high quality sequencing data, we are now able to develop methods to take advantage of this data. My PhD project followed 2 different axes: (1) Using machine learning to discover Drug resistance mutations in HIV (Continuation of the M.Sc project) (2) Search of a function space for sequence transformation functions that improve sequence mapping and alignment.
This PhD was supervized by Rayan Chikhi and funded by the PRAIRIE institute. The manuscript can be found here

Research Engineer, Institut Pasteur, Paris, France 2018–2020

Using machine learning and large datasets to explore the drug resistance landscape in HIV. Supervized by Olivier Gascuel. A large dataset of UK HIV sequences associated with patient RT-treatment status was used to train several machine learning algorithms. Examining the importatnt learned features gave us insight in mutations associated to ART failure. This project resulted in a publication.
Project page here, and M.Sc thesis (in french).

Data Scientist Intern, CapitalData, Paris, France Mar-Sept 2017

As a data scientist intern at Capital data I was tasked with developing a software pipeline to validate data prior to database integration, and detect anomalies usign time series. This was implemented in R and Python. I also, developed reporting-oriented data visualizations of KPIs using BIME and Tableau. Finally statistical AB-testing was also done using python.

Teaching and tutoring, self-employed, Paris, France 2015–2022

Since the beginning of my higher education, I have beedn tutoring middle school, high school and university students in scientific subjects, mainly mathematics. The aim is to turn the students into independents people, more confident in their abilities and more knowledgeable in order to excel in their respective studies.
Since Septemeber 2021, I have been doing oral examination in English for PCSI preparatory class students as the ENCPB in Paris, France.

Education

Masters of Science (M.Sc.), Dep. MIDO, Dauphine université, France 2017–2018

during the IASD program, I followed courses in Deep learning, reinforcement learning, statistical modelling, software engineering and data science. This was a double curriculum with the AgroParisTech IODAA program.

Diplôme d'Ingénieur, AgroParisTech, Paris, France 2014–2018

The first two years of this programm are focused on biology with courses in molecular biology, ecology, health, epidemiology and statistics.
During the final year, I followed the IODAA program, which was done as a double curriculum with Duaphine université, studying machine learning, data science and statistical modelling with a focus on biological problems.

BCPST Preparatory class, École Saint-Hilaire, Paris, France 2012–2014
2 year French preparatory class to prepare a nationally ranked exam for entry in the "Grande Écoles" ranked 201/1967 (Veto), ranked 481/2991 (Agro)

Languages

Mother tongue **French & English**

Other languages[1]

	Understanding		Speaking		Writing
	Listening	Reading	Interaction	Production	
German[2]	B1 Independent	B1 Independent	B1 Independent	B1 Independent	B1 Independent
Mandarin Chinese[3]	A2 Basic	A2 Basic	A2 Basic	A2 Basic	A2 Basic

[1] *Common European Framework of Reference for Languages (CEFR)*
[2] *1 month program at Goethe Institute in Schwäbisch Hall, Germany*
[3] *5 month program at SISU Shanghai, Chian & 1 month program at UIR, Beijing, CN*

Publications (Google scholar, ORCID)

[f] co-first authors

Journal articles

Mapping-friendly sequence reductions: Going beyond homopolymer compression
Luc Blassel, Paul Medvedev, Rayan Chikhi
iScience 25.11, p. 105305
DOI: 10.1016/j.isci.2022.105305

Using Machine Learning and Big Data to Explore the Drug Resistance Landscape in HIV
Luc Blassel, Anna Tostevin, Christian Julian Villabona-Arenas, Martine Peeters, Stéphane Hué, Olivier Gascuel
PLOS Computational Biology 17.8, e1008873
DOI: 10.1371/journal.pcbi.1008873

Drug Resistance Mutations in HIV: New Bioinformatics Approaches and Challenges
Luc Blassel [f], Anna Zhukova [f], Christian J Villabona-Arenas, Katherine E Atkins, Stéphane Hué, Olivier Gascuel
Current Opinion in Virology 51, pp. 56–64
DOI: 10.1016/j.coviro.2021.09.009

Origin, Evolution and Global Spread of SARS-CoV-2
Anna Zhukova, **Luc Blassel**, Frédéric Lemoine, Marie Morel, Jakub Voznica, Olivier Gascuel
Comptes Rendus. Biologies 344.1, pp. 57–75
DOI: 10.5802/crbiol.29

COVID-Align: accurate online alignment of hCoV-19 genomes using a profile HMM
Frédéric Lemoine, **Luc Blassel**, Jakub Voznica, Olivier Gascuel
Bioinformatics (btaa871)
DOI: 10.1093/bioinformatics/btaa871

Preprints

Phyloformer: Fast, accurate and versatile phylogenetic reconstruction with deep neural networks
Luca Nesterenko [f], **Luc Blassel** [f], Philippe Veber, Bastien Boussau, Laurent Jacob
bioRxiv
DOI: 10.1101/2024.06.17.599404

Conference talks

Phyloformer: Fast, accurate and versatile phylogenetic reconstruction with deep neural networks
Luca Nesterenko [f], **Luc Blassel** [f], Philippe Veber, Bastien Boussau, Laurent Jacob
2024
MCEB 2024

Mapping-Friendly Sequence Reductions: Going beyond Homopolymer Compression
Luc Blassel, Paul Medvedev, Rayan Chikhi
2022
RECOMB-SEQ 2022