LUIS PEDRO COELHO
Curriculum Vitæ
April 17, 2017

European Molecular Biology Laboratory (EMBL)
Meyerhofstraße, Heidelberg (Germany)
phone : +49 (0) 6221 387 (8150)
email : luis@luispedro.org
ORCID : 0000-0002-9280-7885
Erdös-Bacon Nr. : 7
Citizenship : U.K. and Portugal

Education

2011 PhD in computational biology, Carnegie Mellon University
Dissertation topic: *Modeling the Space of Subcellular Location Patterns Using Images and Other Sources of Information*, advised by Prof. Robert F. Murphy.

2006 MSc in computer science, Instituto Superior Técnico (Technical University Lisbon)
Dissertation topic: *Bayesian Network Parameter Estimation Using Noisy Observations or Soft Evidence*, advised by Prof. Arlindo Oliveira.

2004 BSc in computer science, Instituto Superior Técnico (Technical University Lisbon)
Graduated top of my class.

Professional Experience

2013–present Postdoctoral researcher at European Molecular Biology Laboratory (EMBL)
Supervisor: Dr. Peer Bork

2012 Postdoctoral researcher at Instituto de Medicina Molecular (Lisbon)
Supervisor: Dr. Musa Mhlanga

Scholarships & Awards

2013 JBS Authors' Choice Award
For most popular article published in The Journal of Biomolecular Screening in 2012.

2012 Siebel Scholar
Awarded annually for academic excellence and demonstrated leadership to 85 top students from the world's leading graduate schools

2007–2011 PhD. Scholarship from Portuguese Science Foundation

2009 Joint CMU-U. of Pittsburgh PhD. in Computational Biology Research Excellence Award

2008 Joint CMU-U. of Pittsburgh PhD. in Computational Biology Academic Excellence Award

2006 Fulbright Fellow

2005 Scholarship from Portuguese Science Foundation

2001 Instituto Superior Técnico (IST) Academic Excellence Award

Highlighted Publications

2015 S. Sunagawa*, **L. P. Coelho***, S. Chaffron* et al., *Structure and Function of the Global Ocean Microbiome* in Science [DOI]

2013 **L. P. Coelho***, T. Peng* et al., *Determining the subcellular location of new proteins from microscope images using local features* in Bioinformatics [DOI]

2009 **L. P. Coelho** et al., *Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms* in Proceedings of IEEE International Symposium in Biomedical Imaging [DOI]

Service & Outreach

2016–present   Associate Editor for the Journal of Open Research Software

2014–present   Online outreach activities
I maintain a blog about science at http://metarabbit.wordpress.com and tweet about science as @luispedrocoelho

2014–2015   Organized Software Carpentry Workshop at EMBL

2012–2013   Member of the Organization of the Lisbon Machine Learning School

2010   Local Committee for Portuguese-American Postgraduate Society National Forum
I headed the local organising committee for the 2010 edition of this annual event. It took place in Pittsburgh and included, as speakers, cabinet-level Portuguese government officials, renowned researchers, artists, as well as participants from all around the US.

2006–present   Peer Review
I have reviewed papers for BMC Bioinformatics, PLoS Computational Biology, Database, IEEE Transactions in Computational Biology and Bionformatics, and IEEE Transactions on Medical Imaging. I have also reviewed grant proposals for the Czech Science Foundation and the French National Research Agency (ANR).

2002–2004   Producer for IST Theatre Group

Teaching & Mentoring

2012–present   Certified Software Carpentry Instructor
I have given lectures on Software Carpentry in Germany, Denmark, Cyprus, Jordan, and Spain.

2012–2013   Supervision of Students at Instituto Superior Técnico (IST, Lisbon)
Co-advised one Master thesis in Computer Science (Paulo Monteiro) and one in Biomedical Engineering (Joana Cruz). During 2012, I was co-mentor to bioinformaticians in Luisa M. Figueiredo's lab as part of a collaboration.

2008–2011   Mentoring Junior Members of Murphy Lab
During my PhD. studies, I had the opportunity to directly supervise and mentor several junior members of the Murphy Lab. This experience includes working with **paid undergraduate programmers** (Nathan Herzing and Jephthah Liddie), one **MSc. student** (Shannon Quinn), **undergraduate students** performing lab work for credit (Jimmy Xu), and **high school students** volunteering over the summer (Peter Webb and Robert Webb).

2012   Programming for Scientists (short course)

2012   Practical Tutorial in Next Generation Sequencing in "Omics for Parasite Biology" Course

2011–2015   Tutorial in Lisbon Machine Learning Summer School

2009   Programming for Scientists
I designed and taught a semester-long course on computer programming for scientists at Carnegie Mellon University.

2008   Teaching Assistant for Laboratory Methods for Computational Biologists (CMU)

2005   Introduction to Computers Course in Cacém
I co-developed and co-taught an introductory course in computer usage in Cacém, an underprivileged neighborhood near Lisbon. This was pro-bono work.

2005   Teaching Assistant for Decision Support Systems (IST)

Language Skills

I am bilingual in **English** and **Portuguese**. I speak and write fluent **German** (I attended a German high-school, obtaining an Abitur; and later spent a year as an exchange student at Technical University of Vienna) and speak fluent **French**. I know basic **Luxembourgish**.

Publications

Google scholar profile: http://scholar.google.com/citations?user=qTYua0cAAAAJ

Peer-Reviewed Research Papers

1. Jaime Huerta-Cepas, Kristoffer Forslund, **Luis Pedro Coelho**, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork *Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper* Molecular Biology and Evolution msx148 (2017) [DOI]
   My contribution: *I contributed to the evaluation of the tool, in particular, in the evaluation of the suitability of the tool to metagenomics data*

2. J. R. Bedarf*, F. Hildebrand*, **L. P. Coelho**, S. Sunagawa, M. Bahram, F. Goeser, P. Bork, U. Wüllner *Microbial, functional and viral gut metagenome shifts in early onset L-DOPA naive PD patients* Genome Medicine 9:39 2017 [DOI]
   My contribution: *I built machine learning models based on metagenomics data.*

3. Jens Roat Kultima, **Luis Pedro Coelho**, Kristoffer Forslund, Jaime Huerta-Cepas, Simone S. Li, Marja Driessen, Anita Yvonne Voigt, Georg Zeller, Shinichi Sunagawa and Peer Bork, *MOCAT2: a metagenomic assembly, annotation and profiling framework* Bioinformatics 32 (16): 2520–2523, 2016 [DOI]
   My contribution: *I participated in the design and implementation of the proposed tool.*

4. Lionel Guidi*, Samuel Chaffron*, Lucie Bittner*, Damien Eveillard, Abdelhalim Larhlimi, Simon Roux, Youssef Darzi, Stephane Audic, Léo Berline, Jennifer Brum, **Luis Pedro Coelho**, Julio Cesar Ignacio Espinoza, Shruti Malviya, Shinichi Sunagawa, Céline Dimier, Stefanie Kandels-Lewis, Marc Picheral, Julie Poulain, Sarah Searson, Tara Oceans Consortium Coordinators, Lars Stemmann, Fabrice Not, Pascal Hingamp, Sabrina Speich, Mick Follows, Lee Karp-Boss, Emmanuel Boss, Hiroyuki Ogata, Stephane Pesant, Jean Weissenbach, Patrick Wincker, Silvia G. Acinas, Peer Bork, Colomban de Vargas, Daniele Iudicone, Matthew B. Sullivan, Jeroen Raes, Eric Karsenti, Chris Bowler, and Gabriel Gorsky, *Plankton Ecological Networks Predict Carbon Export in the Global Ocean* in Nature, 2016 [DOI]
   My contribution: *I performed functional analysis of the metagenomics data.*

5. Renato Pinheiro-Silva, Lara Borges, **Luis Pedro Coelho**, Alejandro Cabezas-Cruz, James J. Valdés, Virgilio do Rosário, José de la Fuente, *Gene expression changes in the salivary glands of Anopheles coluzzii elicited by Plasmodium berghei infection* in Parasites & Vectors 8:485, 2015 [DOI]
   My contribution: *I supervised the first-author, who was, at the time a PhD student, in the statistical analysis of the RNA-seq data.*

6. Shinichi Sunagawa*, **Luis Pedro Coelho***, Samuel Chaffron*, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R. Mende, Adriana Alberti, Francisco M. Cornejo-Castillo, Paul I. Costea, Corinne Cruaud, Francesco d'Ovidio, Stefan Engelen, Isabel Ferrera, Josep M. Gasol, Lionel Guidi, Falk Hildebrand, Florian Kokoszka, Cyrille Lepoivre, Gipsi Lima-Mendez, Julie Poulain, Bonnie T. Poulos, Marta Royo-Llonch, Hugo Sarmento, Sara Vieira-Silva, Céline Dimier, Marc Picheral, Sarah Searson, Stefanie Kandels-Lewis, Tara Oceans coordinators, Chris Bowler, Colomban de Vargas, Gabriel Gorsky, Nigel Grimsley, Pascal Hingamp, Daniele Iudicone, Olivier Jaillon, Fabrice Not, Hiroyuki Ogata, Stephane Pesant, Sabrina Speich, Lars Stemmann, Matthew B. Sullivan, Jean Weissenbach, Patrick Wincker, Eric Karsenti, Jeroen Raes, Silvia G. Acinas, Peer Bork, *Structure and Function of the Global Ocean Microbiome* in Science 348 (6237), 1261359, 2015 [DOI]
   My contribution: *As co-first author, I performed analysis of the data, including linking the metagenomics data to the environmental parameters, as well as the comparison to the human gut (this required extending existing software tools which were unable to scale to this dataset); and helped draft the manuscript.*

7. **Luis Pedro Coelho**, Catarina Pato, Ana Friães, Ariane Neumann, Maren von Köckritz-Blickwede, Mário Ramirez, João André Carriço, *Automatic determination of NET (neutrophil extracellular traps) coverage in fluorescent microscopy images* in Bioinformatics 31 (14): 2364–2370, 2015 [DOI]
   My contribution: *I conceived and implemented the algorithm, and wrote the manuscript.*

8. Ana C. Pena, Mafalda R. Pimentel, Helena Manso, Rita Vaz-Drago, Daniel Pinto-Neves, Francisco Aresta-Branco, Filipa Rijo-Ferreira, Fabien Guegan, **Luis Pedro Coelho**, Maria Carmo-Fonseca, Nuno L. Barbosa-Morais, Luisa M. Figueiredo, *Trypanosoma brucei histone H1 inhibits RNA polymerase I transcription and is important for parasite fitness in vivo* in Molecular Microbiology, 2014 [DOI]
My contribution: *I supervised D.P-N., who was at the time, a bioinformatics MSc student, in the statistical analysis of the RNA-seq data.*

9. Kristoffer Forslund, Shinichi Sunagawa, **Luis P. Coelho**, Peer Bork, *Metagenomic insights into the human gut resistome and the forces that shape it* in Bioessays, 2014 [DOI]
My contribution: *I performed data analysis.*

10. Peter Liehl, Vanessa Zuzarte-Luís, Jennie Chan, Thomas Zillinger, Fernanda Baptista, Daniel Carapau, Madlen Konert, Kirsten K Hanson, Céline Carret, Caroline Lassnig, Mathias Müller, Ulrich Kalinke, Mohsan Saeed, Angelo Ferreira Chora, Douglas T Golenbock, Birgit Strobl, Miguel Prudêncio, **Luis P Coelho**, Stefan H Kappe, Giulio Superti-Furga, Andreas Pichlmair, Ana M Vigário, Charles M Rice, Katherine A Fitzgerald, Winfried Barchet, and Maria M Mota, *Host-cell sensors for Plasmodium activate innate immunity against liver-stage infection* in Nature Medicine 20, 47–53, 2014 [DOI]
My contribution: *I performed statistical analysis of the gene expression data.*

11. Shinichi Sunagawa, Daniel R Mende, Georg Zeller, Fernando Izquierdo-Carrasco, Simon A Berger, Jens Roat Kultima, **Luis Pedro Coelho**, Manimozhiyan Arumugam, Julien Tap, Henrik Bjørn Nielsen, Simon Rasmussen, Søren Brunak, Oluf Pedersen, Francisco Guarner, Willem M de Vos, Jun Wang, Junhua Li, Joël Doré, S Dusko Ehrlich, Alexandros Stamatakis and Peer Bork, *Metagenomic species profiling using universal phylogenetic marker genes* in Nature Methods, 2013 [DOI]
My contribution: *I implemented the stand-alone version of the published metagenomics tool.*

12. **Luis Pedro Coelho**, Joshua D. Kangas, Armaghan Naik, Elvira Osuna-Highley, Estelle Glory-Afshar, Margaret Fuhrman, Ramanuja Simha, Peter B. Berget, Jonathan W. Jarvik, and Robert F. Murphy, *Determining the subcellular location of new proteins from microscope images using local features* in Bioinformatics, 2013 [DOI]
My contribution: *I designed and implemented the proposed algorithm, collected microscopy data (in collaboration with other authors), and wrote the manuscript.*

13. **Luis Pedro Coelho** Mahotas: Open source software for scriptable computer vision, Journal of Open Research Software, vol. 1, 2013 [DOI]
My contribution: *I implemented the underlying computer vision software and the manuscript.*

14. **Luis Pedro Coelho***, Tao Peng*, and Robert F. Murphy, *Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing* in Bioinformatics, vol. 26 (12), pp. i7–i12, 2010 [DOI]
My contribution: *I conceived of and implemented one of the methods presented and wrote the manuscript.*

15. **Luis Pedro Coelho**, Amr Ahmed, Andrew Arnold, Joshua Kangas, Abdul-Saboor Sheikh, Eric P. Xing, William W. Cohen, and Robert F. Murphy, *Structured Literature Image Finder: Extracting Information from Text and Images in Biomedical Literature* in Lecture Notes in Bioinformatics, vol. 6004, pp. 23–32, 2010 [DOI]
My contribution: *With the paper below, this forms a package on the SLIF project, where I enhanced the image analysis pipeline (for the analysis of figures in published scientific literature). I wrote the first draft of this manuscript and helped draft the one below.*

16. Amr Ahmed, Andrew Arnold, **Luis Pedro Coelho**, Joshua Kangas, Abdul-Saboor Sheikk, Eric P. Xing, William W. Cohen, *Structured Literature Image Finder: Parsing Text and Figures in Biomedical Literature* in Web Semantics: Science, Services and Agents on the World Wide Web, vol. 8, pp. 151–154, 2010 [DOI]

Peer-reviewed Conference Proceedings

17. **Luis Pedro Coelho**, Aabid Shariff, and Robert F. Murphy; *Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms* in Proceedings of IEEE International Symposium in

BIOMEDICAL IMAGING, 2009 [DOI]
My contribution: *I acquired (part of) the microscopy data, built the hand-segmented dataset, implemented the methods and wrote the first draft of the manuscript.*

18. Taraz Buck, Arvind Rao, **Luis Pedro Coelho**, Margaret Fuhrman, Jonathan W. Jarvik, Peter B. Berget, and Robert F. Murphy; *Cell Cycle Dependence of Protein Subcellular Location Inferred from Static, Asynchronous Images* in CONFERENCE PROCEEDINGS OF THE IEEE ENGINEERING IN MEDICAL AND BIOLOGY SOCIETY, pp. 1016–1019, 2009 [DOI]
My contribution: *I helped the first author acquire relevant microscopy data.*

19. **Luis Pedro Coelho** and Robert Murphy; *Identifying Subcellular Locations from Images of Unknown Resolution* in BIOINFORMATICS RESEARCH AND DEVELOPMENT, Communications in Computer and Information Science, vol. 13, pp. 235–242, 2008 [DOI]
My contribution: *I implemented the algorithm, ran the tests, and wrote the first draft of the paper.*

20. Amina Chebira, **Luis Pedro Coelho**, Aliaksei Sandryhaila, Stephen Lin, William G. Jenkinson, Jeremiah MacSleyne, Christopher Hoffman, Philipp Cuadra, Charles Jackson, Markus Püschel, Jelena Kovacevick; *An Adaptive Multiresolution Approach to Fingerprint Recognition* in PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2007 [DOI]
My contribution: *I developed and implemented one of the methods presented in the paper.*

21. **Luis Pedro Coelho** and Arlindo Oliveira; *Dotted Suffix Trees: A Structure for Approximate Text Indexing* in STRING PROCESSING AND INFORMATION RETRIEVAL, Lecture Notes in Computer Science, vol. 4209, pp. 329–336, 2006 [DOI]
My contribution: *I developed and implemented the algorithm and wrote the first draft of the manuscript.*

Review Articles

22. **Luis Pedro Coelho**, Estelle Glory-Afshar, Joshua Kangas, Shannon Quinn, Aabid Shariff, and Robert F. Murphy; *Principles of Bioimage Informatics: Focus on machine learning of cell patterns* in LINKING LITERATURE, INFORMATION, AND KNOWLEDGE FOR BIOLOGY, Lecture Notes in Computer Science, vol. 6004, pp. 8–18, 2010 [DOI]

23. Aabid Shariff, Joshua Kangas, **Luis Pedro Coelho**, Shannon Quinn, and Robert F. Murphy; *Automated Image Analysis for High Content Screening and Analysis* in JOURNAL OF BIOMOLECULAR SCREENING, August 2010, pp. 726–734 [DOI]

Books

1. **Luis Pedro Coelho**, Willi Richert; *Building Machine Learning Systems with Python*, Packt Publishing, 2013 (first ed.); 2015 (second ed.)

Invited Talks

1. *Looking at the oceans with computer vision in Python*, invited keynote at PyCon Firenze (Italy), April 2016

2. *Structure and function of the ocean microbiome*, lecture at the course *Biology of ecological systems*, Ecole Nationale Superior (France), December 2015

3. *Life in words and pictures: Sequencing and microscopy for the analysis of microbial communities*, Luxembourg Centre for Systems Biomedicine, September 2015

4. *An analysis of 243 metagenomes collected by the Tara Oceans Projects*, Station Biologique de Roscoff, July 2015

5. *Machine Learning: What it can do for you (and what it cannot)*, BrightTalk online seminar, May 2015

6. *Webcast: Penalized Linear Regression in Python*, O'Reilly Webcasts, October 2014

7. *Python for Computer Vision in Biology and Beyond*, Python San Sebastian, September 2014

8. *Large Scale Analysis of Bioimages Using Python*, International Workshop on Technical Computing for Machine Learning and Mathematical Engineering, Leuven (Belgium), September 2014

9. *Mahotas and the Python Ecosystem for Bioimage Informatics Applications*, European Bioimage Analysis Symposium, Barcelona, October 2013

10. *Organizing the Proteome with Location and Function Topics*, Freiburg Institute for Advanced Studies, April 2013

11. *Modeling Subcellular Location from Images and Other Sources of Information*, Luxembourg Centre for Systems Biomedicine, July 2012

12. *Modeling Subcellular Location from Images and Other Sources of Information*, Priberam Machine Learning Lunch Seminar, June 2012

13. *Modeling Subcellular Location from Images and Other Sources of Information*, EAO Seminar (Instituto Gulbenkian da Ciência), Oeiras, March 2012

14. *Learning Subcellular Location from Images and Other Sources of Information*, KDBIO (Knowledge Discovery and Bioinformatics) seminar, Lisbon, February 2012

15. *Bioimage Informatics: Computer Vision for Biology*, EMBO Practical Course on Microscopy: from single molecules to animals, Pretoria, November 2011

16. *Studying the subcellular location space with bioimages and other data modalities*, University of Delaware, Computer and Information Sciences Department, September 2011

17. *Unsupervised Mixture Pattern Unmixing*, University of Bielefeld International Graduate School of Bioinformatics and Genome Research, July 2008

18. *Proteome-scale analysis and modeling of subcellular location*, 4th CeBiTec Symposium BioImaging, Bielefeld (Germany), 25–27 August 2009

Selected Conference Talks

1. **Luis Pedro Coelho***, Sebastien Colin*, Shinichi Sunagawa, Eric Karsenti, Peer Bork, Rainer Pepperkok, and Colomban de Vargas *Environmental High-content Fluorescence Microscopy (e-HCFM) of Tara Oceans Samples Provides a View of Global Ocean Protist Biodiversity*, Ocean Sciences, New Orleans, 2016

2. Shinichi Sunagawa, **Luis Pedro Coelho**, Samuel Chaffron, Eric Karsenti, Jeroen Raes, Silvia Acinas, Peer Bork *Structure and function of the global ocean microbiome*, ASLO, Granada (Spain) 2015

3. **Luis Pedro Coelho** *Jug: Reproducible Research in Python*, BOSC, Berlin 2013

4. Rita Reis and **Luis Pedro Coelho**; *Using Theatre to Fight HIV/AIDS in Mozambique*, National Conference of the Association for Theatre in Higher Education, Chicago 2011

5. **Luis Pedro Coelho** and Robert F. Murphy; *Determining Resolvable Subcellular Location Categories as a Function of Image Resolution*, 24th ISAC Congress, Budapest 2008