RESUMEN DEL CURSO O TÓPICO

Aspectos generales

Título:	Systems Biology: a conceptual and hands-on introductory course.
Semestre:	2025-1
Sede:	Instituto de Biotecnología, UNAM.
Horario:	Martes y Miércoles de 9:00 a 11 Hrs.
No. sesiones:	40
Duración de la sesión:	2.00
Cupo total:	10
Observaciones:	This proposal for this course follows on an ad hoc shorter course taught by about a month by Jorge Carneiro at IBt in the fall of 2023. This course attracted 18 students of which 12 completed the course. The course was free, not integrated in the graduate school. The 6 students who stopped attending the course justified with the course being demanding given the time they had available or health reasons (dengue epidemic). The ones that completed the course gave excellent feedback on the course and recommended that it would be better if extended to 4 months and integrated in the official postgraduate programme. The submission of this course is a response to the students' recommendation. The proposed course will be in presence and by videoconference given that the course organiser and main instructor, Jorge Carneiro, resides in a different country. He will be at UNAM in the first 3 weeks and also in the last 3 weeks of the course. The first in person period aims to get familiarised with cohort of students, establishing effect communication bonds, and to assist the students in defining their individual project. In the last period, the course organiser will assist more closely the final phase of the individual projects, will wrap up the course and evaluate the students and also the course itself. The session in between will be by videoconference using the Zoom platform. The course will be in English but Jorge Carneiro understands and speaks Spanish. He has chosen a set of speakers who in addition to being outstanding researchers with different degrees of seniority, who can also communicate with students in Spanish if necessary. From previous experience it is important to be able to do so to keep students less fluent in English in the loop.

Tutor responsable

Nombre:	JAIME ARTURO PIMENTEL CABRERA
Entidad:	Instituto de Biotecnología
Email:	arturo.pimentel@ibt.unam.mx
Teléfono:	7772077756

Métodos de evaluación

MÉTODO	CANTIDAD	PORCENTAJE
Assignement on data analysis or modellig.	17	30%
Individual project report.	1	40%
Journal club sessions	1	30%

Integrantes

INTEGRANTE	ROL	HORAS	ACTIVIDAD COMPLEMENTARIA
JAIME ARTURO PIMENTEL CABRERA	Responsable	8.00	
DANIEL PRIEGO	Profesor invitado (Externo)	3.00	
DENIS THIEFFRY	Profesor invitado (Externo)	3.00	
ISABEL DUARTE	Profesor invitado (Externo)	3.00	
JORGE CARNEIRO	Profesor invitado (Externo)	57.00	
PABLO VILLOSLADA	Profesor invitado (Externo)	3.00	
VERONICA GRIENEISEN	Profesor invitado (Externo)	3.00	
		80/80	

Introducción

Systems biology established itself at the core of biological research and biotechnology and is now promising new advances in medical practice and precision medicine. In contrast, systems biology is not yet central in undergraduate biological education and training. This course aims to introduce graduate students to systems biology concepts, language and practice.What is systems biology? What is involved in studying a biological system as a system? Why must one be concerned with the structure, the dynamics and the adaptive dynamics or rewiring of a biological system? The course is planned for about five months and is structured as 8 topics:

T1. An Historical and Conceptual Exploration of Systems Biology; T2. The challenges of complex, heterogeneous and high-dimensional biological data; T3. Systems biology of metabolism; T4. Complex dynamics of membrane potential and ion flux regulation; T5. Gene regulatory networks; T6. Principles of bodybuilding: from pattern formation to embryo body plans; T7. The challenge of integrating cellular and supracellular dynamics in multicellular systems: the examples of the immune system ; T8. The final frontier: systems and precision medicine.

These 8 topics will be explored in two threads, a theoretical thread and a practice thread, which dovetail on each other. The theoretical thread will explore the concepts, hypotheses and theoretical frameworks associated with each topic. Lectures by the course organizer and instructor and occasional invited speakers will alternate with journal clubs in which students present and discuss scientific articles with their peers The theoretical thread will be complemented by a parallel hands-on practice in which students will work on the challenges of high-dimensional data analysis (e.g. multidimensional flow cytometry or single cell RNAseq) and model systems using different formalisms (from ordinary differential equations, stochastic differential equations, partial differential equations and logical-networks). The practical sessions will be based on Jupyter notebooks using either Python or R programming languages depending on the student preference and familiarity. We will resort to Large Language Models, such as ChatGPT, to overcome the steepness of the programming learning curve, so that one can concentrate on assimilating the essential concepts of data analysis and modeling. During the whole course, students will develop an individual small research project and write the corresponding report.

Objetivos

The course aims to familiarize the students with the concepts, scientific language and theoretical frameworks of systems biology and medicine, and to initiate them in the practice of high dimensional data analysis and modeling of complex biological systems. At the end of the course, students are expected to understand the scientific language and concepts of systems biology and to harness the power of quantitative and computational approaches in their research, making their results more transparent and reproducible as well as to use critically LLM as code development tool.

Target Audience

- Researchers who wish to know more about systems biology and to address biological problems at systems level.
- Biology-trained researchers would like to improve their ability to analyze complex high dimensional biological data, and to model and simulate the systems generating such data.
- Researchers who would like to initiate or improve their capacity to use R or python scripts to accelerate standardization, and make more transparent and reproducible the analysis of their data.
- Researchers who would like to deploy Large Language Models, such as ChatGPT, in their research practice.

Temario

T1. Historical and Conceptual Exploration of Systems Biology. Week 1.

Biological systems have been studied since the origins of biology. In late 19th century literature, one often finds whole system views and analogies between biological systems and complex engineered systems, e.g. nervous system and the telegraph network. In some sense one can say that systems biology is a centuries-old science. However, the emergence about two decades ago of new scientific journals, departments and institutes claiming the title of “Systems Biology” suggests the emergence of a new discipline. What characterises modern systems biology? We will briefly overview the history of the systems biology to show that it is a multithreaded convolution of (i) molecular biology with its (in)capacity to handle massive throughput genomics, (ii) principled mathematical and theoretical biology, and (iii) engineering and control of electric, electronic and information systems. The marriage of the three has not always been auspicious.

Articles for Journal Club.

- JC1.1. Lazebnik (2002) Can a biologist fix a radio?--Or, what I learned while studying apoptosis. Cancer Cell 2, 179 10.1016/s1535-6108(02)00133-2
- JC1.2. Cohen (2004) Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. PLoS Biol 2, e439 10.1371/journal.pbio.0020439
- JC1.3, Noble (2010) Biophysics and systems biology. Philos Trans A Math Phys Eng Sci 368, 1125 doi.org/10.1098/rsta.2009.0245
- JC1.4, Westerhoff, Palsson (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22, 1249 10.1038/nbt1020
- JC1.5. Kitano (2002) Systems biology: a brief overview. Science 295, 1662 10.1126/science.1069492
- JC1.6 Auffray, et al. (2003) From functional genomics to systems biology: concepts and practices. C R Biol 326, 879 10.1016/j.crvi.2003.09.033

Practice.

- Validation of students Jupyter, Python or R settings using scripts provided by instructors

T2. The challenges of complex, heterogeneous and high-dimensional biological data. Week 2-3.

Todays biology and medicine are data-rich. It is almost straightforward to measure the transcripts of all the genes in the genome of an organism in many organs, tissues or single cells. These high dimensional transcriptomic data are routinely complemented with measurements of metabolites and proteins (in their different functional forms) or multidimensional images of the cells. The result are complex, high-dimensional multivariate data sets. A typical systems biology approach involves the massively parallel measurements under multiple conditions organised under the guidance of complex experimental designs. The purpose of this lecture is not to enter the details of how these complex data are produced by the different technologies or how they are specifically analysed. Instead, students will discover how awkward and counterintuitive things, such as density and distance, are in the high dimensional spaces where these data dwell. Then students will visit the generic statistical and computational solutions to deal with high dimensional multivariate data like PCA, tSNE, UMAP that are use to reduce dimensionality and visualise the data in low dimensions.

Articles for Journal Club

- JC2.1 La Manno, et al. (2018) RNA velocity of single cells. Nature 560, 494
- JC2.2 Do, Canzar (2021) A generalization of t-SNE and UMAP to single-cell multimodal omics. Genome Biol 22, 130
- JC2.3. Ding, Regev (2021) Deep generative model embedding of single-cell RNA-Seq profiles on hyperspheres and hyperbolic spaces. Nat Commun 12, 2554. 10.1038/s41467-021-22851-4
- JC2.4 Cui, et al. (2024) scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods https://doi.org/10.1038/s41592-024-02201-0

Practice

- P2.1. A primer in low dimensional embedding using synthetic data
- P2.2. Warming up with intuitive mapping of cities problem
- P2.3. Analysis of real data sets from scRNAseq and flowcytometry

T3. Systems biology of metabolism. Weeks 4-6.

The relentless analysis of cellular biochemistry led to a full network of hundreds of thousands of reactions that convert any metabolite into any other metabolite in the cell. The success in this endeavour is patent in the metabolic maps to which we stare at mesmerised in the walls of our biochemistry departments. This static image of the structure of the metabolic network, however, has no dynamics. In these lectures, we will overview the mathematical and computational principles used to analyse and predict the dynamics and regulation of such metabolic networks, from the underlying reaction network. We will first consider the functional forms used to express how the rate of synthesis of a product depend on the activity of the reacting metabolites and catalysts: conservation of mass, mass action, Michaelis-Menten-Henri kinetics, Hill kinetics, generalised mass action, MCA, BST and S-systems. We will conclude with an overview of Flux Balance Analysis (FBA) which has become the workhorse of bioreactor optimisation. As we will see, FBA of genome-scale metabolic models (GEMs) has been surprisingly successful in bridging genomic and metabolic data in bacteria and yeast, despite throwing kinetics and regulation under the carpet.

Articles for Journal Club.

- JC3.1 Monk, et al. (2013) Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci U S A 110, 20338
- JC3.2 Jouhten, et al. (2022) Predictive evolution of metabolic phenotypes using model-designed environments. Mol Syst Biol 18, e10980. 10.15252/msb.202210980

Practice.

- P3.1 Mass action, stoichiometry and chemical kinetics
- P3.2 Modelling enzyme kinetics with differential equations
- P3.3 FBA analysis of the Warburg Effect

T4. Complex dynamics of membrane potential and ion flux regulation. Week 7

The understanding of heart beat dynamics and neuronal systems dynamics are two examples of systems biology success stories. These are rooted in the pioneering work of Hodgkin and Huxley on ion flux dynamics in the excitable squid giant axon. In this lecture we will dig into the quantitative description of membrane potential and ion flux regulation, revisiting the excitable media paradigm introducing the associated dynamic system theory. We will overview the use of this paradigm in several biological systems from neurons to spermatozoa and pollen tubes.

Articles for Journal Club

- JC4.1. Hodgkin, Huxley (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol (Lond) 117, 500
- JC4.2. Fitzhugh (1961) Impulses and Physiological States in Theoretical Models of Nerve Membrane. Biophys J 1, 445.
- JC4.3. Priego-Espinosa, et al. (2020) Modular analysis of the control of flagellar Ca2+-spike trains produced by CatSper and CaV channels in sea urchin sperm. PLoS Comput Biol 16, e1007605

Practice.

- P4.1 Implementation of the Hodgkin-Huxley model
- P4.2 Implementation of the FitzHugh-Nagumo model

T5. Gene regulatory networks. Week 8-11

How can one analyse and predict the behaviour of the large gene regulatory networks? In principle, one can describe the dynamics of the expression levels of each gene product as ordinary differential equations (ODE) or simulate the kinetics using stochastic formalisms such as stochastic differential equations (SDE) or Gillespie-type

simulations. However, these quantitative approaches involve as an absolute prerequisite to define each functional form in the equations' terms and specify values for all the parameters. If this is a daunting task for a small regulatory network, it is unfeasible for large genome-scale networks. Boolean or logical models are well-suited to capture essential dynamical properties of these networks, circumventing the need to specify parameter values. For networks controlling cell fate decisions, cell fates are associated with model attractors — stable states or cyclic attractors — whose identification and reachability properties become particularly relevant. In these lectures, we will overview the challenges and solutions to study gene regulatory networks using quantitative ODE-based models or discrete Boolean/logical dynamics.

Articles for Journal Club

- JC5.1. Cross, et al. (2002) Testing a mathematical model of the yeast cell cycle. Mol Biol Cell 13, 52. 10.1091/mbc.01-05-0265
- JC5.2. Sha, et al. (2003) Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. Proc Natl Acad Sci U S A 100, 975
- JC5.3. Chang, et al. (2006) Multistable and multistep dynamics in neutrophil differentiation. BMC Cell Biol 7, 11
- JC5.4. Fauré, et al. (2009) Modular logical modelling of the budding yeast cell cycle. Mol Biosyst 5, 1787

Practice.

- P5.1 Implementation of a simple autoactivator and inhibitor gene regulatory (AI) network as ODE. Phase plane and bifurcation analysis.
- P5.2 Implementation of the AI network as SDE system. Probability distribution analysis.
- P5.3 Implementation of the AI network as a discrete logical dynamics model.

T6. Principles of body building: from pattern formation to embryo body plans. Weeks 12-13

Mathematical and theoretical biology have been a reliable source of general principles that give insight into biological systems and guide researchers in generating, interpreting, reasoning and inferring consequences of hypotheses about the specific system. Often this is under-appreciated by researchers with a molecular biology training and insufficient training in mathematics. To illustrate this with a concrete systems biology arena, we will overview how the general principles of non-equilibrium dynamics and pattern formation can guide our qualitative and quantitative thinking about animal body plans. Using the simple auto-activator and inhibitor system explored in the previous topic, we will discuss which body plans can and, perhaps more importantly, which body plans cannot be explained based on these principles.

Articles for Journal Club

- JC6.1 Gierer, Meinhardt (1972) A theory of biological pattern formation. Kybernetik 12, 30 10.1007/BF00289234
- JC6.2 Akam (1989) Drosophila development: making stripes inelegantly. Nature 341, 282 10.1038/341282a0
- JC6.3 Marcon, Sharpe (2012) Turing patterns in development: what about the horse part? Curr Opin Genet Dev 22, 578 http://dx.doi.org/10.1016/j.gde.2012.11.013
- JC6.4 Glover, et al. (2023) The developmental basis of fingerprint pattern formation and variation. Cell 186, 940. 10.1016/j.cell.2023.01.015

Practice

- P6.1 Implementation of the AI network as a reaction diffusion system using partial differential equations. Spatial and temporal patterns characterisation and bifurcation analysis.

T7. The challenge of integrating cellular and supracellular dynamics in multicellular systems: the example of the immune system.

Week 14-15

In multicellular systems, the potential structure of the regulatory molecular networks is open-ended and spans sub- and super-cellular levels. The molecular networks within each cell evolve with a dynamics influenced by inputs from intercellular interactions. These molecular networks define if the cell differentiates, divides or dies, and these cellular events, underpin the cell population and tissue dynamics that in turn defines the context-dependent intercellular interactions that feedback on the molecular networks in the cells. Dealing with complex multilevel dynamics is challenging as we will illustrate using the vertebrate immune system. In the vertebrate immune system, tolerance refers to the absence of pathologic autoimmunity. Regulatory T cells, expressing the transcription factor Foxp3, and able to control the activation and clonal expansion of other cells, are essential for tolerance. What defines molecularly a regulatory T cell ? How do regulatory T cells differentiate from precursors? How stable is their differentiation programme ? How do they persist as a population capable of ensuring robust tolerance by interacting with other cell populations? How is the immune repertoire of these cells selected and organised? These are typical systems biology questions. We will overview how different mathematical models — logical network models, ordinary and stochastic differential equations, and stochastic multilevel agent-based simulations — were used to address these questions and uncover different aspects of regulation in the immune systems. The upshot will be that tolerance and autoimmunity are system properties, and their understanding, requires a simultaneous consideration of the gene regulatory network dynamics within the cell together with the dynamics of cell populations.

Articles for Journal Club

- JC7.1. Naldi, et al. (2010) Diversity and plasticity of Th cell types predicted from regulatory network modelling. PLoS Comput Biol 6, e1000912 https://doi.org/10.1371/journal.pcbi.1000912
- JC7.2. Schneider (2011) Tracing personalized health curves during infections. PLoS Biol 9, e1001158 https://doi.org/10.1371/journal.pbio.1001158

Practice

- T7.1 Implementation and analysis of an extension of the crossregulation model of the immune system featuring volatile cellular types.
- T7.2. Modelling disease curves : which Schneider curves can and cannot be modelled?

T8. The final frontier: systems and precision medicine. Week 16

Systems biology has come of age and established itself at the core of biological research. Its basic tenets are now permeating medical research and precision medicine. Big data-based computational models are giving new insights into cancer, resolving individual differences in diagnosis and therapy. In a theoretical lecture we will see that

these are auspicious times for medical practice and research. But like any potential area of development there are many challenges, which are essentially the same challenges of systems biology. As a way a validating the knowledge acquired during the previous lectures, journal clubs and practical assignments, students will be asked to write a New & Views type of review of a typical precision medicine article.

Article for Journal Club

- JC8.1 Cords et al. Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer 2024, Cancer Cell 42, 1–17 https://doi.org/10.1016/j.ccell.2023.12.021

Practice

- P8 Students will write a two page News & Views article about the article JC8.1

Bibliografía

Books

- Uri Alon. 2006 An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC Mathematical and Computational Biology. ISBN 9781584886426
- Bernhard Ø. Palsson. 2012 Systems Biology. Properties of Reconstructed Networks. Cambridge University Press. ISBN 9780511790515

Articles

- Deichmann, et al. (2014) Commemorating the 1913 Michaelis-Menten paper Die Kinetik der Invertinwirkung: three perspectives. FEBS J 281, 435.
- Westerhoff, Palsson (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22, 1249 10.1038/nbt1020
- Roth (2011) Mathematics and biology: a Kantian view on the history of pattern formation theory. Dev Genes Evol 221, 255.
- Ideker, et al. (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2, 343.
- Nielsen (2017) Systems Biology of Metabolism. Annu Rev Biochem 86, 245.
- Alon (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450.
- Hood, et al. (2012) Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J 7, 992. 10.1002/biot.201100306.
- Jaskowiak, et al. (2014) On the selection of appropriate distances for gene expression data clustering. BMC Bioinformatics 15Suppl2, S2.
- Altaf-Ul-Amin, et al. (2014) Systems biology in the context of big data and networks. Biomed Res Int 2014, 428570. 10.1155/2014/428570.
- Carneiro, et al. (2007) When three is not a crowd: a Crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells. Immunol Rev 216, 48.

Workshop: Tooling up for systems biology and medicine
Instructor: Jorge Carneiro
Venue: UNAM, Unidad de Posgrado, CU, Mexico City
Date: November 4-8, 2024

Synopsis

Systems biology established itself at the core of biological research and biotechnology and is now promising new advances in medical practice and precision medicine. In contrast, systems biology is not yet central in undergraduate biological education and training. Biological science graduates struggle to handle the quantitative and computational research tools they need, and are often deterred from engaging in systems biology research. This workshop is a primmer in systems biology and medicine that aims to initiate and prepare graduate students to further study and enter this research field.

Objectives

The objectives of the course are: (i) to introduce the students to some basic concepts, scientific language and theoretical frameworks of systems biology and medicine; (ii) to guide students in the first steps in coding and scripting using Jupyter notebooks; (iii) to learn how to organise and analyse experimental data in programmatic way that is objective, reliable, reproducible, and shareable; (iv) to familiarise the students with the good practices of using Large Language Models and AI to develop their own code and scripts; (v) to help the students to overcome the fear of mathematical formalisms so that they can read modern scientific articles and become more competent for todays quantitative research. At the end of the course, students are expected to be tooled up to discover the scientific language and concepts of systems biology and to start harnessing the power of quantitative and computational approaches in their research, making their results more transparent and reproducible as well as to use critically LLM as code developing tool.

Programme

The course will have the duration of a week and will be structure in daily one-hour theoretical & discussion session followed by two-hour practical session. Theoretical and practical sessions will be back-to-back and aimed at allowing the students to answer the following five questions:

Day 1. What is systems biology ? What is systems medicine? Why should any graduate student be interested in systems biology and medicine today? What are the objectives of the course?

Day 2. Why should one use coding and scripting to organise biological or biomedical data in a programmatic instead of interactive manner?

Day 3. How can one prompt Large Language Models to obtain reliable code, ready to use in ones scripts?

Day 4. How can one overcome the fear of formal, mathematical tools to deploy and improve one's scientific reasoning, objectivity and research quality?

Day 5. How can one integrate data analysis and modelling to get a better insights into the workings of a system?

Target Audience

Researchers who wish to know more about systems biology and want to self-diagnose their competences for systems biology.

Researchers who would like to initiate or improve their capacity to use R or python scripts to accelerate standardise, and make more transparent and reproducible the analysis of their data.
Researchers who would like to deploy Large Language Models, such as ChatGPT, in their research.

Highly recommended for students trained in life sciences and who wish to do the course Systems Biology and Medicine: a conceptual and hands-on introductory course in the second semestre 2025-2

Requirements

Personal computer — Jupyter notebook — R or Python — ChatGPT
Scripts for data analysis and modelling will be coded in R or Python within Jupyter (https://jupyter.org/). Jupyter is a web-based interactive development environment for notebooks, code, and data.
Students are expected to use Jupyter notebooks in their own computer (see how to install at https://jupyter.org/install) or in Google Colab (https://colab.research.google.com/). The former solution allows students to work offline whereas the latter requires a good Internet connection permanently. Python and R can be used in a local installation of Jupyter. Google Colab uses a python kernel only. It is possible to run R within Colab (see how here https://towardsdatascience.com/how-to-use-r-in-google-colab-b6e02d736497) but the error reports are harder to interpret.
Finally, we will resort to AI language model ChatGPT to generate and analyse R or python codes. Students must have registered at ChatGPT and be familiar with basic prompting of this Large Language Model.

About the Instructor

Jorge Carneiro is a theoretical biologist, who made a transition from wet lab biochemical and immunological experimentation to computational and systems biology research. He is especially motivated and experienced in helping biological sciences graduates to try and engage in computational and quantitative research. He got his licenciatura in Biochemistry 1991 and his PhD in Biomedical Sciences from the University of Porto in 1997. In Porto, he gave his first steps in research in the Theoretical Chemistry Group of the Faculty of Sciences and in the Immunology Laboratory of the Abel Salazar Biomedical Sciences Institute. He did his PhD research at the Pasteur Institute in Paris (France). After a short postdoc at the Bioinformatics and Theoretical Biology Group of the University of Utrecht (Netherlands), he became an independent scientist in 1998 at the Instituto Gulbenkian de Ciência (IGC, Portugal), where he animated a research group dedicated to mathematical modelling of biological organisation, with special focus on the immune systems. He is currently affiliated with the Instituto de Tecnologia Química e Biológica António Xavier ITQB NOVA (Portugal). In the last decades, he has promoted scientific education and science-enabling environments for decades. He informally coordinated the theoretical and computational sector of the IGC; he organised and taught postgraduate courses and trained PhD students in Portugal and beyond, he collaborated in the planning and fundraised for the first PhD Programme in Computational Biology eventually becoming its director; he served as Deputy Director for Science of the IGC and he directed the PhD Programme in Integrative Biology and Biomedicine. Currently, he studies the history of Life on Earth to answer a simple but fundamental question: how come Life never got extinct in the last ~4x10^9 years. He believes that answering this question quantitatively may be instrumental to understand, predict and plan our collective future.