Published in IET Systems Biology
Received on 8th March 2010
Revised on 11th September 2010
doi: 10.1049/iet-syb.2010.0032

ISSN 1751-8849

Teaching systems biology

R. Alves[1] *E. Vilaprinyo*[2] *A. Sorribas*[1]

[1]*Departament Ciencies Mediques Basiques, Universitat de Lleida & Institut de Recerca Biomèdica, de Lleida (IRBLleida), Montserrat Roig 2, 25008 Lleida, Spain*
[2]*Evaluation and Clinical Epidemiology Department, Hospital del Mar-IMIM, Doctor Aiguader 88, Barcelona, 08003, Spain*
E-mail: ralves@cmb.udl.es

Abstract: Advances in systems biology are increasingly dependent upon the integration of various types of data and different methodologies to reconstruct how cells work at the systemic level. Thus, teams with a varied array of expertise and people with interdisciplinary training are needed. So far this training was thought to be more productive if aimed at the Masters or PhD level. At this level, multiple specialised and in-depth courses on the different subject matters of systems biology are taught to already well-prepared students. This approach is mostly based on the recognition that systems biology requires a wide background that is hard to find in undergraduate students. Nevertheless, and given the importance of the field, the authors argue that exposition of undergraduate students to the methods and paradigms of systems biology would be advantageous. Here they present and discuss a successful experiment in teaching systems biology to third year undergraduate biotechnology students at the University of Lleida in Spain. The authors' experience, together with that from others, argues for the adequateness of teaching systems biology at the undergraduate level.

1 Systems biology and its applications

The need for an integrated approach in order to understand how organisms work at the molecular level has been recognised as early as in the 19th century by Claude Bernard [1]. In the beginning of the 20th century, this need has been further emphasised by Ludwig von Bertalanffy [2]. Since the technical know-how was not yet available, the only set of tools that provided such an integrated view of molecular biology systems was that derived from mathematics in general and mathematical modelling in particular [3]. This led to the development and application of different systems theories to the study of biological systems [4, 5]. The development of the computer has facilitated the expansion and application of these methods and the blooming of computational biology. With the advent of the human genome project [6], analytical technologies that allow measuring the simultaneous behaviour of the whole transcriptome, proteome and metabolome of a cell type have been developed [7, 8].

The large amount of data that is generated in these experiments, compounded with the non-linearity of interactions between the different elements of the systems, creates a situation where human intellect is clearly overwhelmed. Thus, the use of computers and mathematical models has remained at the centre of most well thought out systems biology studies, because such tools make it possible to sort out, organise and often make sense of these enormous amounts of data.

The impact of computational systems biology in the productivity of biotechnological industries is increasing. In medical biotechnology, the use of in silico methods to predict both the best combination of genes and proteins to target in order to cure or palliate a disease and the best molecules to be used for this job are now becoming increasingly common [9–12]. In production biotechnology, the use of computational systems biology to assist in the rational design of organisms has increased the microorganism productivity by several folds [13–17]. In ecological biotechnology, several applications have used systems biology to design organisms that perform bioremediation tasks [18–20]. This has gone as far as proposing the design of entire microbial communities [18, 21, 22]. The wide scope of methods that are available to and needed in systems biology requires that systems biologists' teams are both fairly extended and multidisciplinary, because at present not many people hold the necessary combined expertise. Owing to the growing need for people that have training in the area, there is currently a shortfall of such professionals. Thus educating a sufficient number of new systems biologists has become a challenge and several suggestions have been made regarding the best way to provide this education [23, 24].

2 Educating systems biologists

Graduate programmes that train students in the different areas of knowledge that contribute to systems biology, such as life sciences, mathematics and engineering, have existed in prestigious institutions for more than 50 years. In fact, one of the earliest examples of a graduate course that was exclusively dedicated to training students for the analysis of molecular biological systems from an

integrative perspective was Michael Savageau's course on biochemical systems theory at the University of Michigan Medical School. This course started in the early 1970s, leading to one of the classical books about systems biology [5].

The inherent interdisciplinary skills needed by a systems biologist more or less led to an initial consensus that the introduction of systems biology as a main stream subject in the biology curricula should be done at the graduate level. Thus MSc and PhD programmes focusing on systems biology are now provided by many universities around the world, and it is fair to say that, in the US, biomedical engineering and bioengineering departments are responsible for a large fraction of the graduate training in such programmes (Table 1, http://rumo.biologie.hu-berlin.de/education/pmwiki/pmwiki.php?n=Main.HomePage or Table I in [25]).

However, with the appropriate approach and the right curricula, systems biology can be profitably introduced at the undergraduate level. In fact, over the past 20 years several Universities have introduced systems biology courses at the undergraduate level in the USA. In Europe, this move was not extensively followed, and only a small number of the European Universities offer such courses. This discrepancy may be due to many factors, including but not limited to the procedural differences between the American system and many of the European systems in founding new departments and hiring new professors. In many countries of the Europeans systems, the undergraduate curricula of biological and engineering degrees are typically much less flexible than in the American system. Professors inherit undergraduate courses that have already been taught for many years by others and have a much higher teaching load than their American counterparts. This situation provides a disincentive to creating and teaching new courses, because those will take up more of their already short research time. With the recent drive towards unifying the university systems in the EU, caused by the application of the Bologna treaty, a redesign of undergraduate study programmes became

Table 1 Samples of systems biology related MSc educational programmes in Europe

University	Degree	Title	Web page	Duration	Country
U. Luxembourg	MSc	integrated systems biology		2 Years	Luxembourg
U. Bielefeld	MSc	genome-based systems biology	http://www.zfl.uni-bielefeld.de/studium/master-as/gbsb	2 Years	Germany
U. Gothenburg	MSc	systems biology	http://www.science.gu.se/english/education/master/systems_biology/	2 Years	Sweden
Tech. U. Denmark	MSc	systems biology	http://studyindenmark.dk/study-programmes/programmes-in-english/systems-biology	2 Years	Denmark
VU University Amsterdam	MSc	bio-molecular sciences	http://www.vu.nl/en/programmes/international-masters/programmes/a-b/biomolecular-sciences-msc/index.asp	2 Years	Holland
U. Aberdeen	MSc	systems biology	http://www.abdn.ac.uk/sysbio/	1 Year	Scotland
U. Evry-Val-d'Essonne/Ecole Centrale de Paris	MSc	systems and synthetic biology	http://www.mssb.fr/	1 Year	France
U. Skövde	MSc	systems biology	http://www.his.se/english/education/master-studies/masters-programmes/systems-biology---one-year-masters-degree/	1 Year	Sweeden
U. Vic	MSc	systems biology	http://www.uvic.es/node/248&wiki=Màster_Universitari_en_Biologia_de_Sistemes	1 Year	Spain
U. Freiburg	MSc	bioinformatics and systems biology	http://www.euroeducation.net/euro/freiburg_university_bioinformatics_systems_biology.htm	2 Years	Germany
Imperial College	MSc	bioinformatics and theoretical systems biology	http://www.findamasters.com/search/showcourse.asp?cour_id=13667	1 Year	England
U. Minho	MSc	bioinformatics	http://www.di.uminho.pt/ensino/mestrados/mestrado-em-bioinformatica	2 Years	Portugal
U. Complutense Madrid	MSc	bioinformatics and computational biology	http://bbm1.ucm.es/masterbioinfo/	1 Year	Spain
U. Pompeu Fabra	MSc	bioinformatics for health sciences	http://www.upf.edu/postgrau/en/masters/biomedicina/bioinfo/presentacio/index.html	2 Years	Spain

doi: 10.1049/iet-syb.2010.0032

necessary. This created a window of opportunity to introduce new subjects in those curricula.

One of the European Universities that took advantage of that opportunity to offer undergraduate courses in systems biology is the University of Lleida, in Spanish Catalunya. This University has provided its biotechnology majors with a third year course in bioinformatics where they receive a rather wide education in computational systems biology. (See Programme of the Course in Supplementary Information.)

3 Recent case study: support for an earlier exposition of students to systems biology education

This course runs for a whole semester (six European Credit Transfer System Units – ECTS) and it is divided into five modules. Owing to the stage the students are in, we argue for a wide scope of subject matters, favouring comprehensiveness over depth. The first module discusses genome sequencing and annotation, what computational methods are used for these endeavours, and how is the general logic that underlies their inner workings. In the second module, the students are taught about methods to predict protein and gene properties and structure. In the third module they are introduced to the concept that proteins need to work with other proteins in order to perform their physiological role. They are then introduced to the concept of omics data and faced with the different omics approaches that are available to systems biologists. In the fourth module they learn about the importance of data integration and how it can be used to perform in silico network reconstruction, sensu using the data to derive causal network topologies for the proteins and genes regulating and executing a given process. In the fifth and final module the students are taught how to use the reconstructed causal network to create a mathematical model that can be used to interrogate the system and generate hypothesis about its inner workings and responses. Two or three of the classes each year are given by invited professors who, in the flesh or through video-conferencing, talk to the students about their work in the relevant area of systems biology.

The course lectures are given in English. This course is divided into a fairly brief theoretical part (approximately 25 h of lectures), intermingled with five tasks that are designed to be performed in groups and in such a way that the concepts that are explained in the lectures have a clear application. Each task builds on the results of the previous ones. Fig. 1 shows the logical sequence of the tasks and Tables 2 and 3 provide more details. The students are organised into groups of three or four people. The problems are tailored individually for each group. First, each group is given a list of ten DNA sequences, out of which five are random ACTG strings and five belong to genes that participate in the same biological process. They are asked to identify the real genes and use their sequence to identify orthologues in the fully sequenced genome of another organism. Once these proteins are identified, they are set a second task, where they need to characterise their proteins and predict their structure. In the third task they are asked to, given their proteins, find orthologues in a set of organisms that is the same for all proteins. The results are then used to study the molecular evolution of the different proteins in the set of organisms, in order to assess if co-evolution by itself is an accurate method to infer functional relatedness between proteins. Their fourth

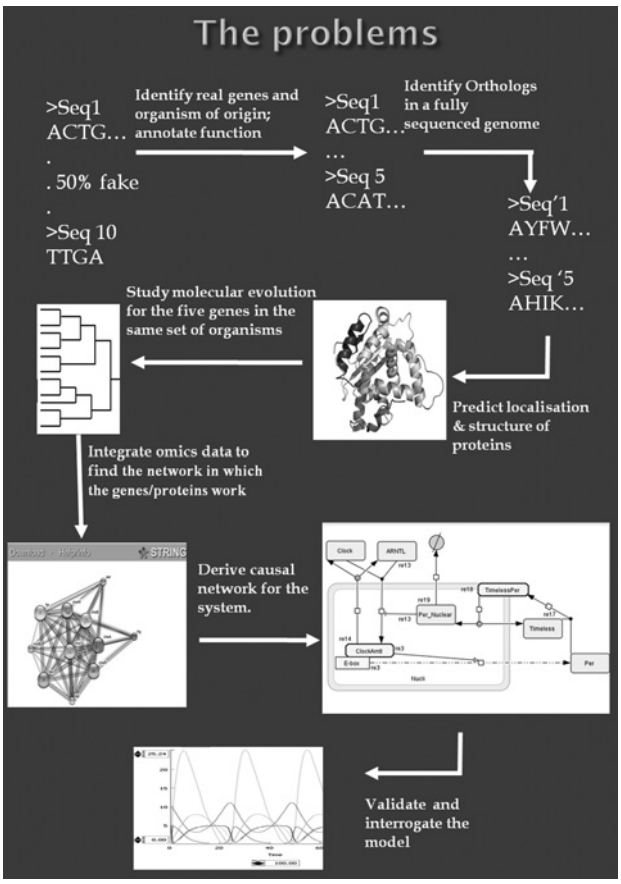

Fig. 1 *Logical sequence of the set of tasks used in the course*

task is to derive a network topology for the process in which their proteins are involved by combining different data from evolutionary and omics studies. Their final task is transforming this network topology into mathematical models that can be used to interrogate their understanding of the process.

An alternative to this set of problems would be to allow the students to choose, design and develop their own projects. If the course was focused on any one aspect of systems biology, this would have been our approach. However, and given that we clearly hoped to give exposure to the students to many different aspects and techniques, we felt that the approach proposed here is more time effective.

The students are evaluated at the end of each task. They execute these tasks in groups of three or four, whose elements remain constant throughout the semester. For each task, they are required to write a paper, with strict formatting and length limits, that mirrors a paper for a scientific journal. During each task, they are strongly encouraged to read manuals and instructions on how to use the programmes they need, and to find out the limitations of each of these programmes. They can easily consult with the teachers when they need to clarify some aspect of the work. In the end, they are also required to write a final paper where the whole story comes together as a complete work. This paper is then discussed individually with the students.

4 Evaluation of the case study

The methodology used in the course is designed to potentiate several aspects of the student's training. First, given the wide scope of subject matters during the course, they are exposed

Table 2 Tasks and programmes for the students

Task	Specific goals for the students	Description	Programmes and web pages
1. gene discovery and annotation	i. learn how to recognise genes from non-coding sequences	i. given 10 DNA sequences. 5 code for genes that participate in the same biological process	local blast ExPASy server KEGG server and ftp
	ii. learn and apply concept of sequence homology	ii. asked to identify the five DNA sequences that code real genes	NCBI server and ftp
	iii. learn to do homology search without a graphical interface	iii. students given an organism with fully sequenced genome	
	iv. recognise limitation of methods	iv. asked to identify the protein in new genome that are likely to be orthologues to those identified in description ii	
2. prediction of protein properties and structure	i. learn to create structural models for proteins	i. given proteins from task 1, predict post-translational modifications and cellular localisation	ExPASy server Protcomp SWISS-MODEL
	ii. learn to predict protein localisation	ii. create structural models for each protein	3D-JIGSAW Robetta
	iii. recognise the limitations of available methods	iii. analyse models	
3. analysis of the molecular evolution of functionally related proteins	i. understand basic concepts of molecular evolution	i. take the proteins from earlier tasks and find orthologues in the same 15 organisms for all of them	NCBI KEGG MEGA4
	ii. understand concepts of co-evolution between cooperating proteins	ii. create a multiple alignment for each set of proteins	PHYLIP
	iii. learn to create multiple alignments	iii. create phylogenetic trees using different methods, compare them and interpret the results	
	iv. learn to create phylogenetic trees using different methods	iv. compare the evolution of the five proteins in the same set of proteins	
4. integration of data to reconstruct protein networks	i. understand the different types of Omics experiments	i. given the set of proteins, use data from co-evolution, Omics experiments and bibliography to infer the network in which the original proteins participate	iHOP STRING NCBI
	ii. understand the problems that underlie data integration		
	iii. use data integration to create a network of proteins that participate in a given process		
5. creation and analysis of a model for the network	i. understand the limitation of interaction networks as predictors for systemic behaviour	i. transform network from task 4 into a causal network	COPASI CELLDESIGNER PLAS
	ii. understand the difficulties of creating a mathematical model of a process	ii. create mathematical model for the causal network. Use both approximate and exact formalisms, depending on the case	
	iii. develop the ability to create, analyse and validate models	iii. interrogate and validate the model	

to many different areas of the field. Second, they are taught to synthesise information and emphasise clearness in their reports. Third, their learning is based on a combination of problem-based learning and classical lectures that appear to give a good equilibrium to their education. For more details on the methods, materials and problems used in the course, readers are directed to http://web.udl.es/usuaris/pg193845/Courses/Bioinformatics_2009/index.htm. Finally, they learn to work effectively in a group and solve problems by themselves.

This is the third year in a row that this course has been run. All students have approved the course in the three years and the lowest grade so far has been six out of ten (Fig. 2). Three independent yearly surveys of the students (by the university services and by the professors) reveal that all students felt the course to be useful and were satisfied with what they learned and how they learned it. However, they also felt that solving the problems took them longer than initially expected. Of the approximately 90 students who finished the course, 10% expressed interest in following up their training in the area. Five per cent have done so by taking short research stays in systems biology labs before they graduate and an additional five have followed up to enter systems-biology-related PhD programmes and/or Masters degrees.

5 Take home message

Overall, this experience supports the notion that it is possible and desirable to start educating systems biologists from an early stage of their career, namely at the undergraduate level, and that the final years of a bachelor's degree may be a good time for introducing students to the subject. Given that this course is given within a degree that is adapted to

doi: 10.1049/iet-syb.2010.0032

Table 3 Systems to reconstruct

System	Organisms	Degree of success
mitochondrial fes cluster biogenesis	*Saccharomyces cerevisiae, Homo sapiens, Pans troglodytes*	all groups were able to create a model and run simulations from the bottom up. All groups were able to create partially valid models
cyanobacterial circadian clock	*Synechococcus elongates*	all groups were able to create a model and run simulations from the bottom up. No group was able to create a validated model
mammalian circadian clock	*Homo sapiens, Pans troglodytes*	all groups were able to create a model and run simulations from the bottom up. One group was able to create a valid model for cell cycle, from the bottom up and using approximate formalisms, with periodic oscillations that ran with periods of 22–25 h
mammalian cell cycle	*Homo sapiens*	all groups were able to create a model and run simulations from the bottom up. No group was able to create a valid model
fungal cell cycle	*S. cerevisiae, Schizosaccharomyces pombe*	all groups were able to create a model and run simulations from the bottom up. No group was able to create a valid model

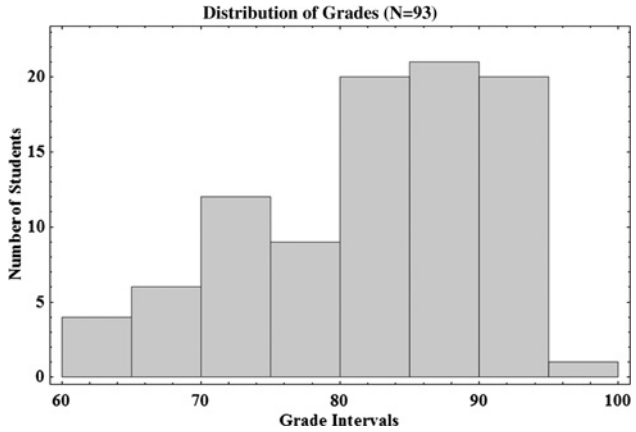

Fig. 2 *Distribution of student grades in the last three years*

the new European Superior Education Space, it is likely that the methodologies and conclusions discussed here are generalisable and applicable within other EU countries.

The results of this course also emphasise the ability of undergraduate students to perform fairy advanced research tasks with little overall supervision. These students are one of the most underused research resources in most Universities and perhaps new institutional policies should be put in place in order to tap this resource.

6 Acknowledgments

Rui Alves was partially supported by the Ministerio de Ciencia e Innovación (MICINN, Spain through the Ramon y Cajal program and grants BFU2007-62772/BMC and BFU2010-17704) and by the FLAD foundation during a short stay at the Computational Biology Collaboratorium. We thank the students of the UdL Bioinformatics course, 2007–2009, for their enthusiasm and feedback. This work would not have been possible without them. Rui Alves would also like to acknowledge David Fell for an invitation to speak about this work at the 2009 FEBS meeting in Prague. This paper is dedicated to the memory of Prof. Ruy E. Pinto, who introduced R. Alves to systems biology as an undergraduate, as he majored in biochemistry from the University of Lisbon in the late 1980s to early 1990s. We thank the referees who, through their comments and suggestions, improved the quality of this paper.

7 References

1 Bernard, C.: 'Introduction à l'étude de la médecine expérimentale', Flammarion, 2008, **1865**
2 Bertalanffy, L.v., Woodger, J.H.: 'Modern theories of development; an introduction to theoretical biology' (Oxford University Press, 1933)
3 Omholt, S.W.: 'Cell biology – foundations of systems biology', *Science*, 2002, **295**, (5563), p. 2220
4 Mesarovic, M.D.: Case Institute of Technology, Systems Research Center: 'Systems theory and biology'. Proc. Third Systems Symp. on Case Institute of Technology, 1968
5 Savageau, M.A.: 'Biochemical systems analysis: a study of function and design in molecular biology' (Addison-Wesley Pub. Co., Advanced Book Program, 1976)
6 Randal, J.: 'The human genome project', *Lancet*, 1989, **2**, (8678–8679), pp. 1535–1536
7 Duncan, M.W.: 'Omics and its 15 minutes', *Exp. Biol. Med. (Maywood)*, 2007, **232**, (4), pp. 471–472
8 Kiechle, F.L., Zhang, X., Holland-Staley, C.A.: 'The-omics era and its impact', *Arch. Pathol. Lab. Med.*, 2004, **128**, (12), pp. 1337–1345
9 Butcher, E.C., Berg, E.L., Kunkel, E.J.: 'Systems biology in drug discovery', *Nat. Biotechnol.*, 2004, **22**, (10), pp. 1253–1259
10 Chang, M.C., Keasling, J.D.: 'Production of isoprenoid pharmaceuticals by engineered microbes', *Nat. Chem. Biol.*, 2006, **2**, (12), pp. 674–681
11 Hopkins, A.L.: 'Network pharmacology', *Nat. Biotechnol.*, 2007, **25**, (10), pp. 1110–1111
12 Vivona, S., Gardy, J.L., Ramachandran, S., *et al.*: 'Computer-aided biotechnology: from immuno-informatics to reverse vaccinology', *Trends Biotechnol.*, 2008, **26**, (4), pp. 190–200
13 'Systems biotechnology allows development of 100% rational amino acid producer', *Biotechnol. J.*, 2007, **2**, (8), p. 927
14 Carothers, J.M., Goler, J.A., Keasling, J.D.: 'Chemical synthesis using synthetic biology', *Curr. Opin. Biotechnol.*, 2009, **20**, (4), pp. 498–503
15 Lee, S.K., Chou, H., Ham, T.S., Lee, T.S., Keasling, J.D.: 'Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels', *Curr. Opin. Biotechnol.*, 2008, **19**, (6), pp. 556–563
16 Petranovic, D., Vemuri, G.N.: 'Impact of yeast systems biology on industrial biotechnology', *J. Biotechnol.*, 2009, **144**, (3), pp. 204–211
17 Wang, Y., Chu, J., Zhuang, Y., Xia, J., Zhang, S.: 'Industrial bioprocess control and optimization in the context of systems biotechnology', *Biotechnol. Adv.*, 2009, **27**, (6), pp. 989–995
18 Cases, I., de Lorenzo, V.: 'Genetically modified organisms for the environment: stories of success and failure and what we have learned from them', *Int. Microbiol.*, 2005, **8**, (3), pp. 213–222

19 French, C.E.: 'Synthetic biology and biomass conversion: a match made in heaven?', *J. R. Soc. Interface*, 2009, **6**, (Suppl 4), S547–558
20 Rylott, E.L., Bruce, N.C.: 'Plants disarm soil: engineering plants for the phytoremediation of explosives', *Trends Biotechnol.*, 2009, **27**, (2), pp. 73–81
21 de Lorenzo, V.: 'Systems biology approaches to bioremediation', *Curr. Opin. Biotechnol.*, 2008, **19**, (6), pp. 579–589
22 de Lorenzo, V.: 'Recombinant bacteria for environmental release: what went wrong and what we have learnt from it', *Clin. Microbiol. Infect.*, 2009, **15**, (Suppl 1), pp. 63–65
23 Tan, T.W., Lim, S.J., Khan, A.M., Ranganathan, S.: 'A proposed minimum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era', *BMC Genomics*, 2009, **10**, (Suppl 3), S36
24 Wolkenhauer, O., Auffray, C., Baltrusch, S., *et al.*: 'Systems biologists seek fuller integration of systems biology approaches in new cancer research programs', *Cancer Res.*, 2010, **70**, (1), pp. 12–13
25 Ideker, T., Winslow, L.R., Lauffenburger, D.A.: 'Bioengineering and systems biology', *Ann. Biomed. Eng.*, 2006, **34**, (7), pp. 1226–1233

doi: 10.1049/iet-syb.2010.0032

Supplementary Materials

Program of the Course

1. Origins & Development of Bioinformatics & Systems Biology.

- Historical development of tools for computational approaches in biology. Data accumulation and expansion of computer access as *leitmotifs* for the spread of bioinformatics. The reductionist vs. integrative paradigm and computational systems biology.

2. General description of Bioinformatics and Computational Biology methods and applications

- A summary of what computational methods can and are needed to do at different levels of biological analysis. From the genome to the ecosystem.

3. Basic tools in Bioinformatics

- Databases and Database technology.
- Resource integration and data-mining.
- Functional classifications and ontologies.

4. Genome sequencing and annotatation

- Bioinformatics of gene discovery and annotation.
- Metagenomics.
- Alternative genetic codes.

5. Extracting information from the primary structure of genes and proteins

- Methods for predicting RNA structure.
- Methods for predicting protein localization and physical-chemical properties.

6. Sequence alignments and molecular evolution

- Sequence alignment methods. Heuristic vs. exact methods. Dynamic programming.
- Basic concepts in molecular evolution. Multiple sequence alignment.
- Phylogenetic trees. Alternative methods for construction. Interpretation.

7. Omics experiments

- Introduction to omics. Omics approaches as tools to analyze integrated systemic behavior.
- Genomics.
- Proteomics.
- Metabolomics.
- Fluxomics.

8. Biological network reconstruction and analysis

- Form the gene to the system.
- Predicting network structure. Using bibliomics and analysis of experiments.
- Predicting network structure. Using comparative evolution methods.
- Predicting network structure. Using docking.
- Predicting network structure. Motif identification.

9. Predicting the behavior of biological networks

- Going from graphs to causal network representations.
- Characterizing the behavior of causal networks: the mathematical model.
- Alternative mathematical models: Finite state models vs. differential equations
- Alternative mathematical models: Homogeneous models vs. models with spatial differentiation.
- Alternative mathematical models: Stochastic models vs. deterministic models.
- Mathematical representations: Mass action, classical enzyme kinetics, and approximated formalisms. When to use each of them.
- Sensitivity analysis and robustness of biological systems.
- Steady state and stability.

10. Biological design principles

- The concept of biological design principles from an engineering perspective. Different types of design principles.
- Design principles in metabolism, gene expression and signal transduction.
- Methods to identify and analyze design principles. Mathematically controlled comparisons.
- Applications in synthetic biology.
- Applications in biotechnology.