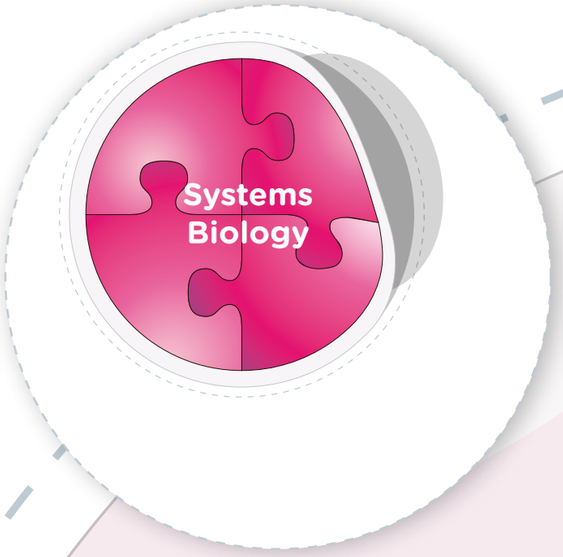
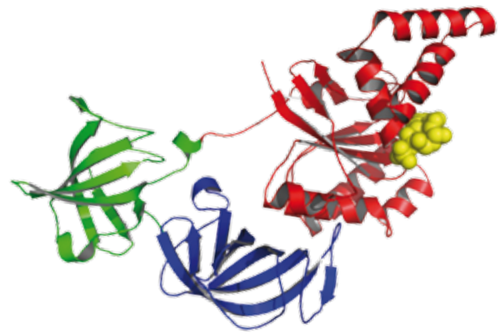


One of the great conceptual novelties of recent years in life sciences research is so-called Systems Biology (and its technological ramification, Synthetic Biology). Systems biology seeks to address the complexity of living systems as such, rather than divide them into smaller parts (unlike the extreme reductionism of traditional molecular biology). The issue is not only understanding, but also modifying and ultimately recreating novel biological systems through the application of a variety of tools and expertise derived from the physical sciences, engineering, biology, computer science and mathematics. Systems biology thus addresses naturally evolved biological systems in a holistic context that considers the emergent properties of layered levels of complexity. On the other hand, synthetic biology tends to focus on the application of the engineering cycle of design, modelling, construction and testing, with a certain emphasis on strong manufacturing concepts including standardization and abstraction. The scientific and technological potential of systems and synthetic biology are immense, in the fields of biomedicine and industrial, agricultural and environmental biotechnology. This has led to the creation of new programmes and centres devoted to the discipline in many of the most respected international research organisations. As a reaction to this challenge, the CNB is leading an ambitious initiative to translate the existing critical mass in these fields into a sound, in-house research area.



Systems Biology

Molecular Environmental Microbiology

LEAD INVESTIGATOR
Víctor de Lorenzo

ASSOCIATED SCIENTIST
Javier Tamames

POSTDOCTORAL SCIENTISTS
Belén Calles
Gonzalo Durante
Esteban Martínez
Pablo Nikel
Raúl A. Platero
Danilo Pérez

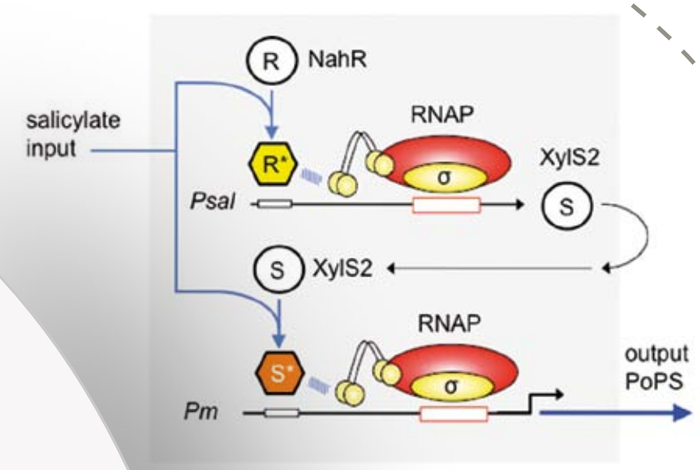
PREDOCTORAL SCIENTISTS
Alejandro Arce
Ilaria Benedetti
Max Chavarría
Juhyun Kim
Rafael Silva

TECHNICIANS
Sofía Fraile

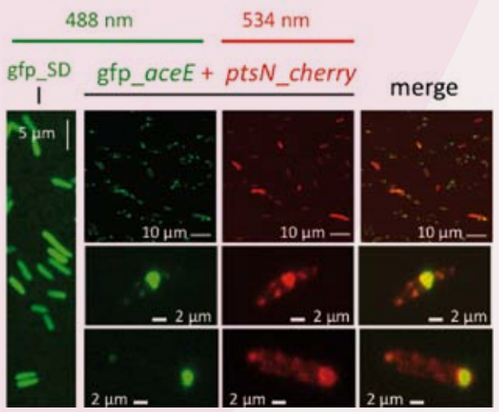
ADMINISTRATIVE ASSISTANT
Inés Merino

Our laboratory is committed to understanding how bacteria that inhabit natural niches sense and process multiple environmental signals into distinct responses –both at the level of single cells and as a community.

Unlike laboratory settings, in which growth conditions can be controlled and changed one at a time, bacteria in the environment must perpetually make decisions between activating metabolic genes for available, frequently mixed C-sources and those for escaping or adapting to physicochemical stress. Our preferred experimental system involves the strain KT2440 of the soil and plant root colonizer *Pseudomonas putida* bearing the plasmid pWW0, which allows growth on toluene and m-xylene as the only C and energy source. The biotechnological side of this biological question is the possibility of programming bacteria for deliberate environmental release, aimed at biodegradation of toxic pollutants or as biosensors for monitoring the presence of given chemicals. Apart from understanding and developing such sensor or catalytic bacteria, their release requires the GMO to be endowed with a high degree of containment and predictability. In this line, our research takes on board the development of novel molecular tools for the genetic analysis and construction of soil microorganisms (mostly *Pseudomonads*) destined for the environment or as catalysis for selected biotransformations. We recently became active in the interface between synthetic biology and environmental microbiology as a source of new tools for addressing some outstanding environmental pollution problems. We are currently developing technologies for deep genetic engineering of *P. putida* and other environmental bacteria. These have allowed both a complete understanding of various catalytic and physiological processes and have opened up exciting opportunities for a radical genomic refactoring of the corresponding biological agents. One of the outcomes involved the design of bacteria able to translate the presence of residues of explosives in soil (which can be traced by following the fate of 2,4 dinitrotoluene) into a luminescent signal. At the same time, we studied the logic programs that underlie many of the metabolic and regulatory networks that endow *P. putida* (and other bacteria) with the ability to endure harsh environmental conditions, and found most of them amenable to modelling with Boolean formalisms. Since digital circuits based on logic operations are the basis of computation, we recently began to develop a suite of basic, connectable Boolean gates using regulatory and metabolic components aimed at programming cells to behave in a predetermined way.



The NahR/XylS2 feed-forward loop. The scheme shows the first synthetic regulatory cascade for amplification of gene expression reported in the literature. The design included the salicylate-responsive activators NahR of and XylS2 from two Pseudomonas strains, and their respective cognate promoters Psal and Pm. Expression of xylS2 was placed under the control of the nahR/Psal system. When cells face the common effector of the two regulators, both the increase in XylS2 concentration and the stimulation of its activity with salicylate act synergistically on the Pm promoter (having PoPS as its output, see above). This type of regulatory layout could be later classified and formalized as a feed-forward loop (FFL).



Co-localization of PtsN and the pyruvate dehydrogenase AceE subunit in the cytoplasm of Pseudomonas putida cells. P. putida MAD2 was either transformed with pVLT_gfp_SD or co-transformed with pIZ_gfp_aceE plus pVLTptsN_cherry, expressing respectively red fluorescence-labelled AceE and green-fluorescence labelled PtsN. Cells were grown in LB with 500 µM IPTG to the late exponential growth phase and analyzed in a confocal laser microscope. Fields shown were then excited at 488 nm and 534 nm for emission of green and red fluorescence, respectively, in order to distinguish the production and location of the corresponding proteins. Note the uniform distribution of GFP in the cytoplasm of the control construct (lane 1), while Ace and PtsN (lanes 2 and 3) localize in distinct foci that can be considerably overlapped in the merged images (lane 4).

SELECTED PUBLICATIONS

de Lorenzo V (2010). Environmental biosafety in the age of Synthetic Biology: Do we really need a radical new approach? **BioEssays** 32:926-31.

de las Heras A, Carreño CA, Martínez E, de Lorenzo, V (2010) Engineering input/output nodes in prokaryotic regulatory circuits. **FEMS Microbiology Reviews** 34:842-65.

Tamames J, de Lorenzo V (2010) EnvMine: A text-mining system for the automatic extraction of contextual information. **BMC Bioinformatics** 11:294.

Silva-Rocha R, de Lorenzo V (2010) Noise and robusteness in prokaryotic regulatory Networks. **Ann Rev Microbiol** 64:257-275.

Páez-Espino D, Tamames J, de Lorenzo V, Canovas D (2009) Microbial responses to environmental arsenic. **BioMetals** 22:117-130.

PATENT

E Castellón, M Chavarría, V de Lorenzo, M Zayat & D Levy. Combinación de una bio-película bacteriana y cristal líquido para la preparación de un dispositivo electro-óptico. Spanish patent. Register number **ES1641.698.**

Systems Biology

Bioinformática de Sistemas

LEAD INVESTIGATOR

Florencio Pazos Cabaleiro

POSTDOCTORAL SCIENTIST

Monica Chagoyen Quiles

PHD STUDENTS

Daniel Lopez Lopez
David Ochoa Garcia
Abhimanyu Singh
Natalia Pietrosemoli

TECHNICIAN

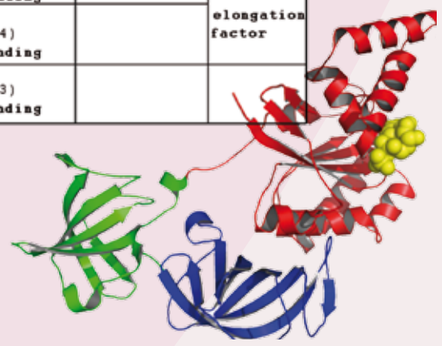
Juan Carlos Sanchez Ferrero

Our group is interested in different aspects of Bioinformatics, Computational Biology and Systems Biology.

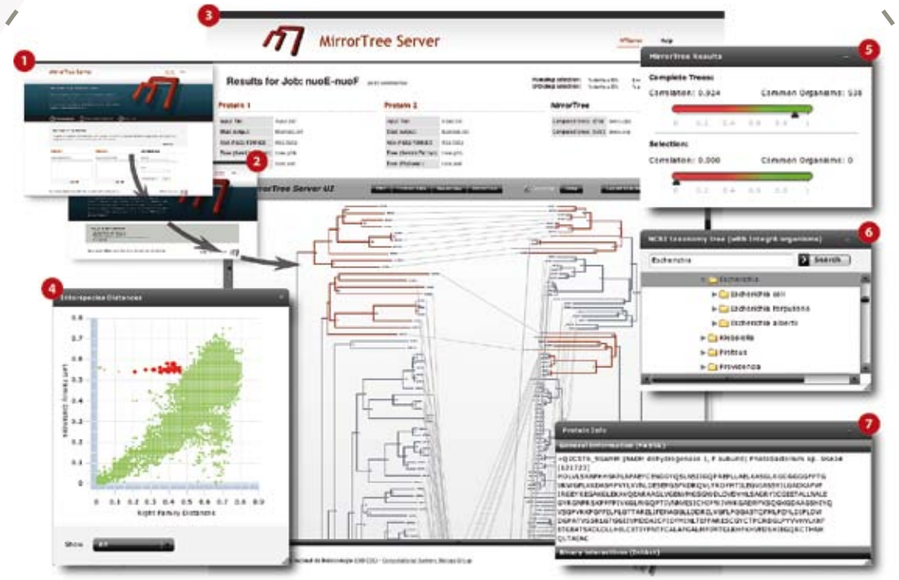
Our goal is to obtain new biological knowledge with an "in-silico" approach which complements the "*in-vivo*" and "*in-vitro*" methodologies of Biology. This mainly involves mining the massive amounts of information stored in biological databases. Besides our lines of scientific research, we also collaborate with experimental groups providing them with bioinformatics support for their specific needs, and participate in different teaching projects.

We have developed evolutionary-based method for predicting sites with some functional importance in protein sequences and structures. Experimental determination of functional/active sites can not cope with the massive stream of new sequences coming from genome sequencing projects. Hence, computational methods are highly demanded for this task. The methods we develop in this area are based on the fact that functional sites are subject to certain evolutionary constraints whose landmarks can be detected on multiple sequence alignments.

Domain	Pfam/Interpro	scop2go	
d1jnyA3	(PF00009) GTP-binding	GTP-binding	elongation factor
d1jnyA1	(PF03144) GTP-binding		
d1jnyA2	(PF03143) GTP-binding		



Results of the "scop2go" method for transferring functional annotations to the domain level for the Sulfolobus solfataricus elongation factor 1α (PDB: 1jny_A). Comparison with the corresponding annotations inferred from Pfam/Interpro, which incorrectly assign the 3 domains as "GTP-binding". The GTP bound to the N-terminal (red) domain (SCOP:d1jnyA3) is shown in yellow.



Screenshots of the interface of the Mirrortree server for studying the co-evolution between protein families.

The biological functions of many proteins can only be explained in the context of their relationships with others. Experimental techniques for the determination of interaction partners are still far from perfect and computational methods for predicting pairs of proteins which interact or are functionally associated have emerged. We have developed evolutionary-based methods for predicting interaction partners which have been accepted and followed by the community. These methods are mainly based on the hypothesis that interacting or functionally related proteins adapt to each other during the evolutionary process (co-evolution). We try to detect the landmarks that this co-evolutionary process left in the sequences and structures of the proteins.

SELECTED PUBLICATIONS

Daniel Lopez & Florencio Pazos. (2009). Gene ontology functional annotations at the structural domain level. **Proteins.** 76(3):598-607.

Juan C. Triviño & Florencio Pazos. (2010). Quantitative global studies of reactomes and metabolomes using a vectorial representation of reactions and chemical compounds. **BMC Syst Biol.** 4:46.

David Ochoa & Florencio Pazos. (2010). Studying the co-evolution of protein families with the Mirrortree web server. **Bioinformatics.** 26(10):1370-1371.

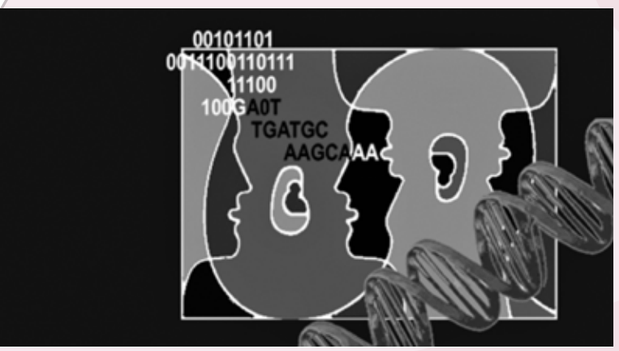
Antonio Rausell, David Juan, Florencio Pazos & Alfonso Valencia. (2010). Protein interactions and ligand binding: From protein subfamilies to functional specificity. **Proc Natl Acad Sci USA.** 107(5):1995-2000.

Monica Chagoyen & Florencio Pazos. (2010). Quantifying the biological significance of gene ontology biological processes - implications for the analysis of systems-wide data. **Bioinformatics.** 26(3):378-384.

Systems Biology

LEAD INVESTIGATOR
Juan F. Poyatos

POSTDOCTORAL SCIENTIST
Matteo Cavaliere

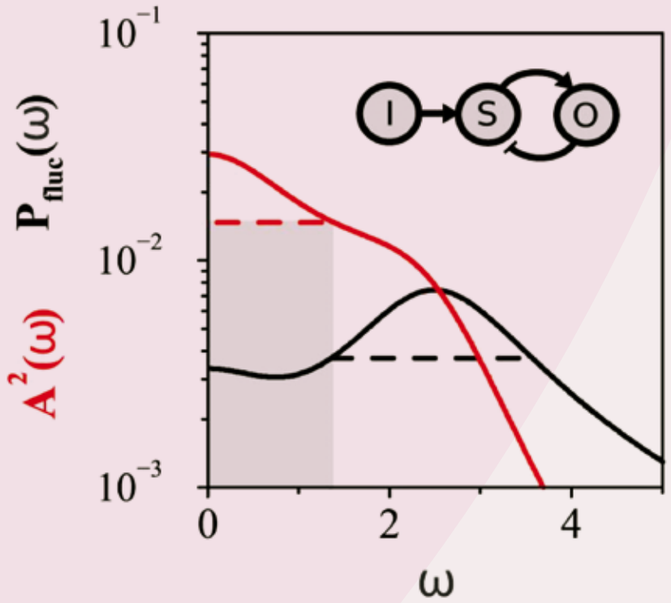


PREDOCTORAL SCIENTISTS
Juan Barrero
Djordje Bajic
Clara Moreno

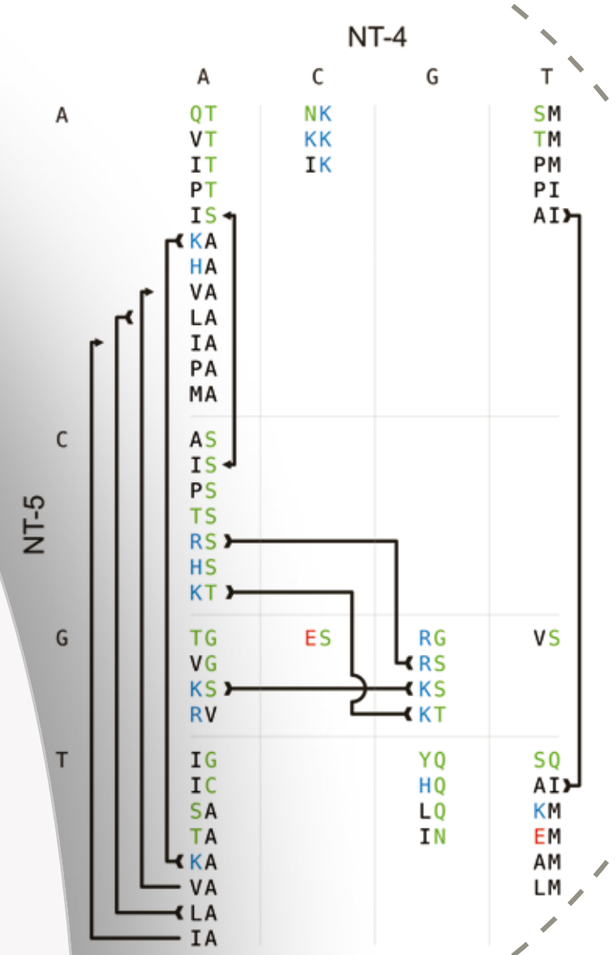
Logic of Genomic Systems

Signals continuously impinge on cells, modifying their behaviour.

This can be understood at three levels. At a population level, signals could modify the distribution of cell classes in a population, influencing core cell processes such as cell renewal in stem cell niches. At the intracellular pathway level, signals are effectively sensed and processed by combinations of genetic circuits. An open question is how these circuits distinguish signal from noise, and how circuit structure might limit this ability. Finally, at the signal response level, transcriptional control is fundamental to activate adequate responses. This implies the ability of transcriptional factors to distinguish specific nucleotide sequences in the genome. In the last two years, we have analysed these three aspects of signal processing in the lab.



Signals can be discriminated from fluctuations when their characteristic frequencies are different from the noise frequency content. In this case, a negative feedback circuit whose fluctuation bandwidth (dashed black line) is always at higher frequencies than the oscillatory bandwidth (red dashed line), allowing for a noise-free frequency regime in transmitted oscillations (grey shaded region).



Presence of a wide-coverage recognition code within the LacI family. AA sequences (AA-15,AA-16) recognising a same sequence of NT (NT-5,NT-4) were grouped. Here, we only considered significant palindromic NT sequences. Recognition degeneracies are represented as unidirectional arrows (asymmetrical intrinsic), bidirectional divergent arrows (symmetrical intrinsic), and bidirectional convergent arrows (extrinsic). Polar (green), basic (blue), acidic (red) and hydrophobic (black) amino acids.

1) Signals and populations. Competition for the survival factor Dpp leads to the proliferation of one class of cells (the winners) at the expense of other (loser) cells; both types exhibit normal growth in homotypic environments. Recent studies in Drosophila demonstrated the role of the dMyc protein in this process. We examined competition in the (Drosophila) ovary stem-cell niche, and showed that differential expression of dMyc triggers competitive interactions. We also presented data in support of the hypothesis that such ongoing competition –between high dMyc stem cells and low dMyc differentiating daughters– increases the efficiency of the differentiation program.

2) Signals and genetic circuits. We considered different types of signals acting on a two-component module to theoretically analyse information processing by genetic circuits. We showed that the presence of feedback in the module imposes a trade-off on amplitude and frequency detection. A direct interaction between the signal and the output species, in a type of feed-forward loop architecture, greatly modifies these trade-offs. Our study emphasised the limits imposed by circuit structure on its stimulus response, and the paradoxical advantage of improving detection with noisy circuit components.

3) Signals and transcriptional control. Transcription factors (TF) commonly act as effectors of cell signal processing by binding DNA sequences adjacent to the response genes whose production they regulate. Could a wide-coverage recognition code between interacting amino acids (of the TF) and nucleotides be found? Our analysis suggested that a set of relatively consistent recognition rules does apply for the extensive LacI family of TF. These rules could ultimately act as a blueprint for the synthetic redesign of TF with new specificities.

SELECTED PUBLICATIONS

Persistent competition among stem cells and their daughters in the Drosophila ovary germline niche. **Development** 136:995.

Guantes R, Estrada J, Poyatos JF (2010) Trade-offs and noise tolerance in signal detection by genetic circuits. **PLoS One** 5:e12314.

Camas FM, Alm EJ, Poyatos JF (2010) Local gene regulation details a recognition code within the LacI transcriptional factor family. **PLoS Comput Biol**, 6:e1000989.