Abstracts

Ard Louis University of Oxford
Algorithmic information theory and GP maps

Many GP maps exhibit strong bias in the number of genotypes per phenotype [1]. This bias in the arrival of variation can dominate evolutionary outcomes [2] and may help explain some kinds of evolutionary convergence [3]. Here we argue that concepts from algorithmic information theory provide a generalised explanation for the bias.

[1] A tractable genotype-phenotype map for the self-assembly of protein quaternary structure, Sam F. Greenbury, Iain G. Johnston, Ard A. Louis, Sebastian E. Ahnert, J. R. Soc. Interface 11, 20140249 (2014)
[2] The structure of the genotype-phenotype map strongly constrains the evolution of non-coding RNA, Kamaludin Dingle, Steffen Schaper, and Ard A. Louis, Interface Focus 5: 20150053 (2015)
[3] Contingency, convergence and hyper-astronomical numbers in biological evolution, Ard A. Louis, Studies in History and Philosophy of Biological and Biomedical Sciences 58, 107 (2016)

Marjon de Vos Wageningen University
Breaking through evolutionary constraint by environmental fluctuations

Suboptimal fitness peaks are generally recognized as causing evolutionary stasis. Here, we show that these constraints can be overcome in an adaptive manner by reconstructing mutational trajectories for a transcription factor and its DNA binding site in variable environments. Cross-environmental tradeoffs, typically associated with evolutionary limitations, are an essential enabling component of this evolutionary mechanism. Our results underscore the importance of characterizing environmental dependencies when studying genetic interactions and provide the clearest indication so far that environmental variability can accelerate evolution hampered by stasis in constant conditions. Given that environmental variations and tradeoffs are ubiquitous, this evolutionary mechanism may be relevant to a wide range of genetically constrained phenotypes and major evolutionary transitions.

Joshua Payne University of Zürich
Exhaustively-enumerated genotype-phenotype maps in transcriptional regulation

Transcriptional regulatory circuits drive the development, physiology, and behavior of organisms across the tree of life. Their phenotypes are gene expression patterns - i.e., when, where, and to what extent the circuit's genes are expressed. These phenotypes can be mapped to genotypes that comprise the coding regions of the circuit's transcription factors, as well as the non-coding regulatory regions that bind these factors. Recent modeling efforts have produced exhaustively-enumerated genotype-phenotype maps for small regulatory circuits, which have provided new insights into the design constraints of such circuits and the extent to which circuit phenotypes can be inferred from circuit designs. At the same time, advances in high-throughput sequencing technologies have facilitated the construction of empirically-derived, exhaustively-enumerated genotype-phenotype maps for an important subcomponent of transcriptional regulatory circuits: transcription factor-DNA interactions. In this talk, I will discuss these recent advances and how they have helped us to better understand how transcriptional regulatory circuits evolve.

Chico Camargo University of Oxford
Gene networks as a GP map

Recent work on macromolecule sequence to structure genotype-phenotype (GP) maps has found many similar system level properties, such as a large bias in the distribution of genotypes mapping to each phenotype, genetic correlations that lead to large robustness and mutation probabilities to new phenotypes that are proportional to the fraction of genotype space that they occupy. Here we study gene regulatory networks using using Boolean threshold models as well as differential equations models. We find that these GP maps have many similar properties to those found for the or macromolecule sequence-to-structure GP maps. Finally, we relate these results to concepts from algorithmic information theory, discussing how these GP map properties can strongly affect evolutionary outcomes.

Michael Stich Aston University
Replicator dynamics on an RNA fitness landscape under the influence of mutation and recombination

The map from RNA sequence to secondary structure has been studied intensively as an example of a relatively simple, yet sufficiently complex genotype-phenotype map. Based on this map, fitness functions can be defined and the resulting evolution of a replicating population can be studied. We review some basic properties of such systems, in particular the search and fixation times for populations evolving towards a given target structure. We then consider some recent results where recombination is included in addition to point mutations. We focus on two situations: first, a population constrained to a neutral network (sequences folding into the same secondary structure). The rules of recombination determine the topological properties of the network which are, however, different from those of a mutational neutral network. Second, we consider a population able to explore the whole sequence space and subjected to a more complicated fitness function.

Jacobo Aguirre CNB-CSIC Madrid
Tipping points in the genetic composition of populations induced by environmental changes

Nature has been exposed during Earth's history to a wide variety of gradual changes in environmental and climate conditions. Recent studies hypothesize that these smooth alterations might be responsible for some of the drastic state shifts that have been reported in our planet's biosphere [1, 2]. The question that ultimate faces this new line of research is whether the local impact of humans could provoke a planetary-scale tipping point in the nearby future.

Our work [3] deals with the possibility of observing such drastic changes in the composition of populations at the genotypic level, e.g. RNA populations. In our study, a graph represents the space of genotypic sequences of length N, where every locus (or nucleotide in the case of RNA) is a state taken from a genetic alphabet of A letters: each node stands for a different sequence and two nodes are connected by an undirected link if they differ in the state of only one locus [4]. We use a modification of the widely known NK model to map a rugged fitness landscape to such sequences. Finally, we model the evolution of the environment with time via perturbations in the fitness landscape.

Our work shows analytically and numerically that not only gradual and monotonic changes in environmental conditions, but also totally random fluctuations of very small amplitude, generically give rise to critical transitions in the genetic composition of populations for a wide range of situations. In summary, our results extend the studies on planetary tipping points to the genotypic scale, where it had not been studied to the date.

[1] Approaching a state shift in Earth's biosphere, A.D. Barnosky, E.A. Hadly, J. Bascompte et al., Nature 486, 52-58 (2012).
[2] Does the terrestrial biosphere have planetary tipping points?, B.W. Brook, E.C. Ellis, M.P. Perring et al., Trends in Ecology & Evolution 7, 396-401 (2013).
[3] Tipping points and early warning signals in the genomic composition of populations induced by environmental changes, J. Aguirre and S. Manrubia, Scientific Reports 5, 9664 (2015).
[4] Evolutionary dynamics on networks of selectively neutral genotypes: Effects of topology and sequence stability, J. Aguirre, J. M. Buldú, and Susanna C. Manrubia, Physical Review E 80, 066112 (2009).

Pablo Catalan University Carlos III, Madrid
toyLIFE, or the importance of being promiscuous

Functional promiscuity is becoming one of the most relevant evolutionary mechanisms in recent years. Most computational models of the genotype-phenotype map show functional promiscuity in varying degrees. In this talk I will show how toyLIFE, a multi-level model of the genotype-phenotype map, contains considerable functional promiscuity, allowing us to explore some of its evolutionary consequences.

Susanna Manrubia CNB-CSIC Madrid
A generic distribution of phenotype sizes results from the organization of biological sequences into sites with varying neutrality

The distribution of phenotype sizes is very broad in all model systems studied to date including, among others, the RNA sequence-to-secondary-structure map, the hydrophobic-polar model for protein folding, or toyLIFE. It has been suggested (Greenbury and Ahnert, 2015) that this broad shape results from the organization of biological sequences into constrained and unconstrained parts. In this contribution we show that the combination of different degrees of neutrality for sequence sites, together with the combinatorics associated to the placement of those sites in different positions along the sequence implies a lognormal distribution of phenotype sizes. Our results are applied to examples of exhaustively folded sequence spaces for RNA (Aguirre et al., 2011; Dingle et al., 2015) and the HP model (Irbäck and Troein, 2002).

References:
Aguirre, J. M. Buldú, M. Stich, S. C. Manrubia. Topological structure of the space of phenotypes: the case of RNA neutral networks. PLoS ONE 6, e26324, 2011.
K. Dingle, S. Schaper, A. A. Louis. The structure of the genotype-phenotype map strongly constrains the evolution of non-coding RNA. J. Roy. Soc. Interface 5, 20150053, 2015.
S. F. Greenbury, S. E. Ahnert. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype-phenotype maps. J. Roy. Soc. Interface 12, 20150724, 2015.
A. Irbäck and C. Troein. Enumerating designing sequences in the HP model. J. Biol. Phys. 28, 1-15, 2002.

Jose Cuesta University Carlos III, Madrid
Evolution on neutral networks accelerates the ticking rate of the molecular clock

Large sets of genotypes give rise to the same phenotype, because phenotypic expression is highly redundant. Accordingly, a population can accept mutations without altering its phenotype, as long as the genotype mutates into another one on the same set. By linking every pair of genotypes that are mutually accessible through mutation, genotypes organize themselves into neutral networks (NNs). These networks are known to be heterogeneous and assortative, and these properties affect the evolutionary dynamics of the population. By studying the dynamics of populations on NNs with arbitrary topology, we analyse the effect of assortativity, of NN (phenotype) fitness and of network size. We find that the probability that the population leaves the network is smaller the longer the time spent on it. This progressive 'phenotypic entrapment' entails a systematic increase in the overdispersion of the process with time and an acceleration in the fixation rate of neutral mutations. We also quantify the variation of these effects with the size of the phenotype and with its fitness relative to that of neighbouring alternatives.

Reference: Evolution on neutral networks accelerates the ticking rate of the molecular clock, Susanna Manrubia and José A. Cuesta, J. Roy. Soc. Interface 102, 20141010 (2015)

Esther Ibañez-Marcelo ISI Foundation, Turin
Surviving evolutionary escape on complex genotype-phenotype networks

We study the problem of evolutionary escape and survival for cell populations with genotype-phenotype map. That is the process whereby a population under sudden changes in the selective pressures acting upon it try to evade extinction by evolving from previously well-adapted phenotypes to those that are favoured by the new selective pressure.

In order to explore these issues, we formulate a population dynamics model, consisting of a multi-type time-continuous branching process, where types are associated to genotypes and their birth and death probabilities depend on the associated phenotype (non-escape or escape). We show that, within the setting associated to the escape problem, separation of time scales naturally arises and two dynamical regimes emerge: a fast-decaying regime associated to the escape process itself, and a slow regime which corresponds to the (survival) dynamics of the population once the escape phenotype has been reached (i.e. conditioned to escape).

We exploit this separation of time scales to analyse the topological factors which determine escape and survival. In particular, the aim is to analyse the influence of topological properties associated to robustness and evolvability on the probability of escape and on the probability of survival upon escape. We show that, while the escape probability depends on size of the neutral network of the escape phenotype (i.e. its degree), the probability of survival is essentially determined by its robustness (i.e. the resilience of the escape phenotype against genetic mutations), measured in terms of a weighted clustering coefficient.

Iain Johnston University of Birmingham
Mitochondrial DNA - a ‘collective’ GP map governing deadly inherited diseases

Mitochondria are our cellular energy stations, producing the ATP that powers our lives. Originally independent organisms that were engulfed and enslaved by our ancestral cells, mitochondria retain their own DNA (mtDNA), which encodes vital parts of their (and so our) energetic machinery. Hundreds of mtDNA molecules exist inside each of our cells, replicating, degrading, and mutating - forming intracellular evolutionary systems. Some mutations in mtDNA cause deadly diseases, but only when a certain cellular proportion of mtDNA molecules contain this mutation. The GP map governing these diseases thus depends on the collective properties of cellular mtDNA populations. I'll talk about our stochastic models for how populations of mtDNA molecules evolve in our cells, and how we can learn from evolutionary history to combat the inheritance of these diseases.

Daphne Ezer Sainsbury Laboratory Cambridge University
Ara-BOX-sis: a computational tool for predicting the regulatory mechanisms of genes downstream of conserved motif boxes

Plants are sessile and must adapt to their environment by adjusting their transcriptional networks in response to cues such as temperature and light. Plant transcription factor (TF) families have undergone massive expansion - plants have significantly more and larger TF gene families than animals and fungi, and many of these expansions are linked to adaptation to environmental stressors (Shiu et al, 2005). Since many TF family members bind to similar or identical sequence motifs, mapping the phenotypic consequences of these family expansions is a major challenge. For instance, G-boxes (CACGTG) are enriched in promoters of genes involved in the circadian clock, photosynthesis, and temperature response. However, many of the 161 annotated bHLH TFs and 71 annotated bZIP TFs can bind to G-boxes.

Determining the TF code - i.e. predicting which TF activates which target with spatial and temporal resolution - is a key goal in plant biology. To address this question, we have performed time course

RNA-seq experiments across different temperatures, within a number of circadian clock and histone remodeller mutant backgrounds. Through analysis of this data with an ensemble of network inference approaches, we identified how the gene expression of bHLH and bZIP transcription factors can be used to predict the gene expression of potential target genes downstream of strong G-box motifs. A number of these predictions are consistent with in vivo TF binding data, suggesting that this may be an effective strategy for predicting functional interactions between TFs and their downstream targets. Furthermore, we have identified additional sequence properties that can predict the putative targets of certain bHLH and bZIP TFs, suggesting mechanisms by which members of TF families can evolve unique sequence specificity. We have integrated these results into an online interactive visualisation tool called Ara-BOX-sis to help researchers explore how bHLH and bZIP gene duplications (genotype) map to changes in the gene expression network responsible for integrating light and temperature signals (phenotype). This approach scales, and we are able to increase the predictive ability of the approach by the addition of extra datasets for both conditions and genotypes. We are currently expanding Ara-BOX-sis to incorporate a similar network analysis of genes controlled by other expanded TF families in plants, like WRKYs and Heat Shock Factors. This will help us understand how TF family expansions may help plants adapt to their environment. We envisage that this approach may be useful in a wide range of organisms where TFs have undergone large-scale expansions.

Richard Goldstein **University College London**
Sequence Entropy and the Absolute Rate of Amino Acid Substitutions

Modelling the rate at which protein sequences change is central to understanding how proteins adapt to their structural, functional, and thermodynamic requirements. It is also key to deciphering the patterns of conservation and variation that reflect evolutionary processes. We approach the problem by building a mechanistic framework for amino acid substitution rates based on the formalisms of statistical mechanics, demonstrating the primacy of sequence entropy. Theoretical models and computer simulations of proteins under selection for thermodynamic stability show that substitutions between amino acids occur when their contributions to stability are nearly identical so the substitution is nearly neutral. Epistatic interactions with fluctuating amino acids at other sites drive stabilities in and out of regions of near neutrality, with the substitution rate determined by the time spent in this neutral region. We derive a theoretical framework for how stabilities are determined, demonstrating that substitution rates and the magnitude of the evolutionary Stokes shift can be predicted from biophysics and the effect of sequence entropy alone.

Carolina Diaz **Okinawa Institute of Science and Technology**
Mutational Robustness in Ribozyme Populations

The mutation-selection equilibrium at high error rates generates a cloud of highly diverse mutant variants, with few dominating genotypes: The quasispecies. This concept has been validated for RNA viruses, viroids, computer models, and with ribozymes – a contribution that I made using the Continuous in vitro Evolution (CE) system. Genotypes have limited tolerance to the effect of the accumulation of mutations before their information is lost. Therefore, at high mutation rates viable genotypes tend to decrease, eventually causing population extinction.

The purpose of my research is to investigate the role of mutational robustness, an important characteristic of quasispecies, in delaying or preventing population extinction. More specifically, I aim to understand the specific role of genotypic connectedness (neutral networks) in the re-acquisition of lost genotypes, and as a mechanism for survival. For this, I have evolved and sequence (Miseq) lineages of ribozymes with the CE system under different mutagenic treatments. Genotypic variants would be determined to built the genetic networks as done previously (Network software).

Phenotypic data would be measured for evolved genetic variants in two ways: (1) Non-functional ribozymes can be distinguished by electrophoresis given its different sequence length compared to

functional ribozymes. This makes it possible to estimate the ratio of active/inactive ribozymes in the population. (2) The catalytic function of frequent mutants would be kinetically assessed. These phenotypic measurements would be mapped onto the genotypic data (either with graph theory, the Louvain Method in Python NetworKit, or other up to date method) to uncover “cluster sub-structures” of genotypic similarities, and to estimate the size/number of neutral networks. Populations with more (or bigger) neutral networks are expected to be more mutationally robust and to correlate with time to survival (e.g., did not become extinct), perhaps even at higher mutation rates.

Tom McLeish University of Durham

The notion of ergodicity in genotype exploration - towards a predictive evolution

We examine the analogy between evolutionary dynamics and statistical mechanics to include the fundamental question of ergodicity—the representative exploration of the space of possible states (in the case of evolution this is genome space). Several properties of evolutionary dynamics are identified that allow a generalization of the ergodic dynamics, familiar in dynamical systems theory, to evolution. Two classes of evolved biological structure then arise, differentiated by the qualitative duration of their evolutionary time scales. The first class has an ergodicity time scale (the time required for representative genome exploration) longer than available evolutionary time, and has incompletely explored the genotypic and phenotypic space of its possibilities. This case generates no expectation of convergence to an optimal phenotype or possibility of its prediction. The second, more interesting, class exhibits an evolutionary form of ergodicity—essentially all of the structural space within the constraints of slower evolutionary variables have been sampled; the ergodicity time scale for the system evolution is less than the evolutionary time. In this case, some convergence towards similar optima may be expected for equivalent systems in different species where both possess ergodic evolutionary dynamics. When the fitness maximum is set by physical, rather than co-evolved, constraints, it is additionally possible to make predictions of some properties of the evolved structures and systems. We propose four structures that emerge from evolution within genotypes whose fitness is induced from their phenotypes. Together, these result in an exponential speeding up of evolution, when compared with complete exploration of genomic space. We illustrate a possible case of application and a prediction of convergence together with attaining a physical fitness optimum in the case of invertebrate compound eye resolution.

Bhavin Khatri University College London & Francis Crick Institute

A simple genotype-phenotype map for protein-DNA binding, sequence entropy and its role in speeding up speciation

Natural selection acts on organismal function, whilst variation arises from mutations; understanding the mapping between the two, between genotype and phenotype, represents a major outstanding challenge for evolutionary theory. Here we present a simple biophysical model of the map from genotype to phenotype for protein-DNA binding that controls gene expression and its consequences for the process of speciation. Using both simulation (Khatri, et al. Genetics, 2015) and theory (Khatri et al., Journal of Theoretical Biology, 2015), we find that a quantity analogous to Boltzmann entropy from statistical mechanics, called sequence entropy (Iwasa, JTB, 1988 & Sella & Hirsh, PNAS, 2005), plays a key role in the dynamics of speciation; at small populations stochastic effects due to genetic drift dominate and those phenotypes with the greatest number of sequences (high sequence entropy) dominate – simply as a consequence of the fact that there are many more sequence pairs that bind poorly than well, means the protein-DNA binding is on average less well adapted for small populations, but still co-evolved to maintain function. When a lineage splits into two, the protein-DNA pair on each lineage now co-evolves independently and speciation arises once the hybrid DNA-binding becomes non-functional. Since the common ancestor pair is less well adapted at small population sizes, a smaller number of mutations is needed before hybrids become non-functional and speciation arises more quickly. In contrast to founder effect models, this model represents a new robust mechanism for more rapid speciation at small population sizes, consistent with large species diversity in small habitats such as Cichlids in the East African Great Lakes, and contrasted with the lower diversity of marine animals, which have large ranges and population sizes.