Evolution of Structural and Functional Complexity in Biology 2012

17-19 September 2012

King's College, Cambridge

Programme

Monday, 17 September

10:00-11:00 **Registration and Coffee**

11:00 **Sebastian Ahnert** University of Cambridge
Opening Remarks

Session 1 **Forms, Shapes and Morphologies**

11:10 **Nick Jones** Imperial College London
Beyond Sequence: Evolutionary Inference About Past Form

12:00 **Rebecca Cotton-Barratt** University of Warwick
Modelling biological form

12:30 **Sam Greenbury** University of Cambridge
Evolutionary properties and processes in a self-assembling system

13:00 **Lunch**

Session 2 **Genotype-Phenotype Maps**

14:00 **Kamaludin Dingle** University of Oxford
Simplicity Bias in Genotype-Phenotype Maps

14:30 **Ard Louis** University of Oxford
Genotype-Phenotype Maps and the Survival of the Frequent

15:20 **Bhavin Khatri** MRC National Institute of Medical Research
Emergent complexity, speciation and a simple genotype phenotype map

15:50 **Coffee**

Session 3 **Genetic Circuits**

16:20 **Julian Hibberd** University of Cambridge
Insights into the evolution of C_4 photosynthesis

17:10 **Ben Williams** University of Cambridge
Characterising the evolutionary pathways generating C_4 photosynthesis

17:40 **End of first day talks**

19:00 **Drinks**

19:30 **Conference Dinner**

Tuesday, 18 September

Session 4: **Dynamics of Regulatory Networks**

10:00 **Olivier Martin** Université Paris-Sud
Dynamical Regulatory Networks: from function to structure and back

10:50 **Coffee**

Session 5: **Protein and RNA structures**

11:20 **Emmanuel Levy** Weizmann Institute
Promiscuous protein-protein interactions: a burden for the cell and a tool for the biologist

12:10 **James Anderson** University of Oxford
Characterising RNA Secondary Structure Space

12:40 **Lunch**

Session 6: **Evolution of collective complexity**

14:00 **Richard Watson** University of Southampton
Algorithmic Principles of Individual and Collective Adaptation

14:50 **Paul Ryan** University of Southampton
Propagule size and the evolution of biological individuality

15:20 **Matthew Turner** University of Warwick
Non-local models of swarming

15:50 **Coffee**

Session 7: **Evolutionary dynamics I**

16:10 **Tobias Galla** University of Manchester
Evolutionary Dynamics in Finite Populations

17:00 **Roland Schwarz** University of Cambridge
Intra-patient mutational landscapes of high-grade serous ovarian cancer exhibit clonal expansion and determine progression-free survival

17:30 **Richard Goldstein** MRC National Institute of Medical Research
Self policing among non-conjugative plasmids: The evolution of copy number control

18:00 **End of second day talks**

Wednesday, 19 September

Session 8: **Multi-scale organisation of biological systems**

9:00 **Lucio Vinicius** University of Cambridge
The evolution of biological complexity by "modularity transfer"

9:50 **Emanuele Cozzo** University of Zaragoza
Stability of Boolean Multiplex

10:20 **Coffee**

Session 9: **Evolutionary Dynamics II**

10:35 **Madan Babu Mohan** MRC LMB
The impact of tissue-specifically spliced segments on protein interaction networks

11:25 **Darka Labavic** Jacobs University
Demographic fluctuations and inherent time scales in a genetic circuit

11:55 **Closing remarks**

Abstracts

James Anderson

Characterising RNA Secondary Structure Space.

With the advent of next-generation sequencing technologies and new methods in transcriptomics, an explosively growing amount of biological RNA data is available in public databases. Key to understanding function and regulatory effects of RNA, structure prediction is still difficult, despite many efforts. At the heart of the problem is unreliability in RNA secondary structure prediction, the best methods gaining about 60% sensitivity on single-stranded RNA sequences. What is more, many structural predictions are relied upon as correct in biological applications. Efforts must continue to be made to better understand the distribution of predictive accuracy and when structure prediction fails. Here we explore the landscape of RNA secondary structure space, particularly with a view to predictive quality, considering where structural signals can be found. Novel phylo-SCFG entropy methods will be presented, and these compared with current PPfold reliability scores to demonstrate various failings in secondary structure prediction. Further factors will be explored including changes in alignment quality and various evolutionary distances, allowing for greater quantitative understanding of accuracy of resulting secondary structure predictions.

Madan Babu

The impact of tissue-specifically spliced segments on protein interaction networks

Alternative inclusion of exons increases the functional diversity of proteins. Among alternatively spliced exons, tissue-specific exons play a critical role in maintaining tissue identity. This raises the question of how tissue-specific protein-coding exons influence protein function. Here we investigate the structural, functional, interaction, and evolutionary properties of constitutive, tissue-specific, and other alternative exons in human. We find that tissue-specific protein segments often contain disordered regions, are enriched in posttranslational modification sites, and frequently embed conserved binding motifs. Furthermore, genes containing tissue-specific exons tend to occupy central positions in interaction networks and display distinct interaction partners in the respective tissues, and are enriched in signaling, development, and disease genes. Based on these findings, we propose that tissue-specific inclusion of disordered segments that contain binding motifs rewires interaction networks and signaling pathways. In this way, tissue-specific splicing may contribute to functional versatility of proteins and increases the diversity of interaction networks across tissues.

Rebecca Cotton-Barratt

Modelling biological form

Computer-based models of biological evolution have typically ignored morphological form. Yet there are compelling reasons to believe that form has an important role to play in the long-term development of the system.

We present a novel computer model of form-driven evolution, modelling form abstractly via hierarchical structures represented by directed acyclic graphs. This model incorporates elements

from a wide variety of recent advances in evolutionary modelling, including adaptive dynamics, genetic algorithms, population dynamics, hypercycles and graph theory.

We see several nontrivial phenomena emerge from this minimal imposed structure, and include analysis of the model when embedded in different evolutionary frameworks.

Emanuele Cozzo

Stability of Boolean Multiplex

Nearly four decades ago, Random Boolean Networks (RBNs) were introduced as a way to theoretically address several scientific challenges regarding the description and dynamics of biochemical networks [1]. Since then, this framework has been successfully applied to model theoretically and computationally the biochemical and genetic control of cells [2] . RBNs consider that each gene of a genetic regulatory network is a node of a directed graph, the direction corresponding to the effect of one gene on the expression of another. Additionally, the nodes can be in one of two states: they are either on (1) or off (0) - i.e. in the case of a gene its target protein is expressed or not. The system so composed evolves at discrete time steps. At each time step nodes are updated according to a boolean rule assigned to each node that is a function of its inputs.

The stability properties of this system is of interest in understanding evolution. The existence of two phases, a chaotic and an ordered one, leads to the hypothesis known as "edge of chaos".[3]

The previous description implicitly assumes that all biochemical signals are equivalent and then collapses information from different pathways. Actually, in cellular biochemical networks, many different signaling channels do actually work in parallel, i.e., the same gene or biochemical specie can be involved in a regulatory interaction, in a metabolic reaction or in another signaling pathway. Therefore, a more realistic set up will be obtained by considering the participation in different pathways as different interconnected layers of interaction, something more consistent with a multiplex network [4,5] representation . Namely, each level in the multiplex would represent the different signaling pathways or channels the element participates in.

In this work, we study the stability of Boolean networks defined at multiple topological layers. In particular, we inspect a Boolean multiplex network model, in which each biochemical specie is a node that participates in one or more layers of interactions.

Capitalizing on a semi-annealed approximation, we analytically and numerically study the conditions defining the stability of the aforementioned system.

We show that, if sub-systems represented by single layers are critical, when coupled, the system as a whole is still critical regardless of the strength of the coupling (expressed in terms of the number of shared nodes). Moreover, a critical (or ordered system) can have layer that, if isolated, would operate in the chaotic phase.

The results are of particular interest in understanding the evolutionary relationship between structural and functional complexity.

References:

[1] S. Kauffman, Curr. Top. Dev. Biol., 6: 145-182, (1971).
[2] L. P. Kadanoff; M. Aldana, S. Coppersmith. Boolean dynamics with random couplings. Springer, (2003)
[3] S. Kauffman, Origins of Order: Self-Organization and Selection in Evolution
[4] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, J.-P. Onnela, Science 328, 876 (2010).
[5] K.-M. Le et. al., N. J. Phys. 14, 033027 (2012).

Kamaludin Dingle

Simplicity Bias in Genotype-Phenotype Maps

Organisms produce their varied structures and functions by processing genetic information, and as such the principles of information theory will in part delimit biological phenomena. Here we make a novel application of algorithmic information theory to genotype-phenotype (G-P) maps, and derive a number of predictions. Specifically, we relate a phenotype's information content (i.e. complexity) to its designability, i.e. the number of genotypes underlying it. We show that high complexity phenotypes will be much less designable, while simpler phenotypes typically will be more designable. Our work predicts that strongly non-uniform phenotypic designabilities should be common in many G-P maps, and in these cases that nearly all genotypes map to a tiny fraction of phenotypes. Additionally, we show that in such contexts there is a bias for simpler phenotypes in the sense that random mutations on average lead to less complex phenotypes, potentially rationalising a number of experimental and database observations. More generally, we usually feel that random alterations to a structure (e.g. a car) can only cause disarray, but our work shows that this intuition is mistaken for many G-P maps: counterintuitvely, random alteration (i.e. mutations) can often lead to more structured and regular phenotypes.

Tobias Galla

Evolutionary dynamics in finite populations

Describing the dynamics of evolutionary processes in finite populations requires modelling approaches beyond the well-known deterministic replicator equations and related ordinary or partial differential equations. Deterministic approaches are valid only for infinite populations and systematically neglect stochastic effect induced by demographic noise. For example, they cannot capture noise-driven phenomena such as fixation, stochastic slowdown or the occurrence of coherent quasi-cycles. In this talk I will give an overview of the origins of demographic noise, the stochastic effects it induces, and I will describe the mathematical tools with which these effects can be predicted analytically. I will focus mostly on evolutionary dynamics, but time allowing I will also discuss stochastic pattern forming processes in spatial systems in ecology.

Richard Goldstein

Self policing among non-conjugative plasmids: The evolution of copy number control

From an evolutionary perspective, why should anyone act responsibly when selfish behaviour provides opportunities for more offspring? Why sacrifice one's own narrow interests for the good of the group? When would policing evolve, and why would anyone submit to the resulting restrictions? This question is not only relevant for us and other higher animals, but also are important topics in the bacterial world. Many bacteria contain plasmids, non-chromosomal DNA segments that replicate independently of the chromosome. The faster they replicate, the better they can out-compete the other plasmids in the cell. But, although they often are beneficial for the bacterium in which they reside, they also impose a burden - an excessive replication rate results in an excessive number of plasmids, lowering the fitness of the bacterium and harming the evolutionary prospects for all of the cell's plasmid residents. In this way, selection occurring at the intracellular level conflicts with intercellular selection. Plasmids have evolved a policing mechanism to control their number, restricting their replication rate as their numbers grow. How would this evolve? Why wouldn't selfish plasmids evolve with abilities to evade or ignore this policing? Computer modelling of agent-based simulations, backed by analytical analyses, can provide insights into the evolutionary dynamics, the transparency of the calculations providing us with a powerful window to examine these issues.

Sam Greenbury

Evolutionary properties and processes in a self-assembling system

Biological organisms possess complicated genotype-phenotype mappings, providing a means to innovate in order to survive in changing environments. Given the complexity of these genotype-phenotype maps in even the simplest organisms, simplified models are of great utility in investigating underlying principles that may govern their evolution.

Our work considers one such simplified model. A set of square-tile building blocks is specified by the genotype, which is transformed into a larger tile structure (polyomino) defined as the phenotype through a self-assembly process. The mapping may be considered analogous to the assembly of protein complexes, where proteins (building blocks) interact to form larger structures (the protein complex).

Despite the simplicity of the polyomino model (PM), non-trivial dynamics are exhibited, providing potential for realistic theoretical investigation of evolutionary systems. Further, the model is amenable to a quantitative definition of complexity for self-assembling systems, allowing the complexity of phenotypes within the model to be measured.

Within our model, the role of complexity, modularity, robustness, evolvability and evolutionary processes, such as gene duplication, are examined through two methods. Firstly, by considering the evolutionary dynamics towards target phenotypes and, secondly, by investigating the structure of the phenotype's neutral network in genotype space.

We present results demonstrating that less complex phenotypes occur more frequently under random evolution. Further, when a specific phenotype is chosen as the target of an evolutionary run, the complexity of the target correlates strongly with adaptation ability. Gene duplication permits the production of phenotypes with greater complexity, whilst further indicating that lower complexity phenotypes evolve more frequently. Finally, we discuss how robustness and evolvability relate within the model.

Our current work consists of attempts to use the model further with respect to understanding the properties discussed above as well as through application to protein complexes, by considering how complexity of a protein complex topology may affect evolutionary rate and neutral component size.

Julian Hibberd

Insights into the evolution of C_4 photosynthesis

The C_4 pathway is complex and involves alterations to the biochemistry, cell biology and development of leaves. Despite this complexity C_4 photosynthesis it is thought to have evolved independently at least 62 times. One particularly important aspect of the C_4 leaf is compartmentation of photosynthesis proteins, typically between mesophyll (M) and bundle sheath (BS) cells. I will address the extent to which this cell-specificity is associated with transcriptional or post-transcriptional mechanisms and provide specific examples of how the regulation of genes has altered as they are recruited into the C_4 pathway. I will then discuss the extent to which distinct lineages of C_4 plants share regulatory circuitry that generates accumulation of photosynthesis proteins in BS or M cells, and whether cis-elements responsible for BS specificity evolved de novo in C_4 lineages or whether they are present in orthologous genes from C_3 species.

Nick Jones

Beyond Sequence: Evolutionary Inference About Past Form

We have refined methods to reconstruct candidate evolutionary trees and candidate ancestral sequences given contemporary (genetic) sequence data. But what if our data is not a set of sequences but something more complex? For example, how can we find the evolutionary relationships between a set of shapes or sounds? I will outline some of our recent attempts to reconstruct past form; in particular attempting to develop inference methods suitable for ancestral inference with function-valued data. I will discuss challenges in attempts to reconstruct past speech sounds and animal forms. If I have time I will also touch on our work to compare different proposed mechanisms for the evolution of protein interaction networks.

Bhavin Khatri

Emergent complexity, speciation and a simple genotype phenotype map

It is well established that natural, selection, mutation and genetic drift constitute the basic processes of evolution. Yet little is understood about the how the complexity of life emerges from these essentially random processes. A missing component is an understanding of how sequences map to function, or genotype-phenotype maps; I will present a simple, yet biophysically motivated genotype-phenotype map for gene expression in development and show that even for very simple gene regulation, a fitness landscape of remarkable complexity emerges. Populations evolve sub-optimal functions due to sequence entropic pressures at low population size, whilst at large population size although fitness dominates, we find that non-critical phenotypes or degrees of freedom exhibit quenched disorder - in other words some phenotypes are conserved even if they are not critical to function. We then examine how population size and sequence entropy interplay to determine the probability of speciation, as two populations evolve independently from a common ancestor. We find that incompatibilities arise between diverging lines as co-evolving sequences become increasingly uncorrelated. In particular, in contrast to classical models of speciation, which do not say anything about the effect of population size, we find incompatibilities are more likely as population size is decreased, as sequence entropy pushes hybrid populations more often into regions of poor fitness.

Darka Labavic

Demographic fluctuations and inherent time scales in a genetic circuit

We focus on a prototypic dynamical unit which consists of only two species interacting in a non-linear way. This unit may be regarded as a coarse-grained description of a genetic circuit. For comparison we first discuss the phase structure of this unit on a coarse-grained level, in a deterministic description. Then we turn to a fully stochastic description, for which we observe quasi- cycles for parameters that correspond to values deeply in the fixed-point regime in the deterministic limit [1]. We shall unravel the effect of demographic fluctuations and fluctuations in the reaction times. The power spectrum will show which source of stochastic behavior is dominant, in particular if the dynamics is very spiky. We compare analytic predictions with Gillespie simulations which come closest to the experiments in vitro. The genetic circuit has applications to all systems in which a self- activating species also activates its own repressor. Both interactions, the self-activation and the repression, need not necessarily be realized by direct links, they can amount to an effective description on a coarse scale with a different number of intermediate steps and different realizations of the very activation or repression. Intermediate steps, however, may introduce additional time scales. In our

ongoing work [2] we therefore analyze the effect of competing time scales on the validity range of the coarse-grained description. As it turns out, what is the appropriate model depends on the ratio of protein decay rates to binding/unbinding rates of transcription factors, which lead to different switching rates of genes to another expression level. This way the inherent time scales also determine the type of bifurcations which the circuit undergoes under variation of a certain parameter. For example, a whole regime with regular oscillations fades away if the gene states change over a time scale that is not short, but comparable to the lifetime of the involved proteins. This may give a hint on a possible origin for the malfunction of oscillatory systems like circadian clocks. Circadian clocks are just one example of systems that are supposed to be described by our genetic circuit.

References:
[1] A. Garai, B. Waclaw, H. Nagel and H. Meyer-Ortmanns, Stochastic Description of a Bistable Frustrated Unit, J. Stat. Mech. (2012) P01009, 28 pages
[2] D. Labavic, H. Nagel, W. Janke and H. Meyer-Ortmanns, Coarse-grained modeling of a genetic circuit in the presence of several inherent time scales, in preparation

Emmanuel Levy

Promiscuous protein-protein interactions: a burden for the cell and a tool for the biologist.

The interior of cells is a highly crowded environment where proteins continuously encounter each other. In this environment, functional protein-protein interactions compete with a much larger number of non-functional, or promiscuous, interactions. We will examine several questions associated with the notion of promiscuous interactions. First, we will ask how widespread they are [1,2,3,4,5]. Second, we will ask if, and how, cells have adapted to minimize their presence [6]. Third, we will see how we can exploit them in a novel strategy to measure local protein concentrations in vivo and with high accuracy.

References:
[1] Levy, Landry, Michnick. How perfect can protein interactomes be? Science Signaling. 2009.
[2] Landry, Levy, Michnick. Weak functional constraints on eukaryotic phosphoproteomes Trends in Genetics 2009
[3] Levy, Landry, Michnick. Signaling through cooperation. Science. 2010
[4] Levy, A simple definition of structural regions in proteins and its use in analyzing interface evolution, J. Mol. Biol. 2010
[5] Levy, Michnick, Landry. Protein abundance as a dominant factor in the promiscuous phosphorylation of proteins. Philos Trans R Soc Lond B Biol Sci. 2012 (in press)
[6] Levy, De, Teichmann. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Submitted.

Ard Louis

Genotype-Phenotype Maps and the Survival of the Frequent

Evolutionary dynamics are driven by the interplay of random mutations, which change genotypes and thus generate variation, and selection, which favours fitter phenotypes. It is well established that many genotype-phenotype (GP) maps have a much larger number of genotypes than phenotypes, leading to large neutral spaces. Moreover, these genotypes are often distributed in a highly inhomogeneous fashion: a relatively small number of "frequent" phenotypes take up the majority of genotypes. We

study the effects of this skew in the GP maps by deriving a microscopic theory of evolutionary dynamics as a function of the topology of the neutral spaces, the mutation rate and population number. We test our theory by comparing to simulations of a randomly distributed GP map, and of a more biologically motivated RNA secondary structure mapping, finding excellent quantitative agreement for a wide range of parameters. Both in the monomorphic and in the polymorphic regimes, skew in the GP map leads to a strong phenotypic bias: a random mutation is much more likely to lead to a frequent phenotype than to a less frequent one. This bias in the arrival of variation has profound consequences for adaptation and evolvability, leading not just to survival of the fittest, but also to survival of the frequent.

Olivier Martin

Dynamical Regulatory Networks: from function to structure and back

Even though any vital cellular process may be tightly controlled, its associated regulatory network varies a lot from species to species. To explore the space of all possible networks or circuits that implement a specified control or function, we use Markov Chain Monte Carlo (MCMC) and biophysical modeling. First, we show that the molecular encoding of the genetic interactions naturally leads to sparse networks. Second, we consider different types of cellular functions; our MCMC sampling then reveals that the structure of the functional networks is shaped by these functions. Finally, by examining the emergent structural features found such as edge usage or network motifs, we uncover the regulatory logic responsible for the functional capabilities of these networks.

Paul Ryan

Propagule size and the evolution of biological individuality

Collective living has evolved many times independently (e.g. evolution of multi-cellularity in animals, plants and fungi - each from uni-cellular ancestors). Clearly, there can be adaptive advantages to group living (economies of scale, division of labour or avoidance of predation). However, any collective life-form must also tackle the public-goods dilemmas which are a necessary part of group living - the so-called 'free rider problem', or Tragedy of the Commons.

The simplest way for a collective life-form to reproduce is by binary fission (as in the colonial choanoflagellates). N/2 is thus the theoretical maximum propagule size. Yet, if we survey the life histories of diverse multi-cellular organisms across the animals, plants and fungi, we very commonly find a uni-cellular stage - zygotes, seeds or spores.

** Why has a minimally small propagule size evolved (independently) so many times, in such diverse taxa? **

I present a computational model in which propagule size is allowed to co-evolve with a social trait (strategy in a public-goods consumption dilemma). I find that, when propagule size is large (i.e. vegetative growth) most genetic variance is within- rather than between- collectives and the unit of selection is primarily the cell. However, if mutation in propagule size is allowed then collectives employing smaller propagules out-compete those employing larger ones (despite the opposing effect of starting life larger). Reducing propagule size denies heritable-variation-in-fitness to the cells and helps cement Darwinian individuality at the collective level. This effect can be used to explain major evolutionary transitions of the 'fraternal' kind.

My model has broad applicability and can also be used to explain the evolution of unitary propagules in eusocial hymenopterans (the singly-mated queen) and colonial marine invertebrates (the protozooid in siphonophores such as nanomia cara).

Roland Schwarz

Intra-patient mutational landscapes of high-grade serous ovarian cancer exhibit clonal expansion and determine progression-free survival

Functional and structural complexity of human cancers is determined by genetic heterogeneity which arises from tumor evolution. The possibility that significant genetic heterogeneity could exist within many cancers was originally proposed by Nowell in 1976 and then demonstrated using cytogenetic methods in 1978 (Dexter et al 1978). However, only recently the technological requirements have been met to rigorously assess intra-tumor heterogeneity. Recent studies using next-generation sequencing and high accuracy SNP CGH arrays now paint an accurate picture of the degree and types of genetic variation present within individual epithelial cancers (Cooke et al 2010 and others). Intra-tumour genomic heterogeneity is a potential prognostic indicator and might be a source of drug resistance and progressive disease. It has been argued that relapsed disease, which is frequently resistant to chemotherapy, is caused by clonal evolution from the initial tumor mass and potentially derived from a low-prevalence subclone of presentation disease (Cooke et al 2011 and others).

A better understanding of tumour heterogeneity is particularly important for high grade serous ovarian cancer (HGSOC), a lethal disease with 5-year overall survival that is <30 %. As compared to other epithelial cancers it is relatively hypersensitive to platinum-based chemotherapy but 70 -80% of cases that initially responded to treatment relapse with platinum-resistant disease. Genomic studies have focused on the identification of early genetic events which include TP53 mutations in virtually 100% of cases, loss of homologous recombination pathway dysfunction predicted in at least 50% of cases (TCGA 2011) and essential gene profiles for cell survival (Marcotte et al 2012).

We have recently proposed that genetic heterogeneity in HGSOC could explain the development of drug resistance through clearance of the treatment-sensitive dominant clone followed by repopulation with intrinsically resistant subclones (Cooke et al 2011). This is supported by data from cell lines derived before and after the development of platinum-resistant disease that show that relapsed tumour shows significant genetic divergence incompatible with simple linear genetic change. The most parsimonious explanation is therefore that resistant lineages were present as a minor subpopulation of the tumour mass at the time of first therapy (Cooke et al 2010). In clinical samples, co-existence of different genetic subpopulations at presentation and functional consequences of intra-tumour genetic heterogeneity have not yet been demonstrated. Further, knowledge of clinically relevant mechanisms of drug resistance is limited. However, if multiple subpopulations exist, dynamic changes in the frequency of genetic subpopulations during and following treatment could demonstrate the selective effect of treatment acting on pre-existing variation in the intrinsic levels of drug resistance present in each population.

To understand the clinical impact of tumor heterogeneity we need to trace its etiology to the evolutionary processes that govern it and finally influence patient survival. To achieve this we gathered 172 samples of HGSOC from 20 patients undergoing neoadjuvant chemotherapy at initial diagnostic biopsy, interval debulking surgery and relapse over multiple metastatic sites. Using SNP CGH arrays, paired-end NGS and Fluidigm sequencing, we infer the evolutionary histories and genomic profiles of the last common ancestor in these 20 cases of HGSOC. We developed tailored algorithms for phylogenetic reconstructions and to quantify temporal and spatial heterogeneity and

degree of clonal expansion. We show that genetic heterogeneity is a strong predictor of patient outcome and provide robust evidence that relapsed disease is a clonal expansion of a low-prevalence subclone that was present prior to therapy. Model tests of the shape of the reconstructed trees show that the majority of cells retain their tumorigenic potential, making clonal evolution the primary source for genetic heterogeneity. Our analyses further allow us to identify driver events by finding common variants between reconstructed precursor genomes across all patients and allow us to ask questions about how mutator phenotypes are reflected in non- homogeneous mutation rates along branches.

Matthew Turner

Non-local models of swarming

Swarming is a conspicuous behavioural trait observed in bird flocks, fish shoals, insect swarms and mammal herds. This is thought to provide an improved collective awareness and offer protection from predators. Most current models for this organisation involve the hypothesis that coordinating information is exchanged between immediate neighbours. However, such local interactions alone are insufficient to explain primary features of the organisation of large flocks of birds. The method by which long-range information is exchanged, thereby controlling density, remains unknown. Existing models incorporate a number of undesirable features: They include interactions between individuals that are not visible to one another, as well as precisely demarcated zones of repulsion/alignment/attraction coupled with "physics-like" interaction potentials. As a result these models have extensive parametric and structural freedom. We do not believe that they represent a promising starting point for the development of more realistic models.

We therefore propose a hybrid projection model in which the behaviour of each individual is controlled by the projected view through the flock seen by each individual, coarse-grained to the level of patches of "dark" (bird) and "light" (sky). We include the term hybrid to signify that it also involves a (metric free) co-alignment with nearest neighbours. Our model reproduces a number of features observed in data, provides a novel mechanism for the regulation of density and can give rise to phenotypes reminiscent of birds, fish and insects, all with a far smaller number of adjustable parameters.

Finally we compare our model with experimental data to show that, in both cases, large flocks self-organise to the maximum density at which a typical individual is still just able to see out of the flock in many directions. Such flocks are marginally opaque - an external observer can also just still see a substantial fraction of sky through the flock. Although seemingly intuitive we show that this need not be the case; flocks could easily be highly diffuse or entirely opaque. The emergence of marginal opacity controls the density of a flock. Larger flocks are correspondingly less dense, as observed, but not previously understood. We show how marginal opacity could give rise to a flock density that scales with a power of the size of the flock. It also provides a mechanism for global interactions with individuals responding to the projection of the flock that they see, motivating our hybrid projection model. This global interaction results in faster information transfer and hence rapid flock dynamics, an advantage over local models. Marginal opacity strongly constrains our understanding of how individuals interact with each other within large swarms. From a behavioural perspective it optimises the information available to each bird while maintaining the protection of a dense, coherent flock.

Lucio Vinicius

The evolution of biological complexity by ‘modularity transfer’.

The scheme of 'major evolutionary transitions' proposed by Maynard Smith and Szathmáry remains highly influential as a model for evolutionary increases in biological complexity. Although insightful, the scheme is flawed: six out of the seven examples by Maynard Smith and Szathmáry themselves fail to satisfy their own definition of a transition, namely the occurrence of both phenotypic aggregation ('the origin of new biological wholes') and informational emergence ('the origin of new biological codes').

Here I propose an alternative view of the macroevolution of complexity. According to Schrödinger, life is based on the 'Principle of (structural) Order from (informational) Order'. I argue that Schrödinger's Principle has evolved, resulting in the process of 'modularity transfer': entities that initially worked as modular phenotypes (proteins, somatic cells, vocalisations) have sometimes evolved into new modular information carriers (transcription factors, neural cells, words), engendering the rare origin of higher levels of biological organisation. In the light of the modularity transfer hypothesis, I re-examine the role of natural selection in the evolution of complexity, the classification of biological information systems, the role of organisational vs. informational criteria in definitions of complexity, and the meaning of human language and material culture.

Richard Watson

Algorithmic Principles of Individual and Collective Adaptation.

Darwinian adaptation does not always lead to biological complexity. Evolution by natural selection, ENS, sometimes increases Darwinian fitness by decreasing organismic complexity, and in other cases biological complexity can increase in systems that are not Darwinian individuals, e.g., ecological complexity. In fact, in an obvious sense, the complexity of an ecosystem containing many interacting entities is greater than that of any one individual it contains – but since an ecosystem is not a Darwinian individual, ecological complexity cannot be adapted by natural selection. Interestingly, however, in some cases ecological complexity can be transformed into organismic complexity. For example, in some of the major transitions in evolution, previously free-living entities became parts within a new reproductive entity (e.g. organelles within eukaryote cells and cells within multi-cellular organisms). Such transitions have been fundamental in the development of adaptive bio-complexity - arguably more so than the incremental tinkering we normally associate with ENS - but accommodating such phenomena in conventional evolutionary theory is problematic.

We have been investigating the algorithmic principles underlying individual and collective adaptation in various domains, including: the evolution of regulatory interactions between genes in a gene network, the formation of social relationships between strategic agents in a social network, and the evolution of ecological dependencies between species within an ecosystem. Our findings suggest that although these domains are very different, the fundamental dynamical interactions between individual and collective adaptation are systematically similar. Specifically, in each case the distributed action of self-interested motives to increase the fitness of each individual component creates an adaptive network at the system level that behaves in a manner directly analogous to connectionist models of associative learning. In such contexts each individual may get less complex over time as it stabilises relationships with others – but as individual relationships stabilise, new adaptive structures emerge from previously disorganised dynamics at higher levels of organisation (e.g., developmental modules, strategic coalitions, symbiotic partnerships). We discuss the conceptual and algorithmic relationships between individual adaptation in networked contexts, associative learning theory and the formation of new evolutionary units. This perspective suggests that a selective pressure to reduce individual complexity at any one scale of organisation is not incompatible with, and may even be central to, the evolution of complexity writ large – but this only makes sense if we appreciate that evolution repeatedly rescales the scope of its own application over time.

Ben Williams

Characterising the evolutionary pathways generating C4 photosynthesis

The majority of plant species use C3 photosynthesis, using an inefficient enzyme to acquire CO2 from the atmosphere. However, several species use C4 photosynthesis, an alternative metabolic pathway that assimilates CO2 with greater efficiency. C4 photosynthesis has evolved on at least 66 independent occasions within flowering plants. This convergence is remarkable, as the C4 phenotype consists of a complex set of biochemical, cellular and anatomical adaptations. As a result of this complexity, inferring the likely evolutionary events that have generated C4 on so many independent occasions provides an exciting challenge within evolutionary biology. In this seminar, I present the first computationally-driven approach towards understanding the evolution of the C4 phenotype, using existing data from 44 independent studies. By modelling C4 evolution as a series of transitions on a phenotype landscape, a new blueprint of the order and diversity of evolutionary pathways generating C4 in multiple taxa is proposed.