

THE INVENTION OF EUKARYOTIC GENE CONTROL. A NOVEL CLASSIFICATION METHOD SHOWS THE DIFFERENCE BETWEEN PROKARYOTIC AND EUKARYOTIC GENETIC CONTROL IS CONTROL LOGIC, NOT CHEMISTRY W. Bains¹ and D. Schulze-Makuch^{2,3}. ¹Department of Earth, Atmospheric and Planetary Science, MIT, 77 Mass. Ave., Cambridge, MA 02139, USA bains@mit.edu. ²School of the Environment, Washington State University, Pullman, WA 99164, USA. ³Center of Astronomy and Astrophysics, Technical University Berlin, 10623 Berlin, Germany

Background and model.

The evolution of life from the simplest, original form to complex, intelligent animal life occurred through a number of key innovations. Here we present a new tool to analyse these key innovations by proposing that the process of evolutionary innovation may follow one of three underlying processes. The processes are;

Random Walk (one critical step is needed, but that occurs randomly, so the innovation can occur randomly across geological time)

Critical Path (where a specific set of events is a precondition for an innovation, such that the innovation occurs rapidly after those steps have been taken, and hence in a determined time window)

Many Paths (where randomly occurring events built to the innovation, but many combinations of such events can lead to the same functional outcome, and so the timing of appearance of the events is constrained)

A biological function may be classified in this way either on the basis of its evolutionary history or the basis of its modern mechanism. Parallel appearance of functions characterising key innovations in this way helps to identify the likelihood that an innovation could arise. Our analysis is based on the occurrence of function in modern biology, rather than specific structure or mechanism. In this paper, we describe methods to classify functional features of modern organisms into these three classes based on the analysis of how a function is implemented in modern biology.

Application to Eukaryotic genetic control. We present the application of our categorization to the evolution of eukaryotic gene control. All complex, multicellular organisms are eukaryotes, and both their genome size and the amount of genome dedicated to genetic control are orders of magnitude larger than those in bacteria or archaea. We argue that there are close chemical similarities between bacterial, archaeal and eukaryal gene control circuits which suggest a Many Paths evolutionary track for all the components of Eukaryal genetics. Specifically, control by protein binding, nucleoprotein formation, structure and modification, short- and long-range nucleoprotein structure modulation, short and long RNA interaction with RNAs and with proteins, RNA splicing and alternative translation modes, have all evolved independently in the three domains of life (as well as being inherited in

Eukarya from their presumed Archaeal ancestor). Prokaryotes also show structural analogues (not homologues) of features thought to be exclusively eukaryotic, such as linear chromosomes and a membrane-bound compartment for the nucleoprotein. In short, all the 'distinguishing' structural and functional features of eukaryotic gene control mechanisms have evolved many times, from non-homologous ancestral structures. The evolution of complex Eukaryotic genetics is a Many Paths process, that is, one that requires a large number of steps but which can be achieved on many different paths.

Implications

If the evolution of all the functional components of eukaryotic gene control is a Many Paths process, and all the components have independently evolved several times in prokaryotes and eukaryotes, why are there no multicellular prokaryotes? We argue that there is one core difference between the logic of gene control in eukaryotes and prokaryotes, and this enabled eukarotes to evolve gigabase genomes dedicated mostly to their own control where prokaryotes could not. This logic step is the difference between whether a gene in a genome is by default 'on' or by default 'off'. Either default can be implemented in the genetics of any of the major domains of life. Given this starting point, the evolution of complex eukaryotic genomes was a highly probable, Many Paths process.