From left to right: Dion Leijnse, Ilija Milutin, Femke van Hout, Jonas Kamminga, Paula Teeuwen. Below: Luca Consoli

The applicability of autocatalytic networks to metabolic evolution

Femke A.H. van Hout, Dion Q. Leijnse, Ilija S. Milutin, Jonas L. Kamminga and Paula C. P. Teeuwen

Supervisor: Dr Luca Consoli

Abstract: *Ever since the importance of autocatalytic networks to the origin of life has been proposed, multiple models to describe autocatalytic sets have been developed. A mathematical formalisation of autocatalytic sets is given by RAF theory, that describes Reflexively Autocatalytic Food-generated networks. In this research proposal we outline experiments to further examine the applicability of RAF theory, by investigating the correlation between RAF parameters and first occurrence of reactions in evolutionary history as found by ancestral state reconstruction.*

Samenvatting: *Een vrij recente benadering om meer te weten te komen over het ontstaan van leven richt zich op zogeheten 'autokatalytische sets'. Dit zijn netwerken van reacties tussen (bio)moleculen, waarvan de producten de reacties weer versnellen. Veel van de reacties die plaatsvinden in simpele bacteriën zijn ook in onze cellen terug te vinden. Vermoed wordt dat we deze reacties van onze gemeenschappelijke voorouder doorgegeven hebben gekregen en dat de reacties waarin we verschillen pas later, in de loop van de evolutie, zijn ontstaan. Met dit voorstel willen wij onderzoeken of autokatalytische sets gebruikt kunnen worden om de leeftijd van reacties te voorspellen. Als dat zo blijkt te zijn, zou het kijken naar de plaats van een reactie in het netwerk een indicatie kunnen geven over hoe oud deze reactie is. We kunnen dan stappen terug maken in de tijd en dichter bij het ontstaan van leven komen.*

Introduction: The origin of life is widely regarded as an important unsolved scientific question. Various hypotheses have been proposed, for example the primordial soup, the RNA world, the lipid world and the iron-sulfide world hypothesis. As all these hypotheses have their supporters as well as their critics, the search for an answer still continues.

Our growing knowledge of how cellular processes form networks gives rise to new perspectives on the question of how complex life could have developed from a lifeless earth. Ever since the importance of autocatalytic networks to the origin of life has been proposed, multiple models to describe autocatalytic sets have been developed [1]. In short, autocatalytic sets are chemical reaction systems in which molecules belonging to the set catalyse the synthesis of other members of the set, as visualized in figure 1. However, not much research has been performed on the applicability of these network models. Only recently has one of these models, the RAF theory, been applied to the *E. coli* metabolism [2]. The RAF theory describes Reflexively Autocatalytic Food-generated networks. These networks have the property that, if the right starting molecules are supplied, all of the reactions in the network are catalysed by at least one molecule and that each molecule is constructed (by a sequence of reactions) from the food set [1]. This RAF theory showed to be able to

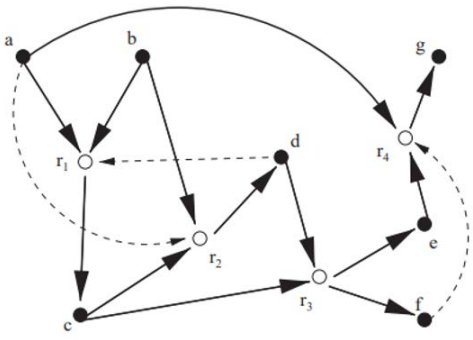

Figure 1*: Schematic representation of a reaction network. Black dots represent molecules, white dots reactions and dashed lines catalysis. Taken from Hordijk and Steel [1]*

describe 98% of the *E. coli* metabolism [2]. This indicates the utility of RAF theory to describe the metabolism of organisms.

In our research proposal we outline experiments to further examine the applicability of RAF theory, in particular for ancestral state reconstruction. Ancestral state reconstruction is the extrapolation back in time from measured characteristics of species to their common ancestors [3]. In figure 2 a visualization of ancestral state reconstruction is shown. As it is generally accepted that one unicellular organism must have preceded all species we know today [4], the quest for finding an answer to the origin of life is often related to finding this common ancestor.

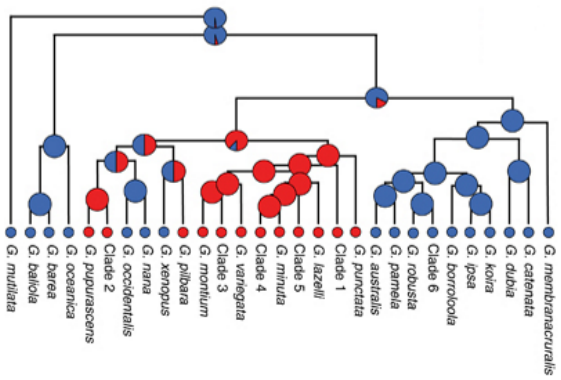

***Figure 2**: Example of a phylogenetic tree, the pie charts indicate probabilities as found by ancestral state reconstruction. Adapted from [5].*

Research Proposal: To further examine the applicability of RAF theory, the correlation between RAF parameters and the age of a reaction will be investigated. Two interesting RAF parameters can be defined: decrease in maxRAF size and height. The maxRAF size is the amount of reactions in the biggest possible RAF set in the reaction network. The decrease in maxRAF size is the number of reactions that drop out of the maxRAF set because at least one of their reactants or one of their catalysts can no longer be made from the food set [2]. The second parameter, the height of a reaction, is the numbers of steps through the shortest path of reactions from the food set to the reaction.

A model group of organisms will be chosen, for which the RAF sets will be generated using available metabolomics data. The computed RAF sets will be used to determine the height and the decrease of the maxRAF size after removal each reaction. The age of each reaction will be estimated by the maximum likelihood estimator of ancestral state reconstruction. To test the correlation between the age and the RAF parameters of each reaction, a Spearman's rank correlation test will be applied.

There are multiple possible scenarios for the statistical correlation between the decrease in maxRAF size and the age of a reaction. The first hypothesis postulates that there is a positive correlation. As older reactions have had more time to get built into the metabolic network, they are expected to be more strongly connected in the network. On the other hand, a negative correlation can be hypothesised as well. Over the course of evolution alternative pathways have been developed in a cell to compensate for environmental influences and to improve the fitness of the organism. Removal of a relatively old reaction will hardly decrease the RAF set since it has other reactions to correct for its absence. Of course, also a combination of both hypotheses is possible, giving no net correlation. The follow-up question then is which kind of reactions are mainly characterised by which correlation.

Outlook: More knowledge will be gained on how reactions are built into the metabolism of a cell and how processes emerged over the course of evolution. If indeed RAF theory can be used to gain more insight into the age of a reaction, this offers a more efficient way to determine the age of reactions compared to the conventional methods that require more computing power and data from phylogenetic relatives.

References

[1] W. Hordijk et al., "Detecting autocatalytic, self-sustaining sets in chemical reaction systems", J Theor Biol 227, 451–461 (2004).

[2] F. L. Sousa et al, "Autocatalytic sets in E. coli metabolism", Journal of Systems Chemistry 6, 4 (2015).

[3] K. E. Omland, "The assumptions and challenges of ancestral state reconstructions", Sys Bio 48, 604–611 (1999)

[4] N. Glansdorff, Y. Xu, and B. Labedan, "The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner", Biology direct 3, 29 (2008).

[5] M. Sistrom et al. (2014). Evaluating evolutionary history in the face of high gene tree discordance in Australian Gehyra (Reptilia: Gekkonidae). Heredity. 113. 10.1038/hdy.2014.6.