

The quest for the origins in evolutionary biology

Marie-Christine Maurel^a and Philippe Grandcolas^b

Institut de Systématique, Evolution, Biodiversité (ISYEB), UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, CP. 50, 57 rue Cuvier, 75005 Paris, France

Abstract. Studying origins in evolutionary biology is an endeavour to reconstruct a chronicle of past events. Traditionally rooted in comparative biology and phylogenetics, these studies can also be conducted with either logical or experimental modelling approaches. The interaction between these two domains of inference – comparative reconstruction of the evolutionary past and modeling/experimenting the evolutionary process – must be encouraged. In both domains, the study of origins needs to be carefully designed to take into account anterior evolutionary stages on which they could depend. Comparative biology and past reconstruction must be performed on simple observational data (natural kinds), not on artificial general classes. And finally, understanding causality relationships calls for correlation approaches that must be optimized with regard to the number of natural replicates.

The term “origin” is often employed when studying how some living organisms or their main characteristics appeared, or even how Life itself appeared. This term is generally considered as unambiguous [1], as it refers to very closely interrelated aspects: the early stage after which the organism or the characteristic appeared, the process that gave rise to the organism and its characteristics, the transition event between the anterior state and the state where the characteristic had already occurred, or even the characters in the anterior stage that allowed the emergence of the feature studied. This quest for the origins – a theme central to evolutionary biology – is a way to focus on special moments of evolution when some critical events occurred. It allows understanding evolutionary transitions that led to either large diversifications or to major qualitative changes and new evolutionary trajectories. This quest is traditionally rooted in systematic, comparative and paleontological studies. Yet many modeling or experimental approaches also strongly contribute to this field.

The present book deals with methodological accounts or case studies at very diverse levels and scales in varied scientific domains of comparative or experimental evolutionary biology. It exemplifies how very different studies finally point to the same kind of prospects, depend on the same general methodological requirements and open similarly wide perspectives. These chapters are issued from the contributions presented during two workshops organized by the authors at the *Institut de Systématique, Evolution, Biodiversité* (Muséum national d'Histoire naturelle, Paris) in 2013 and 2014. These workshops were intended to provide a room for discussion about the study of origins,

^a e-mail: marie-christine.maurel@upmc.fr

^b e-mail: philippe.grandcolas@mnhn.fr

stimulating exchange between scientists from different institutions and students from two Master classes of Systematics, Evolutionary Biology and Molecular Biology. We hope that the following texts will serve the same aim for a wider audience and we are very thankful to those authors who agree that the confrontation of different viewpoints helps to propose better scientific approaches.

1. Depending on the anterior stage

Speaking of origin means that something appeared or changed, and is impliedly from a previous state of being. This is the most general question regarding the study of origins, a question that introduces all others. Evolutionary biology, as all other studies about changes, needs to consider a “zero point”, in other terms, a reference state or a baseline, since it is well known that starting conditions must be considered to understand a change. In order to consider such a previous state or a zero point, the study must be designed to include a documented context and a wide sampling based on the question studied. The study will then consider a larger inclusive taxonomic group, a longer time frame or a wider geographical frame than just the group of organisms, the period or the area on which the study focuses.

This proposal could look self-evident but speaking in terms of sampling strategy for evolutionary comparative studies, it is actually quite uncommon. Evolutionary biology is still too often seen as a non-experimental science, even with the notable exception of small-scale laboratory genetic studies mostly conducted on model organisms [2]. The tradition established in experimental biology of designing experiments with the treatments and samples adequate to answer specific questions should also be carefully followed in comparative biology. It entails finding the appropriate natural replicates of evolutionary events by sampling different groups of taxa. There is still a tendency for comparative studies to focus on the object and just a few things around, although scientific sampling should depend on the question, not on the object. Actually, answering some particular questions can require a very broad sampling window, much larger than the object studied [3]. For example, understanding the origin of Life means considering biogenesis based on DNA and RNA inheritance and metabolic oxidation reactions, and therefore on planets with C, O, N elements, allowing the oxidation process based on oxygen. It brings us to the process of planet formation, much earlier than the four billion years commonly considered to study biogenesis on the Earth.

The term origin could be misleading if it is understood as describing a process of appearance, either sudden or arising from a completely different state, i.e. that is ordinarily sorted in a different class. Big questions such as the origin of Life or the origin of Man, are often treated as if such entities appeared from nothing, because Life, Man or his cultural traits are generally considered as completely different from anterior stages that would belong to another universe. This is for instance why the transition between the chemical world and the biological world is so much questioned [4–7]. This is also why the subject of synthetic Life is so striking as if it is an impossible or unacceptable performance for us to produce Life by ourselves [8, 9]. The same trend also appears to explain the origin of other appealing and exceptional features, though less unique, such as the origin of metazoans in hydrothermal vents, mimetic butterflies, or the island organisms [10–12]. These features have often been considered in a far too much focused context and the study of their origin is much improved in a wider context, wider than the vents, the mimetic systems and the islands, respectively.

These cases show how much a comprehensive logical framework with very clear epistemological bases is needed, especially when confronted with hard situations where ad hoc or irrational explanations (the so-called “beliefs”) are still frequently invoked [13]. In such cases of states of apparently “nothing before”, we must disentangle the characteristics of our objects under study and search whether some of them preexisted in potential relatives [14]. We must also look at the past with interest and modesty and realize that our present day thinking, even though proudly anchored in modern scientific and evolutionary reasoning, sometimes the same difficulties of consistency as much older ones, with respect to the dominant paradigm of the epoch [15].

2. Different kinds of deduction: Reconstructing versus modeling

Evolutionary approaches about origins rely on the reconstruction of historical chronicles [16]. These chronicles about species/cultures and their characteristics can be more or less detailed but they are supposed to include at the least the successive states and the inferred list of the evolutionary events implied. There are two ways to obtain such a chronicle, by inference and deduction.

The first way belongs classically to comparative biology now embodied in the method of phylogenetic analysis, applied to species or to cultures [2, 17–19]. It is based on an inference operation aimed at detecting consistency between the features of organisms – namely the character's states – inherited and modified in different taxa, according to the fundamental principle of descent with modification. Consistent patterns of shares of different states for the studied characters are translated into a hierarchical relationship – the phylogeny – involving varied nested degrees of evolutionary relatedness between the species/cultures. Because it is aimed at building an evolutionary chronicle, this deductive inference operation incorporates diverse process-like premises (e.g., heritability, probabilistic models of nucleotide changes). In spite of sophisticated premises, such an operation is not really a modeling procedure. Its aim is primarily to search for significant consistency among observational data according to basic process hypotheses, not to simulate a process by building a predictive or explanatory logical framework.

There is a second way to search for an evolutionary chronicle. In case of an anterior stage in the evolutionary process belonging to another universe or in case of an auto-destructive or fast process, an approach based on reconstruction can be quite difficult and more speculative. Observational data, even those based on comparative studies between taxa or on traces of past events (fossils, etc.), may simply not be available in these cases (but see [20–22]). The possible alternative approach is then to build a model, either purely formal (logical and most often mathematical) or experimental. This model is aimed at simulating evolution, and exploring the possible processes and chronicles that conducted to the states presently observed. This approach can be more explanatory than just reconstructing a pattern, given the properties of the model, but it must be clear that many different models can lead to predictions consistent with the world as it is observed [23]. For this reason, modeling approaches can be considered exploratory, aimed at provoking understanding or thinking rather than providing an accurate reconstruction in a particular case study. Both reconstructing and modeling are complementary and can stimulate each other.

Several examples can be provided for cases where modeling approaches have been considered particularly valuable:

- To understand the origin of life and in the absence of detailed knowledge on the prebiotic environment and stages, biogenesis experiments are conducted, putting together the materials and the environment supposedly present, four billions years ago, to explore the possible working processes [4–7].
- Some terrestrial regions may be essential to understand the evolution of specific distributions or regional climates but these regions could have been subducted and melted by tectonic plates, or very rapidly eroded in ancient times. Consequently, there might remain no direct evidence of their role, and hence regional tectonic, geological or climatic models may be needed to explore this possibility [25].
- Some evolutionary processes such as social evolution may be fast or punctuated, and intermediate states can be ephemeral or even lacking [26]. Searching for remaining intermediate stages of a gradual process may be a red herring and some cases may be better studied by modeling the transition process and possibly realizing that it can be punctuated or abrupt [27]. This is one of the many reasons why the concept of missing link is misleading [28].

3. Deconstructing the classes

The reconstruction of past processes requires that classes or categories built by the human mind and which cannot be observed and followed as such be deconstructed. Classes or categories are opposed to logical individuals or natural kinds that exist independently of us, according to a well-known duality in science philosophy [24]. In other words, classes must be decomposed in different traits that can be directly observed: No one can observe the classes “Life” or “Species” by themselves but natural kinds such as individual specimens and their development, their reproduction, their behaviour, their phenotypic characters and at a lower scale their genetic characters or molecular replication processes. Many difficulties in the reconstructing approach are often related to the elaboration of general classes that cannot be logically observed and whose changes cannot be traced back or mapped on phylogenetic trees [14].

This is a logical pitfall related to a generalization step that is placed early in the reasoning procedure. Generalizations can possibly be used first as a heuristic tool to suggest general patterns or to explore causality relationships but they need to be deconstructed to conduct specific studies. Then generalizations can possibly be established in a final generalization procedure based on several case studies.

Modeling is more tolerant to the problem of classes than reconstructing. Models do not necessarily search for consistencies among the data observed; they are logical constructions that can stand on their own and, as such, they can deal with various parameters or generality levels depending on their goal [23].

4. The correlation approach and its limitations

Typically, once an evolutionary chronicle is obtained, the factors that caused the succession of events are searched with correlation analyses. Most often, they are built in adaptational explanations, using several candidate factors [11, 23]. But these potential causality factors can be numerous and inter-related, making the statistical analysis difficult. The analysis strategy needs to be adapted in several possible ways.

Increasing the number of samples to allow examining the role of more factors is often very demanding. In comparative evolutionary biology, this requires increasing the number of natural replicates by finding more species or more diverse species to compare when all the presently known or easily available species are already sampled within the clade investigated. Finding more fossils can be a very elusive task, given that the useful periods and locations can first appear to be without traces or remains [20–22]. Yet finding such extant or fossil replicates is part of the design to study evolutionary biology and must be considered as the taxonomic heart of the problem.

Conversely, considering a priori the role of only a few factors allows one to deal with a lower number of samples. Selecting such a low number of factors is a task that can be helped by exploratory modeling approaches. It can however remain speculative and can prevent detection of any indirect causality by pruning some critically related factors.

5. Conclusion

The study of origins is strongly constrained by the very nature of evolutionary biology, a typically non-experimental science where replicates need first to be found in past natural events [2]. True evolution experiments where selective pressures or environments are created may represent valuable explanatory models but they cannot totally replace the study of deep time evolution and of natural evolutionary patterns. Both domains of activities must interact and stimulate each other. One of the expected stimulations of such an exchange is to design study protocols whose sampling is appropriately focused on the question, not the study object.

ORIGINS

We are very grateful to the colleagues who accepted to give a talk at the workshops “Origins” of the Institut de Systématique, Evolution, Biodiversité (UMR 7205 CNRS MNHN UPMC EPHE) and who in addition produced a stimulating chapter for this book. We are grateful to the Labex BCDIV “Diversités biologiques et culturelles: origines, évolution, interactions, devenir” for financing the production of the volume and permitting its publication. We also thank Hélène Citerne who skillfully helped with several chapter translations and Anne-Lise Haenni and Marc-André Selosse who kindly commented on a first draft of this introduction. Thanks a lot to Laura Da Silva and Cécile Anne who produced the cover with slides sent by several authors.

References

- [1] Merriam-Webster.com from <http://www.merriam-webster.com/dictionary/origin> (February 9, 2015)
- [2] R.J. O’Hara, In *New Perspectives on the History of Life: Systematic Biology as Historical Narrative* (M.T. Ghiselin, G. Pinna, eds), *Memoirs of the California Academy of Sciences* **20**, 7–17 (1996)
- [3] P. Grandcolas, E. Guilbert, T. Robillard, C. D’Haese, J. Murienne, F. Legendre, *Cladistics* **20**, 579–582 (2004)
- [4] J.F. Lambert, *BIO Web of Conferences* **4**, 00012 (2015)
- [5] A. Lazcano, *BIO Web of Conferences* **4**, 00013 (2015)
- [6] C. Malaterre, *BIO Web of Conferences* **4**, 00002 (2015)
- [7] M.C. Maurel, *BIO Web of Conferences* **4**, 00014 (2015)
- [8] B. Bensaude-Vincent, *BIO Web of Conferences* **4**, 00015 (2015)
- [9] T. Heams, *BIO Web of Conferences* **4**, 00016 (2015)
- [10] M. Elias, M. Joron, *BIO Web of Conferences* **4**, 00008 (2015)
- [11] P. Grandcolas, *BIO Web of Conferences* **4**, 00010 (2015)
- [12] S. Samadi, *BIO Web of Conferences* **4**, 00007 (2015)
- [13] G. Lecointre, *BIO Web of Conferences* **4**, 00004 (2015)
- [14] P. Grandcolas, R. Nattier, F. Legendre, R. Pellens, *Cladistics* **27**, 181–185 (2011)
- [15] C. Blanckaert, *BIO Web of Conferences* **4**, 00003 (2015)
- [16] R.J. O’Hara, *Biology and Philosophy* **7**, 135–160 (1992)
- [17] W. Hennig, *Phylogenetic systematics* (University of Illinois Press, 1966)
- [18] E.O. Wiley, *Phylogenetics. The theory and practice of phylogenetic systematics* (Wiley-Liss, 1981)
- [19] R. Mace, C.J. Holden, *Trends in Ecology & Evolution* **20**, 116–121 (2005)
- [20] E. Geigl, E.A. Bennett, T. Grange, *BIO Web of Conferences* **4**, 00005 (2015)
- [21] A. Nel, *BIO Web of Conferences* **4**, 00006 (2015)
- [22] M.A. Selosse, C. Strullu-Derrien, *BIO Web of Conferences* **4**, 00009 (2015)
- [23] R. Levins, *American Scientist* **54**, 421–431 (1966)
- [24] M. Mahner, M. Bunge, *Foundations of biophilosophy* (Springer, Berlin, 1997)
- [25] D.N. Reusch, *Geology* **39**, 807–810 (2011)
- [26] C.D. Michener, In *Experimental Behavioral Ecology and Sociobiology* (B. Hölldobler, M. Lindauer, eds), *Fortschritte der Zoologie* **31**, 293–305 (1985)
- [27] N. Eldredge, S.J. Gould, In *Models in Paleobiology* (T.J.M., Schopf ed), 82–115, (Freeman, Cooper and Company, 1972)
- [28] P. Grandcolas, R. Nattier, S.A. Trewick, *Trends in Ecology & Evolution* **29**, 655–663 (2014)