EO: THE EVOLUTION ONTOLOGY

Adam M. Goldstein
American Museum of Natural History, 79th & Central Park West, New York, NY 10024
z_californianus@shiftingbalance.org, http://shiftingbalance.org

Abstract

Ontologies describing organic evolution focus on relationships of descent among lineages: they are aimed at reconstructing phylogeny. No ontology exists describing evolutionary processes in a manner suited for explaining how and why patterns of phylogeny or other forms of organic diversity came to be. The Evolution Ontology, EO, is intended to fill this lacuna in the landscape of bio-ontology. Expected payoffs include research into the structure and function of entities at all levels of organic hierarchy, and the organization of data, biological materials, and literature.

1. Motivations

In addition to reconstructing the history of life on Earth, a central aim of evolutionary biology is to understand how and why that history turned out the way it did turn out. Understanding why requires knowing the general laws and particular circumstances obtained so that, accounting for their stochastic nature, the states of organic diversity that actually obtained ought to have been expected to do so, or were physically necessary: understanding how requires recapitulating, in cases in which organic changes are not physically necessary nor could have been expected to occur, the sequence of events by which states of organic diversity have come to be as they are. The answers to why and how questions about evolution are formulated in terms of the processes of evolution: phylogeny, natural selection, mutation, migration, extinction, and the like. Accordingly, applying tools of computational biology to these questions requires a formal description of these processes. EO is just such a description. The motivations for developing EO are threefold:

1. Annotation of data at all levels of biological investigation with EO terms will facilitate hypothesis generation and testing about fundamental biological questions. No Gene Ontology or MGI Nucleotide database record at present contains information about the processes of evolution, present or past, that effect the structure and prevalence of a gene or other heritable material. Variation under control of natural selection is functionally important; knowing which genes (or other heritable material) have a history of variation under control of natural selection is functionally important; knowing which genes (or other heritable material) have a history of selection brings the researcher toward identifying functionally important entities and understanding their biological role. Some researchers are pursuing these types of studies using genome-wide data.

2. Building computer systems containing data sets by using metadata including EO terms will facilitate discovery of relevant information about evolutionary processes in repositories of otherwise untractable size.

3. Organizing and mining literature using EO terms will facilitate discovery of theories and arguments about evolutionary processes, past and present. Machine indexing and reasoning with ontologies makes mining very large text bases a real prospect.

2. Modeling Processes of Evolution

A good model of processes of organic evolution will describe the following:

1. An entity, such as a species, local population, or family unit, which persists through time, and which varies in some heritable state.

2. A characteristic pattern or time-course of the state of variation during the time period under study.

3. In some cases, the causal background or mechanism by which the characteristic pattern arises.

For instance, natural selection takes place in a species or local population. Variants (alleles) in the genetic material differ in fitness, that is, they differ in their disposition to survive and reproduce. If natural selection occurs, fitter genes will increase in representation over less fit genes. Speciation occurs when there is a biological separation between one population and another that constitutes a change in kind. In contrast with natural selection, speciation need not occur, like natural selection, by a single kind of mechanism, i.e., fitter genes having greater success in survival and reproduction due to a stronger disposition to survive and reproduce for both of the latter.

3. Semantics

Continuants & Occurrents

Desiderata for a model of evolutionary processes EO are met in EO by the introduction of several new universals (figs. 1 and 2).

Continuant

DependentContinuant

SpecificallyDependentContinuant

Quality

<VaryingState>

IndependentContinuant

MaterialEntity

ObjectAggregate

<OrganicLineage>

Site

<GeographicPlace>

Figure 1: EO continuants (surrounded with **) in the Basic Formal Ontology, v. 1.1, at http://www.ifomis.org/efo/home

<OrganicLineage>

is the evolving entity, such as a species, population, or phyllum: it persists throughout the entire period during which it is changing. <VaryingState> is some property of this lineage which can vary through time: the number of species in a phyllum or their geographic distribution; the frequency of an allelic difference; or heterozygosity. <Site> is a place or location.

Figure 2: EO occurrences (surrounded with **) in the Basic Formal Ontology, v. 1.1; see caption to fig. 1 for reference.

Relations

HasVaryingState
takes, as arguments, an instance of OrganicLineage and VaryingState, for instance, to describe a population whose allele frequencies vary. OccursInEvolutionaryContext places an evolutionary process in which an instance of OrganicLineage participates in relationship to the conditions affecting evolutionary change. This can be used to state, for instance, that speciation occurs in geographic isolation. HasParticipant expresses the claim that a particular population of, say, Geospiza fortis, undergoes natural selection.

<OrganicLineage>

hasVaryingState <VaryingState>

<EvolutionaryProcess> OccursInEvolutionaryContext

<EvolutionaryProcess>

hasStart <StartOfEvolutionaryProcess>

<EvolutionaryProcess> hasEnd <EndOfEvolutionaryProcess>

<OrganicLineage> hasGeographicPlace <GeographicPlace>

<EvolutionaryProcess> hasParticipant <OrganicLineage>

Figure 3: EO relations, hasParticipant is taken directly from the Relation Ontology (at http://www.obofoundry.org/efo); others are unique to EO.