

This is the preprint version. The final version is available as:

Trends in Genetics 26: 99–102

<http://dx.doi.org/10.1016/j.tig.2009.12.012>

## **An ontology to clarify homology-related concepts**

Julien Roux<sup>1,2</sup>, Marc Robinson-Rechavi<sup>1,2</sup>

<sup>1</sup>Université de Lausanne, Département d'Ecologie et d'Evolution, Quartier Sorge, 1015  
Lausanne, Switzerland.

<sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland.

*Corresponding author:* Robinson-Rechavi, M. ([marc.robinson-rechavi@unil.ch](mailto:marc.robinson-rechavi@unil.ch))

Although homology is a fundamental concept in biology, and is one of the shared channels of communication universal to all biology [1], it is difficult to find a consensus definition [2]. The interpretations of homology have changed as biology has progressed.. New terms have been introduced into the literature, such as paramorphism [3], with mixed success. In addition, different research fields operate with different definitions of homology, for example the mechanistic usage of Evo-Devo [4] is not strictly historical, and would not be acceptable in cladistics. This makes a global understanding of homology complex, whereas the integration of evolutionary concepts into bioinformatics and genomics is increasingly important. We propose an ontology organizing homology and related concepts, which might provide a solution, and we hope it will also facilitate the integration and sharing of knowledge among the community.

## **The problem: the concept of homology is divided by specialized usage**

[1]The lack of a consensus definition of homology does not prevent us from perceiving and recognizing homologies in practice. Scientists have long been trying to understand the underlying cause of homology [1, 2, 5]. Several working definitions exist in specific fields of research. One example is the concept of homology based on common descent, applied at the molecular level. Many terms describing specific evolutionary histories of sequences, such as orthology or paralogy (Figure 1), are commonly used in genetics and molecular evolution. But the abundance of terms has become another hurdle to a good understanding of homology related concepts for biologists [1]; most of them are redundant or very specialized.. Importantly this terminological confusion can also hinder large-scale studies: in comparative and evolutionary biology, with the exponential increase of data available, the use of high-

throughput computational tools is now generalized. There is a need for a bioinformatics framework to deal with the multiplicity of concepts related to homology.

## **Towards a solution: an ontology of homology related terms**

An ontology can provide such a framework. Ontologies are increasingly being used for data integration in biology [6] and can provide an efficient way to organize knowledge. Based on definitions from the literature, we have reviewed and organized terms related to the concept of homology into an ontology with an emphasis on the terms in modern use. This accounts for 65 terms plus 67 synonyms. The HOM ontology is presented according to Open Biological Ontologies Foundry principles [7] (<http://www.obofoundry.org/>), including a definition of each term and key references. The relationships between the terms are explicit, with some concepts as sub-classes of others (Figure 1). An overview of the type of information gathered is shown in Table 1; the full details can be obtained from the following website <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=HOM>.

### **Similarity as root**

An important choice when developing an ontology is the choice of the root (i.e. the most general term) because this defines the domain of application of the ontology. The root of the HOM ontology is ‘similarity’, or ‘sameness’. To quote Stevens: “without some similarity, we should not even dream of homology” [8]. We define it as a relation between biological objects that resemble or are related to each other sufficiently to warrant a comparison. ‘Homology’ is thus a sub-class of similarity. Another is ‘homoplasmy’ (or ‘analogy’, but the use of this term is ambiguous in the literature), describing similarity due to independent evolution. These two concepts are traditionally considered as disjoint (although see Ref. [9]), and are defined as such in HOM.

Other sub-classes of similarity are independent of a homology hypothesis: ‘homocracy’ is the relation between two structures that share homologous patterning genes [10] and ‘functional equivalence’ is used to state that two structures share the same function.

## **Working definitions of homology**

We propose a broad definition of homology, which encompasses the definitions proposed so far and can be seen as a common denominator or minimal approach: 'similarity that results from common evolutionary origin' [5].

Three different operational definitions, which are not disjoint [4], are gathered under this broad umbrella: (i) 'Historical homology' is the notion of similarity due to common descent [5]. (ii) 'Biological homology', fitting evo-devo usage, is process-oriented and more mechanistic, focusing on establishment and individualization of structures through common developmental processes [4]. It accommodates repeated parts of the same organism ('iterative homology') and sexually differentiated parts of individuals (e.g. testis and ovaries). (iii) 'Structural homology' refers to the traditional criteria of homology focused on similarity with regard to selected structural parameters (sometimes called 'idealistic homology' [1, 4]).

## **Multiple inheritance**

An ontology can represent complex concepts by encoding multiple inheritance: a term can be a sub-class of more than one other term. Examples where homology statements do not translate smoothly between multiple levels of organization (e.g. anatomical structures and genes) are easily represented. For example, 'deep homology' is a sub-class both of homoplasy and of homocracy, because it involves anatomical structures that result from independent evolution and yet share the expression of homologous patterning genes [11].

## **Availability**

The HOM ontology is available at [www.obofoundry.org](http://www.obofoundry.org). Interactive views are available at the Bioportal (<http://bioportal.bioontology.org/ontologies/40983/>, see Figure 1) or the Ontology Lookup Service at EBI (<http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=HOM>).

## **Concluding remarks**

Discussions related to the concept of homology have led to repeated confusion. Like discussions on the terms ‘species’ or ‘gene’, it is not clear whether a better understanding will simply emerge from future advances in biology. Indeed, what makes the concept intrinsically difficult to outline is probably the complexity of living organisms and their evolution. As West-Eberhard puts it: “evolution makes a mess of homology” [12].

In this context, we feel that the most helpful solution is to order and clarify existing concepts. This should provide an evolvable tool for computational studies, and a framework for future conceptual advances (i.e. proposals for new terms should be set in relation to existing concepts).

## **Acknowledgments**

We acknowledge funding from Etat de Vaud and Swiss National Science Foundation grant 116798. We thank Frederic Bastian and other members of the lab for discussions; the members of the OBO foundry mailing list for their suggestions; and the National Center for Biomedical Ontology for authorization of using their visualization tool in Figure 1. We apologize to the many colleagues whose work could not be cited because of space limitations; all citations relevant to specific terms are included in the HOM ontology.

**Table 1. Example of data represented in HOM ontology for ‘paralogy’ and ‘latent homology’**

	Example 1	Example 2
<b>Id</b>	HOM:0000011	HOM:0000057
<b>Name</b>	Paralogy	Latent homology
<b>Definition</b>	Historical homology that involves genes that diverged after a duplication event.	Parallelism that involves morphologically very similar structures, occurring only within some members of a taxon and absent in the common ancestor (which possessed the developmental basis to develop this character).
<b>is_a</b>	HOM:0000007 historical homology	HOM:0000005 parallelism HOM:0000058 syngeny
<b>References</b>	Fitch WM (2000) Homology: a personal view on some of the problems. Trends in Genetics 16:227-231. DOI:10.1016/S0168-9525(00)02005-9 Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst. zool. 19(2): 99-113. PMID:5449325 Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. Annual	Rutishauser R and Moline P (2005) Evo-devo and the search for homology (“sameness”) in biological systems. Theory in Biosciences 124:213-241. DOI:10.1007/BF02814485 Hall BK (2007) Homoplasy and homology: Dichotomy or continuum? Journal of Human Evolution. 52:5, 473-479.

	<p>Review of Genetics, 39: 309-338.</p> <p>DOI:10.1146/annurev.genet.39.073003.114725</p>	<p>DOI:10.1016/j.jhevol.2006.11.010</p> <p>Sanetra M et al. (2005) Conservation and co-option in developmental programmes: the importance of homology relationships. <i>Frontiers in Zoology</i> 2:15. DOI:10.1186/1742-9994-2-15</p> <p>de Beer G (1971). <i>Homology, an unsolved problem</i>. London, Oxford University Press. ISBN:0199141118</p>
<b>Cross-references</b>	<p>SO:0000854 paralogous_region</p> <p>SO:0000859 paralogous</p> <p>SO:paralogous_to</p>	
<b>Comment</b>		Used for structures in closely related taxa
<b>Synonyms</b>		<p>Apomorphic tendency (exact)</p> <p>Cryptic homology (exact)</p> <p>Homoiology (related)</p> <p>Homoplastic tendency (related)</p> <p>Re-awakening (related)</p> <p>Underlying synapomorphy (exact)</p>

## References

- 1 Kleisner, K. (2007) The Formation of the Theory of Homology in Biological Sciences. *Acta Biotheoretica* 55, 317-340
- 2 Brigandt, I. (2003) Homology in comparative, molecular, and evolutionary developmental biology: The radiation of a concept. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 299B, 9-17
- 3 Minelli, A. (2000) Limbs and tail as evolutionarily diverging duplicates of the main body axis. *Evol Dev* 2, 157-165
- 4 Wagner, G.P. (1989) The Biological Homology Concept. *Annual Review of Ecology and Systematics* 20, 51-69
- 5 Hall, B., ed (1994) *Homology: the hierarchical basis of comparative biology*. Academic Press
- 6 Bard, J.B. and Rhee, S.Y. (2004) Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 5, 213-222
- 7 Smith, B., et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech* 25, 1251-1255
- 8 Stevens, P. (1984) Homology and phylogeny: morphology and systematics. *Systematic botany* 9, 395-409
- 9 Hall, B.K. (2007) Homoplasy and homology: Dichotomy or continuum? *Journal of Human Evolution* 52, 473-479
- 10 Nielsen, C. and Martinez, P. (2003) Patterns of gene expression: homology or homocracy? *Development Genes and Evolution* 213, 149-154
- 11 Shubin, N., et al. (1997) Fossils, genes and the evolution of animal limbs. *Nature* 388, 639-648
- 12 West-Eberhard, M. (2003) *Developmental plasticity and evolution*. Oxford University Press, USA

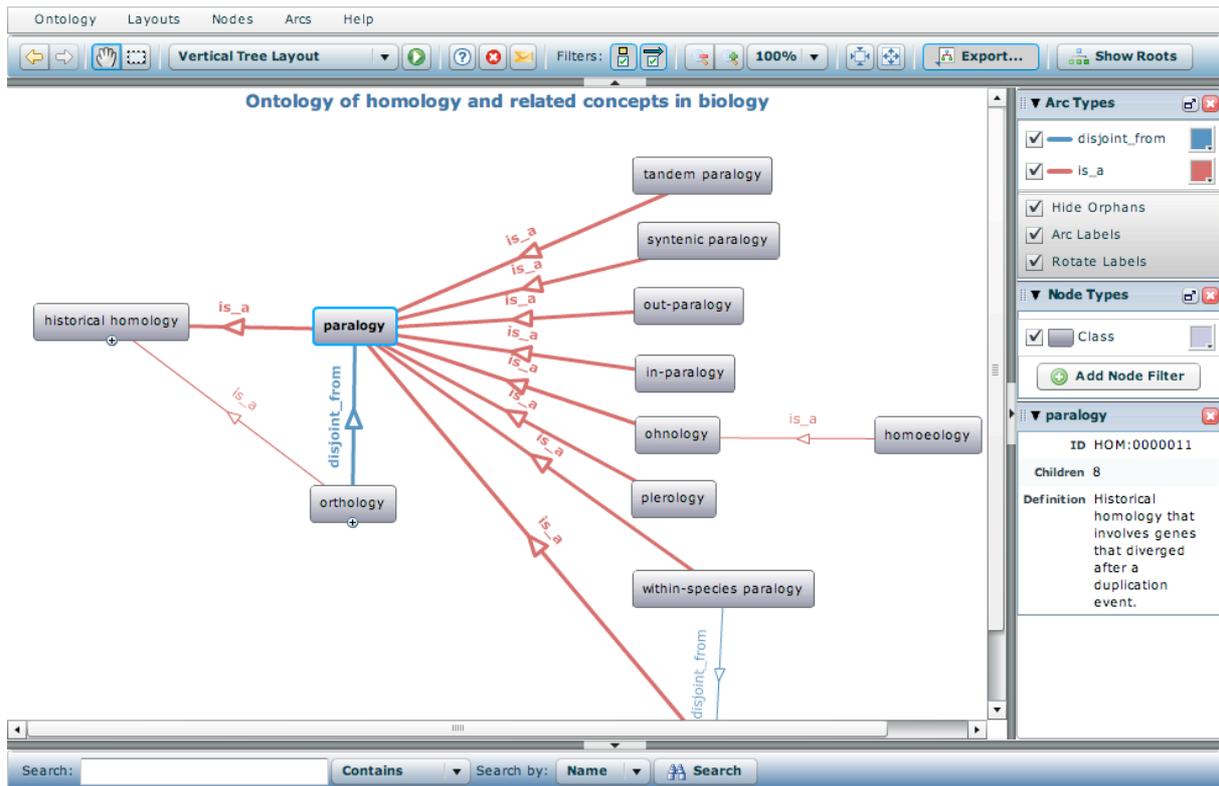


Figure 1. A partial view of the ontology of homology and related concepts (HOM). The concepts related to the concept of ‘paralogy’ are displayed. Boxes represent terms, arrows represent relations between the terms. The relation ‘is\_a’ denotes that one term is a sub-class of another. Courtesy of the National Center for Biomedical Ontology. Copyright © 2005–2009, Stanford University. <http://biportal.bioontology.org> and <http://keg.cs.uvic.ca/ncbo/flexviz/FlexoViz.html>