

# Bad habits about homology

Marabotti A<sup>(1)</sup>, Facchiano A<sup>(1)</sup>.

<sup>(1)</sup> Laboratory of Bioinformatics and Computational Biology,  
Institute of Food Science, CNR, Avellino

## Motivation

The scientific community has discussed for a long time the use and misuse of the term “homology”. In 1987, an interesting debate started by a Letter to the Editor of *Cell* <sup>(1)</sup> and commentaries in other prestigious journals <sup>(2-4)</sup> outlined the need for a careful use of the term “homology”, stating that its precise meaning is “having a common evolutionary origin” and that this is a concept of quality and cannot be “quantified” <sup>(2)</sup>. However, the common experience shows that “homology” is often wrongly used instead of “similarity” in articles describing a comparison of protein or nucleic acid sequences. Since sequence comparison is a leading issue for bioinformaticians, it is important that they use this term properly. To awaken the community on this issue, we made a survey of the literature to check if this term is correctly used, 20 years and more after that debate.

## Methods

We searched the PubMed archive for articles published in 2007, which have the keyword “homology” in their abstract or title, by excluding those cases in which “homology” is part of a gene or protein name (e.g. Bcl-2 homology domain) or indicates a procedure (e.g. “homology modelling”). With the same criteria, we performed an analogous search in the abstracts of articles published in 1986, one year before the debate.

## Results

We noticed that in 2007, the term “homology” was still used incorrectly in 43% of the selected abstracts. In 264 abstracts, the term “homology” was associated with a percentage value, in 452 abstracts it was used together with terms such as “high”, “low” and so on. In 94 abstracts, the term “homology” was associated to “significant” and finally, another 28 abstracts report a misleading association indicating somehow a quantitative evaluation of the “homology”. In 1986, the term “homology” was incorrectly used in about 51% of abstracts analyzed. Looking at the different types of errors, we noticed that the frequency of the expression “percentage of homology” was more or less the same as in 2007, whereas the frequency of expressions like “high homology”, “low homology” and so on, was higher in 1986 than it was in 2007. We checked for the 20 journals in which the abstracts containing “homology” appeared more frequently in 1986. They account for 67% of the analyzed abstracts of that year. The same journals in 2007 represented only 17% of the abstracts containing “homology”. Regardless of the reasons for this difference, we noted that the percentage of errors is decreasing in almost all of them, but often it is just a small decrease. A further investigation about was also performed comparing the misuse of the term “homology” in the English abstracts of articles in other languages with the whole number of articles. Our simple search with PubMed does not include the whole text, so we cannot exclude that in the article itself the authors used “homology” correctly, or that many other similar mis-uses occurred. However, our results suggest that the lessons of the 1987 debate have not been fully acknowledged: despite a small improvement in the term’s usage after 20 years, it appears that this particular bad habit dies hard. To improve the quality of publications, more journals should add in their guidelines some specific suggestions for the correct use of the terms and also ensure that such errors do not get past the copy editors, and the scientific community should support initiatives for the education of young researchers, especially from emerging countries. Last but not least, researchers should always read and learn from the past lessons, and get themselves used to checking the formal correctness of their language before submitting an article.

## References:

- <sup>1)</sup> Reeck GR, de Haen C, Teller DC, Doolittle RF, Fitch WM, Dickerson RE, Chambon P, McLachlan AD, Margoliash E, Jukes TH and Zuckerkandl E: “Homology” in proteins and nucleic acids: A terminology muddle and a way out of it. *Cell* (1987), 50, 667.
- <sup>2)</sup> Lewin R. When does homology mean something else? *Science* (1987), 237, 1570.
- <sup>3)</sup> Kimelberg HK. “Homology” controversy. *Science* (1987), 238, 1217. <sup>4)</sup> Aboitiz F. Nonhomologous views of a terminology muddle. *Cell* (1987), 51, 515-516. A related article is currently in press in *Trends in Biochemical Sciences* (DOI information: 10.1016/j.tibs.2008.12.001)

**Contact :** angelo.facchiano@isa.cnr.it