

Bad habits about homology

Anna Marabotti and Angelo Facchiano

Laboratory of Bioinformatics and Computational Biology Institute of Food Science, Italian National Research Council, Via Roma 64, 83100 Avellino, Italy

Bioinformatics

E-mails: anna.marabotti@isa.cnr.it; angelo.facchiano@isa.cnr.it

What does the term 'homology' mean?

The precise meaning of this term is 'having a common evolutionary origin'. However, it is often wrongly used instead of 'similarity'. This misleading use is frequent in articles describing a comparison of protein or nucleic acid sequences. This bad habit has been discussed over the years, and probably the maximum peak of discussion was reached in 1987 when an interesting debate started by a letter to the Editor of Cell [1] and commentaries in other prestigious journals [2-4] outlined the need for a careful use of the term homology in the literature, stating that homology is a concept of quality and cannot be 'quantified' [2].



after that debate? We searched the PubMed archive for articles published in 2007 that 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 are shown on the right.



In the table we report 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'. The percentage of errors is decreasing in almost all of them, whereas in the other journals the percentage of errors in 1986 and 2007 is unchanged.

Journal title	n. of abstracts containing "homology" in	n. of errors in abstracts containing	% of errors in 1986	n. of abstracts containing "homology" in	n. of errors in abstracts containing	% of errors in 2007	Abstracts of articles written in foreign languages containing 'homology' and global percentage of errors found:		
	1986	"homology" in 1986		2007	"homology" in 2007		1084	2007	
Biochem Biophys Res Commun	41	27	65.8	25	10	40.0	1900	2007	
Biochemistry	27	15	55.5	22	10	45.4			
Biochem J	29	15	51.7	8	4	50.0		70 25	
Biochim Biophys Acta	18	8	44.4	16	8	50.0	10 5	/8 25	
Cell	25	13	52.0	1	0	0			
Eur J Biochem/FEBS J	27	16	59.2	12	4	33.3			
EMBO J	64	33	51.6	5	2	40.0			
FEBS Lett	42	23	54.8	9	5	55.5			
Gene	62	38	61.3	24	17	70.8			
J Bacteriol	54	22	40.7	28	10	35.7			
J Biol Chem	119	67	56.3	60	16	26.7		4	
J Mol Biol	32	17	53.1	29	9	31.0			
J Virol	38	21	55.3	16	7	43.7			
Mol Cell Biol	47	18	38.3	7	2	28.6			
Mol Gen Genet	24	14	58.3	0	0	n.d.			
Nucleic Acids Res	113	57	50.4	29	3	10.3			
Nature	36	19	52.8	6	2	33.3	+ (4)		
Proc Natl Acad Sci USA	156	90	57.7	24	5	20.8			
Science	17	9	52.9	2	0	0		v 4	
Virology	36	19	52.8	9	6	66.7			
Others	485	213	43.9	1634	717	43.9	% of errors: 32%	% of errors: 62%	

Our analysis indicates that the lessons of the 1987 debate have not been fully acknowledged or applied: despite a small improvement in the usage of the term after 20 years, it seems that this particular bad habit dies hard. To improve the quality of publications, 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 copyeditors. 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. The results of this analysis have been published [5] with the aim of awaken the research community on this subject. Acknowledgements

The research activity of the laboratory is supported by internal funds and by project contracts no. 527/A/3A/5 and ACC2/R7.11. The laboratory is a member of the CNR-Bioinformatics network (http://www.cnr-bioinformatics.it), Rete Nazionale di Bioinformatica Oncologica (http://www.rnbio.it) and Rete Regionale di Bioinformatica e Biologia Computazionale (http:// bioinformatica.isa.cnr.it/BBCC).

- [3] Kimelberg, H.K. (1987) 'Homology' controversy. Science 238, 1217
 [4] Aboitiz, F. (1987) Nonhomologous views of a terminology muddle. Cell 51, 515-516

[5] Marabotti A, Facchiano A. When it comes to homology, bad habits die hard. Trends Biochem Sci. 2009, in press. Epub Jan 30. PMID: 19181528

^[1] Reck, G.R. et al. (1987) 'Homology' in proteins and nucleic acids: a terminology muddle and a way out of it. Cell 50, 667 [2] Lewin, R. (1987) When does homology mean something else? Science 237, 1570