

Comment

Homologuephobia

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
E-mail: petsko@brandeis.edu

Published: 8 February 2001

Genome Biology 2001, **2**(2):comment1002.1-1002.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/2/comment/1002>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

I might as well start by admitting that I don't like the preferred spelling of the word homologue. That 'ue' at the end seems gratuitous to me, rather like the pretentious 'e' that sometimes shows up on the names of expensive business establishments ('Gift Shoppe'; 'The Olde Tavern'). But every time I type 'homolog', my word processor flags it as an incorrect spelling, flaunting its smug superiority at me by underscoring the word in wavy red lines.

Well, maybe I can't spell homolog (although for the purpose of this column I am going to spell it the way I want to), but I can at least use it correctly. And that seems to be a rare thing these days. One of the downsides of genomics is that it has caused biologists to use a lot of new words, some that have recently been made up and others that were not in common use before. The former add unnecessary and undesirable jargon to our discourse; the latter muddy the waters by being frequently misused.

The creation of new terms seems to be an irreversible trend, but I wish it could be stopped. Genomics is best carried out by multidisciplinary teams, but meaningful communication between scientists of different backgrounds is not aided by the use of jargon words that are not easy to understand from their context. Medicine is famous for this, of course, but at least physicians have the excuse of wanting to build a wall of mystery around their profession to provide themselves with the distance and authority they believe they need to deal with patients effectively. Scientists have no such justification; in fact, they should eschew anything that separates them from the public, who, after all, pay for their research. The rationale I hear most often is that of economy of expression, and I concede that brevity is often desirable, but not at the expense of ease of understanding. Do we really lose so much time and word-space by substituting 'programmed cell death' for 'apoptosis', a word no one is even sure how to pronounce?

A physicist who wants to enter biology has to learn a new way of thinking; do we have to make them learn a whole new

language too? The words 'ortholog' and 'paralog' (note that my spelling is at least consistent) in my view add nothing to our subject, and by replacing easily understandable simple phrases they cause us to forget our assumptions and substitute the appearance of erudition for an attempt to be clear to everybody. I fail to see that ours is a better world for the invention of the term 'proteomics' either, especially since it seems to mean different things to almost everyone who is trying to do it. And why on earth do we need 'metabolomics', which doesn't even sound nice, or 'transcriptome' (which is clearly a dense book of American academic records)? Instead of wasting their time deciding to replace perfectly good units like the atmosphere and the kilocalorie and the Angstrom with daft ones like the Pascal and the kiloJoule and the nanometer, the International Union commissions on nomenclature should be substituting plain English expressions for the unruly mob of new terms that have descended upon us.

But even worse than a silly new word is an old one that is seldom used correctly. It happens with phrases all the time. Shakespeare never wrote 'to gild the lily'; he wrote 'to gild refined gold, to paint the lily'. Bogart never said 'Play it again, Sam'; he said 'You played it for her, you can play it for me. Play it'. No real harm is done by that sort of thing. But the misuse of a technical term can obscure meaning. No word provides a more compelling an example of this problem than 'homolog'. Biologists look at a sequence and say 'protein X is 43% homologous to protein Y'. Well, it's not. The two sequences can be 43% identical or they can be 43% similar, but they can't be 43% homologous. There is no such thing as percent homology. The meaning of homologous is 'related by divergent evolution from a common ancestor'. That's the only thing it means. You can't be partially homologous: that would be like being partially dead, or partially pregnant. You're either homologous or you're not.

If one gives a percent homology when talking about the relationship between two sequences, the reader or listener has

no idea whether what is meant is identity or similarity - and the difference matters a lot. Two sequences that are 43% identical clearly belong to homologous proteins; two sequences that are 43% similar may be less than 20% identical, a gray area in which proteins are not obviously descended from a common ancestor. In such ambiguous cases only structural similarity can confirm the evolutionary connection between sequences, but we won't know whether we need it if we say they are 43% homologous. We must reserve the word homolog for those cases where its precise meaning applies; if we don't, we lose the distinction between numerical relationship of sequences and the underlying genetic history of the proteins.

Functional relationship is another problem altogether, which is why I find the words 'ortholog' and 'paralog' more harmful than helpful. When we say that two molecules perform the same function, we imply either that each has only one function or that we know all of their functions and that they all overlap. I doubt we can be certain of either situation for even a handful of proteins in the genomes of higher organisms. Experimental replacement of one gene by the other in an organism is no guarantee either, for the two proteins may share a common function but have other functions that differ. The modular nature of protein construction makes this sort of situation quite likely. In the end, the real problem with obscure new words and misused old ones is the same: they make it seem as though we understand more than we do. And if genomics has one key thing to teach us, it's that we actually understand very little.