In the beginning there was babble...

Why do we need a unified nomenclature? To some extent we do not. Does it matter if we refer to that spherical fruit as an orange and you refer to it as an apple? In this case it is probably not a matter of life and death, but it can be annoying. In other cases, it can be critical. The German physician Robert Koch realized the importance of knowing exactly what microbe he was working with when he developed some of the techniques we still use today for isolating individual colonies. For example, it can be critical to know whether you have an infection of Staphylococcus aureus vs. Mycobacterium tuberculosis. Furthermore, it is important that we agree on the name. After all, you do not want your physician prescribing an inappropriate antibiotic simply because he/she was mistaken about the causative agent—“Oh, was it Mycobacterium? I’m sorry, I always get that confused with Mycoplasma.” Although not necessarily a matter of life and death, the same concerns apply to our research.

Consider a topic closer to home, the fungal ATG genes.1 At one time there were at least 10 different names being used to identify these genes. For example, APG1, AUT3, CVT10, GSA10, PAZ1 and PDD7 all refer to the Saccharomyces cerevisiae gene we now call ATG1. To be honest, this aura of confusion was actually quite handy for a long time as it kept competitors out of the field—they did not know what we were talking about. However, it got to the point where those of us working with these genes were having a hard time keeping them straight, so we had to agree on a unified nomenclature. The way this happened deserves a brief mention. The handful of labs working on fungal autophagy essentially got together at the first Gordon Research Conference on Autophagy in Stress, Development and Disease, and decided on a tentative name. Importantly, we checked various possibilities against the Saccharomyces Genome Database (the organization that is responsible for officially maintaining gene nomenclature in yeast) to be sure we were not choosing a name that was already being used to designate other genes. Unfortunately, checking with the official organization does not seem to happen very often in the case of human or mouse genes.

For human genes there are many examples that are confusing, such as p38. Are we referring to AHSA1, AIMP2 or MAPK14, all of which come up in a gene search through the HGNC database? The corresponding gene products have very different functions (an activator of heat shock 90kDa protein ATPase homolog 1, aminoacyl tRNA synthetase complex-interacting multifunctional protein 2, or mitogen-activated protein kinase 14). Furthermore, multiple genes are aliased as p38 in the NCBI Gene database (AHSA1, AIMP2, CRK, GRAP2, HRB87F, MAPK1, MPK2, NURF38, RNF19A, SYp and so on, and this does not even get into the variations such as p38-2, p38a, p38Beta, p38beta2, p38delta, p38gamma, etc.) and a further range of “p38” proteins can be found in UniProtKB (including POLDIP2 and RPP38). For p55 there are four possible genes noted with this as an alias by HGNC (ERG, FSCN1, PIK3R3 and PSMD12), and for p85 there are five possibilities (ARHGEF7, PIK3R1, PIK3R2, PPP1R12C and PPP1R13B). Even when we refer to p62 there is the potential for confusion as this designation is used to refer not only to SQSTM1, but also to KHDRBS1 and NUP62. Of course a gene that is named as “p” followed by a number is prone to potential confusion, and that is one reason that the HGNC recommends against using this type of designation (in addition to the fact that the molecular mass can vary between isoforms and species, making this type of name particularly meaningless). However, there are many examples that are not limited to the “p” nomenclature, such as the name CAP that corresponds to at least six different genes. We can see yet another example in autophagy: Bif-1 (Bax-interacting factor 1)/endophilin B1 associates...
with BECN1/Beclin 1/beclin1/beclin-1 via its interaction with UVRAG. A search in HGNC indicates that BIF1 is a synonym for ZBTB24 (zinc finger and BTB domain containing 24). However, Bif-1 is also referred to as SH3GLB1 (SH3-domain GRB2-like endophilin B1), which is encoded by a completely different gene than ZBTB24. Thus, Bif-1 is not the same as BIF1, but even if you work in the field of autophagy you might find this confusing. We suggest that this is not simply a matter of semantics; a misunderstanding of this nature can lead to a tremendous waste of time (and hence funds) spent working on the wrong gene/protein.

There are also reasons of self-interest that should prompt authors to use the official nomenclature. For example, some online journal articles are now being linked directly to databases.2 The ability to do this automatically will depend on the use of the correct gene/protein names, and should enhance the visibility of the corresponding papers, potentially resulting in increased citations. In general, accurate curation of data requires the use of standardized nomenclature among the various model organisms.3

Table 1. Examples of model system nomenclature

| Gene      | WT       | Mutant       | Protein   |
|-----------|----------|--------------|-----------|
| Human     | ULK1     | ULK1<sup>−/−</sup> | ULK1     |
| Mouse     | Ulk1     | ulk1         | Ulk1     |
| Rat       | Ulk1     | ulk1         | Ulk1     |
| Chicken   | Ulk1     | Ulk1<sup>−/−</sup> | ULK1     |
| Xenopus   | ulk1     | ulk1         | Ulk1     |
| Zebrafish | ulk1α    | ulk1α        | Ulk1α    |
| Caenorhabditis | unc-51<sup>atg-1</sup> | unc-51<sup>atg-1</sup> | UNC-51/ATG-1 |
| Arabidopsis | ATG1α | atg1α     | ATG1α     |
| Yeast     | ATG1     | atg1<sup>c</sup> | atg1<sup>c</sup> |

<sup>1</sup> atg-1 is an alias; however, in this case the “other name” may actually help avoid confusion by clearly identifying the gene as a homolog of ATG1. Most of the C. elegans <i>atg</i> genes that have non-<i>atg</i> designations have “other names” that incorporate the yeast <i>ATG</i> nomenclature. For example, <i>lgg-1</i>/<i>atg-8</i>-1 and <i>bcc-1</i>/<i>atg-6</i>. We recommend that authors use both names at least for the initial time the gene is mentioned in a paper.3 Mutants can also be expressed by specific allele designations, as in <i>unc-51(e369)</i> or more generally as <i>unc-51(lof)</i> to indicate loss of function. Allele designations typically take the form of the gene name followed by a dash and an allele number, as in <i>atg-17</i>. This is one reason that a protein-protein interaction such as Atg12–Atg5 should not be abbreviated as “Atg12–5”.

There are also reasons of self-interest that should prompt authors to use the official nomenclature. For example, some online journal articles are now being linked directly to databases.2 The ability to do this automatically will depend on the use of the correct gene/protein names, and should enhance the visibility of the corresponding papers, potentially resulting in increased citations. In general, accurate curation of data requires the use of standardized nomenclature among the various model organisms.3

Table 2. Nomenclature guidelines and gene search/database URLs

| Model System | URL                                                                 |
|--------------|----------------------------------------------------------------------|
| Human        | [http://www.genenames.org/guidelines.html](http://www.genenames.org/guidelines.html) |
| Mouse        | [http://www.informatics.jax.org/mgihome/nomen/gene.shtml](http://www.informatics.jax.org/mgihome/nomen/gene.shtml) |
| Rat          | [http://www.informatics.jax.org/mgihome/nomen/gene.shtml](http://www.informatics.jax.org/mgihome/nomen/gene.shtml) |
| Chicken      | [http://projects.roslin.ac.uk/chickmap/nomenclature.html](http://projects.roslin.ac.uk/chickmap/nomenclature.html) |
| Xenopus      | [http://www.xenbase.org/gene/static/geneNomenclature.jsp](http://www.xenbase.org/gene/static/geneNomenclature.jsp) |
| Zebrafish    | [https://wiki.zfin.org/display/general/ZFIN+Zebrafish+Nomenclature+Guidelines](https://wiki.zfin.org/display/general/ZFIN+Zebrafish+Nomenclature+Guidelines) |
| Arabidopsis  | [http://www.arabidopsis.org/portals/nomenclature/guidelines.jsp](http://www.arabidopsis.org/portals/nomenclature/guidelines.jsp) |
| Caenorhabditis | [http://www.wormbase.org/about/userguide/nomenclature](http://www.wormbase.org/about/userguide/nomenclature) |
| Yeast        | [http://www.yeastgenome.org/help/community/nomenclature-conventions](http://www.yeastgenome.org/help/community/nomenclature-conventions) |
or rat proteins should be in all uppercase letters, despite the fact that few authors follow this convention.

References
1. Klionsky DJ, Cregg JM, Dunn WA Jr, Emr SD, Sakai Y, Sandoval IV, et al. A unified nomenclature for yeast autophagy-related genes. Dev Cell 2003; 5:539-45; http://dx.doi.org/10.1016/S1534-5807(03)00296-X

Note that we are not trying to establish a new convention, but rather are asking that authors abide by the already established conventions for these model systems. Now, that is not asking too much, is it?

2. Rangarajan A, Schedl T, Yook K, Chan J, Haenel S, Otis L, et al. Toward an interactive article: integrating journals and biological databases. BMC Bioinformatics 2011; 12:175; PMID:21595960; http://dx.doi.org/10.1186/1471-2105-12-175

3. Hirschman J, Berardini TZ, Drabkin HJ, Howe DA. A MOD(ern) perspective on literature curation. Mol Genet Genomics 2010; 283:415-25; PMID:20221640; http://dx.doi.org/10.1007/s00438-010-0925-8