Gene Nomenclature

Genes should only be named (what geneticists formally call a gene symbol, for example *ODA11* or *RBCS2*) if one of the following is true: (1) A function or involvement in a specific biological process is associated with a publication. In this case, a PubMed ID (PMID) or other citation should accompany the gene symbol, which should be included in the Phytozome Description. (2) A gene is associated with a high-throughput screen or global study that may include proteomes of organelles or transcriptomes under defined conditions.

If the above criteria are not met, then a gene symbol should not be created. Genes without an assigned symbol should be referred to by their locus ID, since every locus has a unique and stable ID available in Phytozome.

The Cre name (Cre03.g187350, for example) should be used for: 1.) Genes encoding proteins with poor similarity to sequences in other organisms or for which the naming would be only based on a single conserved domain. 2.) Genes should not be named on the basis of similarity to proteins involved in a process that does not exist (or has not been shown) in *Chlamydomonas.* For example, the protein encoded by Cre02.g116900 displays high similarity to small hydrophilic plant seed proteins in *Arabidopsis*. In the absence of seed production, this protein clearly cannot perform this function in *Chlamydomonas*, and therefore should not be named after the *Arabidopsis* gene *ATEM1*.

Nomenclature rules for nuclear, chloroplast and mitochondrial genes

Gene nomenclature guidelines have been established by the *Chlamydomonas* community, but are not always strictly followed. Here are the basic rules that you should follow in your Sourcebook chapter.

1. Avoid the use of superscripts and subscripts.
2. Avoid Cr or C in the name of a gene if it stands for *Chlamydomonas* (*CNK1-CNK11*; the acronym was Chlamydomonas nek kinase or *CALK* for Chlamydomonas aurora-like kinase). *ACA2* is fine as C is for Calcium.
3. Acronyms should be 3 letters if possible, but less than 5 letters. Acronyms with two letters (*AC17*, for example) do not need to be changed.

	Nuclear	***Mitochondria***	***Chloroplast***
Gene and wild-type allele	All upper case, with number 3-5 letters for new genes and a number Italics *STA8*	Lower case 3 letters and a number Italics *nad6*	Lower case 3 letters and a capital letter Italics *petB*
Mutant allele	All lower case with number and allele number Italics *sta8-1*	All lower case with number and allele number Italics *nad6-1*	Lower case with number for allele Italics *petB-1*
Protein	All caps, Roman STA8	First letter cap, Roman Nad6	First letter cap Roman PetB
Phenotype	First letter cap, Roman with minus sign Sta8-	First letter cap, Roman with minus sign Nad6-	First letter cap, Roman with minus sign PetB-
Epitope tag 3'UTR Promoter Other	Name, one colon and tag Italics If the construct is not at the endogenous location, use TG for transgene *STA8:HA-TG* *Endogenous tag* If the tag is inserted by genome editing, no need for TG notation *STA8:HA*	Name, one colon and tag Italics *nad6:HA*	Name, one colon and name of tag (2-4 letters) Italics *petB:HA*

	STA8: RBSC3'UTR -TG *STA8:HSP70PRO-TG*		
Gene disruption by inserting a selectable marker	Name, two colons Italics *sta8::aphVIII* *sta8::ARG7* Since *aphVIII* is a bacterial gene, it is lower case. *ARG7* is a *Chlamydomonas* gene and it is all caps as above		*petB::aadA*
Transgene (unlinked)	*STA8-TG*	*NA*	*NA*
Genotypes	String with semicolons and spaces, italics *act2-1; fla4-1; pf23-1* *sta8::aphVIII; STA8-TG*	String with semicolons, italics *nad6-1; nad4-2*	String with semicolons and spaces, italics *trnI-1; petB::aadA*