

Tips for making your published data about Arabidopsis genes more FAIR (Findable, Accessible, Interoperable and Reusable)

To ensure that your Arabidopsis gene information is accessible here are some helpful pre-publication guidelines for authors. These guidelines here are written to be specific for Arabidopsis genes, but are based on a general set of guidelines and principles (see references). This makes it easier for TAIR curators to curate your published data (which will increase its visibility) and for other researchers to find and reuse your data.

Include AGI Locus Identifiers for genes

To ensure that the genes described in your paper are unambiguously identified, include the systematic locus identifier for that locus. If you have identified a new gene that does not yet have an AGI locus ID, please **contact TAIR curators PRIOR to publishing your gene**. TAIR, and other resources, use text mining to associate publications to biological entities in databases (e.g. genes and proteins). In the absence of a unique identifier such as an AGI locus code or UniProt ID, text mining software cannot distinguish between [CCR1](#) and [CCR1](#). To ensure that your published data can be curated and accurately linked to a database record, use the AGI locus ID.

Do not reuse gene symbols

To avoid the problem of different genes being referred to by the same symbolic name, before you publish check to see if the name is in use. Check the [gene symbol registry at TAIR](#), and search PubMed, ePubMed, Google Scholar to see if that symbolic name is in use for another Arabidopsis gene. If the name is not in use, please [register the gene symbol](#).

Follow gene, protein and allele nomenclature standards

There is an [established nomenclature for genes, proteins and alleles for Arabidopsis thaliana](#). For example, alleles are lowercase and distinguished by a dash and number (abc2-1, abc2-2). Again, check and make sure that the name is not already in use.

Naming and re-naming T-DNA insertion lines

There is a difference between a stock/germplasm and the specific T-DNA insertion that is causal for a phenotype. Therefore when referencing an allele please include the specific polymorphism and not just the name of the ABRC/NASC stock because many ABRC/NASC T-DNA stocks contain multiple insertions. Before naming your allele, check to see if it already has a name in TAIR or in the literature. At TAIR, we will update the allele name with the newly published name. If you have an allele/polymorphism/phenotype that does not already exist in TAIR, please [send us the information](#).

Use standard formats for reporting data (if available)

To ensure that data is reusable and interoperable (e.g. can be integrated across platforms/tools) community based initiatives have been established to define data standards. For example the Minimal Information about any Sequence [MIxS standards](#) or Minimal Information about a Plant Phenotyping Experiment ([MIAPPE](#)) standards. If you are not sure that a standard exists, check the [FAIR Sharing](#) website or [contact a curator](#).

Include metadata

Metadata is data about your data that is necessary to understand and reuse that data. Metadata may be as simple as a read me file explaining file contents. Metadata about experiments may include things like stock IDS for sample germplasms, descriptions of environmental /experimental conditions that use controlled vocabularies. Without metadata, your data won't be reusable.

Submit your data to an appropriate repository

To ensure that your data is findable and reusable, you should submit it to a permanent repository. Many journals will specify the preferred repository for specific data types such as sequence, expression data, proteomic data, etc... If no data type specific repository exists there are generic repositories such as [Dryad](#) and [Zenodo](#) that will house your data and give it a digital object identifier (DOI). Contact us if you are unsure where your data should go.

Publish data in machine readable formats

Machine readable means that the data can be 'read'/ingested by a computer program such as a script that will import a tab delimited file. Exporting an Excel table into a PDF format will render the data inaccessible to machines especially if the PDF is an image. Remember to use standard formats and include metadata.

References:

Wilkinson, M., et al., (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. DOI:[10.1038/sdata.2016.18](#)

Reiser, L., et al., (2018) FAIR: A Call to Make Published Data More Findable, Accessible, Interoperable, and Reusable. Molecular Plant.

DOI:[10.1016/j.molp.2018.07.005](#)