



## BRIEF NOTE

doi: 10.1111/age.13159

## Improved MeSH analysis software tools for farm animals

Sabrina T. Amorim\* , Koki Tsuyuzaki† ,  
Itoshi Nikaido†  and Gota Morota\* 

\*Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; †Laboratory for Bioinformatics Research, RIKEN Center for Biosystems Dynamics Research, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Accepted for publication 18 November 2021

## Background

The Medical Subject Headings, also known as MeSH, comprise a controlled life sciences vocabulary maintained by the National Library of Medicine for indexing journal articles. These headings are manually curated at the National Center for Biotechnology Information to be used in the MEDLINE database accessible from PubMed. Mapping MeSH IDs to Entrez Gene IDs can turn it into a powerful resource for enrichment analysis. We developed MeSH annotation Bioconductor packages for over 80 species coupled with a MeSH enrichment analysis package that initially became available in April 2014 as Bioconductor 2.14.<sup>1</sup> We reported the first detailed MeSH enrichment analysis of cattle, swine<sup>2</sup> and chicken<sup>3</sup> accompanied by reproducible R code for MeSH analysis. A unique aspect of MeSH is that both the quantity and quality of annotations are driven by new knowledge that is disseminated via peer-reviewed scientific articles.<sup>1</sup> It may therefore be viewed as a community-driven annotation project that is expected to improve over time with the publication of more scientific articles. In October 2021, the MeSH framework underwent a major change starting from Bioconductor 3.14 that resulted in the previous R code becoming dysfunctional. The objective of this study was to report improvements in MeSH annotations since its release in 2014 and to provide an update to the animal genetics community regarding the new usage of MeSH Bioconductor packages.

## Results

## Improvement in MeSH annotations

A suite of MeSH analysis Bioconductor packages has been closely maintained and updated at least once a year to keep up with the latest biomedical literature, including animal genetics. Table 1 compares the total number of genes annotated by MeSH terms between pre-Bioconductor 3.14 and Bioconductor 3.14 for cattle, swine and chicken. The number of MeSH annotated genes was obtained from the previous literature.<sup>2,3</sup> An improvement in MeSH annotations was observed for all three species, as

**Table 1** Total number of genes annotated with MeSH in cattle, swine and chicken

Species	Pre-Bioconductor 3.14	Bioconductor 3.14
Cattle	17 310 <sup>1</sup>	39 273
Swine	15 786 <sup>1</sup>	23 672
Chicken	9043 <sup>2</sup>	11 133

<sup>1</sup>Bioconductor 3.0.<sup>2</sup>Bioconductor 3.2.

the numbers of annotated genes increased by 127, 50 and 23% for cattle, swine and chicken respectively. This suggests that the findings reported in inferential studies (e.g. GWAS and RNA-seq analysis) were successfully incorporated in MeSH annotations. Because genes of farm animals are progressively being annotated in MeSH, downstream MeSH enrichment analysis is a powerful approach for aiding biological interpretations as well as generating biological hypotheses.

## New MeSH enrichment analysis framework

Previously, the responsibility of installing Bioconductor annotation packages that mapped MeSH IDs and Entrez Gene IDs for each species to perform enrichment analysis (e.g. MeSH.Bta.eg.db for cattle) lay with the end user. Bioconductor 3.14 saw a major change, where instead of downloading and installing these packages, all annotation data are now stored in and can be accessed from a cloud server via the AnnotationHub Bioconductor package.<sup>4</sup> One of the primary reasons behind this change is data reproducibility, a critical aspect of scientific research. Cloud storage allows developers and users to simultaneously maintain different versions of their data, thereby making provisions for old data to be archived. Additionally, the AnnotationHub style allows ease of practice during package installation by reducing the Bioconductor packages that need to be installed. The BioFileCache Bioconductor package additionally supports the feature wherein a dataset once downloaded from AnnotationHub is stored as a cache file and is not retrieved from the server again. Implementation of such modern features in recent Bioconductor saves time otherwise spent on downloading packages and data. The R code for MeSH enrichment analysis and creation of farm animal annotation data under the new AnnotationHub framework is available in Appendix S1 and in a reproducible manner on Docker Hub (Appendix S2).

## References

- 1 Tsuyuzaki K. *et al.* (2015) *BMC Bioinformatics* **16**, 1–17.
- 2 Morota G. *et al.* (2015) *Animal Genetics* **46**, 381–387.
- 3 Morota G. *et al.* (2016) *G3 Genes|Genomes|Genetics* **6**, 2447–2453.
- 4 Morgan M. & Shepherd L. (2021) *AnnotationHub: Client to access AnnotationHub resources*. R package version 3.0.2.

Correspondence: Gota Morota (morota@vt.edu)

### Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix S1.** R code to perform MeSH enrichment analysis and to create farm animal annotation data under the AnnotationHub framework.

**Appendix S2.** Instructions on running the MeSH docker image available on Docker Hub.

---