Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 286 - 289

ONLINE MENDELIAN INHERITANCE IN ANIMALS (OMIA) – LOOKING TO THE FUTURE

I. Tammen[1], M. Mather[2], D. Vanichkina[2], J. Nothman[2], Z. Li[2], S. Jufri[2] and F.W. Nicholas[1]

[1]The University of Sydney, Sydney School of Veterinary Science, Sydney, NSW 2006, Australia
[2]The University of Sydney, Sydney Informatics Hub, Darlington, NSW 2008, Australia

SUMMARY

Online Mendelian Inheritance in Animals (OMIA) is a freely available curated knowledgebase that contains information and facilitates research on inherited traits and diseases in animals. For the past 27 years, OMIA has been used by animal geneticists, breeders, and veterinarians worldwide as a definitive source of information. Recent increases in curation capacity and funding for software engineering support have resulted in software upgrades and commencement of several new initiatives, which include the review of variant information and links to human diseases caused by orthologous genes, and the introduction of phenotype and breed ontologies. We provide an overview of current information and recent enhancements to OMIA and discuss how we are expanding the integration of OMIA into other resources and databases via the use of ontologies.

INTRODUCTION

OMIA (https://omia.org) is a freely available, curated, online knowledgebase which provides users with up-to-date summary information on the known harmful and beneficial variants in animals, together with background information on known inherited disorders and beneficial traits. OMIA is modelled on and reciprocally hyperlinked to Online Mendelian Inheritance in Man (OMIM, https://www.omim.org), and provides further links to PubMed and Gene records at the National Center for Biotechnology Information and the European Bioinformatics Institute's Ensembl.).

OMIA focuses on traits and diseases ('phenes') with confirmed or suspected Mendelian modes of inheritance. However, several phenes with unknown or complex modes of inheritance and phenes caused by somatic mutations, genetic modifications or genome editing are also included. Furthermore, OMIA highlights 'landmark' papers (reporting major advances) and lists reviews and papers describing genetic maps and reference genomes. While most OMIA entries are for the major domesticated animal species, more than 370 (mainly vertebrate) animal species have entries in OMIA. Information about humans and model organisms such as mouse, rat, and zebrafish are not included, as they have dedicated species-specific resources.

Since the beginning of 2021, the curation team has increased from one (FWN) to two main curators (FWN and IT) and bequest funding has enabled software engineering support (MM). In this paper we provide an overview of OMIA data and a summary of recent software updates, major enhancements to likely causal variant tables and OMIA-OMIM hyperlinks, and the launch of Pioneers of Mendelian Inheritance in Animals (PMIA). We provide an update on current initiatives that focus on the use of ontologies to expand the interoperability of OMIA with other resources such as the Anstee Hub for Inherited Diseases in Animals (AHIDA, https://ahida.sydney.edu.au/app/home).

MATERIALS AND METHODS

OMIA software upgrade. Since August 2010, the OMIA database and website have been using Django software, a high-level Python Web framework. In July 2021, software packages were upgraded from Python 2 to Python 3 (https://www.python.org/) and from Django 1.9 to 3.2

(https://www.djangoproject.com/), and coding was reviewed and refined with the aim of future-proofing OMIA and improving homepage response times. Furthermore, search options were refined to increase the number of fields that are searched in a 'quick' search to improve user experience.

Likely causal variant tables. 'OMIA variant ID' and 'Source of Genetic Variant' fields were added to variant tables to provide unique numerical variant identification and to facilitate inclusion of variant information for genome-modified or edited variants. In collaboration with many colleagues (see OMIA's acknowledgement page for details: https://omia.org/acknowledgements/) variant information for cats, dogs, cattle, sheep, horse, pigs and goats has been reviewed and updated to Human Genome Variation Society (HGVS) nomenclature (https://varnomen.hgvs.org/). New site administration tools were introduced to facilitate automated 'liftover' of variant information to newer reference genome assemblies, and to enable export of variant information in variant call format (VCF) for submission to the European Variant Archive (EVA, https://www.ebi.ac.uk/eva/).

Review of OMIA-OMIM hyperlinks. Since 1997, OMIA has been reciprocally hyperlinked to OMIM. Links to OMIM are created by OMIA curators when new phenes are entered into OMIA. In the past, this focused on adding OMIM phenotype identifiers (IDs), while OMIM gene IDs were rarely added. OMIM automatically downloads OMIA phene IDs that have an OMIM ID link once a week and updates OMIM accordingly. In OMIA, separate fields for 'OMIM phene' and 'OMIM gene' hyperlinks were recently created, and we reviewed OMIA-OMIM hyperlinks for all OMIA phenes for which a likely causal variant has been identified in at least one species. OMIM links were confirmed, deleted, or added.

Pioneers of Mendelian inheritance in animals (PMIA). In 2022, PMIA was added to OMIA, accessible from the home page. This project comprises a series of commentaries on papers that illustrate the early discoveries of Mendelian inheritance in animals.

Integration of phenotype, disease and breed ontologies. OMIA previously included a home-grown list of 20 'phene categories' that could be used in 'Advanced Search' and in the 'Browse' page to create category-specific phene lists, but many OMIA phenes did not have a phene category specified. To allow for comprehensive categorisation, OMIA's 20 phene categories have been replaced with 28 major biological system headers from the Mammalian Phenotype (MP) Ontology (Smith and Eppig 2009) and two headers from the Mondo Disease Ontology (Mondo, https://mondo.monarchinitiative.org/). The MP ontology headers are included in Mondo, a global disease ontology that aims to harmonise disease definitions across the world. To facilitate this, the 'category' field has been included in phene-species pages and each phene has been linked to at least one category by a curator (IT). In addition, a new field enables inclusion of hyperlinks between OMIA disease entries and the corresponding homologous disease in Mondo.

Recognising the need to replace OMIA's home-grown breed list with a computable comprehensive list of standardised breed names, the OMIA team instigated the creation of the Vertebrate Breed Ontology (VBO, https://github.com/monarch-initiative/vertebrate-breed-ontology) in a project led by the Monarch Initiative (https://monarchinitiative.org/), with key personnel funded by the University of Colorado, in collaboration with colleagues from Iowa State University and with FAO colleagues responsible for the Domestic Animal Diversity Information System (DAD-IS). Curation tools relating to 'breed' in OMIA have been updated to allow inclusion of hyperlinks to VBO.

RESULTS AND DISCUSSION

OMIA is a globally used knowledgebase. Google Analytics user data for 2022 identified 41,803 users (94,579 sessions) from 163 countries. Until recently, curation was predominately conducted single-handedly by one curator, and limited funding restricted access to urgently needed software upgrades and modifications. Increased curation capacity and bequest funding support to upgrade and refine the underlying software is improving curator and user experiences and has resulted in the

commencement of several innovations while maintaining ongoing curation activities. In February 2023, OMIA included information on 2,327 phenes across 377 species, contained 4,336 phene-species entries and included a total of 29,453 references. Core statistics for key livestock species are summarised in Table 1.

Table 1. Summary of OMIA information relating to key livestock species (22/3/2023)

	Dog	**Cattle**	**Cat**	**Pig**	**Sheep**	**Horse**	**Chicken**	**Goat**	**All**
Total phenes	863	628	404	355	300	259	239	102	4336
Mendelian phenes	397	297	136	133	117	61	135	24	1778
Mendelian phenes with at least one likely causal variant known	336	204	103	65	54	48	56	17	1012
Likely causal variants known	491	269	171	72	86	105	71	30	1486

During 2022 the daily automated PubMed literature search resulted in 17,653 hits, of which 719 papers were identified to be added to OMIA. Additional references were added as part of other curation activities. We are currently trialling other literature search strategies to reduce the number of false-positive 'hits', including use of the machine learning tool LitSuggest (Allot *et al.* 2021) and an AI-based tool developed in house from Microsoft's PubMedBERT (Gu *et al.* 2021).

Likely causal variant tables. We reported the introduction of variant tables in OMIA in 2018 (Tammen and Nicholas 2018) and indicated that the ultimate aim would be to provide an EVA ID for all variants to reduce the need to standardise and update variant information in OMIA. However, EVA does not accommodate all types of variants, very few authors of OMIA-relevant papers submit variant information to EVA, and new EVA IDs are allocated infrequently. With the help of many colleagues (https://omia.org/acknowledgements/) we reviewed and standardised historic variant information in OMIA using HGVS nomenclature, with the aim to report location information based on a recent reference genome assembly where possible. In October 2021, variants listed in OMIA that were lacking an EVA ID but had standardised location information were submitted to EVA using a new OMIA pipeline for export of variant information in VCF for submission to EVA. The need for more standardised nomenclature for variants has been widely discussed to ensure greater transparency in relation to DNA testing. To this end, OMIA numerical variant IDs are now presented in the first column of all OMIA variant tables. An OMIA variant ID provides a unique unchanging ID for each likely causal variant, including those complex variants for which there is no HGVS nomenclature or no EVA ID. Review papers have started to include OMIA variant IDs in their tables.

Review of OMIA-OMIM hyperlinks. For phenes in OMIA that had at least one causal variant identified the review of OMIM links resulted in confirmation of 607 OMIM links, addition of 683 OMIM links and deletion of 46 OMIM links. Most of the added OMIM links were OMIM gene IDs (n=493), as these were in the past not routinely added to OMIA. OMIA currently lists 2424 models of human traits based on links to OMIM. 2119 OMIM entries have a link to OMIA. The revision of OMIA-OMIM hyperlinks will facilitate comparative-medicine-related research approaches. However, a large list of OMIA phenes without known likely causal variants have not yet been reviewed, as it is more speculative to identify homology between human and animal phenes if the underlying genetic cause is unknown.

Pioneers of Mendelian Inheritance in Animals. PMIA was first announced on Mendel Day (8 March) in 2022, and launched as part of OMIA on the 8th of July 2022, two weeks before the

bicentenary of Mendel's birth on the 22nd. Currently PMIA includes detailed commentaries by FWN on 15 papers that illustrate the early discoveries of Mendelian inheritance in animals.

Integration of phenotype, disease and breed ontologies. Three major current projects relate to the introduction of phenotype, disease and breed ontologies to OMIA. Ontologies are controlled vocabularies that represent knowledge both by their meaning and their relationship to each other and provide unique numerical identifiers to enable advanced computational analysis. We aim to harmonize breed and disease definitions in OMIA in a computer-accessible format, thus enabling integration with other global online resources and integration with the submission portal of AHIDA, (Tammen *et al*. 2021) which is currently under development.

So far, home-grown OMIA phene 'categories' have been replaced with 28 MP 'major biological system headers' and 2 Mondo categories. These categories are visible on phene-species pages in addition to visibility in advanced searches and in the OMIA browse page (https://omia.org/browse/). At least one category has been added to each OMIA phene, so that it is now possible, e.g., to search OMIA for all entries categorised as 'pigmentation phenotype' (MP:0001186). Further curation work is needed to add multiple categories as required.

Breed information in OMIA phene-species entries and in variant tables has been replaced with links to the VBO. VBO is based on FAO's Domestic Animal Diversity Information System (DAD-IS) breeds list and has been updated (especially for cat and dog breeds) with information from other international organizations, communities, and experts.

Finally, in a second collaboration with the Monarch Initiative, we are working towards integrating OMIA information into Mondo. So far, a new field has been created to enable addition of hyperlinks to Mondo, we have commenced adding Mondo links in OMIA and are currently discussing how to integrate OMIA information into Mondo.

CONCLUSION

Our vision for the future is that in addition to summarising information about inherited conditions in animals, OMIA becomes a global repository for standardised information on likely causal variants for diseases to allow transparent delivery of DNA diagnostics, and in linkage with the currently under-development Anstee Hub for Inherited Diseases in Animals, becomes a tool that enables semiautomated diagnosis for rare or emerging inherited conditions in animals.

A key remaining challenge is how best to harness automation and engage a wider contribution to curation efforts to ensure sustainability for the next 25 years. We always welcome feedback on current information presented in OMIA.

ACKNOWLEDGEMENTS

The Acknowledgements tab on the OMIA home page provides a detailed account of the contributions of the many people who have contributed to the development of OMIA since its inception. Recent software engineering support has been possible due to support of the Sydney Informatics Hub, a core research facility of the University of Sydney, and funding from the Ronald Bruce Anstee Bequest to the Sydney School of Veterinary Science.

REFERENCES

Allot A., Lee K., Chen Q., Luo L. and Lu Z. (2021) *Nucleic Acids Res*. **49:** W352.

Gu Y., Tinn R., Cheng H., Lucas M., Usuyama N., Liu X., Naumann T., Gao J. and Poon H. (2021) *ACM Trans. Comput. Healthcare* **3:** 1.

Smith C.L. and Eppig J.T. (2009) *Wiley Interdiscip. Rev. Syst. Biol. Med*. **1:** 390.

Tammen I. and Nicholas F.W. (2018) *Proc. of the 11th WCGALP*, Auckland, New Zealand.

Tammen I., Wade C., Waud B., Gimeno M., Velie B.D. O'Rourke B. and Nicholas F.W. (2021) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **24:** 383.