The Case for Standardising Gene Nomenclature across Vertebrates

Fiona M. McCarthy[1]*, Tamsin E.M. Jones[2]*, Anne E. Kwitek[3], Cynthia L. Smith[4], Peter D. Vize[5], Monte Westerfield[6], Elspeth A. Bruford[2,7]**

***These authors contributed equally to this work**

1 The Chicken Gene Nomenclature Committee (CGNC), School of Animal and Comparative Biomedical Sciences, University of Arizona, AZ, USA

2 HUGO Gene Nomenclature Committee (HGNC), European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK

3 Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI, USA

4 Mouse Genome Database, The Jackson Laboratory, Bar Harbor, ME, USA

5 Xenbase, Departments of Biological Sciences and Computer Science, University of Calgary, Calgary, Alberta, Canada

6 ZFIN, Institute of Neuroscience, University of Oregon, Eugene, OR, USA

7 Department of Haematology, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, UK

**** To whom correspondence should be addressed**

Key words: gene nomenclature; vertebrate genomics; oxytocin; arginine vasopressin

Abstract: Standardized gene nomenclature supports unambiguous communication and identification of the scientific literature associated with genes. To support the increasing number of annotated genomes that are now available for comparative studies, gene nomenclature authorities coordinate the assignment of approved gene names that can be readily propagated across species. Theofanopoulou et al. (Theofanopoulou et al. 2021) propose a new nomenclature for the genes encoding oxytocin and arginine vasopressin and their receptors. Rather than changing to a different nomenclature system, we propose minor updates to the current approved nomenclature of these vertebrate genes to better reflect their evolutionary history. We call on authors, journal editors and reviewers to help support communication and indexing of gene-related publications by working with existing gene nomenclature committees and ensuring that standardized gene nomenclature is routinely used.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2022 doi:10.20944/preprints202109.0485.v2

Standardized gene nomenclature. Standardized gene nomenclature provides a common language for the biomedical community, and beyond. Gene nomenclature refers to both the full gene name and the unique gene symbol; often aliases (or synonyms) used in published literature are also recorded to facilitate disambiguation, indexing and text mining. In vertebrates, gene nomenclature committees focus on species that represent key classes, including mammals (Bruford et al. 2020; Smith et al. 2020; Blake et al. 2021), birds (Howe et al. 2021; James-Zorn et al. 2015), fish (Howe et al. 2021), and amphibians (Howe et al. 2021), and coordinate their efforts to ensure that approved gene names are assigned consistently across species. This standardized nomenclature is widely disseminated through all the major genomic resources and model organism databases. Notably, this approach takes into account genetic and evolutionary similarities in addition to function, exactly as proposed by Theofanopoulou et al. (2021) and many genes are named based on their homologs in yeast, fly and other non-vertebrates. Gene nomenclature groups work closely with community experts (Olender et al. 2020), researchers, clinicians, bioinformaticians and biocurators to ensure that the approved gene names and symbols are informative, non-redundant and broadly applicable across diverse biological fields of study. One rationale cited for the newly proposed nomenclature system of Theofanopoulou et al. is to create a universal nomenclature system that can be consistently used across vertebrates. However, such a system is already established by the existing vertebrate nomenclature authorities (Table 1).

Table 1. Comparison of approved and proposed symbols for the oxytocin and arginine vasopressin ligand and receptor genes. Newly approved symbols are indicated with *.

Approved symbol from joint nomenclature committees	**Theofanopoulou et al. proposed symbol**
OXT	OT
AVP	VT
OXTR	OTR

 doi:10.20944/preprints202109.0485.v2

AVPR1A	VTR1A
AVPR1B	VTR1B
AVPR2 (aliased as AVPR2A*)	VTR2C
*AVPR2B**	VTR2B
AVPR2C / AVPR2L*	VTR2A

Revising gene nomenclature in the light of new data. Theofanopoulou et al. propose a new nomenclature system for the genes encoding oxytocin, arginine vasopressin and their receptors based on their evolutionary analysis of these genes in the context of newly sequenced, high quality genomes generated by the Vertebrate Genomes Project (VGP) (Rhie et al. 2021). While we share their desire to ensure gene nomenclature reflects evolutionary relationships, we believe the existing approved nomenclature, first established in vertebrates 30 years ago, is already largely representing these relationships (Table 1). Instead, only minor updates are needed in some species to better reflect the orthology and paralogy between these genes (Supplementary Information). We consider many factors when making nomenclature decisions: structure and function of genes and gene products, evolutionary history (including consideration of gene synteny), current and historical nomenclature usage, utility of nomenclature as search terms (including avoiding symbol clashes with other genes across the tree of life), levels of support for nomenclature updates in the research community, and concordance with nomenclature guidelines in multiple model systems (see Supplementary Information). Additionally, the current remit of the HUGO Gene Nomenclature Committee (HGNC) includes a commitment to move towards gene symbol stability in humans (Bruford et al. 2020), especially for genes that are clinically relevant, which includes the genes encoding oxytocin and arginine vasopressin and their receptors. Confusion about gene nomenclature in the medical literature could have serious negative consequences for patient safety (Braschi et al. 2021).

 doi:10.20944/preprints202109.0485.v2

The benefits and downsides of gene nomenclature changes. Major revisions to approved nomenclature are considered when the benefits clearly outweigh the downsides. Benefits can include the correction of incorrect or misleading gene nomenclature, better representation of evolutionary relationships, standardizing nomenclature throughout a gene family and providing nomenclature that can be used across all vertebrate species. Theofanopoulou et al. argue that the nomenclature of the oxytocin and arginine vasopressin genes and their receptors merits an update for all of these benefits. We believe that the existing approved nomenclature does not merit major revision as it is widely used, is not incorrect or misleading in the vast majority of vertebrate species, largely represents evolutionary relationships (with only minor additions needed to represent sub-clades in the AVPR2 subfamily), and has long been standardized across species (see Supplementary Information). The drawback is the introduction of additional identifiers in databases and the literature, increasing the risk of confusion to researchers and readers. Unfortunately the potential for confusion has already been exemplified in a recent publication by Ocampo Daza et al. (2021), who disagreed with Theofanopoulou et al.'s assignment of ABC suffixes in the AVPR2/VTR2 subfamily and therefore used the same symbols to refer to different genes.

Revising gene nomenclature on a large scale. Theofanopoulou et al. argue that their study acts as a model for gene nomenclature revision in the context of large scale vertebrate sequencing projects including the VGP. Their stated intent is to completely revise vertebrate gene nomenclature, including human gene nomenclature, to fully reflect evolutionary histories that are revealed by large scale sequencing projects. We are concerned that the authors may not fully appreciate the level of disruption that would be caused by major revisions to gene nomenclature on this scale. It is worth noting that the gene family analysed in their study is relatively simple with regard to its evolutionary history, and to perform such an analysis for every vertebrate gene family is an inconceivably large task. Given that over 40 years and millions of dollars of public funding have been invested into the current standardized nomenclature projects, we propose that an overhaul of the entire system would not be a prudent use of the limited resources we have in genomics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2022 doi:10.20944/preprints202109.0485.v2

The importance of applying standardized gene nomenclature in scientific journals. Requiring scientists to consistently use approved nomenclature avoids confusion and supports search indexing. While an increasing number of scientific journals mandate the use of standardized gene nomenclature, this requirement is not always clearly stated or strictly enforced for authors – citing the approved gene symbol and its associated gene ID should be compulsory in all journals. Nature's instructions to authors states that authors can "use their preferred terminology" for genes and proteins, which enables authors to publish novel nomenclature without first checking with the relevant nomenclature authority. If all journals, and especially influential ones such as Nature, would insist authors consult with nomenclature committees when suggesting updates much confusion could potentially be avoided. Unequivocally communicating about genes facilitates research and development in all biological and clinical fields.

We assert that the changes suggested by Theofanopoulou et al. to the official vertebrate gene nomenclature would cause considerable confusion with little perceivable benefit.

Our analysis of their study (Supplemental Information) demonstrates how the integration of genomic data from a broader range of species can help us to update and improve an already established nomenclature with only minor modifications. Theofanopoulou et al. call for collaboration between the gene nomenclature committees and genomic initiatives, which we whole-heartedly support. While our previous attempts to initiate collaborations with large multi-genome sequencing consortia have proved unsuccessful, we sincerely hope that future contact will prove more fruitful. We continue to encourage researchers and communities to collaborate with the gene nomenclature committees when proposing nomenclature updates.

Competing Interests Statement

The authors declare no competing interests.

Authors' Contributions

Study conceptualization: EAB

Analysis of gene nomenclature and evolutionary relationships: TEMJ & EAB

Manuscript writing and reviewing: Initial writing by FMM, TEMJ & EAB with input from all authors

Acknowledgements

T.E.M.J and E.A.B. are currently funded by the National Human Genome Research Institute (NHGRI) grant U24HG003345 and Wellcome Trust grant 208349/Z/17/Z. F.M.M. is supported by the US Department of Agriculture National Institute of Food and Agriculture projects ARZT-1370570-R50-131 and ARZT-1370520-R50-130. A.E.K. is supported by National Institutes of Health (NIH) grants from NHGRI (U24HG010859) and the National Heart Lung and Blood Institute (NHLBI; R01HL064541). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health nor any of the funders.

References

Blake, Judith A., Richard Baldarelli, James A. Kadin, Joel E. Richardson, Cynthia L. Smith, Carol J. Bult, and Mouse Genome Database Group. 2021. "Mouse Genome Database (MGD): Knowledgebase for Mouse-Human Comparative Biology." *Nucleic Acids Research* 49 (D1): D981–87.

Braschi, Bryony, Ruth L. Seal, Susan Tweedie, Tamsin E. M. Jones, and Elspeth A. Bruford. 2021. "The Risks of Using Unapproved Gene Symbols." *American Journal of Human Genetics* 108 (10): 1813–16.

Bruford, Elspeth A., Bryony Braschi, Paul Denny, Tamsin E. M. Jones, Ruth L. Seal, and Susan Tweedie. 2020. "Guidelines for Human Gene Nomenclature." *Nature Genetics* 52 (8): 754–58.

Howe, Douglas G., Sridhar Ramachandran, Yvonne M. Bradford, David Fashena, Sabrina Toro, Anne Eagle, Ken Frazer, et al. 2021. "The Zebrafish Information Network: Major Gene Page and Home Page Updates." *Nucleic Acids Research* 49 (D1): D1058–64.

James-Zorn, Christina, Virgillio G. Ponferrada, Kevin A. Burns, Joshua D. Fortriede, Vaneet S. Lotay, Yu Liu, J. Brad Karpinka, Kamran Karimi, Aaron M. Zorn, and Peter D. Vize. 2015. "Xenbase: Core Features, Data Acquisition, and Data Processing." *Genesis* 53 (8): 486–97.

Ocampo Daza, Daniel, Christina A. Bergqvist, and Dan Larhammar. 2021. "The Evolution of Oxytocin and Vasotocin Receptor Genes in Jawed Vertebrates: A Clear Case for Gene Duplications Through Ancestral Whole-Genome Duplications." *Frontiers in Endocrinology* 12: 792644.

Olender, Tsviya, Tamsin E. M. Jones, Elspeth Bruford, and Doron Lancet. 2020. "A Unified Nomenclature for Vertebrate Olfactory Receptors." *BMC Evolutionary Biology* 20 (1): 42.

Rhie, Arang, Shane A. McCarthy, Olivier Fedrigo, Joana Damas, Giulio Formenti, Sergey Koren, Marcela Uliano-Silva, et al. 2021. "Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species." *Nature* 592 (7856): 737–46.

Smith, Jennifer R., G. Thomas Hayman, Shur-Jen Wang, Stanley J. F. Laulederkind, Matthew J. Hoffman, Mary L. Kaldunski, Monika Tutaj, et al. 2020. "The Year of the Rat: The Rat Genome Database at 20: A Multi-Species Knowledgebase and Analysis Platform." *Nucleic Acids*

 doi:10.20944/preprints202109.0485.v2

Research 48 (D1): D731–42.
Theofanopoulou, Constantina, Gregory Gedman, James A. Cahill, Cedric Boeckx, and Erich D. Jarvis. 2021. “Universal Nomenclature for Oxytocin-Vasotocin Ligand and Receptor Families.” *Nature* 592 (7856): 747–55.