Xenopus Gene Naming Guidelines

1. Overview

- Gene/RNA names and symbols are lowercase italics (*pax6*)

- Proteins symbols are first letter capital, not italics (Pax6)

- Should not start with X, Xt, Xl for Xenopus species

- Official Xenopus gene names and symbols are found on Xenbase gene pages and are based on human gene nomenclature (http://www.genenames.org/)

- Orthology to human genes are assigned by synteny

- laevis homeologs (duplicate genes) are designated ".a" and ".b"

- Legacy gene name/symbols that are no longer available are recorded as synonyms (*Xbra* is a synonym of *t*)

- For gene name questions please email the Xenbase gene name coordinator [Joshua Fortriede](#)

> **Example:**
> Gene Name: *beta-carotene oxygenase 2*
> Gene Symbol: *bco2*
> RNA Symbol: *bco2*
> Protein Symbol: Bco2

2. Gene Names

2.1. Detailed Gene Nomenclature

- Xenopus gene names and symbols are identical to human gene names whenever possible. A full description of the nomenclature rules used by the [Human Genome Nomenclature Committee (HGNC)](#) can be found at http://www.genenames.org/.

- Orthology assignments are based primarily on synteny and requires more than a Blast result in order to apply the human gene name. Data for 12,000 tropicalis gene models has been generated by Dan Rokhsar see http://www.metazome.net

- Orthology assignments should be approved by the HGNC, the Xenopus Gene Nomenclature Committee and communicated by Xenbase staff.

- In cases where mammalian gene names reference an original Xenopus name (*chordin-like*), the Xenopus name will be retained

(*chordin*).

- Gene names should not start with any characters or words in order to identify the gene as being Xenopus (e.g. X, Xt, Xl, Xenopus, tropicalis, laevis).

- Gene names are lower case and italics, and should only contain Latin letters and Arabic numbers. Greek letters should be spelled out (β -> beta), and Roman numerals should be changed to Arabic equivalents (IV -> 4).

     Example: *beta-carotene oxygenase 2*

- Punctuation should only be used if the human gene name uses punctuation (except paralogs / homeologs as described below).

- When identity is uncertain be cautious. Use a temporary symbol or name such as " caudal type homeobox 2 [provisional]" until more information is available, at which time the name would be changed and the [provisional] tag removed.

- Pioneer species names should not be used. For example, in some species nanos3 is known as "nanos homolog 3 (Drosophila)". In Xenopus it would simply be named "nanos homolog 3".

- Xenbase administers Xenopus gene nomenclature.

- When there is no human ortholog of a new Xenopus gene or when the human gene name is provisional, new gene names will be based on consultations with the HGNC, the Xenopus gene nomenclature committee, and the requesting parties. Gene name requests should be sent to the Xenbase gene name coordinator Joshua Fortriede.

.2 Gene Families and Paralogs

- Gene families are a set of genes that were formed by duplication of a single ancestral gene. Genes within gene families usually have similar biological functions.

- When naming genes in gene families, a root word should be used to identify the gene as being a member of the gene family. Gene family members should be assigned increasing numerical identifiers, in keeping with HGNC policies.

     Example: *nodal homolog 1, nodal homolog 2*

- Some exceptions to this are necessary due to rare legacy names that have a different format. e.g. *nkx2-1, nkx2-2*

- **Paralogs**: When a gene is duplicated in Xenopus tropicalis relative to mammals the duplicated genes should be tagged with a ".1" or ".2"

    Example: *bmp7.1* and *bmp7.2*

- **Xenopus laevis homeologs** (duplicate laevis genes that arose from a genome polyploidization event) should be indicated with a ".a" or ".b" to identify unique genes. It is preferential to have all duplicated genes contained within chromosomal segments to use the same ".a" or ".b" designation.

    Example: *laevis wnt3* homeologs are *wnt3.a, wnt3.b,*
                    laevis wnt3a homeologs *are wnt3a.a, wnt3a.b*

- **Pseudogenes** are non-functional DNA sequences that are similar in structure to normal genes. Xenopus pseudogene names should be given the next interger within the gene family name, and designation "pseudogene" should be appended to the end of the gene name. HGNC pseudogene naming guidelines will be applied.

    Example: *fer-1-like 4 pseudogene*

- **Note:** Genes that are pseudogenes in one species may be expressed in other species.

3 Gene Symbols

- Gene symbols are lower case and italics, and should only contain Latin letters and Arabic numbers (unless specified below).
- Gene symbols are identical to human gene symbols whenever possible.
- Should not start with X, Xt, Xl for Xenopus species.
- Symbols are short-form representations (or abbreviations) of the descriptive gene name. Symbols should also be at least three characters long, with the first character being a letter.
- Gene symbols should have no spaces and punctuation should

only be used if the human equivalent uses punctuation (except for paralogs or laevis homeologs as described above).

- Gene symbols must be unique and should avoid matching common words or abbreviations in order to avoid problems with database searching (e.g. DNA, EGTA, PBS, CAN, GET...).

    Example: *bco2*

- Symbols for genes in gene families should contain a base or root "word", followed by increasing numerical identifiers.

    Example: *nodal, nodal1, nodal2, nodal3*

4 RNA Symbols
- RNA symbols are the same as gene symbols in **lowercase and italics** and match human symbol nomenclature.
- Latin Letters and Arabic Numbers only.

    Example: *bco2*

- **RNA splice variants:** RNA strands that arise from splice variants of genes should use the same gene symbol as the gene, followed by –v and increasing numerical identifiers.

    Example: *fzd4-v1*

5 Protein Symbols
- Protein names and symbols are exactly the same as the gene name and symbol but have the **first letter uppercase, and are not italics.**
- The word "protein" or additional terms are not included.

    Example: Bco2

- Protein variants arising from alternative spliced variants of genes should use the symbol as the alternative transcript, including the–v and increasing numerical identifiers.

    Example: Fzd4-v1

Please address all comments or questions to the nomenclature administrator at Xenbase [Joshua Fortriede.](#)

Xenopus Gene Nomenclature Committee (2013)

Enrique Amaya, University of Manchester, UK
Julie Baker, Stanford University, USA
Ira Blitz, University of California, Irvine, USA
Dale Frank, Technion – Israel Institute of Technology, Israel
Mike Gilchrist, MRC National Institute for Medical Research, UK
Matt Guille, EXRC, University of Portsmith, UK
Richard Harland, University of California, Berkeley, USA
Marko Horb, Marine Biological Laboratory, USA
Mustafa Khokha, Yale School of Medicine, USA
Hajime Ogino, Nara Institute of Science and Technology, Japan
Nicolas Pollet, Institute of Systems & Synthetic Biology, France
Masanori Taira, University of Tokyo, Japan
Gert Veenstra, Nijmegen Center for Molecular Life Sciences, Netherlands
Peter Vize, University of Calgary, Canada
Aaron Zorn (Chair), Cincinnati Children's Hospital, USA