

Proposed standard nomenclature for the α - and β -globin gene families

Gabriela Aguilera^{1,2*}, Joseph P. Bielawski^{3,4} and Ziheng Yang¹

¹Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, England

²Linnaeus Centre for Bioinformatics, University of Uppsala, BMC, Box 598, 75124, Uppsala, Sweden

³Department of Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada

⁴Department of Mathematics & Statistics, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada

(Received 26 December 2005, accepted 12 September 2006)

The globin family of genes and proteins has been a recurrent object of study for many decades. This interest has generated a vast amount of knowledge. However it has also created an inconsistent and confusing nomenclature, due to the lack of a systematic approach to naming genes and failure to reflect the phylogenetic relationships among genes of the gene family. To alleviate the problems with the existing system, here we propose a standardized nomenclature for the α and β globin family of genes, based on a phylogenetic analysis of vertebrate α and β globins, and following the Guidelines for Human Gene Nomenclature.

Key words: α - β -globin gene family — Standardized Nomenclature —

In the α and β globin families, there have been extensive gene duplications and functional divergences (Hardison 1998), shown in Figs 1 and 2, respectively. The structure of these gene families reflects both early duplication events and recent tandem duplications within organismal lineages (i.e., *en bloc* duplications) (Efstratiadis et al. 1980, Czelusniak et al. 1982, Proudfoot et al. 1982, Goodman et al. 1987, Flint et al. 1988).

As has been the custom with many gene families, the different globin genes were given names following the order of letters in the Greek alphabet. However, the nomenclature system adopted for the α and β globin families does not reflect their complex evolutionary history through gene duplication and often makes it difficult to distinguish between paralogs and orthologs.

Many problems affect the current nomenclature system of the α and β globin gene families. In β globins, some of the names given to these genes are assigned with respect to functional analogy (developmental stage during which they are expressed), instead of relating to evolutionary origin. For example, the so-called γ globins in artiodactyls are expressed during the fetal stage of development even though they originate from the adult-expressed β globin gene, and therefore should be referred to as β globins (Schimenti and Duncan 1985). In both

the α and β globins the current nomenclature does not account for gene linkage order on the chromosome, and the tandem organization of the cluster, which arose due to *en bloc* duplications, is not reflected by superscripts assigned to different genes (Tables 1 and 2, respectively). A particular point of confusion stems from some orthologous β globin genes having a different name in different species. For example, β globin is called β in primates and γ in cows (Hardies et al. 1984, Schimenti and Duncan 1985). A similar situation occurs in α globins, where the mammalian ζ globin and the avian π globin are orthologous but have different names (Cooper et al. 2006). For both α and β globins there is an inconsistent use of subscripts and superscripts (Tables 1 and 2, respectively). For example, in the case of β globins, linkage order in rodents is given by cardinal numbers whereas in artiodactyls both roman numbers and letters are used (Hill et al. 1984). β globin pseudogenes are also afflicted by an inconsistent system, as such genes are sometimes preceded by the letter ψ , and in other cases referred to by their assigned Greek letter, as in the case of δ or η (Goodman et al. 1984, Hardies et al. 1984, Hardison and Margot 1984, Hayasaka et al. 1992). This inconsistency makes it difficult to distinguish functional genes from nonfunctional copies. Nomenclature problems are most severe in rodent and artiodactyl β globins, as both groups have undergone *en bloc* duplications, resulting in tandem copies of paralogous genes (Hill et al. 1984, Schimenti and

Edited by Norihiro Okada

* Corresponding author. E-mail: gabriela.aguilera@lcb.uu.se

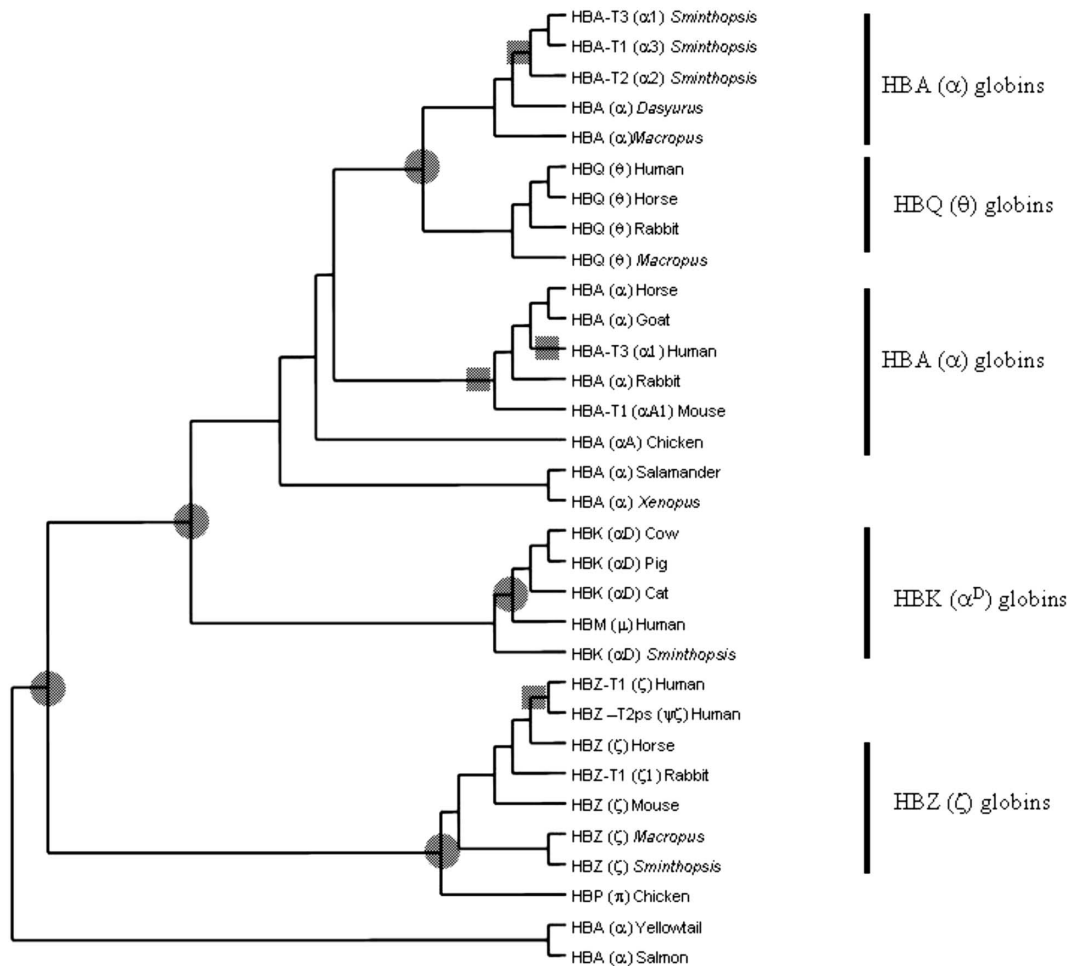


Fig. 1. Proposed phylogeny showing the evolutionary history of the vertebrate α globin gene family through gene divergence and *en bloc* duplication. The tree represents the phylogeny recently proposed for vertebrate α globin genes by Cooper et al. (2006). We used the α globin genes from Yellowtail and Salmon as outgroup in order to root the tree. Genes are assigned both the traditional name, in parenthesis, and the new name under our proposed nomenclature system. Circles represent gene duplication events and squares show *en bloc* duplication events.

Duncan 1985). The current nomenclature system for the β globin family is thus acutely confusing.

In an attempt to alleviate the problem, we propose a revised symbolic system for the genes of the vertebrate α and β globin families, which follows the Guidelines for Human Gene Nomenclature (Wain et al. 2002). In our proposed revision, HB stands for hemoglobin; HBB for the β globin gene in hemoglobin and HBA for the α globin gene in hemoglobin. Furthermore, E is for ϵ , H is for η , K is for κ , D for δ , G for γ , W for ω , R for ρ , P for π , M for μ , Q for θ , and Z for ζ . The lowercase letters "ps" at the end of the symbol designate a pseudogene. The symbol -T followed by a number indicates that the gene is part of a known tandemly duplicated gene block and the number corresponds to the linkage order within the block according to the 5' to 3' orientation. Previously approved symbols already exist for human and mouse genes, which are consistent with our proposed system for globins (see <http://www.gene.ucl.ac.uk/nomenclature> for the Human Nomenclature

Database, and <http://www.informatics.jax.org/> for the Mouse Genome Informatics). Three substantial modifications should be noted. First, artiodactyl genes previously referred to as γ in cow, sheep, and goat now have symbols that identify them as part of the β globin clade, and are thus changed to HBB (Table 1). We make this change because we believe the names should accurately reflect their evolutionary origins (Fig. 2) rather than a function analogous to other globins which are members of the γ clade. Second, we assign mouse genes β h0 and β h1 the symbols HBE-T2 and HBG-T1, respectively, identifying them as part of the ϵ and γ clades to which they belong, and indicating their linkage order in their respective tandem arrays on the chromosome (Table 2). Third, we assign the name HBK (K for κ), to the globin lineage α^D , which has been shown to be an ortholog in birds, reptiles and mammals (Cooper et al. 2006) (Table 1). We retain the traditional names, α (HBA) and β (HBB) globin, for the entire gene families. We list the

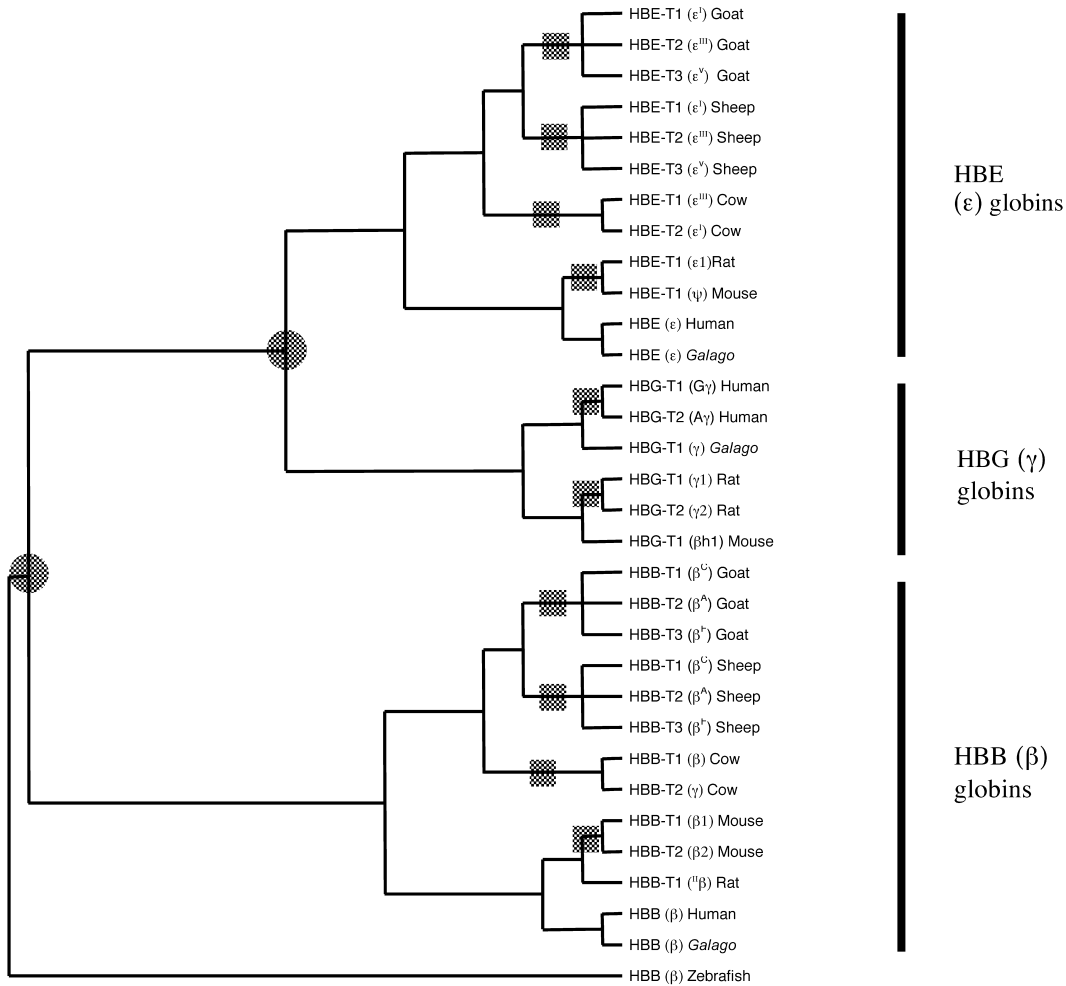


Fig. 2. Proposed phylogeny showing the complex evolutionary history of the eutherian β globin gene family through gene divergence and *en bloc* duplication. The tree represents the accepted phylogeny for eutherian β globin genes (Hardison 1998). In the cases where gene relationships were not resolved we collapsed the branches (*e.g.*, globin genes originated by *en bloc* duplication events share a single node). We used the β globin gene from Zebrafish as outgroup in order to root the tree. Genes are assigned both the traditional name, in parenthesis, and the new name under our proposed nomenclature system. Circles represent gene duplication events and squares show *en bloc* duplication events.

Table 1. Revised Nomenclature for the α globin gene family

Species	Existing symbols	Proposed symbols
Fish	α , β	HBA, HBB
Amphibians	α , β	HBA, HBB
Aves	α^D , α^A , π	HBK, HBA, HBP
<i>Sminthopsis</i>	ζ , α^D , $\psi\alpha3$, $\alpha2$, $\psi\alpha1$, θ , ω	HBZ, HBK, HBA-T1ps, HBA-T2, HBA-T3ps, HBQ, HBW
<i>Didelphis</i>	ζ , ζ , α^D , $\alpha3$, $\alpha2$, $\alpha1$, θ , ω	HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ, HBW
<i>Macropus</i>	ζ , ζ , α^D , $\alpha1$, θ , ω	HBZ-T1, HBZ-T2, HBK, HBA-T1, HBQ, HBW
Rabbit	$\zeta1$, α , θ	HBZ-T1, HBA, HBQ
Mouse	ζ , $\alpha A1$, θ	HBZ, HBA-T1, HBQ
Rat	ζ , $\alpha A1$, θ	HBZ, HBA-T1, HBQ
Artiodactyls	ζ , α^D , α , θ	HBZ, HBK, HBA-T1, HBQ
Cat	ζ , α^D , θ	HBZ, HBK, HBQ
Human	ζ , $\psi\zeta$, α^D , μ , $\psi\alpha3$, $\alpha2$, $\alpha1$, $\theta1$	HBZ-T1, HBZ-T2ps, HBK, HBM, HBA-T1ps, HBA-T2, HBA-T3, HBQ-T1

Table 2. Revised Nomenclature for the β globin gene family

Species	Existing symbols	Proposed symbols
Zebrafish	α , β	HBA, HBB
<i>Xenopus</i>	α , β	HBA, HBB
Chicken	ρ , β , ϵ	HBR, HBB, HBE
Marsupial	ϵ^M , β^M , ω	HBE, HBB, HBW
Rabbit	ϵ , γ , $\psi\delta$, β	HBE, HBG, HBD _{ps} , HBB
Mouse	ψ , $\beta h0$, $\beta h1$, $\beta h2$, $\beta h3$, $\beta 1$, $\beta 2$	HBE-T1, HBE-T2, HBG-T1, HBD-T1, HBD-T2, HBB-T1, HBB-T2
Rat	$\epsilon 1$, $\gamma 1$, $\gamma 2$, $\psi\gamma 3$, $\psi\beta\rho$, $^{\text{II}}\beta$	HBE-T1, HBG-T1, HBG-T2, HBG-T3 _{ps} , HBB _{ps} , HBB-T1
Goat	ϵ^I , ϵ^{II} , $\psi\beta^X$, β^C , ϵ^{III} , ϵ^{IV} , $\psi\beta^Z$, β^A , ϵ^V , ϵ^{VI} , $\psi\beta^Y$, β^F , γ	HBE-T1, HBH-T1, HBD-T1 _{ps} , HBB-T1, HBE-T2, HBH-T2, HBD-T2 _{ps} , HBB-T2, HBE-T3, HBH-T3, HBD-T3 _{ps} , HBB-T3, HBB-T4
Sheep	ϵ^I , ϵ^{II} , $\psi\beta^I$, β^C , ϵ^{III} , ϵ^{IV} , $\psi\beta^{\text{II}}$, β^A , ϵ^V , ϵ^{VI} , $\psi\beta^{\text{III}}$, β^F , γ	HBE-T1, HBH-T1, HBD-T1 _{ps} , HBB-T1, HBE-T2, HBH-T2, HBD-T2 _{ps} , HBB-T2, HBE-T3, HBH-T3, HBD-T3 _{ps} , HBB-T3, HBB-T4
Cow	ϵ^{III} , ϵ^{IV} , ψ^{III} , β , ϵ^I , ϵ^{II} , ψ^I , ψ^{II} , γ	HBE-T1, HBH-T1, HBD-T1 _{ps} , HBB-T1, HBE-T2, HBH-T2, HBD-T2 _{ps} , HBD-T3 _{ps} , HBB-T2
<i>Galago</i>	ϵ , γ , $\psi\eta$, δ , β	HBE, HBG, HBH _{ps} , HBD, HBB
Human	ϵ , G_γ , A_γ , $\psi\eta$, δ , β	HBE, HBG-T1, HBG-T2, HBH _{ps} , HBD, HBB

former and newly proposed symbols for the individual genes in Tables 1 and 2. We believe the standardized nomenclature proposed here for the α and β globin genes is more transparent with regard to the complex orthologous and paralogous relationships among α and β globins. Moreover, we expect this system will lead to easier and more precise communication of results of α and β globin research. An important advantage of the proposed nomenclature system for the α and β globins is that it conveys information about the linkage order in the chromosome. This information is relevant when new globin genes are identified, as the location in which they are found, relative to the other linked globins, can clarify the origin of the genes and their homologous relationships. For instance, the linkage of the α and β globins in a single chromosome in fish and amphibians provides evidence that the two gene families evolved by a tandem duplication event (Jeffreys et al. 1980). In the linked cluster of α and β globins, α globins typically reside on the 5' end and β globins are located on the 3' end of the cluster. In birds and mammals, the two gene families are unlinked in separate clusters including differentially expressed genes (Hardison 1998).

A mechanism for the generation of the present day α and β globin clusters in birds and mammals has been proposed that involves a possible translocation or chromosome duplication of the locus containing both gene clusters, followed by the silencing of the linked α or β globin genes (Jeffreys et al. 1980, Hardison 2001). Support for this hypothesis would come from the identification of possible "fossil" α globins on the 5' end of the β

globin cluster, or the discovery of "fossil" β globins on the 3' end of the α globin cluster (Wheeler et al. 2004). The recent discovery of the ω globin gene in marsupials (Wheeler et al. 2001), which is a β -like globin gene present in the α globin cluster, would arguably be an example of such a "fossil" gene (Wheeler et al. 2004). As more hemoglobin genes are identified from different vertebrate species, it is likely that more linked members of the α and β globin clusters will appear. In order to have a clearer picture of the origin and homologous relationships of these genes, a nomenclature system should be adopted that explicitly makes reference to their linkage order in the globin cluster. It is possible that updates on the order and numbering of genes within the cluster might become necessary if new genes are identified that disrupt the previously known linkage order.

We note that the proposed nomenclature system for the α and β globins may help to make communication easier, particularly in studies of these genes at the gene family level, where homologous relationships can be obscured by an inconsistent nomenclature. However, other nomenclatures may also apply if globins are studied at, for example, the level of protein structure. In this case, it may be more convenient to divide globin proteins into structural classes. There are two structural classes that encompass the plethora of globins discovered to date, namely, the 3/3, and the 2/2 helical sandwich structures (Vinogradov et al. 2005). Furthermore, there are three major globin lineages: (i) the 3/3 plant and metazoan globins, single-domain globins, and flavohemoglobins; (ii) the 3/3 globin-coupled sensors and protoglobins found in

bacteria, and (iii) the bacterial, plant and ciliate 2/2 globins (Vinogradov et al. 2006). Our proposed nomenclature is not inconsistent with such a structure-based system; indeed, we predict that such systems would be complementary, as they portray information relevant to different levels of evolutionary divergence (i.e., deep versus shallow) among members of the globin family.

With the advent of mass sequencing efforts it has become essential that gene names take into account the complexity of gene family architecture and evolutionary relatedness across species, as well as gene expression and gene interactions. Having a common basis for assigning names and symbols to genes in gene families facilitates data mining through automated algorithms, thus aiding a comprehensive depiction of gene relationships within a larger cellular and genomic context. Also, a comprehensive and explicit nomenclature system is essential for reliable gene identification and annotation. We have entered the era of automated analysis of large-system data, e.g., genomics, metagenomics, proteomics, etc.; hence, revision of those nomenclature systems known to obscure evolutionary and functional relationships among genes is warranted, if not already overdue.

G. A. was supported by a grant from the Mexican Council of Science and Technology (CONACYT); J. P. B. and Z. Y. were supported by grant 31/G14969 from the Biotechnology and Biological Sciences Research Council (BBSRC, UK). JPB was partially supported by a start-up grant from the Genome Atlantic centre of Genome Canada, and by an NSERC Discovery Grant (DG298394).

REFERENCES

- Cooper, S. J. B., Wheeler, D., De Leo, A., Cheng, J-F., Holland, R. A. B., Marshall Graves, J. A., Hope, R. M. (2006) The mammalian α^D -globin gene lineage and a new model for the molecular evolution of α -globin gene clusters at the stem of the mammalian radiation. *Mol. Phylogenet. Evol.* **38**, 439–448.
- Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M. L., Venta, P. J., Tashian, R. E. (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* **29**, 297–300.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., Proudfoot, N. J. (1980) The structure and evolution of the human beta-globin gene family. *Cell* **21**, 653–668.
- Flint, J., Taylor, A. M., Clegg, J. B. (1988) Structure and evolution of the horse zeta globin locus. *J. Mol. Biol.* **199**, 427–437.
- Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L. (1984) The eta-globin gene. Its long evolutionary history in the beta-globin gene family of mammals. *J. Mol. Biol.* **180**, 803–823.
- Goodman, M., Czelusniak, J., Koop, B., Tagle, D., Slightom, J. (1987) Globins: A case study in molecular phylogeny. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 875–890.
- Hardies, S. C., Edgell, M. H., Hutchison, C. A. III (1984) Evolution of the mammalian beta-globin gene cluster. *J. Biol. Chem.* **259**, 3748–3756.
- Hardison, R. C. and Margot, J. B. (1984) Rabbit globin pseudogene psi beta 2 is a hybrid of delta- and beta-globin gene sequences. *Mol. Biol. Evol.* **1**, 302–316.
- Hardison, R. C. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.* **201** (Pt 8), 1099–1117.
- Hardison, R. C. (2001). Organisation, evolution and regulation of the globin genes. In: Disorders of hemoglobin (eds.: M. H. Steinberg, B. G. Forget, D. R. Higgs, and R. L. Nagel), pp 95–116. Cambridge University Press, Cambridge.
- Hayasaka, K., Fitch, D. H., Slightom, J. L., Goodman, M. (1992) Fetal recruitment of anthropoid gamma-globin genes. Findings from phylogenetic analyses involving the 5'-flanking sequences of the psi gamma 1 globin gene of spider monkey *Ateles geoffroyi*. *J. Mol. Biol.* **224**, 875–881.
- Hill, A., Hardies, S. C., Phillips, S. J., Davis, M. G., Hutchison, C. A. III, Edgell, M. H. (1984) Two mouse early embryonic beta-globin gene sequences. Evolution of the nonadult beta-globins. *J. Biol. Chem.* **259**, 3739–3747.
- Jeffreys, A. J., Wilson, V., Wood, D., Simons, J. P., Kay, R. M., Williams, J. G. (1980) Linkage of adult α - and β -globin genes in *X. laevis* and gene duplication and tetraploidization. *Cell* **21**, 555–564.
- Proudfoot, N. J., Gil, A., Maniatis, T. (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* **31**, 553–563.
- Schimmenti, J. C. and Duncan, C. H. (1985) Structure and organization of the bovine beta-globin genes. *Mol. Biol. Evol.* **2**, 514–525.
- Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W., Povey, S. (2002) Guidelines for human gene nomenclature. *Genomics* **79**, 464–470.
- Wheeler, D., Hope, R., Cooper, S. B., Dolman, G., Webb, G. C., Bottema, C. D., Gooley, A. A., Goodman, M., Holland, R. A., (2001) An orphaned mammalian beta-globin gene of ancient evolutionary origin. *Proc. Natl. Acad. Sci. USA* **98**, 1101–1106.
- Wheeler, D., Hope, R., Cooper, S. B., Gooley, A. A., Holland, R. A., (2004) Linkage of the β -like ω -globin genes in an Australian marsupial supports the chromosome duplication model for the separation of globin gene clusters. *J. Mol. Evol.* **58**, 642–652.
- Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Guertin, M., Gough, J., Dewilde, S., Moens, L., Vanfleteren, J. R. (2005) Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc. Natl. Acad. Sci. USA* **102**, 11385–11389.
- Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L., Vanfleteren, J. R. (2006) A phylogenomic profile of globins. *BMC Evol. Biol.* **6**, 31.