Complex evolution of *Fbxl21* gene in mammals

G.A. Shilovsky

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoy Karetny per. 19, building 1, Moscow, 127051, Russia;
Lomonosov Moscow State University, A.N. Belozersky Institute of Physico-Chemical Biology, Vorobjovy Gory, Build. 'A', Moscow, 119899, Russia;
Lomonosov Moscow State University, Faculty of Biology, Leninskie Gory, 1, building 12, Moscow, 119234, Russia

O.A. Zverkov, L.I. Rubanov, A.V. Seliverstov, V.A. Lyubetsky

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoy Karetny per. 19, building 1, Moscow, 127051, Russia, `lyubetsk@iitp.ru`

Introduction. The *Fbxl21* gene is largely expressed in the house mouse hypothalamus [1, 2] and its product, the F-box and leucine-rich repeat protein 21, is involved in circadian regulation [3, 4]. We have found species with its pseudogenes including *Gorilla gorilla* and *Homo sapiens* [5, 6]. In humans, this pseudogene is transcribed and is associated with certain diseases [7].

Although other genes involved in circadian regulation, e.g., *Clock*, *Cry1*, *Cry2*, and *Fbxl3*, are known in many animal species and are highly conserved, the *Fbxl21* gene is missing in the platypus *Ornithorhynchus anatinus* and marsupials (*Monodelphis domestica*, *Notamacropus eugenii*, *Phascolarctos cinereus*, *Sarcophilus harrisii*, *Vombatus ursinus*). The placental genomes lacking this gene include the nine-banded armadillo *Dasypus novemcinctus* (Xenantra), African bush elephant *Loxodonta africana* (Afrotheria), giant panda *Ailuropoda melanoleuca* and California sea lion *Zalophus californianus* (Carnivora), European rabbit *Oryctolagus cuniculus* (Lagomorpha), northern greater galago *Otolemur garnettii* (Strepsirrhini), sooty mangabey *Cercocebus atys*, and green monkey *Chlorocebus sabaeus* (Haplorrhini). We have found sequence changes in the gene in certain species including chimpanzees *Pan* spp. larger than could be expected from the mutation rates in related species. In continuation of our previous work [5, 6], here we consider evolutionary changes in the *Fbxl21* gene.

The highest mammalian encephalization quotient calculated from the phylogenetic position of species is observed in primates, cetaceans, and proboscideans [8]. Moreover, the relative weight gain of the neonatal brain peaks in the human and elephants [9], which, among other

things, supports neoteny in humans [10, 11]. The relative cerebral weight gain is much lower in chimpanzees than in humans [12, 13], while the neonatal brain averages ~90% of the adult brain weight in most mammals. The genes linked to human primary microcephaly are subject to positive selection in great apes [14] and cetaceans with high encephalization quotients [15].

Recently, our hypothesis that the loss of the *c-Answer* gene in warm-blooded animals considerably enlarged the forebrain relative to cold-blooded and simultaneously decreased the regenerative capacity [16] has been experimentally confirmed. Thus, elimination of one or a few genes can induce substantial phenotypic changes with no apparent association.

Results. A large-scale genomic alignment was generated. The human *FBXL21P* pseudogene neighbors the conserved *IL9* and *LECT2* genes, the latter is on the complementary strand and overlaps *FBXL21P*. In many other placentals, the orthologs of *IL9* and *LECT2* are colocalized. In particular, these genes are next to each other in the elephant *Loxodonta africana*, panda *Ailuropoda melanoleuca*, seal *Zalophus californianus*, rabbit *Oryctolagus cuniculus*, galago *Otolemur garnettii*, mangabey *Cercocebus atys*, green monkey *Chlorocebus sabaeus*, and western gorilla *Gorilla gorilla*.

In the genomic alignment, the *FBXL21P* pseudogene corresponds to the armadillo gene ENSDNOG00000052722 encoding two long intergenic noncoding (linc) transcripts, while *LECT2* was conserved. *FBXL21P* corresponds to the rabbit gene ENSOCUG00000030764 encoding two linc transcripts. In the mountain hare *Lepus timidus* [17], *FBXL21P* corresponds to the locus with stop codons; apparently, a pseudogene. In the gorilla, *FBXL21P* corresponds to the pseudogene ENSGGOG00000014124. On the other hand, *FBXL21* orthologs neighboring the *LECT2* gene were found in birds, e.g., the Eurasian blue tit *Cyanistes caeruleus*.

The evolutionary trend of the *Fbxl21* gene was evaluated by the analysis of *dN/dS* values describing the rate of amino acid variation within the great apes as compared to the mouse outgroup. The same trend is revealed by a more accurate rate analysis of pairwise *dN/dS* relations between the great apes divided into *dN/dS* for the mouse and the ape.

The dN/dS values indicate a significant increase in the positive selection pressure on the *FBXL21* gene in the greater apes compared to the mouse ortholog. At the same time, this

gene ceased to notably change in the *Pan* genus. The Table presents a fraction of the data obtained; the scientific names of species are given in the rest of the abstract.

Species	*dN/dS*	Target % id	Query % id
chimpanzee – bonobo	0.15873	99.77 %	99.77 %
chimpanzee – orangutan	0.56032	92.84 %	81.02 %
chimpanzee – gibbon	0.73840	93.09 %	93.52 %
bonobo – orangutan	0.46016	93.10 %	81.25 %
bonobo – gibbon	0.59928	93.32 %	93.75 %
orangutan – gibbon	0.45084	83.41 %	96.02 %
mouse – chimpanzee	0.09696	84.72 %	79.57 %
mouse – bonobo	0.09351	84.95 %	79.78 %
mouse – orangutan	0.05316	89.66 %	73.48 %
mouse – gibbon	0.08310	87.10 %	82.17 %

FBXL21 was also considered in Carnivora, 12 species and/or subspecies. These include Felidae: *Panthera leo*, *Panthera pardus*, *Panthera tigris*, *Felis catus*, and *Lynx canadensis*; Herpestidae: *Suricata suricatta*; Canidae: *Canis lupus familiaris basenji*, *Canis lupus dingo*, and *Vulpes vulpes*; Mustelidae: *Neovison vison*; and Ursidae: *Ursus maritimus* and *Ursus thibetanus*. The *dN* variations are minor within these families and between them, and the rate of nonsynonymous substitutions generally corresponds to the taxonomic classification. The greatest variation is observed between *Suricata suricatta* and Caniformia (canids, mustelids, and bears).

Similarly, *FBXL21* was considered in primates, 19 species. These include Strepsirrhini, 3 species: *Microcebus murinus*, *Propithecus coquereli*, and *Prolemur simus*; and Haplorrhini, 16 species: *Carlito syrichta*, *Callithrix jacchus*, *Saimiri boliviensis*, *Cebus capucinus*, *Aotus nancymaae*; Old World monkeys *Mandrillus leucophaeus*, *Theropithecus gelada*, *Macaca fascicularis*, *Macaca mulatta*, *Macaca nemestrina*, *Piliocolobus tephrosceles*, *Rhinopithecus bieti*, and *Rhinopithecus roxellana*; and apes *Nomascus leucogenys*, *Pan paniscus*, and *Pan troglodytes*. The sequences within the Strepsirrhini suborder are much closer to each other compared to those in Haplorrhini. However, the variation within Strepsirrhini is higher than that within Haplorrhini. The high variation rate in Strepsirrhini agrees with the *FBXL21* loss in

the galago *Otolemur garnettii*. The *dS* and *dN* values largely correspond to the taxonomy. However, significant alterations occurred in the common ancestor of the chimpanzee and bonobo (*Pan* spp.) including indels in the conserved region of the protein. The gibbon sequence is much closer to those of Cercopithecidae than of chimpanzees.

As concerns Afrotheria, *FBXL21* was found in the lesser hedgehog tenrec *Echinops telfairi* and rock hyrax *Procavia capensis* but not in *Loxodonta africana*.

Overall, the *FBXL21* gene acquired significant modifications in the Homininae ancestor after gibbons separated from other apes; no such modifications were observed in chimpanzees. Among hominids, *FBXL21* independently pseudogenized only in the gorilla and humans.

Discussion. What underlies the loss of *FBXL21*? It is not improbable that its significant modifications or elimination can lead to cerebrum enlargement via the increased puberty age and other neotenic characters. For instance, the long period of development until sexual maturity is common for the human, apes, and elephants. This period in the elephant *Loxodonta africana* varies from 9 to 19 years (with a high dispersion), which can also be attributed to the quantitative changes of this gene up to pseudogenization and loss. Comparison of the giant panda lost the *FBXL21* gene with bears demonstrates a slightly delayed immaturity period in the panda. Cubs of the Asian black bear *Ursus thibetanus* mature at the age of three years as against 4-8 years in the giant panda. At the same time, the size and body weight of these bears are similar. The polar bear *Ursus maritimus* sows have their first litter at the age of 4-8 years; however, polar bears are much larger and live under severe conditions [18].

On the other hand, the absence of *FBXL21* in the leporids *Oryctolagus cuniculus* and *Lepus timidus* does not conform to these observations.

Conclusions. The data obtained suggest the following scenarios of *FBXL21* evolution in placentals: 1) it was independently lost in many species; however, the loss is observed only in the human and gorilla among hominids; 2) the gene started to significantly change in the Homininae ancestor after the separation of gibbons and other apes. In particular, significant differences are observed in chimpanzees, although without notable variation within the genus; 3) the loss or substantial modification of the gene could predispose an increased age of sexual maturity (as against related species).

Methods. The method described elsewhere [5, 6] was used to search for genomic alignments and pseudogenizations. Besides, the search for genomic alignments was realized using tools of Ensembl release 102 [19]. The unannotated genome of the mountain hare *Lepus timidus* [17] (GCA_009760805.1) was analyzed by the alignment of amino acid sequences against tblastn translations of DNA. Nucleotide sequences were aligned by Clustal in the MEGA X software [20]. *dS* and *dN* were calculated using the method implemented in PAMLX [21, 22].

The study was funded by the RFBR according to the research project No. 18-29-13037. The computations were performed at the Joint Supercomputer Center of the Russian Academy of Sciences.

1. P.J. Bonthuis et al. (2015) *Cell Reports*, **12**:979–991.
2. C. Gregg et al. (2010) *Science*, **329**:643–648.
3. H. Dardente et al. (2008) *PLoS One*, **3**:e3530.
4. A. Hirano et al. (2013) *Cell*, **152**:1106–1118.
5. L.I. Rubanov et al. (2019) *BioData Mining*, **12**:20.
6. L.I. Rubanov et al. (2020) *Life*, **10**:192.
7. X. Chen et al. (2008) *Am J Med Genet B Neuropsychiatr Genet*, **147B**:1231–1237.
8. A.M. Boddy et al. (2012) *J Evol Biol*, **25**:981–994.
9. J. Shoshani et al. (2006) *Brain Research Bulletin*, **70**:124–157.
10. V.P. Skulachev et al. (2017) *Physiol Rev*. **97**:699–720.
11. V.P. Skulachev et al. (2020) *Aging (Albany NY)*, **12**:5566–5584.
12. P. Gunz et al. (2020) *Science Advances*, **6**:eaaz4729.
13. H. Kaplan et al. (2000) *Evolutionary Antropology*, **9**:156–185.
14. S.H. Montgomery (2011) *Molecular Biology and Evolution*, **28**:625–638.
15. S. Xu et al. (2017) *BMC Evol Biol*, **17**:206.
16. D.D. Korotkova et al. (2019) *Cell Reports*, **29**:1027–1040.e6.
17. J.P. Marques et al. (2020) *Genome Biol Evol*, **12**:3656–3662.
18. A.J. Welch et al. (2014) *Genome Biol Evol*, **6**:433–450.
19. A.D. Yates et al. (2020) *Nucleic Acids Research*, **48**:D682–D688.
20. S. Kumar et al. (2018) *Molecular Biology and Evolution*, **35**:1547–1549.
21. B. Xu, Z. Yang (2013) *Mol Biol Evol*, **30**:2723–2724.
22. Z. Yang (2007) *Mol Biol Evol*, **24**:1586–1591.