BGRS/SB-2020

The 12th International Multiconference

Bioinformatics of Genome Regulation and Structure/ Systems Biology

BGRS/SB-2020
NOVOSIBIRSK, RUSSIA
06–10 JULY, 2020

https://bgrssb.icgbio.ru/2020/

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences

Novosibirsk State University

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE/SYSTEMS BIOLOGY (BGRS/SB-2020)

The Twelfth International Multiconference

Abstracts

06–10 July, 2020
Novosibirsk, Russia

Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020) :
The Twelfth International Multiconference (06–10 July 2020, Novosibirsk, Russia); Abstracts /
Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Novosibirsk State University. – Novosibirsk: ICG SB RAS, 2020. – ISBN 978-5-91291-051-7.

Organizers

Institute of Cytology and Genetics, SB RAS (ICG SB RAS)

Novosibirsk State University (NSU)

Ministry of Science and Higher Education of the Russian Federation (Minobrnauki of Russia)

The Vavilov Society of Geneticists and Breeders

State Scientific-Research Institute of Physiology and Basic Medicine (PhBMRI)

Vavilov Journal of Genetics and Breeding

Evolution of proteins involved in response to ROS

Vassily Lyubetsky
IITP RAS
Moscow, Russia
lyubetsk@iitp.ru

Gregory Shilovsky
IITP RAS; MSU
Moscow, Russia
gregory_sh@list.ru

Alexandr Seliverstov
IITP RAS
Moscow, Russia
slvstv@iitp.ru

Oleg Zverkov
IITP RAS
Moscow, Russia
zverkov@iitp.ru

Lev Rubanov
IITP RAS
Moscow, Russia
rubanov@iitp.ru

Abstract — **Original software was used to specify the evolution of transcription factors Nrf2 and Bach1 in Deuterostomia. The transcription factors are antagonistically involved in the response to reactive oxygen species (ROS). The original algorithm shows that *Bach* emerged by duplication of *Nfe2*, an ortholog of *Nfe2l2*, in the chordate ancestor. At the N-terminus, the copy was provided with the BTB domain from a gene orthologous to the European lancelet *Branchiostoma lanceolatum* gene *BL03038_cuf1* followed by a domain typical for the Zinc finger C2H2 superfamily.**

Keywords — Nrf2, anti-ageing program, anti-ROS; Bach1, ageing program, pro-ROS; ROS

Introduction

In many vertebrates, Nrf2 (Nfe2l2), Bach proteins (Bach1 and Bach2), as well as Keap1, β-TrCP, c-Myc, and GSK3b are components of the regulatory network (e.g., in *Mus musculus*) that is, among other functions, associated with regulation of the level of reactive oxygen species (ROS). In particular, the relationship between Nrf2 and Bach1 is critical for the regulation of heme oxygenase 1 expression, which provides for heme degradation. The functioning of this network is related to *species-specific lifespan* as well as to *many human diseases*. Nrf2 activates about 200 genes; Nrf2 and Bach1 are antagonists in controlling ROS levels, which is also due to the competition for ARE DNA-binding sites. These proteins usually function by complexing with each other as well as with other proteins (e.g., Maf) and DNA. The BTB (POZ) domain is common in zinc-finger transcription factors. Bach proteins are the only BTB proteins with the bZIP domain. Bach has the BTB domain at the N-terminus and the bZIP domain at the C-terminus. Nrf2 also has the bZIP domain at the C-terminus. Phosphorylation of Tyr-486 in the mouse Bach1 inactivates the protein. The heme-binding sites have been identified for Bach as well as the sites of Bach homodimerization. In Bach, the heme-binding regions are not similar, and, in particular, differ by the number of regulatory cysteine-proline (CP) motifs. The following evolutionary assumptions have been proposed: the bZIP family descends from a single eukaryotic gene; the common ancestor of the *Bach* genes existed in chordates before the divergence of vertebrates since it exists in *Ciona* spp.; two *Bach* genes emerged in gnathostomes, apparently, after their separation from lampreys. Invertebrates have a single ortholog to four vertebrates genes (*Nfe2*, *Nfe2l1*, *Nfe2l2*, and *Nfe2l3*). Our work contains results about the evolution of *Nrf2* and *Bach* genes and corresponding proteins. Here we present the results of the computer-aided search for *Nrf2* and *Bach* in deuterostomes using the conditions specified below.

Results

1) Here we consider all deuterostomes represented in Ensembl and GenBank. The *Bach* gene has been identified in tetrapods, cartilaginous and bony fishes, cyclostomes, and ascidians using a local alignment with the human BACH1 (NP_001177.1), the presence of the BTB and bZIP-Maf domains, satisfactory alignment of the secondary structure, and high similarity of the 3D structure at the N- and C-terminal regions with the human BACH1. According to *these criteria*, the *Bach* gene is missing in lancelets, hemichordates, and echinoderms. Similarly, the *Nrf2* gene has been identified in deuterostomes using local alignment with the human NFE2L2 (NP_006155.2), the presence of the bZIP-Maf domain, a satisfactory alignment of the secondary structure, the absence of some other domains (kelch-type beta-propeller for kelch-like proteins; BTB for Bach; bZIP-Jun for Jun and similarly for other bZIP proteins), and high 3D structure similarity with NFE2L2 at the C-terminus coupled with a significant difference from other human bZIP proteins at the N-terminus. *Another* important reason for dividing all found proteins into the Bach and Nrf2 groups as well as other groups is the splitting of the unrooted tree into the clades of Bach and Nrf2 separated with a nearly 100% support. Rooting the tree with echinoderms makes it incongruent with the species tree only for Nrf2 in *Oikopleura dioica* and *Ciona* spp., and these proteins are questionable in other respects. Similarly, we recognize the Bach and Nrf2 clades in the bigger tree of bZIP proteins.

Analysis of the search results. Among invertebrate deuterostomes, Bach proteins have been found only in *Ciona intestinalis* and *C. savignyi* (tunicates diverged early from other chordates). At least one Bach protein has been found in all vertebrates. It looks like *Bach* emerged in the common ancestor of vertebrates and tunicates (and was possibly lost in appendicularians), after which it diverged into *Bach1* and *Bach2* in the common ancestor of cartilaginous and bony fishes. In fishes, the *Bach* genes were many times duplicated and lost; their genomes commonly have many paralogs that are orthologous to *Bach*. E.g., the huchen *Hucho hucho* has four genes orthologous to *Bach2*. The *Bach1* gene is represented by a single copy in all tetrapods. The alignment of Bach proteins indicates considerable conservation of each of them in most mammals. Within Euarchontoglires, Bach1 demonstrates only minor changes in most rodents including the Damaraland mole-rat (DMR) *Fukomys damarensis*. The only exception is the C-terminal region of all Bach1 isoforms in the naked mole-rat (NMR) *Heterocephalus glaber*. One Bach1 isoform has an extended deletion, although a short region upstream of the C-terminus is conserved. Another isoform has a long insertion in the same region. The full-length Bach1 protein of the NMR usually shares the dispensable amino acids with the DMR but not the mouse, which agrees with their taxonomical position. No significant differences in Bach1 have been revealed in primates. Beyond Euarchontoglires, the Bach1 sequences of the sloth *Choloepus hoffmanni*, tenrec *Echinops telfairi*, shrew *Sorex araneus*, dolphin *Tursiops truncatus*, flying fox *Pteropus vampyrus* are similar to those in primates and most rodents. Conversely, other

representatives of Laurasiatheria as well as the nine-banded armadillo *Dasypus novemcinctus* have insertions or deletions in Bach1 in the corresponding region of the NMR. The Bach1 of the hedgehog *Erinaceus europaeus*, elephant *Loxodonta africana*, and hyrax *Procavia capensis* in Bach1 has a very long C-terminal deletion covering both conserved and variable regions. On the other hand, one of the Bach2 isoforms in primates *Macaca mulatta*, *M. nemestrina*, and *Pan troglodytes* has an N-terminal extension not observed in humans.

2) Here we consider all species represented in RefSeq and Ensembl supplemented by those in GenBank. The Bach1 protein of cartilaginous fishes lacks the Tyr-486. The Bach1 in tetrapods has conserved the functionally significant phosphorylated tyrosine and the neighboring amino acids remained largely unchanged, although D→S substitutions and proline loss in the CP motifs are observed in long-lived rodent DMR and NMR. In the Australian ghostshark *Callorhinchus milii*, cloudy catshark *Scyliorhinus torazame*, whale shark *Rhincodon typus*, and brownbanded bamboo shark *Chiloscyllium punctatum*, this tyrosine is replaced with phenylalanine. The Bach in *Ciona* spp. has preserved this tyrosine unlike some of the neighboring amino acids. The predicted heme-binding sites in Bach1 of tetrapods insignificantly differ from those in mouse except the species described below. These sites include the 223-LCPKYR-228 (C→G in marsupials, while in the platypus *Ornithorhynchus anatinus* it is the same as in the mouse; or C→Y in the common wall lizard *Podarcis muralis*); 300-QCPAEQ-305, which considerably changed or disappeared in most mammals; 435-ECPWLG-340 (conserved in all tetrapods); 463-NCPFIS-468 (the cysteine is conserved in tetrapods, and I→M in placentals including DMR, NMR and the common degu *Octodon degus*); 494-PCPYAC-499 (conserved in all tetrapods except the bearded dragon *Pogona vitticeps* and platypus); and 648-DCPLSF-653 (conserved in almost all tetrapods). Bach in *Ciona* spp. has only two conserved CP sites, and their positions differ from those involved in the heme-dependent regulation in human and mouse. A similar pattern is observed for Bach2. The 368-ACPFNK-373 heme-binding site is present in most Myomorpha but not in the Upper Galilee Mountains blind mole rat (GMR) *Nannospalax galili*; the ACPFDK (N→D) site are found in other rodents (the North American beaver *Castor canadensis*, guinea pig *Cavia porcellus*, the GMR, and other mole-rats, DMR and NMR) and nearly all tetrapods; the ACPSDK (F→S) site is found in the Philippine tarsier *Carlito syrichta*; ACSFDK (P→S), in flying foxes; VCPFDK (A→V), in the barbed agama; ACPLDK (F→L), in marsupials and the three-toed box turtle *Terrapene carolina triunguis* (with a negligible aging rate); ACPFEK, in birds, saltwater crocodile *Crocodylus porosus*, and Chinese alligator *Alligator sinensis*; ACPVEK, in the western clawed frog *Xenopus tropicalis*; ACPLNR, in the West Indian Ocean coelacanth *Latimeria chalumnae*; no this site was found in the platypus. Other fishes have the P→S substitution and some other modifications of the site. Overall, Bach2 in flying foxes differ from that (commonly ACPFDK) in other chiropterans. The 498-SCPVPI-503 site is strictly conserved in all species starting from cartilaginous fishes to mammals. The 505-VCPRSP-510 site is also strictly conserved within the same taxonomic range excluding the *L. chalumnae* with the C→Y substitution. The 602-SCPVQD-607

site is conserved in the *L. chalumnae* and all tetrapods. This site is missing in the European cattle *Bos taurus*, while *Bos mutus* shares it with all other tetrapods. This site is missing in the bluespotted mudskipper *Boleophthalmus pectinirostris* as well as in other fishes. The sites notably differ in the *C. milii*, and whale shark *Rhincodon typus* from those in tetrapods. The 728-YCPVLI-733 site is only found in all rodents but not in other taxa. However, most vertebrates including the European rabbit *Oryctolagus cuniculus* and *C. milii* have a different site, YCPVLR (I→R). The Chinese softshell turtle *Pelodiscus sinensis* demonstrate singular modifications in the YFPVLR (C→F) site; the *X. tropicalis*, in YCPVLQ (R→Q); the *R. typus*, in FCPVFR (Y→F, L→F). The evolution of the insertion in the N-hook downstream of the MSLSE motif at the N-terminus in Bach1 has been studied. The evolution of response to mitochondrial ROS will be discussed [1].

Discussion

The genomic rearrangement that gave rise to the *Bach* remains an open problem. The Nfe2 is most similar to the Bach in early-diverging deuterostomes. One can think that *Bach* emerged by duplication of *Nfe2*, an ortholog of *Nfe2l2*, in the chordate ancestor. The BTB domain essential in Bach is missing in Nfe2; however, it exists in dozens of ancient proteins (e.g., the BTB-ZF family) that could provide it for the ancestral Bach. Possible BTB sources include a gene orthologous to the European lancelet *Branchiostoma lanceolatum* gene *BL03038_cuf1* with the domain at the N-terminus followed by a domain typical for the Zinc finger C2H2 superfamily (IPR036236). The bZIP domain specific for the C-terminal regions of Bach proteins is not found in BL03038_cuf1. Thus, *Bach* could emerge as a chimeric gene. Anyhow, such an evolutionary scenario is optimal for the reconstruction of the genomic structures carried out by the original algorithm. The absence of Bach proteins in *Branchiostoma* spp. agrees with the proposed closer phylogenetic similarity between vertebrates and tunicates rather than between vertebrates and lancelets. The absence of Bach in *Oikopleura dioica*, which is relatively close to *Ciona* spp., can be attributed to its neoteny. Indeed, like other appendicularians, adult *O. dioica* has a discrete body and tail and preserves the notochord throughout its life, while the body structure of *Ciona* spp. substantially changes in development. The loss or substantial change of *Bach2* genes in the tuatara *Sphenodon punctatus* and the tortoise *Chelonoidis abingdonii* is accompanied by a long species-specific lifespan. The changes in the NMR Bach1 untypical for rodents can also be related to the unusually long lifespan.

REFERENCES

[1] M. Y. Vyssokikh, S. Holtze, O. A. Averina, K. G. Lyamzaev, A. A. Panteleeva, M. V. Marey et al., "Mild depolarization of the inner mitochondrial membrane is a crucial component of an anti-aging program," Proc. Natl. Acad. Sci. U.S.A., Mar 2020, 201916414.