

Cloning the interferon regulatory factor 1 gene in lungfish (*Protopterus annectens*) and its molecular evolution among sarcopterygians

GAN XiaoNi^{1,2,3}, WANG DengQiang², WANG XuZhen² & CHEN XinWen^{1*}

¹ Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China;

² Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China;

³ Graduate University of Chinese Academy of Sciences, Beijing 100039, China

Received December 20, 2010; accepted February 28, 2011; published online May 10, 2011

Sarcopterygians is an important vertebrate clade that includes crossopterygians and tetrapods. Crossopterygians are lobe-finned fish that include lungfish and coelacanth. Tetrapods include amphibians, reptiles, avians and mammals. To compare the interferon regulatory factor 1 (*irf-1*) gene structure and to explore phylogenetic relationships among sarcopterygians, we cloned the cDNA sequence of *irf-1* from lungfish and compared it with *irf-1* orthologs in other sarcopterygian species. The lungfish is a primitive sarcopterygian that occupies a very important position in vertebrate phylogeny. Interferon regulatory factors (IRFs) are a family of proteins involved in innate immunity. To date, 11 IRF family members have been reported. All IRFs share homology in the first 115 amino acids, which encompasses a DNA binding domain containing a characteristic repeat of 5 tryptophan residues separated by 10–18 amino acids. IRF-1 and IRF-2 were the first members of this family to be reported and they have a very important role in innate immunity. However, studies of the *irf-1* and *irf-2* genes are mostly confined to mammals; very few non-mammalian *irf-1* genes have been reported. Consistent with the *irf-1* gene sequences already published, the first 345 nucleotides of lungfish *irf-1* are highly conserved. At the carboxyl terminal a C-terminal transactivating region motif and an interferon associated domain (IAD2) were identified. 417 million years separate the present from the closest common ancestor of lungfish and tetrapods; however, the *irf-1* genes among sarcopterygians are highly conserved and have very obvious phylogenetic relationships. Also the interrelationship tree of sarcopterygians, based on IRF-1 amino acid sequences, is identical with trees produced using other data, such as morphological characteristics or mitochondrial gene sequences.

IRF-1, innate immunity, Sarcopterygii, molecular evolution, phylogeny

Citation: Gan X N, Wang D Q, Wang X Z, et al. Cloning the interferon regulatory factor 1 gene in lungfish (*Protopterus annectens*) and its molecular evolution among sarcopterygians. Chinese Sci Bull, 2011, 56: 1782–1786, doi: 10.1007/s11434-011-4455-6

Interferon (IFN) is a multifunctional cytokine that has a crucial role in innate immunity in animals. Interferon regulatory factor (IRF) has a very important role in interferon synthesis. IRF was first identified as a factor that regulates the transcription of the *IFN* gene and of IFN-induced genes [1]. IRF also has other metabolic functions, such as in antiviral defense, immune regulation and growth control [2]. To date, 11 members of the IRF family have been reported, mostly in mammals. IRF-1 regulates the virus-induced enhancer-like element in the human interferon- β (*IFN- β*) gene

and is a transcriptional activator of the type I IFN gene and of the IFN-induced genes [3,4]. Another factor, structurally similar to IRF-1, was identified and termed IRF-2. IRF-2 is a repressor of transcription in IFN synthesis [5] and also has other important metabolic functions, for example, IRF-2 is transcript activating in the transcription process of histone H4, VCAM-1, gp91phox and interleukin-7, also it is the transcript catalyst for class II transactivator [6–10]. IRF has a DNA binding domain (DBD) and an IRF association domain 1 (IAD1) or IAD2. Its first 115 amino acids are high homologous, and the tryptophan cluster consists of 5 tryptophans. On its carboxyl terminal region, there is lower

*Corresponding author (email: chenxw@wh.iov.cn)

homogeneity but higher variations when comparing with N-terminal region, such as DBD. All IRF members contain IAD-1, except for IRF-1 and IRF-2, which contain IAD-2. The IADs are an important criterion for the relationships among family members and other factors. IRF-1 and IRF-2 interact with IRF-8 through IAD-2 [11,12].

To compare *irf-1* gene structure and to explore the phylogeny of *irf-1* in crossopterygians and tetrapods, the lungfish (*Protopterus annectens*) *irf-1* cDNA was cloned and sequenced. The published phylogeny of sarcopterygians is based on mitochondrial and ribosomal gene sequences. The phylogeny of sarcopterygians, based on a gene involved in innate immunity, has not been reported. In the present study, we obtained the complete cDNA sequence of lungfish *irf-1* using the RACE method. Phylogeny trees were then constructed using the *IRF-1* gene.

1 Materials and methods

1.1 Sample collection and identification

Samples of African lungfish (*Protopterus annectens*) were obtained from an aquatic pet market in Guangzhou, China. Specimens were confirmed in the Institute of Hydrobiology, Chinese Academy of Sciences.

1.2 Extraction of RNA, synthesis of cDNA, gene amplification and sequencing

Total RNA was extracted using a Trizol reagent kit (TaKaRa, China) from a combination of heart, liver, kidney, spleen and intestine tissues. The first cDNA strand was synthesized using an M-MLV 1st strand kit (Invitrogen, USA) in accordance with the manufacturer's instructions. Primers for the amplification of homologous fragments and RACE are listed in Table 1. Initially, we used degenerate primers (Lf-IRF-1F1 and Lf-IRF-1R1) to amplify a conserved fragment. Degenerate primers were designed according to published IRF-1 sequences. PCR was performed in a total volume of 50 μ L that included 100 ng cDNA, 5 μ L 10 \times buffer (TaKaRa), 1 μ L each primer (10 μ mol/L), 2 μ L dNTPs (2.5 μ mol/L each) (TaKaRa), 2.0 U *Taq* DNA polymerase (TaKaRa) and double distilled H₂O to the final

volume to 50 μ L. PCR cycling parameters were: 94°C for 3 min; followed by 34 cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 1 min; and a final cycle of 72°C for 7 min. A control reaction without cDNA template was performed. To identify and clone a full-length cDNA, 3'RACE and 5'RACE were performed using gene specific primers and adaptor primers. RACE primers, 3'Lf-IRF-1-RACEouter inner and 3'Lf-IRF-1-RACEinner, were designed based on the amplified conserved fragment. PCR cycle parameters used were as above except that the annealing temperature was 54°C. 3 RACE was performed using the First Choice RLM-RACE kit (Ambion, USA). Primer sequences are listed in Table 1. PCR cycle parameters were: 94°C 3 min followed by 35 cycles of 94°C for 30 s, 54°C for 30 s and 72°C for 1 min 40 s, and a final cycle at 72°C for 7 min. All PCR products were cloned into pMD18-T vector and sequenced. DNA sequencing was performed using the generic primer, M13.

1.3 Alignment and phylogenetic analysis of the *irf-1* gene sequence

The *IRF-1* gene sequences downloaded from GenBank are listed in Table 2. The multiple alignments of sequences were performed using Clustal X [13]. The editing of DNA sequences and protein translation were performed using SEAVIEW software [14]. Neighborhood joint (NJ), maximum likelihood (ML) and Bayesian analysis (BA) methods were used to construct the phylogenetic trees. In the ML and BA analyses, the models and parameters of nucleotide evolution were estimated using the Model-test software [15]. Each clade on the tree was evaluated with 1000 bootstrap samples. The software used in this study included PAUP 10.4b, MrBayes 3.12 and Mega 4.0 [16–18]. Grass carp (*Ctenopharynodon idellus*), Mandarin fish (*Siniperca chuatsi*) and snakehead (*Channa argus*) were assigned as out-groups.

2 Results

2.1 Characteristics of the lungfish *irf-1* gene

The lungfish *irf-1* gene has an open reading frame (ORF) of

Table 1 Primers for PCR amplification and RACE

Name of primer	Sequence of primer	Application
Lf-IRF-1F1	5'-CARATHCCNTGGATGCAAYGC-3'	IRF-1 conserved fragment
Lf-IRF-1R1	5'-GCYTCCAKGTYTTNGGRTC-3'	IRF-1 conserved fragment
Lf-IRF-1F2	5'-ATGCCNGTVGARMGRATGMG-3'	IRF-1 conserved fragment
Lf-IRF-1R2	5'-GCATTCCAAGTATTAGGTTCT-3'	IRF-1 conserved fragment
3'Lf-IRF-1-RACEouter	5'-GCTTGCAATTTCCAGAGATGGGCG-3'	3'RACE
3'Lf-IRF-1-RACEinner	5'-CTGGGCGGTACAAACCTGGAGAG-3'	3'RACE

Table 2 Gene accession numbers of genes used in alignment and phylogenetic analyses

No.	Species	Latin name	GenBank No.
1	Snakehead	<i>Channa argus</i>	EF067849
2	Mandarin fish	<i>Siniperca chuatsi</i>	AY647431
3	Grass carp	<i>Ctenopharynodon idellus</i>	GU9970981
4	Human	<i>Homo sapiens</i>	NM_002198
5	Sheep	<i>Ovisaries</i>	AF3319701
6	Mouse	<i>Mus musculus</i>	NM_001159396
7	Rat	<i>Rattus norvegicus</i>	NM_012591
8	Chicken	<i>Gallus gallus</i>	NM_245015
9	African claw frog	<i>Xenopus laevis</i>	NM_001089781
10	West African lungfish	<i>Protopterus annectens</i>	HQ873639

942 bp, encoding 314 amino acids, and a 3-UTR of 399 bp. The stop codon is TAG. The putative IRF-1 peptide sequence contains a nuclear localization sequence (NLS), C-terminal transactivating region and an IAD-2. This sequence was submitted to GenBank with the accession number HQ87-3639.

2.2 Comparative analysis of IRF-1 sequences in sarcopterygians

Although there is variation in *irf-1* gene size among the different sarcopterygian species, their ORFs are of similar length. The alignment of *IRF-1* gene sequences of sarcopterygians (mammals, avian, reptiles, amphibians, crossopterygian) and actinopterygii as an out-group show that the first 345 nucleotides, encoding 115 amino acids, are very conserved and that the other regions, such as the IAD-2, have higher levels of variation. The alignment of the putative IRF-1 amino acid sequences also show high levels of conservation, especially the DBD and the 5 tryptophan residues repeats in the first 115 amino acids.

2.3 Molecular evolution and phylogenetic analysis

The comparative analysis of IRF-1 amino acid sequences (Figure 1) indicated variation between tetrapods and crossopterygians. Among tetrapods, avians and reptiles have a very high identity of the IRF-1 amino acid sequence, as do amphibians and lungfish (crossopterygian). In lungfish, the first 115 amino acid residues are conserved. Besides a conserved DBD in lungfish, on the carboxyl terminal region we also identified a conserved C-terminal transactivating region and IAD2.

The phylogenetic trees constructed using different methods showed a similar topological structure. The NJ tree, based on amino acid sequences with 1000 replicated bootstraps, is shown in Figure 2. When 3 actinopterygian species were assigned as an out-group, all the included sarcopterygian species form a monophyletic group, lungfish as the

representative crossopterygian was located on the basal clade. The obtained clades were supported with high bootstrap values. Among tetrapods, the African claw frog is a primitive amphibian and is located on the base, followed by the avians (chick) and then the mammals. All clades obtained >90% bootstrap support. The same tree based on mitochondrial cytochrome *b* gene sequences is similar to that based on *irf-1* sequences. Because the IRF-1 amino acid sequence from the Anolis genus of lizards obtained from GenBank is very short, we did not use it in phylogenetic analysis.

3 Discussion

Because of the high level of sequence diversity and a large intron, to design PCR primers and amplification of the IRF gene from genomic DNA of diverse species is very difficult. At present most of the *IRF* genes reported are mammalian, with few *IRF* genes reported in fish and primitive tetrapods. The sequence of the lungfish *IRF-1* gene gives us a complete dataset of IRF-1 in sarcopterygians, so we have *IRF-1* gene sequences from all the representative members of sarcopterygians. The fossil record indicates that the most recent common ancestor of lungfish and tetrapods is 417 million years old. However, over such a long time scale, the IRF-1 ORF has maintained a high level of conservation, indicating that IRF-1 has a very stable secondary structure of protein and a very important function. Also the involvement of IRF-1 and IRF-2 in innate immunity can be traced back to the time when osteichthyan species originated.

Sarcopterygians are very important vertebrates that include crossopterygians and tetrapods. In addition to the fossil record, the recent species of crossopterygians include two species of coelacanth and three species of lungfish. However, their molecular phylogeny based on innate immune genes is not known. Both the fossil record and morphological data indicated tetrapods originated from the early crossopterygians. The most primitive tetrapods may be the amphibian genus, *ichthyostega*. Reptiles are a huge group derived from amphibians that underwent rapid development in the Jurassic period. Avians constitute a branch of special reptiles, as are dinosaurs. Mammals are derived from another branch of reptiles. But the molecular phylogeny evidence has not provided conclusive evidence of interrelationships among avians, reptiles and mammals. Based on 18S rRNA gene sequences, Xia et al. [19], Hedges et al. [20] and Bininda-Emonds et al. [21] explored the phylogeny of tetrapods using different gene markers and character sets. Our results, based on *irf-1* gene sequences, show similar phylogenetic trees to those in the above studies. In our tree, lungfish were located on the base and form a sister group with tetrapods with very high bootstrap support. Following this are amphibians, avian and mammals. Only a very short

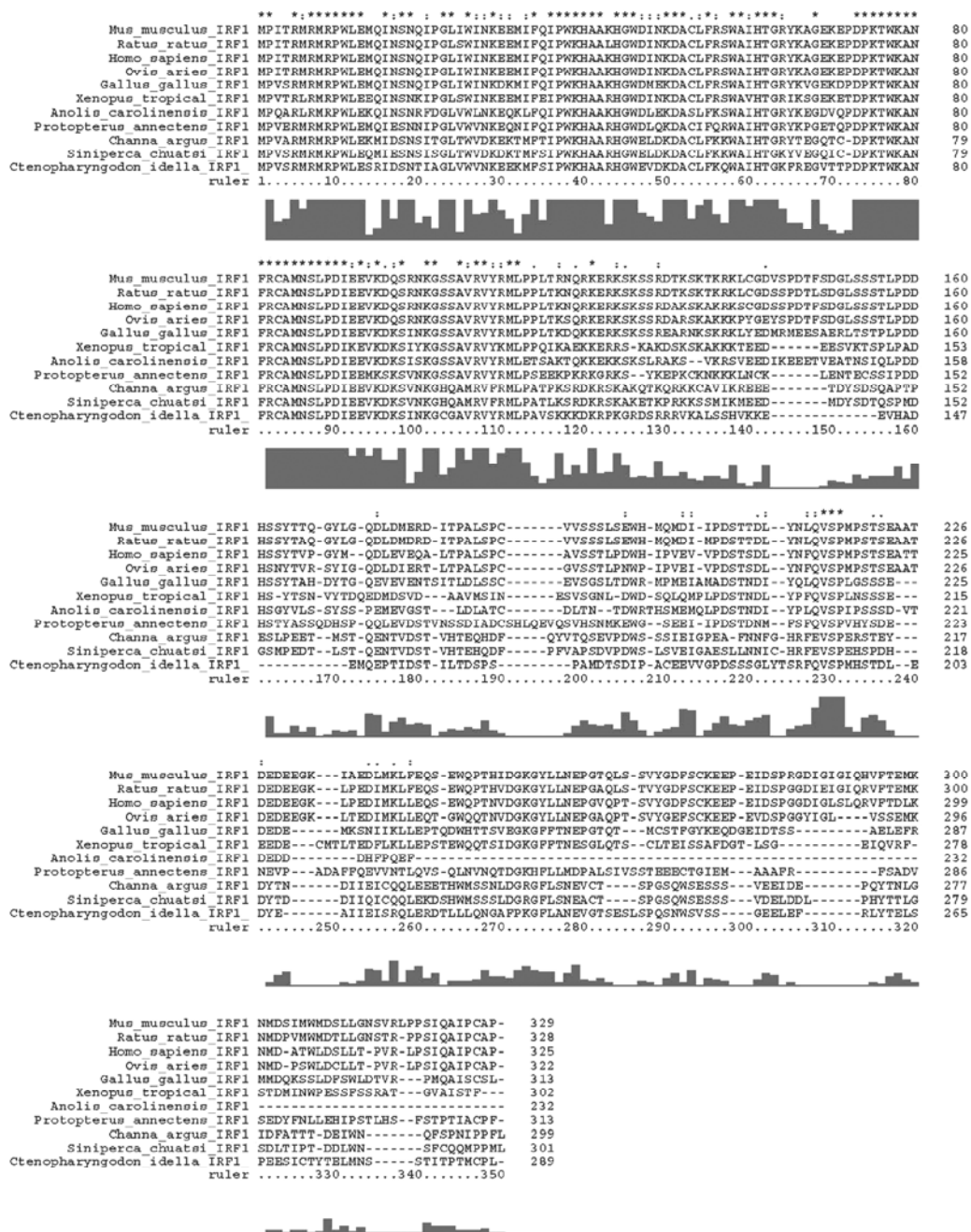


Figure 1 Alignment of IRF-1 amino acid sequences in sarcopterygians.

IRF-1 gene sequence was available for the *Anolis* genus of lizards; therefore we did not include it in the phylogenetic analysis. Compared with full length *IRF-1* gene cDNA, the phylogenetic tree based on only the first 345 nucleotides, which codes for the DNA binding domain, obtained higher bootstrap support for each clade on the tree. We also found that trees based on functional gene sequences yielded higher

bootstrap support than those based on mitochondrial gene sequences. Our research also found that trees based on putative amino acid sequences can obtain higher bootstrap support than those based on DNA sequences. Finally, it is worth noting that because *IRF-1* has such an important function and has remained stable over such a long time scale, it is a very good phylogenetic marker.

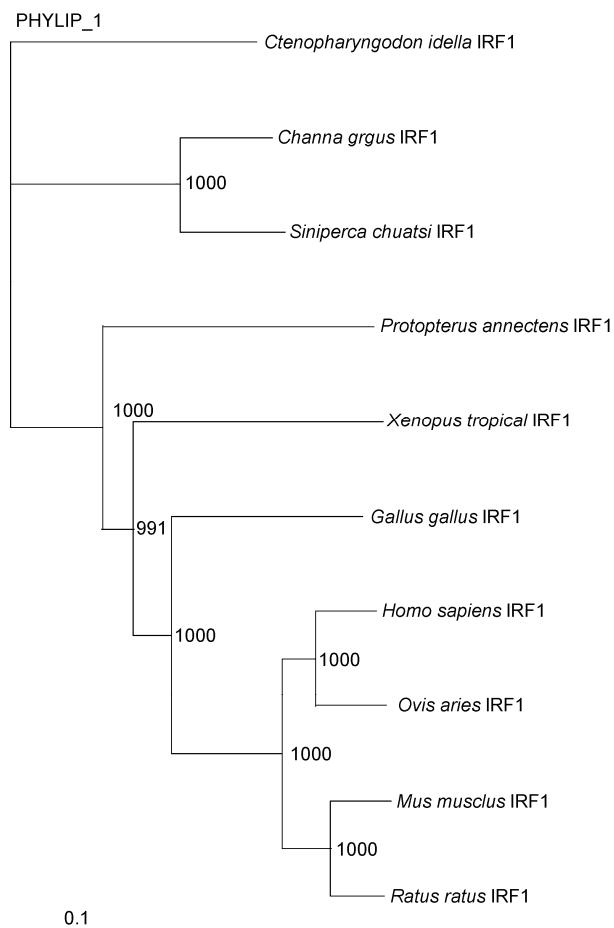


Figure 2 Phylogenetic tree based on the amino acid sequence of sarcopterygian IRF-1s. Genetic distance is estimated using the Kimura 2 parameter model. The tree is constructed using N-J methods, clades are estimated using bootstrap with 1000 resamples. *Protopterus annectens*, lungfish.

This work was supported by the National Key Basic Research Program of China (HCV2009CB522500).

- Mamane Y, Heylbroeck C, Genin P, et al. Interferon regulatory factors: The next generation. *Gene*, 1999, 237: 1–14
- Barnes B, Lubyova B, Pitha P M. On the role of IRF in host defense. *J Interferon Cytokine Res*, 2002, 22: 59–71
- Miyamoto M, Fujita T, Kimura Y, et al. Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to

- IFN- β regulatory elements. *Cell*, 1988, 54: 903–913
- Kroger A, Koster M, Schroeder K, et al. Activities of IRF-1. *J Interferon Cytokine Res*, 2002, 11: 5–14
- Harada H, Takahashi E, Itoh S, et al. Structure and regulation of the human interferon regulatory factor 1 (*IRF-1*) and *IRF-2* genes: Implications for a gene network in the interferon system. *Mol Cell Biol*, 1994, 14: 1500–1509
- Vaughan P S, Azia F, van Wijnen A J, et al. Activation of a cell cycle regulated histone gene by the oncogenic transcription factor IRF-2. *Nature*, 1995, 377: 362–365
- Luo W, Skalnik D G. Interferon regulatory factor-2 directs transcription from the gp91phox promoter. *J Biol Chem*, 1996, 271: 23445–23451
- Jesse T L, LaChance R, Iademarco M F, et al. Interferon regulatory factor-2 is a transcriptional activator in muscle where it regulates expression of vascular cell adhesion molecular-1. *J Cell Biol*, 1998, 140: 1265–1276
- Xi H, Blanck G. The IRF-2 DNA binding domain facilitates the activation of the class II transactivator (CII TA) type IV promoter by IRF-1. *Mol Immunol*, 2003, 39: 677–684
- Oshima S, Nakamura T, Mamiki S, et al. Interferon regulatory factor-1 (IRF-1) and IRF-2 distinctively up-regulate gene expression and production of interleukin-7 in human intestinal epithelial cells. *Mol Cell Biol*, 2004, 24: 6298–6310
- Schaper F, Kirchhoff S, Posern G, et al. Functional domains of interferon regulatory factor 1 (IRF-1). *Biochem J*, 1998, 335: 147–157
- Meraro D, Hashmueli S, Koren B, et al. Protein-protein and DNA-protein interactions affect the activity of lymphoid-specific IFN regulatory factor. *J Immunol*, 1999, 162: 6468–6478
- Thompson J D, Gibson T J, Plewniak F, et al. The Clustal X windows interface: Flexible strategies for 938 multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25: 4876–4882
- Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, 2010, 27: 221–224
- Posada D, Crandall K A. Modeltest: Testing the model of DNA substitution. *Bioinformatics*, 1998, 14: 817–818
- Swofford D L. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), ver 4. Sinauer Associates, Sunderland, Massachusetts, 2002.
- Ronquist F, Huelsenbeck J P. MrBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 2003, 19: 1572–1574
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*, 2004, 5: 150–163
- Xia X H, Xie Z, Kjer K M. 18S ribosomal RNA and tetrapod phylogeny. *System Biol*, 2003, 52: 283–295
- Hedges S B, Moberg K D, Maxson L R. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Mol Biol Evol*, 1990, 7: 607–633
- Bininda-Emonds O R, Jeffery J E, Sánchez-Villagra M R, et al. Forelimb-hindlimb developmental timing changes across tetrapod phylogeny. *BMC Evol Biol*, 2007, 7: 182

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.