HRE Finder: A Tool for Quarrying Hypoxia-Response Element in Genomic Sequences

Article *in* International Journal of Current Microbiology and Applied Sciences · July 2017

DOI: 10.20546/ijcmas.2017.607.190

8 authors, including:

Iliyas Rashid
University of California, San Diego
15 PUBLICATIONS **102** CITATIONS
SEE PROFILE

Naresh Sahebrao Nagpure
National Bureau of Fish Genetic Resources
163 PUBLICATIONS **2,005** CITATIONS
SEE PROFILE

Prachi Srivastava
AMITY University Uttar Pradesh Lucknow
52 PUBLICATIONS **215** CITATIONS
SEE PROFILE

Ajey Kumar Pathak
National Bureau of Fish Genetic Resources
48 PUBLICATIONS **542** CITATIONS
SEE PROFILE

Int.J.Curr.Microbiol.App.Sci (2017) *6*(7): 1580-1586

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 6 Number 7 (2017) pp. 1580-1586**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2017.607.190

HRE Finder: A Tool for Quarrying Hypoxia-Response Element in Genomic Sequences

Iliyas Rashid[1,2], Naresh Sahebrao Nagpure[3], Prachi Srivastava[2], Ajey Kumar Pathak[1]*, Ravindra Kumar[1], Vishwamitra Singh Baisvar[1], Mahender Singh[1] and Basdeo Kushwaha[1]

[1]Molecular Biology and Biotechnology Division, ICAR- National Bureau of Fish Genetic Resources, Lucknow- 226 002, Uttar Pradesh, India
[2]AMITY Institute of Biotechnology, AMITY University Uttar Pradesh, Lucknow-226 028, Uttar Pradesh, India
[3]Fish Genetics and Biotechnology Division, ICAR- Central Institute of Fisheries Education, Mumbai-400 061, Maharashtra, India
**Corresponding author*

Keywords

Database, Hypoxia inducible factor, Hypoxia response element, Hypoxia, Transcription, Gene, Enhancer, Search algorithm.

Article Info

Accepted:
19 June 2017
Available Online:
10 July 2017

A B S T R A C T

During hypoxia, the hypoxia inducible factor (HIF) binds to hypoxia-response elements (HRE) and induces the expression of hypoxia-inducible genes (HIG). A functionally active HRE typically contains an adjacent hypoxia ancillary sequence (HAS) in its downstream within the space of 7-15 nt. Here, a computational tool 'HRE Finder' is discussed using an example data set. The tool provides the ability for mining canonical HREs and predicting functional HREs in the upstream region of genes, which can be an indispensable tool for genome-wide analysis and identification of hypoxia inducible genes in vertebrates.

Introduction

Hypoxia is a condition of oxygen deficiency in the body tissues that affects proper metabolic activity, alters the expression of several genes and causes various abnormalities in different animals, ranging from invertebrates to mammals (Guillemin and Krasnow, 1997). Some genes are known to have a role in hypoxia tolerance, and their expression mechanism is induced by a transcriptional factor known as Hypoxia Inducible Factor (HIF). HIF is a heterodimeric transcription factor composed of an HIF-α (1α, 2α and 3α) and an HIF-β (1β, 2β and 3β) subunits (Wang and Semenza, 1995). A wide range of mammalian cells were well studied for the transcriptional responses to hypoxia and zebrafish have been used as a model for the study of the pathway in hypoxic condition (Smith *et al.*, 2006; van Rooijen *et al.*, 2011). HIF recognizes Hypoxia

Responsible Elements (HREs), a transcription binding site (TFBs), within the promoters of a large number of genes following hypoxic induction and binds to HREs of the gene and enhances the expression of the gene (Wenger *et al.,* 2005). HREs are normally distributed in the genome and their frequent occurrences have been reported in the enhancer or promoter region and 3' UTR of a gene (Fordel *et al.,* 2004). The list of confirmed HIF target genes is ever expanding.

The functionally active HREs were observed in the promoter region of more than 100 mammalian genes involved in different biological processes (Manalo *et al.,* 2005; Semenza, 2010). The frequent occurrence of functional HREs involving in transcriptional response was observed in the 1kb upstream stream from the transcription start site while more than 70% hypoxia responsive genes were found without undetectable HIF-binding site in proximal promoters of a gene (Benita, 2006). Further, the frequently distributed HREs (5'-RCGTG-3') are functionally activated when its downstream contains an adjacent hypoxia ancillary sequence (HAS) in the space of 7-15 nucleotide (Kajimura *et al.,* 2006; Liu *et al.,* 1995).

HRE and HAS are five nucleotide motifs and their consensus patterns [(5'-RCGTG-3') and {5'-CA(G|C)(A|G)(T|G|C)-3'}] were reported in several hypoxia-inducible genes in mammals and IGFBP1 gene in fish (Kimura *et al.,* 2001; Kajimura *et al.,* 2006). Approaches were applied to identify the hypoxia-inducible factor binding sites and target genes in genome-wide analysis (Greenald *et al.,* 2015; Ortiz-Barahona *et al.,* 2010). There are available numbers of software tools and packages like MEME Suite and MOODS (Korhonen *et al.,* 2009; Bailey *et al.,* 2009) for predicting conserved motifs and all types of TFBs in the noncoding genomic sequences. Unfortunately, at present, there is not even a single tool is available specifically dedicated for the finding of HRE sites and identifying functional HRE in the upstream region of the genes. Keeping this in view, a program 'HREFinder' was developed in Perl for mining all canonical HREs in the upstream sequence input by the user and predicting functional HREs.

Materials and Methods

Algorithm implementation

The widely accepted Boyer–Moore string search algorithm (Boyer and Moore, 1977) was adopted here for pattern searching in a Perl program 'HREFinder'. The pattern based algorithm is case-sensitive, hence all input sequences converted first in the upper-case string to ensure a proper match in appropriate pattern finding.

An exact match of the pattern (conserved) and approximate match with fixed number of variations between them (consensus) were applied in this algorithm. This program uses the 5'-RCGTG-3' motif to mine all canonical HREs in the query sequence and produced the result in table 1. Further, it uses the approaches of the Kimura's (2001) and Kajimura's (2006) for the identification of the functional HRE with an adjacent HAS motif, 5'-CA(G|C)(A|G)(T|G|C)-3', in the space of 7-15 nt of the downstream of HRE. Additionally, this program predicts all the possible canonical functional HREs with the presence of HAS consensus motif in its downstream. After a number of successful exercises, HREFinder was found is a useful tool for identification of HRE and functional HRE sites in the upstream query sequence. The download folder contains (i) a main program 'HREFinder. pl', (ii) two sub-directories 'Test Data Set' and 'Results Test Data Set' (iii) a 'readme' file, which describes the working of the program.

Preparation of test data

A test analysis of the 'HREFinder' was initialized by preparing the query dataset. In this process, a collection of 5000 nt (-4900 to +100) long upstream sequence of some reported HIG such as EPO, VEGF, IGFBP1, LdhB, HO-1, ALDOa and ENO1 from different species, like human, cattle, mouse, zebrafish, Atlantic salmon and takifugu, were used for analysis. The information on these selected genes, like gene id, gene name, gene symbol, gene location on genomic sequence, orientation, chromosome numbers and genomic sequences, was downloaded from NCBI (NCBI Resource Coordinators, 2017). A summary of the prepared dataset used for the test analysis is presented in the supplementary table 1. The 5000 nt long upstream sequences of each gene were parsed using the defined gene location on its respective genomic DNA with the help of an in-house developed Perl program 'upstream parser'. The upstream sequences for the genes defined on the negative orientation were parsed using reverse complement methods. The full methodology of parsing upstream from genomic sequences and mining HRE is presented in figure 1. Two types of example data files were prepared in FASTA format of original upstream, *i.e.* (i) a single file of all sequences, and (ii) files for the individual sequence set of the homologous genes. The other input and their corresponding output files of example dataset are stored in respective sub-directories of the tool.

Results and Discussion

This program takes an input file of sequences in FASTA format and simultaneously identifies the canonical HRE on both the strands of the query sequences and generates three output files: (i) a table of all the identified HREs, (ii) list of all possible functional HRE, and (iii) HREs distribution and their frequencies in the set of 500 nt long consecutive fragments covering entire length of input sequences. Figure 2 illustrates the output produced by an example query of the upstream sequences set for gene VEGF. It presents canonical HRE frequencies in each 500 nt long successive sequences covering the entire length (Figure 2-A) and predicted functional HRE (Figure 2-B). Similarly, example queries and results files for other gene sets given in table 1, are available in the respective sub-directories of online available tool. This tool is useful to the scientific community for undertaking genome-wide mining of HRE and functional HRE sites in the upstream sequences of vertebrates.

Fragments of 3-5kb upstream from TSS including few hundred bases of downstream was used for the analysis of the upstream region to detect promoter and TFBs of a gene (Greenland *et al.*, 2015; Ortiz-Barahona *et al.*, 2010). In this process, a collection of 5000 nt (-4900 to +100) long upstream sequence of genes were used to prepare our test data set. The pattern based search algorithm was implemented in a web application (Rashid *et al.*, 2017) to analyze fishes data for functional HRE prediction and found appropriate output. Here, the algorithm is implemented in Perl program with extended analytical activity and available in form of an stand-alone downloading tool for analyzing of the genomic data of all vertebrates with a wider scope of the utilities. The HREFinder is applicable for genome-wide mining of hypoxia binding sites and predicts functional HRE in the varieties of vertebartes.

There are several tools and packages for TFBs prediction are available on the web. These tools rely upon a set of aligned promoter sequences to calculate position weight matrices (PWMs) for the identification of the conserved pattern in the no coding region of genes.

Table.1 List of the different genes which upstream sequences were taken to prepare test dataset for analysis

No	Animal name	GeneID	Symbol	chromosome	Genomic nucleotide accession	Start position on the genomic accession	End position on the genomic accession	orientation
1.	*Bos taurus*	509566	ALDOA	25	AC_000182.1	26470486	26483394	minus
2.	*Bos taurus*	281141	ENO1	16	AC_000173.1	45401684	45416470	plus
3.	*Bos taurus*	280784	EPO	25	AC_000182.1	36409504	36411892	Minus
4.	*Bos taurus*	513221	HMOX1	5	AC_000162.1	73980776	73987841	Plus
5.	*Bos taurus*	282259	IGFBP1	4	AC_000161.1	76720425	76725309	Minus
6.	*Bos taurus*	281275	LDHB	5	AC_000162.1	88962610	88981219	Plus
7.	*Bos taurus*	281572	VEGFA	23	AC_000180.1	17255515	17270515	Plus
8.	*Danio rerio*	336425	aldoaa	3	NC_007114.6	39425126	39435194	Plus
9.	*Danio rerio*	334116	eno1a	23	NC_007134.6	22729939	22751294	Plus
10.	*Danio rerio*	100004455	epoa	7	NC_007118.6	21624135	21652055	Minus
11.	*Danio rerio*	791518	hmox1a	3	NC_007114.6	25880611	25886721	Minus
12.	*Danio rerio*	317638	igfbp1a	20	NC_007131.6	6818827	6822654	Minus
13.	*Danio rerio*	30497	ldhba	4	NC_007115.6	16840135	16846928	Minus
14.	*Danio rerio*	30682	vegfaa	16	NC_007127.6	4459314	4488660	Minus
15.	*Homo sapiens*	226	ALDOA	16	NC_000016.10	30053090	30070420	Plus
16.	*Homo sapiens*	2023	ENO1	1	NC_000001.11	8861000	8879092	Minus
17.	*Homo sapiens*	2056	EPO	7	NC_000007.14	100720800	100723700	Plus
18.	*Homo sapiens*	3162	HMOX1	22	NC_000022.11	35381067	35394214	Plus
19.	*Homo sapiens*	3484	IGFBP1	7	NC_000007.14	45888360	45893668	Plus
20.	*Homo sapiens*	3945	LDHB	12	NC_000012.12	21635342	21657971	Minus
21.	*Homo sapiens*	7422	VEGFA	6	NC_000006.12	43770209	43786487	Plus
22.	*Mus musculus*	11674	Aldoa	7	NC_000073.6	126795234	126809510	Minus
23.	*Mus musculus*	13806	Eno1	4	NC_000070.6	150236726	150248873	Plus
24.	*Mus musculus*	13856	Epo	5	NC_000071.6	137483020	137485816	Minus
25.	*Mus musculus*	15368	Hmox1	8	NC_000074.6	75093618	75100593	Plus
26.	*Mus musculus*	16006	Igfbp1	11	NC_000077.6	7197787	7202546	Plus
27.	*Mus musculus*	16832	Ldhb	6	NC_000072.6	142490249	142507957	Minus
28.	*Mus musculus*	22339	Vegfa	17	NC_000083.6	46016993	46032377	Minus
29.	*Salmo salar*	100380798	aldoa	ssa03	NC_027302.1	51093248	51105749	Minus
30.	*Salmo salar*	106608895	epo	ssa07	NC_027306.1	19515329	19529184	Plus
31.	*Salmo salar*	100195468	ldhb	ssa17	NC_027316.1	34597191	34608811	Minus
32.	*Salmo salar*	106584182	LOC106584182	ssa02	NC_027301.1	35143675	35159397	Minus
33.	*Takifugu rubripes*	101071185	aldoa	1	NC_018890.1	8770282	8773425	Plus
34.	*Takifugu rubripes*	101069815	eno1	3	NC_018892.1	5889540	5895914	Minus
35.	*Takifugu rubripes*	101070079	hmox1	17	NC_018906.1	11058878	11060574	Minus
36.	*Takifugu rubripes*	101074885	igfbp1	22	NC_018911.1	14120	17689	Minus
37.	*Takifugu rubripes*	101079766	ldhb	18	NC_018907.1	242417	245457	Plus
38.	*Takifugu rubripes*	101073237	vegfa	16	NC_018905.1	5097117	5108133	Plus

Fig.1 Parsing upstream sequences of genes in different orientation (A) parsing upstream from genomic sequences for the gene localized on positive strand (B) parsing upstream from genomic sequences for the gene localized on negative strand and (C) Mining HRE from both strands of the upstream sequence

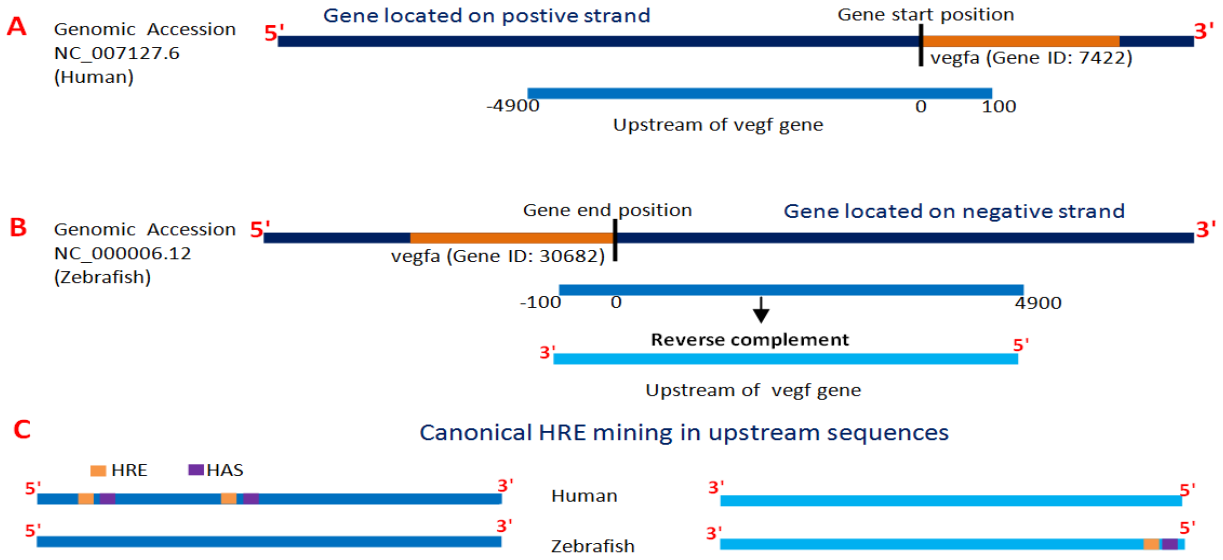
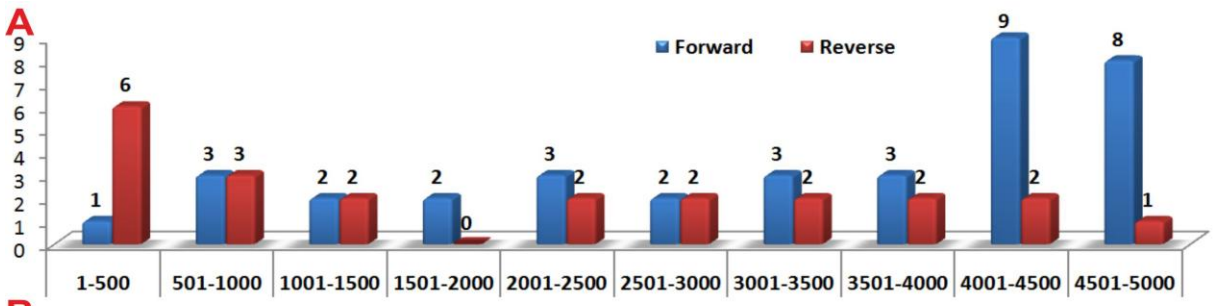


Fig.2 Output of HREFinder program for upstream VEGF gene sequences: (A) HRE frequencies in each 500 nt long consecutive fragments, and (B) predicted functional HREs



B

S.No.	Query definition	Location	HRE sequence	HRE (A|G)CGTG)	HAS (CA(C|G)(A|G)(T|G|C))
Functional HRE in original strand					
1	>VEGFA, *Bos taurus*	1183..1213	ACGTGGGATCTTAGTTCCCAGAGCAGGGAT	ACGTG	CAGAG
2	>VEGFA, *Bos taurus*	3930..3960	ACGTGGGCTCCAACAGGTCCTCTTCTCTCT	ACGTG	CAGGT
3	>VEGFA, *Bos taurus*	4193..4223	ACGTGGGACCACTCGAGTCCCAGACGTGCC	ACGTG	CAGAC
4	>VEGFA, *Homo sapiens*	3925..3955	ACGTGGGCTCCAACAGGTCCTCTTCCCTCC	ACGTG	CAGGT
5	>VEGFA, *Homo sapiens*	4635..4665	GCGTGTCTCTGGACAGAGTTTCCGGGGGCG	GCGTG	CAGAG
6	>Vegfa, *Mus musculus*	3977..4007	ACGTGGGTTTCCACAGGTCGTCTCACTCCC	ACGTG	CAGGT
7	>vegfa, *Takifugu rubripes*	3331..3361	ACGTGCAGGATTGTGCAGATGCAGGGGAGT	ACGTG	CAGAT
8	>vegfa, *Takifugu rubripes*	4570..4600	ACGTGTCCGGCTGCACACGCTTTTATTATA	ACGTG	CACGC
9	>vegfaa, *Danio rerio*	4966..4996	GCGTGTTAACATCCACGGCAGTACGAACCA	GCGTG	CACGG
Functional HRE in reverse strand					
1	>Vegfa, *Mus musculus*	726..756	GCGTGTTTCTGACATACACACACACCCAAA	GCGTG	CACAC
2	>vegfa, *Takifugu rubripes*	410..440	GCGTGTGCAGCCGGACACGTATACCAACCG	GCGTG	CACGT

Further, many tools like Match (Kel *et al.*, 2003) and ConTra (Hooghe *et al.*, 2008) use PWMs library collected in the TRANSFAC (Wingender *et al.*, 2001) and JASPAR (Bryne *et al.*, 2008) databases for the identification of the TFBs in the query sequence. These databases do not provide availability nearly 80% of the scoring matrices of TFBs in the public release. On the other hand, several reports available on the hypoxia binding site and functional HRE in the upstream of hypoxia inducible genes in the vertebrates. Using this information in the pattern matching algorithm, HREFinder tool was developed specifically for functional HRE findings within a sequence or set of sequences, this tool also applicable for searching the canonical HRE in the hypoxia responsive genes in the genome-wide survey of vertebrates. This feature with specificity for finding the functional HRE in the upstream sequence of gene differentiates it from most other software tools that provide conserved site across a set of sequences.

Conflict of interest

The authors declare no conflict of interest

Acknowledgment

Authors are thankful to the Director, ICAR-NBFGR, Lucknow, for providing necessary facilities to carry out this work and to the CABin, ICAR-IASRI, New Delhi, for financial assistance. We are also thankful to Mrs. Reeta Chaturvedi and Mr. Ravi Kumar, ICAR-NBFGR, Lucknow, for providing assistance in an image preparation.

References

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, 37: W202-8.

Benita, Y., Kikuchi, H., Smith, A.D., Zhang, M.Q., Chung, D.C., Xavier, R.J. 2009. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.*, 37: 4587-602.

Boyer, R.S., Moore, J.S. 1977. A fast string searching algorithm. *Communications of the ACM*, 20: 762-772.

Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., Sandelin, A. 2008 JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.*, 36: D102-106.

Fordel, E., Geuens, E., Dewilde, S., De Coen, W., Moens, L. 2004. Hypoxia/ischemia and the regulation of neuroglobin and cytoglobin expression. *IUBMB Life*, 56: 681-7.

Greenald, D., Jeyakani, J., Pelster, B., Sealy, I., Mathavan, S., van Eeden, F.J. 2015. Genome-wide mapping of Hif-1α binding sites in zebrafish. *BMC Genomics*, 16: 923.

Guillemin, K., Krasnow, M.A. 1997. The hypoxic response: huffing and HIFing. *Cell*, 89: 9-12.

Hooghe, B., Hulpiau, P., van Roy, F., De Bleser, P. 2008. ConTra: a promoter alignment analysis tool for identification of transcription factor binding sites across species. *Nucleic Acids Res.*, 36: W128-32.

Kajimura, S., Aida, K., Duan, C. 2006. Understanding hypoxia-induced gene expression in early development: in vitro and in vivo analysis of hypoxia-inducible factor 1-regulated zebra fish insulin-like growth factor binding protein 1 gene expression. *Mol Cell Biol.*, 26: 1142-55.

Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., Wingender, E. 2003. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, 31: 3576-9.

Kimura, H., Weisz, A., Ogura,,T, Hitomi, Y., Kurashima, Y., Hashimoto, K., D'Acquisto, F., Makuuchi, M., Esumi, H. 2001. Identification of hypoxia-inducible factor 1 ancillary sequence and its function in vascular endothelial growth factor gene induction by hypoxia and nitric oxide. *J. Biol. Chem.*, 276: 2292-8.

Korhonen, J., Martinmäki, P., Pizzi, C., Rastas, P., Ukkonen, E. 2009. MOODS: fast search for position weight matrix matches in DNA sequences. *Bioinformatics*, 25: 3181-2.

Liu, Y., Cox, S.R., Morita, T., Kourembanas, S. 1995. Hypoxia regulates vascular endothelial growth factor gene expression in endothelial cells. Identification of a 5' enhancer. *Circ. Res.*, 77: 638-43.

Manalo, D.J., Rowan, A., Lavoie, T., Natarajan, L., Kelly, B.D., Ye, S.Q., Garcia, J.G., Semenza, G.L. 2005. Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. *Blood*, 105: 659-69.

NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, 45: D12-D17.

Ortiz-Barahona, A., Villar, D., Pescador, N., Amigo, J., del Peso, L. 2010. Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and *in silico* binding site prediction. *Nucleic Acids Res.*, 38: 2332-45.

Rashid, I., Nagpure, N.S., Srivastava, P., Kumar, R., Pathak, A.K., Singh, M., Kushwaha, B. 2017. HRGFish: A database of hypoxia responsive genes in fishes. *Sci Rep.*, 7: 42346.

Semenza, G.L. 2010. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. *Oncogene*, 29: 625-34.

Smith, T.G., Brooks, J.T., Balanos, G.M., Lappin, T.R., Layton, D.M., Leedham, D.L., Liu, C., Maxwell, P.H., McMullin, M.F., McNamara, C.J., Percy, M.J., Pugh, C.W., Ratcliffe, P.J., Talbot, N.P., Treacy, M., Robbins, P.A. 2006. Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. *PLoS Med.*, 3(7): e290.

Van Rooijen, E., Santhakumar, K., Logister, I., Voest, E., Schulte-Merker, S., Giles, R., van Eeden, F. 2011. A zebrafish model for VHL and hypoxia signaling. *Methods Cell Biol.*, 105: 163-90.

Wang, G.L., Semenza, G.L. 1995. Purification and characterization of hypoxia-inducible factor. *J. Biol. Chem.*, 270: 1230-7.

Wenger, R.H., Stiehl, D.P., Camenisch, G. 2005. Integration of oxygen signaling at the consensus HRE. *Sci. STKE.*, 2005: re12. Review.

Wingender, E., Chen, X., Fricke, E., Geffers, R.,Hehl, R., Liebich, I., Krull, M., Matys, V., Michae, H., Ohnhäuser, R., Prüss, M., Schacherer, F., Thiele, S., Urbach, S. 2001. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, 29: 281-3.

How to cite this article:

Iliyas Rashid, Naresh Sahebrao Nagpure, Prachi Srivastava, Ajey Kumar Pathak, Ravindra Kumar, Vishwamitra Singh Baisvar, Mahender Singh and Basdeo Kushwaha. 2017. HRE Finder: A Tool for Quarrying Hypoxia-Response Element in Genomic Sequences. *Int.J.Curr.Microbiol.App.Sci.* 6(7): 1580-1586. doi: https://doi.org/10.20546/ijcmas.2017.607.190