Bulgarian Journal of Agricultural Science, 20 (No 3) 2014, 571-575
Agricultural Academy

GENETIC DIVERSITY AND DIFFERENTIATION OF HEAT SHOCK TRANSCRIPTION FACTORS AMONG SPECIES

J. MA[1], X. LI[1, 2], R. ZHOU[1], L. LI[1], Y. WU[3], J. WANG[4] and Q. YANG[5]
[1] *Agricultural University of Hebei, College of Animal Science and Technology, Baoding 071001, China*
[2] *Hebei Normal University of Science and Technology, Qinhuangdao 066600, China*
[3] *Xingtai University, Department of Biochemistry, Xingtai 054001, China*
[4] *Animal Husbandry Station, Agriculture and Animal Husbandry Bureau of Tangshan City, Tangshan 063000, China*
[5] *Animal Disease Central of Xingtai City, Xingtai 054001, China*

Abstract

MA, J., X. LI, R. ZHOU, L. LI, Y. WU, J. WANG and Q. YANG, 2014. Genetic diversity and differentiation of heat shock transcription factors among species. *Bulg. J. Agric. Sci*., 20: 571-575

The genetic diversity and differentiation of *Hsf*1, *Hsf*2 and *Hsf*4 among different species were analyzed based on their nucleotide and amino acid sequences. The *Hsf*4 had more polymorphic sites and total number of mutations in unit length, single polymorphic sites, nucleotide diversity and haplotype diversity than *Hsf*1and *Hsf*2. There were significant differences of amino acid diversity among three peptides. The HSF4 had more amino acid diversity than HSF1 and HSF2. The genetic differentiation among three genes was significant and the *Hsf*4 gene was more obvious than *Hsf*1 and *Hsf*2 gene. The positive selection affected on *Hsf*1, while the purifying selection on *Hsf*2 and *Hsf*4. The *Hsf*2 had different evolution relationship for *Sus scrofa*, *Bos taurus* and *Cricetulus griseus*.

*Key words: Hsf*1, *Hsf*2, *Hsf*4, diversity and differentiation, positive selection

Introduction

Cells and tissues can generate a sequence of emergency reactions and activities when organisms respond to circumstances threatening. Heat shock proteins (HSPs) willaccumulate in adverse situation, which can help proteins related assembling, folding, intracellular transport and degradation and preventing organisms from adverse situation as molecule chaperones (Hartlet et al., 2002). In this case, HSFs play important roles in stress resistance, development and longevity. The HSFs are significant member in signal transduction reaction of regulating activity of some relevant genes including *HSPs* gene (Morimoto, 1998). At present, some researches have verified that HSFs family included four different types in animal cells, HSF1, HSF2, HSF3 and HSF4 (Nakia et al., 1997). The HSF1 is activated by stress and existed organisms extensively. The action mechanism of HSF1 was researched in a great number of organizations. The studies of HSF2, HSF3 and HSF4 were less than HSF1. The HSF2 lacks intrinsic stress responses, and the mechanism is not known (Sandqvist et al., 2009). But some researches have demonstrated that the HSF2 can tolerate thermal stimulus, which participates in the regulation and control of embryonic development. So the HSF2 is deemed to be sensitive for the signal transduction of organism growth, development and differentiation (Liu et al., 1997; Tanabe et al., 1998). Among all of the HSFs in vertebrates, the HSF3 has species specificity, and only exists in avian species (Morimoto, 1998). So the comprehensive bioinformatics analysis of coding regions of *Hsf*1*, Hsf*2 and *Hsf*4 gene, as well as HSF1, HSF2 and HSF4 peptides were carried out in order to investigate their genetic diversity and differentiation among different species.

Materials and Methods

Source of sequences

A total number of 75 sequences with complete coding regions from 8 species were got from GenBank of NCBI (http://www.ncbi.nlm.nih.gov/), which were *Homo sapiens*,

E-mail: lixianglongcn@yahoo.com; majianqing2006@126.com

Mus musculus, *Cricetulus griseus*, *Rattus norvegicus*, *Pongo abelii*, *Sus scrofa*, *Pan troglodytes* and *Bos Taurus* (Table 1).

Analysis methods

All of the sequences were aligned first using Clustal W program in BioEdit (version 7.0.5). The complete coding region of *Hsf*1 (1125bp), *Hsf*2 (1557bp) and *Hsf*4 (1079bp) was selected for analysis. Then the DnaSP software (version 5.10) and SPSS statistical software were used to analyze the genetic diversity, including polymorphic site (S), number of haplotypes (H), haplotype diversity (Hd), nucleotide diversity (Pi), average number of nucleotide substitutions per site between populations (Dxy), average net nucleotide substitutions per site between populations (Da), total number of mutations (Eta) , singleton variable sites (SP), number of synonymous substitutions sites (SS) and nosynonymous substitutions sites (NSS) for *Hsf*1, *Hsf*2 and *Hsf*4 gene. The polymorphic sites and the total number of mutations in unit length were obtained by division from the value of polymorphic site (S) and total number of mutations (Eta) with the length of sequences. Ratios of non-synonymous substitution (Ka) and synonymous substitution (Ks) and mean genetic distance of different genes and were computed using MEGA5.0 software (Nei-Gojobori method (Jukes-Cantor) model). The method of *Z* statistical test was used to verify Ka and Ks. The phylogenetic tree among species based on Dxy was constructed using unweighted pair group method with arithmetic mean (UPGMA) in MEGA5.0. The amino acid diversity was analyzed using SPSS statistical software, including the number of amino acid (N'), molecular weight (W), theoretical PI (PI), instability index (II), aliphatic index (AI) , grand average of hydropathicity (GRAVY).

Results

Nucleotide diversity of *Hsf*1, *Hsf*2 and *Hsf*4 gene

The polymorphic sites, haplotypes and nucleotide diversity of *Hsf*1, *Hsf*2 and *Hsf*4 gene among different species were

Table 1
Sequences of *Hsf*1, *Hsf*2 and *Hsf*4 in different species

Species	Gene	Sample size	Accession number
Homo sapiens	*Hsf1*	10	AB463721.1, AK290975.1, NM_005526.2, DQ893859.2, BC014638.1, BT007351.1, AY8902699.1, AY890268.1, M64673.1, AK222497.1
	Hsf2	13	AK316377.1,NM_001135564.1, NM_004506.3, AK294624.1, EU446897.1, BC128420.1, BC121051.1, BC121050.1, BC112323.1, M65217.1, AB463722.1, DQ492684.1, NG_029607.1,
	Hsf4	8	NM_001040667.2,AB029348.1,AB463724.1, NM_001538.3,BC153061.1,BC146446.1,D87673.1, NG_009294.1
Mus musculus	*Hsf1*	4	NM_008296.2, BC013716.1, X61753.1, BC094064.1
	Hsf2	4	NM_008297.3, BC018414.1, X61754.1, AK052270.1
	Hsf4	8	NM_001256042.1, BC138130.1, AB029349.1, AF160966.1, NM_001256044.1, NM_011939.3, AB029350.1, AF160965.1
Cricetulus griseus	*Hsf1*	3	XM_003507662.1, XM_003507663.1, XM_003507664.1
	Hsf2	1	XM_003504214.1
	Hsf4	2	XM_003508539.1, XM_003508538.1
Rattus norvegicus	*Hsf1*	2	NM_024393.1, X83094.1
	Hsf2	2	NM_031694.2, AF172640.1
	Hsf4	1	NM_001106177.1
Pongo abelii	*Hsf1*	1	XM_002819555.1
	Hsf2	2	XM_002817309.1, XM_002817308.1
	Hsf4	2	XM_002826521.1, XM_002826522.1
Sus scrofa	*Hsf1*	1	NM_001243819.1
	Hsf2	1	XM_003121229.3
	Hsf4	2	XM_003126939.2, XM_003126940.2
Pan troglodytes	*Hsf1*	1	AK306589.1
	Hsf2	1	XM_003311424.1
	Hsf4	2	XM_001161258.2, XM_001161177.2
Bos taurus	*Hsf1*	2	NM_001076809.1, FJ358599.1
	Hsf2	1	NM_001083405.1
	Hsf4	1	NM_001191202.1

shown in Table 2. More nucleotide diversity of *Hsf*4 gene was found than that of *Hsf*1 and *Hsf*2.

Amino acid diversity of HSF1, HSF2 and HSF4 peptide

The amino acid diversity of HSF1, HSF2 and HSF4 peptide was shown in Table 3, including number of amino acid (N'), molecular weight (W), theoretical PI (PI), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) of HSF1, HSF2 and HSF4 peptide. It was obvious that there were significant differences of amino acid diversity among three peptides.

Characteristics of substitution of *Hsf*1, *Hsf*2 and *Hsf*4 gene

The synonymous substitution rate (Ks), non-synonymous substitution rate (Ka) and the ratios of Ka and Ks (ω) of *Hsf*1, *Hsf*2 and *Hsf*4 gene were listed in Table 4. The results of *Z* statistical test showed that there was significant difference between Ka and Ks for *Hsf*1 gene at 0.01, and for *Hsf*2 and *Hsf*4 gene at 0.05, indicating the authenticity of ratio of Ka and Ks (ω) of *Hsf*1, *Hsf*2 and *Hsf*4 gene.

Genetic differentiation of *Hsf*1, *Hsf*2 and *Hsf*4 among species

The number of nucleotide substitutions per site between populations (Dxy), the net nucleotide substitutions per site between populations (Da) and the mean genetic distance of *Hsf*1, *Hsf*2 and *Hsf*4 gene was shown in Table 4. We noticed that the *Hsf*4 gene had highest Dxy (0.143), Da (0.130) and mean genetic distance (0.177) among three genes.

Phylogenetic tree of *Hsf*1, *Hsf*2 and *Hsf*4 among species

The phylogenetic trees based on Dxy of *Hsf*1, *Hsf*2 and *Hsf*4 among species were shown in Figures 1, 2 and 3 respectively. We observed that the similar evolution relationship of *Hsf*1, *Hsf*2 and *Hsf*4 existing among most of species, but the difference of *Hsf*2 was found for *Sus scrofa*, *Bos taurus* and *Cricetulus griseus*.

Table 2
Polymorphic sites and DNA polymorphism in different genes

Gene	N	L (bp)	S (s / bp)	Eta (Eta / bp)	Pi	Hd
*Hsf*1	24	1125	0.263	0.312	0.094	0.848
*Hsf*2	25	1557	0.215	0.235	0.050	0.893
*Hsf*4	26	1079	0.646	0.764	0.135	0.914

N, number of sequences; L, length of sequence; S, number of polymorphic sites in unit length; Eta, total number of mutations in unit length; Pi, nucleotide diversity; Hd, haplotype diversity

Table 3
Amino acid diversity of HSF1, HSF2 and HSF4 peptide

Peptide	N'	W	PI	II	AI	GRAVY
HSF1	528.13 ± 28.73^{a}	57226.34 ± 2910.98^{a}	5.11 ± 0.14^{Aa}	58.82 ± 4.66^{Aa}	80.75 ± 2.14^{A}	0.39 ± 0.03^{Aa}
HSF2	545.75 ± 63.55^{Aa}	61185.10 ± 6578.87^{Aa}	4.78 ± 0.20^{Ab}	54.32 ± 2.33^{Ab}	87.00 ± 2.01^{B}	0.56 ± 0.03^{B}
HSF4	472.38 ± 32.46^{Bb}	51023.58 ± 3439.89^{Bb}	5.79 ± 0.43^{B}	65.71 ± 3.17^{B}	84.13 ± 0.80^{C}	0.39 ± 0.04^{Aa}

N', number of amino acid; W, molecular weight; PI, theoretical PI; II, instability index; AI, aliphatic index; GRAVY, grand average of hydropathicity. The superscripts with different lowercase and capital for corresponding index indicate significant difference at 0.05 and 0.01 level in the same column, respectively.

Table 4
Characteristics of substitution and genetic differentiation of *Hsf*1, *Hsf*2 and *Hsf*4

Gene	Ka	Ks	ω	Mean Genetic Distance	Average Dxy	Average Da
*Hsf*1	0.118^{A}	0.037^{B}	3.189	0.105	0.117 ± 0.051^{Aa}	0.116 ± 0.051^{Aa}
*Hsf*2	0.016^{a}	0.228^{b}	0.070	0.054	0.082 ± 0.038^{B}	0.081 ± 0.038^{B}
*Hsf*4	0.151^{a}	0.227^{b}	0.665	0.177	0.143 ± 0.050^{Ab}	0.130 ± 0.051^{Aa}

Dxy, The number of nucleotide substitutions per site between populations; Da, the net nucleotide substitutions per site between populations; ω, ratios of non-synonymous substitution (Ka) and synonymous substitution (Ks) (ω=Ka/Ks). The superscripts with different lowercase and capital for average Dxy and Da indicate significant difference at 0.05 and 0.01 in the same column, respectively. The superscripts with different lowercase and capital for Ka and Ks indicate significant differences at 0.05 and 0.01 in the same row, respectively.

Discussion

In this study, a total of 697 polymorphic sites of *Hsf*4 gene were found among different species, and only 296 and 334 for *Hsf*1and *Hsf*2. The number of mutations of *Hsf*1, *Hsf*2 and *Hsf*4 was 351, 365 and 824 respectively. The single polymorphic sites of *Hsf*4 gene (391) was obviously higher than that

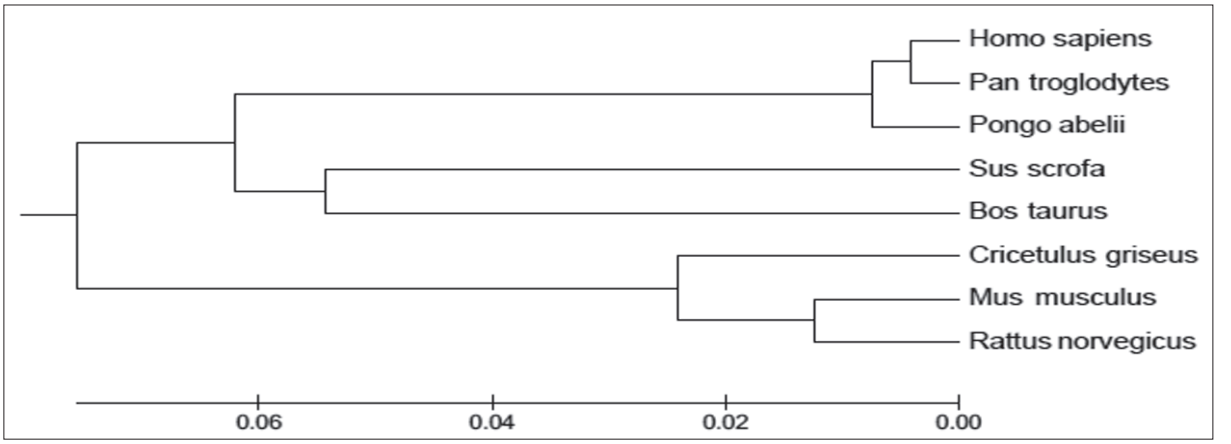

Fig. 1. Phylogenetic tree of *Hsf*1 among species

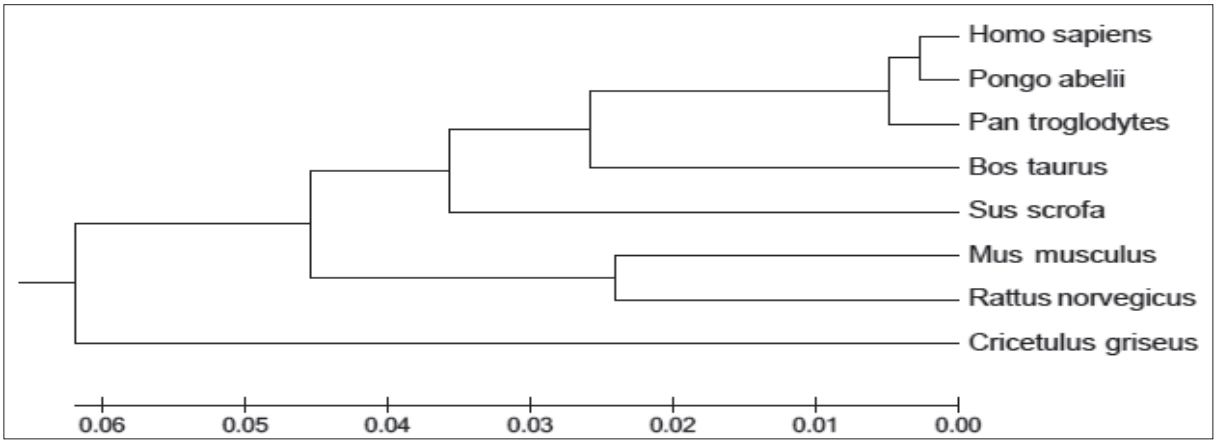

Fig. 2. Phylogenetic tree of *Hsf*2 among species

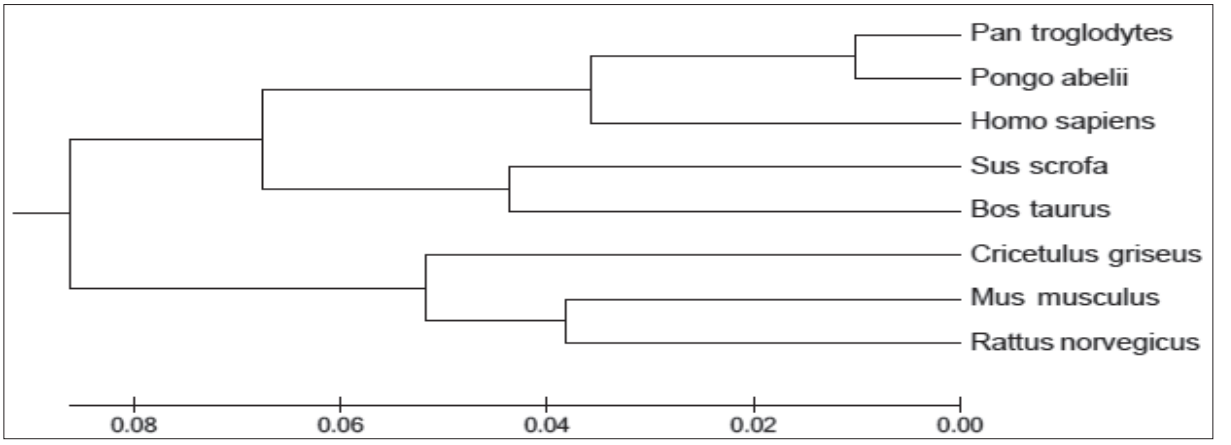

Fig. 3. Phylogenetic tree of *Hsf*4 among species

of *Hsf*1 (44) and *Hsf*2 (133). The *Hsf*4 gene also had higher nucleotide diversity (0.135), haplotype diversity (0.914), the polymorphic sites (0.646) and the total number of mutations (0.764) in unit length (Table 2).

For amino acid diversity, the significant differences of all indexes existed between HSF2 and HSF4 on the level of $P < 0.01$. The significant difference of number of amino acid and molecular weight between HSF1 and HSF4 was also observed ($P < 0.05$), as well as theoretical PI, instability index, aliphatic index and grand average of hydropathicity on the level of $P < 0.01$. The significant difference of theoretical PI and instability index on the level of $P < 0.05$, and aliphatic index and grand average of hydropathicity on the level of $P < 0.01$ between HSF1and HSF2 was also found, except for the number of amino acid and molecular weight. Therefore it could be concluded that the HSF4 had more amino acid diversity than that of the HSF1 and the HSF2.

Generally, it is important to calculate Ks, Ka and ω of genes in the process of molecular evolution. The synonymous substitutions are generally neutral in evolution while non-synonymous substitutions can occur under strong selective pressure. The ratio of Ka/ Ks (ω) was considered as a useful means for quantifying the impact of natural selection on molecular evolution (Kimura. 1983; Ohta. 1992). A value of $\omega = 1$ denotes a neutral mutation, ω less than 1 purifying selection which describes selection against new variants, while ω greater than 1 denotes positive selection (adaptive molecular evolution) in that non-synonymous mutations offer fitness advantages to the protein (Yang et al., 2000). In this paper, the value of ω was greater than 1 for *Hsf*1 gene, while less than 1for *Hsf*2 and *Hsf*4 gene. It revealed that the *Hsf*1 gene was affected by positive selection, while both *Hsf*2 and *Hsf*4 gene were affected by purifying selection. We also noticed that the ω value (0.665) of *Hsf*4 was obviously higher than that (0.070) of *Hsf*2 due to its higher Ka (0.151).

The significant differences of Dxy and Da existed between *Hsf*2 and *Hsf*4 gene on the level of $P < 0.01$. This case was also existed between *Hsf*1 and *Hsf*2 gene. The significant difference of Dxy between *Hsf*1 and *Hsf*4 gene was also found ($P < 0.05$). So the genetic differentiation among three genes was significant and the *Hsf*4 gene was more obvious than *Hsf*1 and *Hsf*2 gene.

And for the phylogenetic tree of *Hsf*1, *Hsf*2 and *Hsf*4 among species, we found the *Sus scrofa* and *Bos taurus* were in same cluster for *Hsf*1 and *Hsf*4 (Figure 1 and Figure 3), while they were separated each other for *Hsf*2 although representing close relationship (Figure 2). The same case of *Cricetulus griseus* was also found in Figure 2 for *Hsf*2. The *Cricetulus griseus* was clustered with *Mus musculus* and *Rattus norvegicus* together for *Hsf*1 and *Hsf*4 (Figure 1 and Figure 3), while it was separated from *Mus musculus* and *Rattus norvegicus* for *Hsf*2 (Figure 2).

Conclusions

In conclusion, it could be concluded that the *Hsf*4 gene had more nucleotide diversity and amino acid diversity than those of *Hsf*1 and *Hsf*2. There was significant genetic differentiation among *Hsf*1, *Hsf*2 and *Hsf*4, and the *Hsf*4 differentiated more obviously among three genes. During the long-term evolution, the *Hsf*1 gene was affected by positive selection, while both *Hsf*2 and *Hsf*4 gene were affected by purifying selection. The *Hsf*4 had more positive selection pressure than *Hsf*2. The *Hsf*1, *Hsf*2 and *Hsf*4 had similar evolution relationship among species, but the *Hsf*2 represented different relationship for *Sus scrofa*, *Bos taurus* and *Cricetulus griseus*.

Acknowledgements

We thank the helps from the fellow in the laboratory on the study. This work was supported by the National Natural Science Foundation of China (31172196), College Innovation Team Leader Training Programme of Hebei Province (Project Number, LJRC 004) and the Earmarked Fund for Hebei Layer Innovation Team of Modern Agro-industry Technology Research System.

References

Hartl, F. U. and M. Hayer-Hartl, 2002. *J. Sci.* **295**: 1852-1858.

Kimura, M., 1983. Cambridge University Press. The Neutral Theory of Molecular Evolution. p. 367. London. UK.

Liu, X. D., P. C. Liu, N. Santoro and D. J. Thiele, 1997. Conservation of Stress Response: Human Heat Shock Transcription Factors Functional Substitute for Yeast HSF. *EMBO J.*, **16**: 6466-6477.

Morimoto, R. I., 1998. Regulation of the Transcription Response: Cross Talk Between A Family of Heat Shock Factors. Molecular Chaperones and Negative Regulators. *Genes. Dev.*, **12**: 3788-3796.

Nakai, A., M. Tanabe, Y. Kawazoe, J. Inazawa, R. I. Morimoto and K. Nagata, 1997. HSF4, A New Member of the Human Heat Shock Factor Family Which Lacks Properties of a Transcriptional Activator. *J. Mol Cell Biol.*, **17** (1): 469-481.

Ohta, T., 1992. The Nearly Neutral Theory of Molecular Evolution. Annu. Rev. Ecol. Syst., **23**: 263-286.

Sandqvist, A., J. K. Björk, T. M. Akerfel, Z. Chitikova, A. Grichine, C. Vourc'h, C. Jolly, T. A. Salminen, Y. Nymalm and L. Sistonen, 2009. Heterotrimerization of Heat-Shock Factors 1 and 2 Provides a Transcriptional Switch in Response to Distinct Stimuli. *J. Mol Biol Cell.*, **20:**1340-1347.

Tanabe, M., Y. Kawazoe, S. Takeda, R. I. Morimoto, K. Nagata and A. Nakai, 1998. Disruption of the HSF3 Gene Results in the Severe Reduction of Heat Shock Gene Expression and Loss of the Motolorance. *EMBO J.*, **17**: 1750-1758.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *J. Genetics. May;* **155** (1): 431-449.

Received July, 2, 2013; accepted for printing February, 2, 2014.