Bioinformatics analysis of Growth Hormone1 Gene (GH1) in Several Species

Ayman Mahmoud Sabry[1,3] Manal. M. Said[2], Nabil. S. Awad[1] and Adel El-Trass[1]

[1]Biotechnology and Genetic Engineering Unit, Scientific Research Deanship, Taif University,
[2]Dept. of Biotechnology, Faculty of Science ,Taif University
[3]Cell Biology Department, National Research Center, Dokki, Giza, Egypt
amsabry@gmail.com

Abstract: Much attention has been focused on the study of Growth hormone 1 (GH1) due to its important role in immune function and bone turnover, in addition to its well-documented influences on stature, muscle mass, lipid and carbohydrate metabolism and postnatal growth. The aim of the present study is to investigate GH1 gene's evolution and differentiation within and among species. A total of 32 sequences with the complete CDS of the GH1 gene and the amino acid sequences belonging to 20 species were obtained from GenBank and analyzed where differentiation within and among species was also studied. The results showed that most of the species have the stop codon TAG with variation of TGA for *Gallusgallus, Meleagrisgallopavo* and *Monodel-phisdomestica* where *Xenopuslaevis* has TAA as stop codon. The length of GH1 gene with complete CDS varies greatly, from 369 to 654bp, due to deletion, insertion or stop codon mutation resulting in elongation. Observed genetic diversity was higher among species than within species. Differentiation of the GH1 gene was obvious among species, and the clustering result was consistent with the taxonomy in the National Center for Biotechnology Information.
[Ayman Mahmoud Sabry, Manal. M. Said, Nabil. S. Awad, and Adel El-Trass. **Bioinformatics analysis of Growth Hormone1 Gene (GH1) in Several Species.** *Life Sci J* 2013;10(2):1202-1206] (ISSN:1097-8135). http://www.lifesciencesite.com. 166

Keywords: GH1gene, Sequence analysis, UPGMA dendrogram

1. Introduction

Growth hormone (GH) plays an important role in immune function and bone turnover, in addition to its well-documented influences on stature, muscle mass, lipid and carbohydrate metabolism and postnatal growth (Kaplan,1999). The specificity of GH action lies in promoting the homodimerisation of its cell surface receptor (GHR), resulting in the induction of post-receptor signaling pathways (Waters *et al.*, 2006). Human GH synthesis is directed by the pituitary-expressed GH1 gene, which is located on chromosome 17q23 within a gene cluster that includes three paralogous placental expressed genes (CSH1, CSH2 and GH2). The control of GH1 gene expression is regulated by the pituitary-expressed transcription factor, PIT1, which drives GH expression by binding not only to the GH1 proximal promoter, but also to a locus control region (LCR) located between 14.5 kilobases (kb) and 32kb upstream of the GH1 gene (Procter *et al.*, 1998).

Members growth hormone family are pituitary growth hormone gene (GH1), placental growth hormone gene (GH2), two chorionic somatomammotropin genes (CSH1 and CSH2) and a chorionic somatomammotropin-like gene (CSHL1) (Chen *et al.*, 1989). Evolutionary geneticists (e.g. (Chen *et al.*, 1989; Krawczak *et al.*,1999)) believe that this gene family has evolved through three successive duplications. The premier duplication resulted in to a pre-GH and a pre-CSH gene. These two genes were then duplicated to produce to the GH1, CSH1, GH2 and CSH2 genes. Eventually, a CSH1 du- placation gave rise to CSHL1. These six genes are present in chimpanzee and rhesus monkey. The six genes were likely present before the divergence of great ape and Cercopithecidae, about 30 million years ago and the human lineage lost one (Chen *et al.*, 1989; Krawczak *et al.*, 1999; Revol DeMendoza *et al.*, 2004; González Alvarez *et al.*, 2006; Steiper & Young, 2006). Given evolutionary history of these genes, it could be assumed that their nucleotide sequences would be quite different from each other. However, all five human growth hormone genes, and their flanking regions, share from 91 to 99% nucleotide identity (Chen *etal.*,1989). This unexpectedly high degree of sequences similarity in a large part due to gene conversions between these genes.

GH1 gene offers some prospects for evolutionary studies as this gene is expressed in different tissues and is believed to have different functions. Such that it could be expected that conservation between the growth hormone gene family could alter their function or expression or both. Moreover, due to the growth variation among and within species as well as conserved genetic mechanism of growth in different species.

In the present study, 32 complete CDS (coding

sequence) of GH1 from 20 species were studied to investigate its evolution and differentiation within and among species.

2. Material and Methods

A total of 32 sequences with the complete CDS of the GH1 gene belonging to 20 species were obtained from GenBank (Table1). All the sequences were aligned using the CLUSTAL-W program implemented in BioEdit (version7.0.5) (Hall, 1999). DnaSP (version5.10.01) software was used to analyze the haplotype diversity (Hd), the average number of nucleotide differences, (Tajima, 1983), the nucleotide diversity (π), synonymous nucleotide diversity (π_S), nonsynonymous nucleotide diversity (π_a) with the Jukes and Cant or correction, the polymorphic site(s), the singleton variable sites (SP), and the parsimony informative sites (PIP) for each species, and the average number of nucleotide substitutions per site between species (D_{xy}) (Lynch & Crease,1990). The phylogenetic tree among species was constructed using the unweighted pair group method with the arithmetic mean (UPGMA) implemented in Mega 5 software (Tamura *et al.*, 2007).

3. Results and Discussion

Sequence length Variation among species

The length of GH1 gene with complete CDS varies greatly among species, ranging from 369 to 654bp (Table1). Species with close relationships according to the taxonomy in NCBI have similar length of GH1 gene CDS. *Cyprinuscarpi, Daniorerio, Oncorhynchusmykiss* and *Salmogairdneri*, for example. The longest length was 654bp which was observed for *HomoSapienes, Bostaurus, Ovisaries, Macacamulatta, Pan Troglodytes* and *Papiohamadryas*. The shortest length (627bp) was found in *Xenopuslaevis*, this sequence is 27bp shorter than the longest length reported in this study, this shortening in length might be as result in mutation in the stop codon observed in *Xenopuslaevis*. Similar observation was also reported by Kang *et al.* (2008) in lactoferrin gene.

Table1:GH1 gene sequences of 15 species

Species	N	Length (bp)	Accession Number
Homosapiens	6	654	NM000515.3,AY613431.1,AY613432.1 BC075012.2,BC075013.2,BC090045.1
Homosapiens	2	609	BC062475.1,NM022559.2
Homosapiens	1	534	NM022560.2
Homosapiens	1	369	NM022561.2
Bostaurus	1	654	NM180996.1
Ovisaries	1	654	NM001009315.3
Macacamulatta	1	654	NM001047155.1
Pantroglodytes	1	654	NM001197164.1
Papiohamadryas	1	654	HM103904.1
Rattusnorvegicus	2	651	NM001034848.2,BC166872.1
Musmusculus	1	651	NM008117.2
Canislupus	1	651	NM001003168.1
Equuscaballus	1	651	NM001081948.1
Caviaporcellus	1	651	NM001172859.1
Feliscatus	1	651	NM001009337.1
Susscrofa	1	651	NM213869.1
Gallusgallus	1	651	NM204359.1
Meleagrisgallopavo	1	649	XM003213025.1
Monodelphisdomestica	1	648	NM001032993.1
Cyprinuscarpi	1	633	AJ640135.1
Daniorerio	2	633	NM001020492.2,BC116501.1
Oncorhynchusmykiss	1	633	NM001124689.1
Salmogairdneri	1	633	M22731.1
Xenopuslaevis	1	627	NM001085615.1

Sequence length Variation within species

Within species length variation was only observed within Homosapiens species, where 11 sequences were studied (Table1). These sequences comprised 6 sequences of 654bp long, where only 2 nucleotide transitional substitutions were observed. One of which was for sequence AY613431.1 at position 116 (T replaced C), this sequence was associated with GH deficiency in heterozygous state. The other nucleotide substitution was for sequence AY613432.1 at position nucleotides 545 (A replaced with G), this sequence was associated with

idiopathic short stature. Both sequences BC062475.1 and NM0 22559.2 have nucleotide length of 609bp where this shortage resulted from deletion of nucleotide sequences between 172 and 216GAAGAAGCC-TATATCCCAAAGGAACAGAAGTATTCATTCCTG CAG. Sequence NM022560.2 has only 534bp where this sequence is missing exon 3 generating an isoform (3), which has an internal deletion relative to the predominant 22-kDa isoform (1) Guevara-Aguirre *et al.* (2012). Sequence NM022561.2 variant (4) is missing exons 3 and 4 generating an isoform (4) which has an internal deletion relative to the predominant 22-kDa isoform(1).

Variation of stop Codon

Two kinds of stop codon mutation were found in the GH1 gene among species, where no mutations were found within different species, (Table 2). Most species use TAG as stop codon for GH1 gene. However, mutant stop codons were found in Gallusgallus, Meleagrisgallopavo and Monodelphisdomestica where TGA was as stop codon, but Xenopuslaev is uses TAA. The presence of these mutations might gave rise to stop codon usage bias among species for different genes (Ghosh, 2000; Higgs & Ran, 2008).

It could be inferred that this kind of mutation might be related to the differentiation of species. Similar conclusion was also reported by Kang *et al.*(2008) on lactoferrin gene. Moreover, it also could be inferred that stop codon mutation might be responsible about sequence length variation.

Table2:Variationin stop Codon

Species	Acc.No.	Length(bp)	Stop Codon
Gallusgallus	NM204359.1	651	TGA
Meleagrisgallopavo	XM003213025.1	649	TGA
Monodelphisdomestica	NM001032993.1	648	TGA
Xenopuslaevis	NM001085615.1	627	TAA

Polymorphism and Genetic Diversity within and among species

The alignment of 32 sequences with region of 669bp and containing gaps was carried out using Bio Edit. The results of the DnaSp analysis indicated that the selected region (1–669) of the 32 sequences from different species have 263 sites excluding sites with gaps (406). There are 64 invariable sites and 199 variable sites include 21-singletone variable site and 178 parsimony informative sites. The nucleotide diversity (π=0.24806) and the average number of nucleotide differences (K=65.244) for all sequences are higher than the was found for *Homosapiens* (π=0.00054 and k=0.2). Only *Homosapiens, Ratnorvegicus* and *Daniorerio* species provided informative data for within species analysis (i.e. NCBI has more than one sequences for each species). However *Ratnorvegicus* and *Daniorerio* have only 2 sequences. Polymorphic information and haplotype diversity of GH1 gene for informative species are presented in table 3.

Table3: Genetic diversity of the GH1 gene Species[a] Diversity parameter[b]

	h	H_d	K	π	π_S	π_a	S	SP	PIP
Homosapiens	2	0.2	0.2	0.00054	0.00054	0.0	1	1	0
Ratnorvegicus	1	0.0	0.0	0.000	0.0	0.0	0	0	0
Daniorerio	1	0.0	0.0	0.000	0.0	0.0	0	0	0

[a] only HomoSapienes has effective data

[b] h, Number of haplotypes; H_d, haplotype diversity; K, average number of nucleotide differences; π, nucleotide diversity; π_S, synonymous nucleotide diversity; π_a non synonymous nucleotide diversity; S, number of polymorphic sites; SP, singleton variable sites; PIP, parsimony informative sites.

DNA Divergence and Phylogenetic Analysis

Phylogenetic analysis can be used to detect gene conservation events because conservation between paralogous genes (reveal history of gene family) will often cause them to group together rather than with their orthologous genes members in other related species (Drouin *et al.*,1999; Graur & Li, 2000).

In this study we built the phylogeny tree based on sequence data. The evolutionary history was inferred using the UPGMA method (Sneath &R.R., 1973). The optimal tree (Fig.1) with the sum of branch length=2.01983401 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (300replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*,

2004) and are in the units of the number of base substitutions per site. The divergence time among different species was also labeled in million years ago (MYA) on the scale par. The analysis involved 20 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 515 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

This tree contains two main groups, one composed of marines species (i.e. rainbow trout, zebra fish and common carp), where the second group composed of the rest of studied species (i.e. mammals, avian, and amphibians).

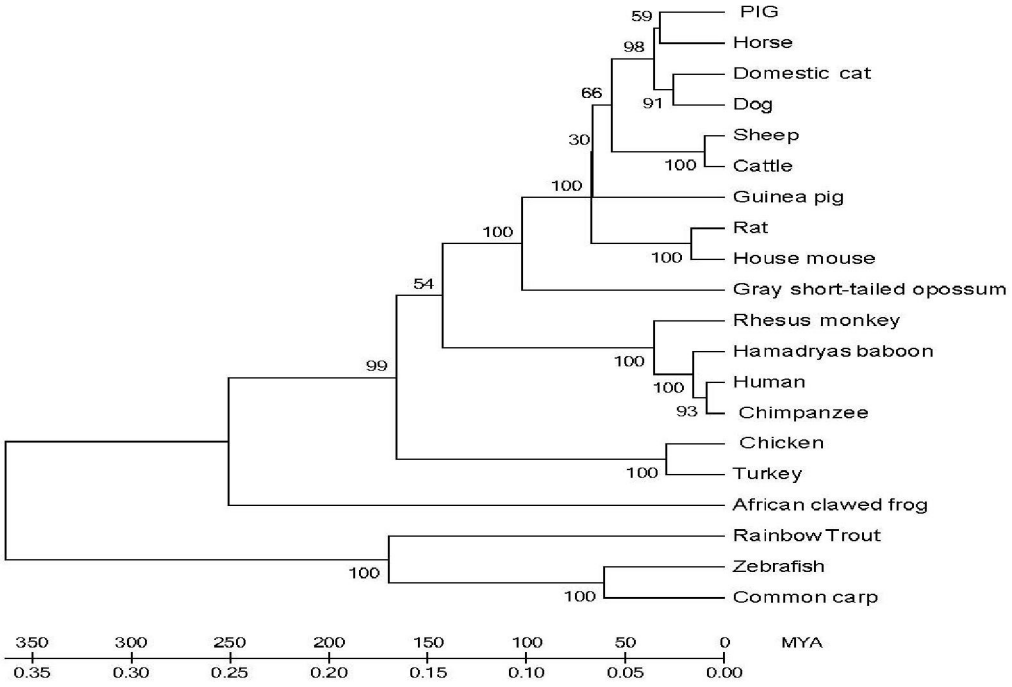

Figure 1: phylogeny tree of the gh1 gene among. the scale represents 1% sequence divergence and the numbers at the nudes are the percent bootstrap support.

The topology of the phylogenetic tree showed that zebra fish and common carp GH1 gene sequences are more conservative compared to rainbow trout. The divergence time between the later and former 2 species was about 175MYA. These species were grouped with 100% bootstrap.

In the second section of the phylogenetic tree the African clawed frog grouped alone and separated from the rest of species by long branch. Based on this phylogenetic analysis African clawed frog has evolutionary time of 250 MYA. Chicken and Turkey were grouped together with short branches and 100% bootstrap indicates the higher at of conservation of GH1 gene in these avian species. The same topology was also observed for sheep and cattle species, Norwegian rat and House mouse species domestic cat and Dog species, and Pig and horse species. The bootstrap ranged for ranged from 91–100% except for Pig and Horse species the boots trap was only 59%. The evolutionary time ranged between 15–25 MYA for all these species.

Rodent species were clustered together in one eudicots where Guinea pig and gray short-tailed opossum species were separated by long branches from both Norwegian rat and house mouse. The divergence time between Guinea pig and gray short-tailed opossum species were about 100MYA.

The homindidae species were also clustered together in one eudicots, where Baboon, Human and Chimpanzee were separated by shorter branch compared to Rhesus monkey. This result is in agreement with the results of Petronella &Drouin (2011). Differentiation of the LF gene was obvious among species, and the clustering result was consistent with the taxonomy in the National Center for Biotechnology Information.

References

1. Chen,E.Y., Liao,Y.C., Smith,D.H., Barrera-Saldaña, H.A., Gelinas,R.E., &

Seeburg, P.H.1989. The human growth hormone locus: nucleotide sequence, biology, and evolution.*Genomics*,**4**,479–497.
2. Drouin,G.,Prat,F.Ell,M.,&Clarke,G.D.P.1999 .Detecting and characterizing gene conversions between multigene family members. *Mol. Biol.Evol*,**16**,1369–1390.
3. Felsenstein, J.1985.Confidence limits on phylogenies: Anapproachusing the bootstrap. *Evolution*, **39**,783–791.
4. Ghosh,T.2000. Studies on codon usage in Entamoeba histolytica. *Int. J Parasitol*, **30**,715–722.
5. González Alvarez, R,Revolde Mendoza, A., Esquivel Es-cobedo, D., Corrales Félix, G., Rodríguez Sánchez, I,González, V., Dávila,G.,Cao,Q., deJong,P.,& Fu, Y. X. BarreraSaldaña, H.A. 2006. Growth hormone locus expands and diverges after the separation of New and Old World Monkeys.*Gene*,**380**,38–45.
6. Graur, D., & Li,W.H.2000. *Fundamentalsofmolecularevolution*. 2edn. Sinauer Associates, Sunderland.
7. Guevara-Aguirre, J., Guevara-Aguirre, M., Hwa, V., Procel, P., Saavedra, J., Ostrer, H., Fang, P.and Rosenfeld, R.G., Kerns, S., & Rosenbloom, A.L.2012. Intrauterineand postnatal growth failure with normal GH/IGF1 axis and insulin-resistant diabetes in aconsanguineouskinship. *Eur. J. Endocrinol.*, **166**(3),521–529.
8. Hall,T.A.1999.BioEdit:auser-friendlybiologicalsequencealignment editor and analysis program for Windows 95/98/NT. *NucleicAcidsSym- posiumSeries*,**41**,95–98.
9. Higgs,P.G.,&Ran,W.2008.Coevolution of Codon Usage and tRNA Genes Leads to Alternative Stable States of Biased Codon Usage. *Mol. Biol.Evol.*,**25**(11),2279–2291.
10. Kang,J.F.,Li,X.L.,Zhou,R.Y.,Li,L.H.,Fen g,F.G., & Guo, X.L. 2008. Bioinformatics Analysis of Lactoferrin Gene for Several Species. *Biochem Genet*, **46**, 312–322.
11. Kaplan,S.L.1999. Hormonal regulation of growth and metabolic effects of growth hormonein: Handbook of *Physiology*. Vol.V. Oxford University Press, NY.
12. Krawczak,M.,Chuzhanova,N.A.,&Cooper,D. N.1999.Evolution of the proximal promoter region of the mammalian growth hormone gene. *Gene*, **237**,143–151.
13. Lynch,M.,&Crease,T.J.1990.The analysis of population survey data on DNA sequence variation. *Mol Biol Evol*,**7**,377–394.
14. Petronella,N.,&Drouin,G.2011.Geneconversion s in the growth hormone gene family of primates: Stronger homogenizing effects in the Hominidae lineage. *Genomics*, **98**,173–181.
15. Procter, A.M., Phillips, J.A.,3rd, & Cooper, D. N. 1998. The molecular genetics of growth hormone deficiency. *Hum.Genet.*, **103**, 255–272.
16. Revol De Mendoza, A., Esquivel Escobedo, D., MartínezDávila, I., & Barrera Saldaña, H. 2004. Expan- sion and divergence of the GHlocus between spider monkey and chimpanzee. *Gene*, **336**,185–193.
17. Sneath,P.H.A.,&R.R.,Sokal.1973.*NumericalT axonomy*. SanFran- cisco:Freeman.
18. Steiper, M. E., & Young, N. M. 2006. Primatemoleculardivergence dates. *Mol. Phylogenet.*, **41**,384–394.
19. Tajima,F.1983. Evolutionary relationship of DNA sequences infinite populations. *Genetics*, **105**, 437–460.
20. Tamura, K., Nei, M., & Kumar, S. 2004. Prospectsforinferringverylargephylogeniesbyus ingtheneighbor-joiningmethod. *Proceedings of the National Academy of Sciences (USA)*, **101**, 11030–11035.
21. Tamura,K., Dudley,J.,Nei,M.,& Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version4.0. *Molecular Biology Evolution*, **24**(8), 1596–1599.
22. Tamura,K., Peterson,D., Peterson,N., Stecher, G., Nei, M., & Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Like lihood, Evolutionary Distance, and Maximum Parsi-mony Methods. *Molecular Biologyand Evolution*, **28**(10),2731–2739.
23. Waters,M.J.,Hoang, H.N., Fairlie, D.P., Pelekanos, R.A., &Brown, R.J.2006.New insights into growth hormone action.*J.Mol. Endocrinol*, **36**,1–7.