CLONING AND SEQUENCE ANALYSIS OF THE GROWTH HORMONE GENE IN INDIAN ELEPHANTS

BHOSALE R. A.

Thesis submitted in partial fulfilment of the requirement for the degree of

Master of Veterinary Science

Faculty of Veterinary and Animal Sciences
Kerala Agricultural University, Thrissur

2009

Department of Animal Breeding, Genetics and Biostatistics
COLLEGE OF VETERINARY AND ANIMAL SCIENCES
MANNUTHY, THRISSUR-680651
KERALA, INDIA

Dedicated To

My Beloved family

And

Respected Teachers

DECLARATION

I hereby declare that this thesis, entitled **"Cloning and sequence analysis of the growth hormone gene in Indian elephants"** is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title, of any other University or Society.

Mannuthy
Date: 18/05/09

Bhosale R.A.

CERTIFICATE

Certified that this thesis, entitled **"Cloning and sequence analysis of the growth hormone gene in Indian elephants"** is a record of research work done independently by **Dr. Bhosale R.A.** under my guidance and supervision and it has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to him.

Mannuthy

Date: 18.05.2009

Dr. T.V. Aravindakshan
(Chairman, Advisory Committee)
Professor,
Centre for Advanced Studies in
Animal Genetics and Breeding,
College of Veterinary and Animal
Sciences, Mannuthy, Thrissur

CERTIFICATE

We, the undersigned members of the Advisory Committee of **Dr. Bhosale R.A.** a candidate for the degree of Master of Veterinary Science in Animal Breeding, Genetics & Biostatistics, agree that this thesis entitled **"Cloning and sequence analysis of the growth hormone gene in Indian elephants"** may be submitted by **Dr. Bhosale R.A.** in partial fulfilment of the requirement for the degree.

Dr. T. V. Aravindakshan
(Chairman, Advisory Committee)
Professor
Centre for Advanced Studies in Animal Genetics and Breeding
College of Veterinary and Animal Sciences,
Mannuthy, Thrissur

Dr. K. V. Raghunandanan,
Director,
Centre for Advanced Studies in
Animal Genetics and Breeding,
College of Veterinary and
Animal Sciences, Mannuthy
(Member)

Dr. Stephan Mathew,
Professor,
Department of Animal Breeding,
Genetics & Biostatistics.
College of Veterinary and
Animal Sciences, Mannuthy.
(Member)

Dr. G. Krishnan Nair,
Professor & Head,
Department of Veterinary Microbiology,
College of Veterinary and
Animal Sciences. Mannuthy
(Member)

24/07/09
(A. SUBRAMANIAM)
External Examiner

ACKNOWLEDGEMENTS

The art blossom, science grows; it is a joy to be alive.

This work is the summit of the past two year's fortitude, which I went through in the College of Veterinary and Animal Sciences, Mannuthy, Thrissur. The ambience of the place I was in, the kind of people I have met and interacted with and the academic environment in the institute provided a right momentum for my personality development and my first journey through the corridors of science.

I am much fortunate to be a member of Centre for advanced studies in Animal Genetics and Breeding.

To accolade oneself, words can barely explicit the depth of my candid gratitude to all those who were bounteous during this noble deed.

With a deep sense of gratitude, I prevail this opportunity to accentuate my thankfulness for my advisor, ***Dr. T.V. Aravindakshan,*** *professor, Centre for Advanced Studies in Animal Genetics and Breeding for enlightened guidance at each step of planning and execution of research, constant encouragement, valuable advice and generosity with personal involvement. I shall never forget his valuable guidance, which has helped me in moulding my thesis to what it is. All of which make him unforgettable personality in my life. I feel proud to be a student of such personality.*

I wish to express my sincere gratitude to ***Dr. K. V. Raghunandanan,*** *Director, Centre for Advanced Studies in Animal Genetics and Breeding for his shrewd suggestions, relentless efforts, indispensable guidance, constructive counsel, critical appreciation and timely help in pursuit of this study. I shall forever remain an admirer of his interest in work, his depth of knowledge, approach to the subject, love for students and constant co-operation throughout my entire stay at this college.*

I am also grateful to ***Dr. Stephan Mathew,*** *Professor, Department of Animal Breeding, Genetics & Biostatistics, for his valuable suggestion, guidance, perceptiveness and timely help during my research work.*

I sincerely thank ***Dr. G. Krishnan Nair****, Professor & Head, and* ***Dr. Koshy John****, Associate Professor, Department of Veterinary Microbiology for providing research facilities in the Division, timely suggestion and eager interest.*

I happily and sincerely express my thanks to ***Dr. K.C. Raghavan,*** *Professor, Centre for Advanced Studies in Animal Genetics and Breeding, COVAS, Mannuthy and* ***Dr. A.P. Usha,*** *Professor and Head, Department of Animal Breeding and Genetics, College of Veterinary and Animal Sciences, Pookot, Wayanad, for their altruistic attitude towards me, moral support, encouragement and nontrivial help.*

I am obliged, thankful and grateful to ***Dr. Rajeev T.S.*** *professor, Dept. of Extension for all the help, inspirations and co-operations provided by him for blood collection from the Indian Elephants.*

My special thanks to ***Dr. K.A. Bindu,*** *Associate professor, Centre for Advanced Studies in Animal Genetics and Breeding for being so kind and friendly, incessant encouragement and affection.*

I record with pleasure the help from ***Dr. T. V. Raja, Dr. Radhika G.*** *and* ***Naicy Thomas,*** *Assistant professor, Centre for Advanced Studies in Animal Genetics and Breeding for their timely suggestion and support during the course of study.*

I feel fortunate to have pleasant company of my seniors like ***Dr. Rajeev M., Dr. Rojan, Dr. Jimcy Joseph*** *and especially of* ***Dr. Arpitha*** *for their support during difficult times in the period of study.*

I wish to record my heartily appreciation to ***Ms. Simi, Mr. Arun*** *and* ***Mr. Binoy*** *whose affectionate friendship, mental support and timely advice helped me in overcoming the hardships faced during the period of study and research work.*

Life would have been dull and monotonous without friends and their uninhibited expression of even the most absurd thoughts that lightened my heavier moments. I cherish the sweet memories of batch-mates ***Dr. Lali*** *and* ***Dr. Rani.***

I really enjoyed the company of my junior friends ***Dr. Mohammad, Dr. Aparna, Dr. Sudina and Dr. Remmya.***

It seems very difficult to express in words what I owe to my friends ***Vijaysinh, Dinkar, Nitin, Albert, Balaji, Dinesh, Kiran, Girish, Guruprasad, Harshad, Prem, Navnath, Ashwin, Prathamesh, Sunil*** *and* ***Gajanan*** *for their sprit buoyant with abiding courage and pleasure through their help and moral support during my golden days of research work at COVAS, Mannuthy, Thrissur.*

I express my heartfelt gratitude to my friends ***Drs. Kanaran, Unnikrishnan, Ayub, Girish, Binoy, Selvakumar, Senthil Sr., Pramod, Sreejith, Subin, Jinesh, Aslam, Bibu, Vishnu, Senthil*** *& many more for their support & back up.*

I would like to acknowledge the staff and laborers of ***Vechur, Goat & Rabbit farm*** *for their moral support & warmth.*

I express my sincere thanks to all ***staff members*** *of the Department of Animal Breeding and Genetics.*

I am grateful to ***Dr. E. Nanu,*** *Dean, College of Veterinary and Animal Sciences, Mannuthy, for the facilities provided for this research work.*

I would like to say a big 'thank- you' to my ***Gojai, Bhau-Akka, Aai, Nana, Pappa, Mammi, Mama, Mami, Manubhau, Ashaakka, Rohinitai, Gita, Hrisheekesh, Tanuja, Vijaysinh & Vikramsinh*** *who provided me a constant source of love, affection and Happiness.*

Last but not the least, my sincere solicitude to the ***Almighty God,*** *who helped and gave me mental as well as physical strength to contend any situation during this great endeavour.*

Bhosale R.A.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

Introduction

1. INTRODUCTION

The Indian elephant goes back to the mythologies. As the *Deva* and the *Asura* churned the oceans during 'sagar manthan' for *Amrit* that would make them immortal, there surfaced the 'navratnas' (nine jewels). One of these jewels was the elephant. The elephant is therefore considered absolutely precious to be preserved and protected like the way jewels are. Hindus worship Ganesha, the elephant headed God of wisdom and the white elephant is so venerated by the Buddhists of South East Asia. The *Airavat* was the chosen carriage for the God of all Gods 'Indra'.

Elephants come under family Elephantidae, order Proboscidae and class Mammalia. They were once classified along with other thick skinned animals in a now invalid order, Pachydermata. There are three living species: the African bush elephant (*Loxodonta africana)*, the African forest elephant (*Loxodonta cyclotis*), until recently known collectively as the African elephant and the Asian elephant (*Elephas maximus*). The Asian elephant is smaller than the African. It has smaller ears and typically, only the males have large external tusks.

The Indian elephant (*Elephas maximus indicus)* is one of the three subspecies of the Asian elephant. Karyotypes of the Indian elephants showed diploid chromosome number of 56 (54 autosomes and one pair sex chromosomes) and they exhibited XX/XY sex chromosome pattern in which the cow elephant possessed XX sex chromosome while tusker and makhna XY (Sakthikumar *et al.*, 1992). The other two subspecies of the Asian elephant are E. *m. sumatranus* in Sumatra and *E. m. maximus* in Sri Lanka.

Elephants are mammals and the largest land animals alive today. The elephant's gestation period is 22 months, the longest of any land animal. At birth it is common for an elephant calf to weigh 120 kilograms. An elephant may live as long as 70 years, sometimes longer. The largest elephant ever recorded was

shot in Angola in 1956. This male weighed about 12,000 kg, with a shoulder height of 4.2 m, one meter taller than the average male African elephant.

The Asian elephant is now a species listed as endangered in the International Union for Conservation of Nature's (IUCN) red data list. Because of habitat encroachment and the demand for ivory, the Indian elephants are now found only in highly fragmented areas of south and north east of India and have experienced a dramatic decline in numbers. Elephants are symbols of wisdom in Asian cultures and are famed for their memory and high intelligence. It is the most majestic and elegant creature in this planet. It is a symbol of India for many and has held an integral space in our history and culture since time immemorial. It forms about half of the elephant species in the world, a population which is facing the threat of extinction. Despite these unique features about elephants, little is known about its genome, its organization and structure. Very few genes or DNA markers have been identified or studied in elephants.

Growth hormone (GH) is a peptide hormone produced in the anterior pituitary. It stimulates the growth of vertebrates. According to National Center for Biotechnology Information (NCBI) sequence viewer (Accession # P20392) growth hormone (or *somatotropin*) of African savanna elephant is a protein hormone of about 190 amino acids, synthesized and secreted by cells called somatotrophs. Growth hormone is a major participant in the control of several complex physiologic processes, including growth and metabolism and it is also of considerable interest as a drug used in both humans and animals. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete insulin like growth factor-1 (IGF-1). It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.

A study on nucleotide and amino acid sequence analysis would throw light on the elephant growth hormone gene architecture and the extent of functional and structural similarity with that of other species. The ease and

automation of DNA sequencing have greatly facilitated the characterization of the genes. A wider knowledge of both the structure and function of elephant growth hormone gene will provide the potential for genetic modification.

The proposed study is aimed to determine the sequence, nature and patterns of exon-intron length distributions and the protein structure of growth hormone gene and to compare its features with that of other species. The inter-species comparison of the growth hormone protein sequence will also give an idea about the evolutionary status of the animal. With these points in view this study has been taken up with the following objectives:

(1) To clone the growth hormone (GH) gene of Indian elephants.

(2) To identify the nucleotide sequence of the cloned gene.

(3) To compare the sequence with other mammalian species.

Review of Literature

2. REVIEW OF LITERATURE

2.1 GROWTH HORMONE

Growth hormone (GH) is also called Somatotrophin or Somatotrophic hormone. GH is a peptide hormone produced in the anterior pituitary. It stimulates the growth of the vertebrates. It should be first recognized that GH is one of the many factors involved in growth.

GHs can be conveniently placed into two main categories: those controlling the growth of the body as a whole and those which control the growth of specific organs. In current usage the term "growth hormone" is reserved for a preparation obtained from the anterior pituitary, which is concerned with general body growth. In addition to this substance, there are a number of other hormones secreted by the pituitary, placenta and gonads, which are concerned with the growth of specific organs (Cole, 1938).

Li *et al.* (1945) described a method, for isolating a protein from the anterior lobe of ox pituitary, which could be shown to have the biological characteristics of the GH.

Chicken growth hormone (cGH), a polypeptide hormone synthesized in and secreted by the pituitary gland, is involved in a wide variety of physiological functions such as growth, body composition, egg production, ageing and reproduction (Stephen *et al.*, 2001).

GH, prolactin, the fish hormone, somatolactin and related mammalian placental hormones, including placental lactogen, form a family of polypeptide hormones that share a common tertiary structure. They produce their biological effects by interacting with and dimerizing specific single transmembrane-domain receptors. The receptors belong to a super family of cytokine receptors with no intrinsic tyrosine kinase, which use the Jak-Stat cascade as a major signaling pathway (Forsyth and Wallis, 2002).

2.1.1 Growth Hormone Gene

Barta *et al.* (1981) isolated the rat GH gene on a cloned 11.4-kb *EcoRI*-generated DNA fragment from a phage library of chromosomal DNA. The structural gene sequence, ~ 2.1 kb long, was identified by hybridization to the corresponding cloned rat GH cDNA and was shown to contain four intervening sequences.

Woychik *et al.* (1982) isolated the gene coding for bovine GH from a bovine genomic library. The nucleotide sequence of the coding regions of the gene was found to be identical with that of a nearly full-length GH cDNA clone. The gene sequence was ~ 1.8 kb in length and contains four intervening sequences.

Orian *et al.* (1988) isolated the GH gene from ovine genomic library using the cDNA clone for bovine GH pG23 as a probe. The complete DNA sequence was determined and the amino acid sequence was also predicted. The exon-intron junctions in the sequence were determined by analogy with the bovine GH genomic sequence. The ovine GH and bovine GH genes were 97.5 per cent homologous in the coding regions.

Kioka *et al.* (1989) isolated the goat GH (gGH) from a genomic library using bovine GH cDNA as a probe. The gGH gene was located on a 5.8-kb *EcoRI-HindIII* fragment using the data of Southern hybridization analysis and the gGH gene and its flanking region were completely sequenced. The nucleotide sequence homologies (expressed in per cent) of the promoter region of the gGH gene were 98 with the bovine, 90 with the porcine, 74 with the human and 72 with the rat genes.

Das *et al.* (1996) isolated and sequenced the mouse GH (mGH) gene. The mGH gene consisted of five exons and four introns, as was observed in other mammalian species. The second intron in mGH was much smaller than its rat counterpart, thus being similar in size to human, bovine and porcine GH. As

expected, the mGH promoter showed a higher degree of homology with rat, as compared to the other mammalian species like pig, cattle and human.

Lagziel *et al.* (1996) analysed the single-strand conformation polymorphisms (SSCP) at bGH locus in *Bos taurus* and *Bos indicus*. The SSCP haplotypes differed qualitatively in these cattle, confirming the previously proposed long evolutionary separation of these cattle sub races.

Lioupis *et al.* (1997) have cloned and characterised the GH gene of red deer (*Cervus elaphus*), using genomic DNA and a polymerase chain reaction technique. The deduced sequence for the mature GH from red deer was identical to that of bovine GH, indicating that the burst of rapid evolution of GH that occurred in Artiodactyla might have been completed before the divergence of Cervidae and Bovidae and suggesting that the rate of evolution during this burst might have been greater than previously estimated.

Lioupis *et al.* (1999) have cloned and characterised part of the GH gene of the Eurasian mole rat (*Spalax ehrenbergi*), using genomic DNA and a PCR technique. The sequence of the entire coding region, 5' untranslated region (UTR), most of the 3' UTR and part of the promoter region was described. The overall organisation of the mole rat GH gene was similar to that of GH genes from other mammals.

2.1.2 Growth Hormone Protein

Hartree (1966) developed a method for separation and partial purification of four protein hormones from acetone-dried human pituitary glands. The major portions of growth hormone, follicle-stimulating hormone, luteinizing hormone and thyroid stimulating hormone activities were obtained in separate fractions. Although the yields of the individual hormones were similar to those obtained by other investigators, the procedure was simpler and a cleaner separation of the hormones had been achieved.

Bellair (1972) reported the amino acid sequence of the C-terminal 68 residues of ovine GH. The degrees of homology (expressed in per cent) between this region of ovine with bovine, porcine, human GH and ovine prolactin were 99, 90, 62 and 22, respectively.

Li *et al.* (1987) purified GH from elephant pituitary glands. They reported that GH had 191 amino acids with two disulfide bridges and a single *Trp* residue. The somatotrophin activity was only 15 per cent as compared with the bovine hormone in the radioreceptor-binding assay.

Yamano *et al.* (1988) isolated the cDNA that encoded goat GH (gGH), from a goat pituitary cDNA library. The cDNA, about 880 base pairs long, had a coding sequence, 5'- and 3'-untranslated regions and a poly (A) chain. The cDNA could encode a polypeptide of 217 amino acids. The amino acid sequence homology (expressed in per cent) between gGH and the sequences of bovine GH, rat GH and human GH were 99, 83 and 66, respectively.

The rabbit GH gene was similar to other mammalian GH, being comprised of five exons and four introns. As in rodents and artiodactyls, the rabbit GH occurred as a single gene, with no evidence for a cluster of GH-like genes, as was found in primates. The amino acid sequence of rabbit GH was similar to that of pig GH but differed markedly from the available sequences of ruminant and primate GH (Wallis and Wallis, 1995).

2.1.3 Structure of Growth Hormone Gene

Wallis (1969) reported that Ox pituitary GH possessed two N-terminal residues (either *Phe* or *Ala*) and therefore two types of polypeptide chains. The molecular weight of the hormone had been found in recent studies to be 20-26,000 Da.

Chen *et al.* (1970) reported that the porcine GH appeared to be a single polypeptide chain with *Phe* at both the NH_2-terminal and COOH-terminal

residues. The amino acid composition resembled that of bovine, canine and human GH. Preliminary observations of the molecular weight of the monomer of porcine GH by sedimentation equilibrium in the presence of 5.5 M guanidine hydrochloride have yielded an estimate of 22,000 ±1,500 Da.

Santome *et al.* (1971) reported the amino acid sequence of the bovine GH molecule. They have established that the true molecular weights of the hormones obtained from bovine, porcine and ovine species were similar and close to 21,000 Da. The location of one disulphide bridge in bovine GH amino acid sequence was established to be between residues 53 and 161. The disulphide bridge near the C-terminal end of the molecule had already been identified. The single *Trp* residue was in position 86. The three *His* residues were in positions 20, 28 and 167. Thirty per cent of the molecules had *Val* replacing *Leu* 124.

Conde *et al.* (1973) isolated pure equine GH from frozen pituitary glands by a new procedure, which avoided extremes of pH during the extraction. The molecular weight of equine GH was determined to be 20200 Da by sedimentation equilibrium analysis. The amino acid composition shows considerable analogy with those of bovine, ovine and porcine GHs.

The GHs isolated so far are all simple proteins. Their composition is entirely accounted for by their content of amino acids. Their general properties are those of class of globulins. They are soluble in dilute acid or alkali and are precipitated nearly completely by half-saturation of their solutions with ammonium sulphate at pH 7. The solubility is strongly depressed by low concentrations of sodium chloride at pH 3 or at pH 4. This may be attributed to the tendency of the hormone to aggregate under these conditions (Wilhelmi, 1974).

Barsh *et al.* (1983) determined the structure of the human GH gene cluster over a 78 kb region of DNA by the study of two overlapping cosmids.

There were two GH genes interspersed with three chorionic somatomammotropin genes, all in the same transcriptional orientation.

Einarsdottir *et al.* (2002) cloned and sequenced the cDNA of GH from Atlantic halibut (*Hippoglossus hippoglossus*) and its amino acid sequence deduced from the cDNA sequence. The mature *Hippoglossus hippoglossus (hh*GH) protein consisted of 186 amino acids. Comparison with other flatfish species as well as a species from a different order, revealed that the sequence similarities of the mature *hh*GH with that of the barfin flounder, the Japanese flounder, the sole and the pufferfish were 99.5, 81.7, 74.2, and 65.2 per cent, respectively.

2.1.4 Sequence Comparison of Growth Hormone

Bewley and Li (1970) compared the primary structures of human pituitary GH and sheep pituitary lactogenic hormone with regard to similarities in their sequences. Three homologous segments in each molecule, comprising approximately 45 per cent of either peptide chain, have been found suggesting that these two proteins have apparently evolved from a common ancestor molecule.

According to Conde *et al.* (1973) the amino acid sequence of equine GH was quite similar to those of bovine, ovine and porcine GHs. The N-terminal and C-terminal amino acid residues of equine GH were the same as those of bovine, ovine and porcine. The structure found for the C-terminal bridge was quite homologous to that in other GHs but the other bridge was totally different. It was found that both disulfide loops were of the expected size and with homologous sequence to those in all other GHs.

Wilhemi, (1974) observed that there is one less *Arg* in the ring of the human GH molecule, as well as three substitutions as compared to ovine and bovine GH, which were identical in sequence. In human and bovine GH, the

second disulphide bond is formed between the half *Cys* residues at positions 53 and 165.

Varma *et al.* (1990) cloned the cDNA for buffalo *(Bubalus bubalis)* GH and determined the sequence through RT-PCR approach. The nucleotide sequence of bubaline GH cDNA was in a single reading frame coding for a protein of 191 residues, comprising a putative signal sequence of 27 amino acids. Homology comparison of the sequence with other mammalian GH cDNAs showed a very high degree of evolutionary conservation. Bubaline GH sequence shared a homology (expressed in per cent) of 99.5, 99.5, 98.6, 87.6 and 61.9 with those of ovine, caprine, bovine, porcine and human, respectively at amino acid level.

Zhao *et al.* (1996) showed that an open reading frame of the carp GH gene contained 630 bp which coded for a polypeptide of 210 amino acids including 22 amino acids of the signal peptide and 188 amino acids of the mature GH. Nucleotide sequence and amino acid sequence of carp GH gene were the same as Koren's carp GH cDNA in the coded region. Compared with Chao's carp GH cDNA, the homology of nucleotide sequence and amino acid sequence for carp GH gene was 95.6 and 96.7 per cent, respectively, in the coded region.

Xu *et al.* (2001) estimated the molecular weight of Sea perch GH to be 19,200 Da. The complete sea-perch GH amino acid sequence of 187 residues was determined. Alignment of sea-perch GH with those of other fish GHs revealed that sea-perch GH was most similar to advanced marine fish such as tuna, gilthead sea bream, yellowfin porgy, red sea bream, bonito and yellow tail with 98.4, 96.2, 95.7, 95.2, 94.1 and 91 per cent sequence identity, respectively.

According to Venugopal *et al.* (2002), the GH encoding cDNA of Mrigal consisted of 1146 nucleotides, which coded for a protein of 210 amino acids. It had two putative glycosylation sites at 135 and 187 positions. The *Cys* residues, which were important for the disulfide bond formation were located at 49, 161,

178 and 186 positions. In addition it had an extra *Cys* residue at position 123, like the GH proteins of carps. Overall, the Mrigal GH protein was longer than that of Rohu by three residues and shorter than that of Catla by one residue.

2.2 INDIAN ELEPHANTS

The Indian elephant has been a part of Indian culture, mythology and ethos from time immemorial. Historically, one of the most exhaustive writings in elephants is that of Abul-Fazl Allami in his An-Akbari outlining the management of thousands of elephants in the stables of Akbar the great. His son Emperor Jahangir is said to have had over 12000 elephants in the imperior Mughal fighting force. A lot of traditional wisdom influences the treatment of Indian elephant in captivity. Today elephants have become an integral part of the social scenario of India.

Indian kings made use of elephants from very early times partly for ceremonial display and partly for military purposes. In the latter respect they may be said roughly to have corresponded to heavy artillery before the days of gunpowder. The oldest Indian treatise, the *'Kautiliya Arthashstra'* includes the oldest data on Elephantology (Edgerton, 1931).

The Asian elephant *(Elephas maximus),* is the second largest terrestrial mammal mostly found in India and enjoys distribution in Bangladesh, Srilanka, Burma, Thiland, Cambodia, Vietnam, Laos, Southern China, Malaya, Borneo and Sumatra. Within Indian limits, Asian elephants are seen mostly in two belts: firstly in the foothills of Himalayas including areas of south of Brahmaputra and secondly, in the belt extending northwards from Nagercoil to Shimoga in the Western Ghats. Elephants have been elevated as the state animal of Kerala (Surendran, 1992).

2.2.1 Molecular Studies on Indian Elephant

Hungerford *et al.* (1966) obtained karyotypes of both Asian and African elephants from cultured somatic cells and reported that the diploid chromosome

number in both species was 56 and the chromosomes of both were morphologically identical except for the sex chromosomes.

According to Saktikumar *et al.* (1992) karyotypes of the Indian tusker, cow elephant and makhna have the diploid chromosome number of 56 comprising of 54 autosomes and a pair of sex chromosomes.

Fernando and Lande (2000) reported genetic evaluation and behavioral study of social organization in the Asian elephant. They demonstrated a substantial difference in the complexity and structure of Asian elephant social groupings from that described for African savanna elephant. Genetic sampling of individuals from the four groups in Ruhuna National Park and three other groups in surrounding areas, and analysis through PCR amplification and sequencing of mitochondrial DNA (mtDNA) from dung supported the matriarchal nature of female groups and the lack of inter-group transfer of females.

Fernando *et al.* (2000) reported the first genetic analysis of free-ranging Asian elephants. They sampled 118 elephants from Sri Lanka, Bhutan, North India, Laos and Vietnam by extracting DNA from dung and sequenced 630 nucleotides of mtDNA, following PCR amplification, including part of the variable left domain of the control region. Comparison with African elephant sequences indicated a relatively slow molecular clock in the *Proboscidae*, with a sequence divergence of ~ one per cent/Myr. Seventeen haplotypes were identified within Asian elephants, which clustered into two well-differentiated assemblages with an estimated Pliocene divergence of 2.5±3.5 Myr ago.

Fleischer *et al.* (2001) sequenced mtDNA to assay genetic variation and phylogeography across much of the Asian elephant's range. They assessed variations in part of the mtDNA control region (CR) and adjacent tRNA genes in 57 Asian elephants from seven countries, viz., Sri Lanka, India, Nepal, Myanmar, Thailand, Malaysia and Indonesia. Asian elephants had typical levels of mtDNA

variation and coalescence analyses suggested that their populations were growing in the late Pleistocene.

Fernando *et al.* (2003) compared DNA of Borneo elephants to that of elephants from across the range of the Asian elephant, using a fragment of mtDNA, including part of the hypervariable d-loop and five autosomal microsatellite loci. They found that Borneo's elephants were genetically distinct, with molecular divergence indicative of a Pleistocene colonisation of Borneo and subsequent isolation. Their genetic distinctiveness makes them one of the highest priority populations for Asian elephant conservation.

Selective poaching of Asian elephant males for ivory has resulted in highly female-biased adult sex ratios, necessitating regular monitoring of population structure and demography. Vidya *et al.* (2003) demonstrated that the molecular sexing from dung-extracted DNA based on ZFX–ZFY fragment amplification and ZFY-specific *BamHI* site restriction could be applied to estimate sex ratios of free-ranging Asian elephants, in addition to or instead of field demographic methods.

Vidya *et al.* (2005) examined the population genetic structure of Asian elephants across India, which harboured over half the world's population of this endangered species. mtDNA control region sequences and allele frequencies at six nuclear DNA microsatellite markers obtained from the dung of free-ranging elephants revealed low mtDNA and typical microsatellite diversity.

Gupta *et al.* (2006) have developed a technique based on molecular markers to determine the carcass of an elephant and also if it was of a male. Using DNA sequence information from Genbank, they have developed two primer pairs: one for the mtDNA and the other for the sex-determining region of Y chromosome (SRY) gene of the Indian elephant. After PCR amplification of known elephant DNA, they found that the mtDNA was common in both males and females, whereas the SRY-specific amplicon was observed only in the male.

2.3 MOLECULAR CLONING

Molecular cloning refers to the procedure of isolating a defined DNA sequence and obtaining multiple copies of it *in vivo*. Cloning is frequently employed to amplify DNA fragments containing genes but it can be used to amplify any DNA sequences such as promoters, non coding sequences, chemically synthesized oligonucleotides and randomly fragmented DNA. Cloning is utilized in a wide array of biological experiments and technological applications such as large scale protein production.

Roskam and Rougeon (1979) have cloned and sequenced almost the complete cDNA of human GH. The nucleotide sequence obtained confirmed the already known protein sequence and they were also able to predict the sequence of a precursor region of 26 amino acids.

Miller *et al.* (1980) have cloned cDNA coding for bovine GH (bGH) into the *PstI* site of plasmid *pBR322* by the dC x dG tailing technique and transformed into *E. coli*. A recombinant plasmid containing bGH cDNA was identified by hybridization with cloned Rat GH cDNA.

Sekine *et al.* (1985) isolated cDNA clones encoding chum salmon (*Oncorhynchus keta*) GH (sGH) from a cDNA library prepared from chum salmon pituitary gland poly $(A)^+$ RNA. Synthetic oligodeoxyribonucleotide mixtures based on amino acid residues 23-28 of sGH were used as hybridization probes to select recombinant plasmids carrying the sGH coding sequence.

Hu *et al.* (2000) cloned the frog (*Xenopus laevis)* pituitary adenyl cyclase-activating peptide (PACAP) or Growth Hormone Releasing Hormone (GHRH) which appeared to regulate several neuro endocrine functions in the frog.

Horikawa *et al.* (2001) cloned the ovine and bovine pituitary GHRH receptors to provide information about species differences in GHRH receptors useful for studies of receptor- ligand binding properties and receptor function.

Venugopal *et al.* (2002) have developed a modified rapid amplification of cDNA ends (RACE) strategy for cloning highly conserved cDNA sequences. Using this modified method, the GH encoding cDNA sequences of *Labeo rohita, Cirrhina mrigala* and *Catla catla* have been cloned, characterized and over expressed in *E. coli.*

According to Jeong and Lee (2003) experiments for the over production of recombinant proteins begin by the cloning of a gene for the desired protein into a multiple copy plasmid under the control of a strong inducible promoter. This approach has been successful in producing large amounts of some proteins but has often resulted into slow growth, increased stress response and eventually cessation of growth.

Deane and Woo (2006) have cloned and characterized the pituitary GH gene of Silver Sea bream (*Sparus sarba*) and found it to be 615 base pairs encoding a protein of 204 amino acids.

According to Toogood *et al.* (2006) the regulation of GH in birds was controlled by a receptor, which was cloned from chicken pituitary cDNA and was found to be having 61 per cent amino acid sequence identity to the Human pituitary GHRH receptor.

2.3.1 Isolation of Genomic DNA

The increasing use of genetic methods in clinical laboratories with large series of samples requires methods for rapid and efficient DNA isolation. Also many studies such as linkage mapping, characterization of gene and DNA polymorphism requires isolation of genomic DNA from a large number of animals. Deoxyribose nucleic acid is fundamental for polymerase chain reaction (PCR) based methods. So an efficient DNA isolation method should give high quality DNA in sufficient amounts and should be fast and simple.

Blin and Stafford (1976) described a method for isolation of high molecular weight DNA from eukaryotes. This procedure allowed preparation of DNA from a variety of tissues such as calf thymus or human placenta and from cells which were more difficult to lyse until now.

Sambrook *et al.* (1989) described that the standard way to remove proteins from nucleic acid solution was to extract first with phenol-chloroform and then with chloroform. This procedure takes advantage of the fact that deproteinisation is more efficient when two different organic solvents are used instead of one.

Parzer and Mannhatter (1991) have developed a DNA isolation method which allowed the isolation of high quality DNA from the frozen citrated blood and cell lines within 90 min. This procedure avoided the use of toxic organic reagents and tedious extraction steps.

Wu *et al.* (1995) described a procedure for obtaining genomic DNA from animal tissues that minimized manipulations of the samples and required no organic solvent extraction and high speed centrifugation. Micrograms to milligrams of DNA could be easily isolated from each tissue sample which could be used directly for both PCR and Southern blot analysis.

Aravindakshan *et al.* (1998) compared three methods of DNA extraction, viz., the guanidine hydrochloride method, the high salt method and phenol chloroform method and demonstrated that both high salt and phenol chloroform methods produced good yields of high molecular weight DNA from cattle white blood cells, whereas the guanidine hydrochloride method failed to yield clean DNA.

2.3.2 Yield and Purity of DNA

Douglas *et al.* (1992) described a rapid, nontoxic method for the purification of DNA from human leukocyte. Preliminary experiments, which

tested different methods of DNA purification, indicated that digestion of protein with proteinase K was unnecessary. This led to the development of a simple procedure involving lysis of the cells in SDS followed by extraction with 6M NaCl.

Ahmad *et al.* (1995) reported a modification of the widely used standard proteinase K-phenol extraction method for improving the yield and purity of DNA from frozen blood samples by an initial trypsinisation of whole blood before cell lysis to obtain lymphocytic nuclei and subsequent DNA purification. They reported an increased total yield of DNA as well as improved purity with pretrypsinised blood samples.

Clegg *et al.* (1997) described an effective protocol of extracting high yields of microbial community DNA from humidified soils. Evaluation using three soils yielded up to 30 microgram DNA per gram dry soil with absorbance ratio at 260:280 nm of 1.6-2.0.

Aravindakshan *et al.* (1998) reported that the mean yields of DNA extracted from 10 ml of whole blood of cattle and buffalo by phenol and high salt methods were 394.50/446.16 and 344.26/432.83 microgram, respectively. The ratio of optical densities at 260 and 280 nm was consistently between 1.75 and 1.91, indicating good deproteinisation.

Kalia et al. (1999) reported a modified procedure for extraction of high quality genomic DNA that was rapid, simple, biologically non hazardous and generally applicable to pathogenic bacteria and consistently high yields of genomic DNA were obtained from 465 isolates representing over 30 clinically important bacterial species.

2.3.3 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a technique widely used in molecular biology. It derives its name from one of its key components, a DNA

polymerase used to amplify a piece of DNA by an *in vitro* enzymatic replication. PCR entails enzymatic amplification of specific DNA sequences using two oligonucleotide primers that flank the DNA segment to be amplified. The rapid production of large quantities of a specific DNA sequence took a leap forward with the development of the PCR.

Polymerase chain reaction, now a common and indispensable technique was developed by Mullis *et al.* (1986), is used in medical and biological research laboratories for a variety of applications.

Saiki *et al.* (1988) used a thermostable DNA polymerase in an *in vitro* DNA amplification procedure due to which single copy genomic sequences were amplified by a factor of more than 10 million with very high specificity. The enzyme isolated from *Thermus aquaticus* (Taq), enabled DNA amplification to be performed at higher temperatures and significantly improved the specificity, yield, sensitivity and the length of the product amplified.

Innis and Gelfand (1990) pointed out that "time of temperature" was the main reason for loss of activity of Taq DNA polymerase. They suggested that reducing the denaturation time to 30 s for short template sequences and increasing the denaturation temperature up to 96 ^{0}C might prolong the activity of the enzyme.

Yap and Mc Gee (1991) suggested the possibility of reducing the denaturation temperature after about 10 rounds of amplification as the mean length of target DNA was decreased. For templates of 300 bp or less, denaturation temperatures may be reduced to as low as 88°C. For templates with 50 per cent (G+C) content as many as 40 cycles could be carried out without much decrease in enzymes efficiency.

Hongbao (2005) used a peltier heat pump to quickly heat and cool the DNA and used Taq polymerase for the synthesis of DNA.

White (2005) suggested that non specific products in PCR could be minimized by using a "hot start" and a variable annealing temperature. The various other factors in the optimization of the PCR included the buffer, magnesium chloride, deoxyribonucleotides, design of primers, thermostable DNA polymerase, DNA template, reaction volume, denaturation time, annealing temperature and extension time.

2.3.3.1 Design of Oligonucleotide Primers

A primer is a strand of nucleic acid that serves as a starting point for DNA replication. They are required because the DNA polymerase enzyme that catalyses the reaction can add new nucleotides only to an existing strand of DNA. The polymerase starts replication at 3' end of the primer and copies the opposite strand.

According to Lexa *et al.* (2001), the success of the PCR strategy was highly dependent on the small synthetic oligonucleotides that hybridize to the DNA sequences. These short nucleotides known as the forward and reverse primers, function in pairs and amplify a specific DNA sequence and more importantly, anneal exclusively to that DNA target locus.

2.3.3.1.1 Annealing Temperature and Primer Design

Primer length and sequence are of critical importance in designing the parameters of a successful amplification. The melting temperature of DNA duplex increases with its length and (G+C) content.

A simple formula for calculation of the Tm is

$$Tm = 4\ (G+C) + 2\ (A+T)\ ^{o}C$$

Thus, annealing temperature chosen for PCR depends directly on the length and composition of the primer (s). One should aim at using an annealing

temperature (T*a*) about 5°C below the lowest T*m* of the pair of primers to be used (Innis and Gelfand, 1990).

Rychlik *et al.* (1990) stated that if the T*a* was increased by 1°C every other cycle, specificity of amplification and yields of products <1 kb in length were both increased. One consequence of having too low T*a* was that one or both primers will anneal to sequences other than the true target, as internal single base mismatches or partial annealing might be tolerated, leading to non-specific amplification and consequent reduction in the yield of the desired product.

A consequence of too high T*a* is that too little product will be made, as the likelihood of primer annealing is reduced. Another and important consideration is that a pair of primers with very different T*a* values may never give appreciable yields of unique product and may also result in inadvertent "asymmetric" or single strand amplification of the most efficiently primed product strand.

2.3.3.1.2 Primer Length

The optimum length of a primer depends upon its (A+T) content and the T*m* of its partner. Sixteen base sequence will statistically be present only once in every 4^{16} bases ≈ 4 billion. Consequently 17-mr or longer primers are routinely used for amplification from genomic DNA. Too long primer length may mean that even high annealing temperatures are not enough to prevent mismatch pairing and non-specific priming (Rybicki, 2001).

2.3.3.1.3 Primer Specificity

According to Wu *et al.* (1991), primer specificity was an important parameter in PCR primer design. The primer should bind to the target sequence only and not anywhere else to ensure the amplification of the intended fragment alone, i.e. the target sequence should occur only once in the template.

2.3.3.1.4 Hairpin and Self Dimer

Rychlik (1995) described the importance of avoiding primer complimentarity and primer dimer formation. The hardest part in PCR primer design was to avoid primer complimentarity, especially at the 3' ends which might result in folding of the primer in half and form the so called hairpin structure, which is stabilized by the complimentary base pairing. Primer- dimer formation causes problem to PCR reactions as it may act as a competitor to the amplification of the target DNA.

According to Vallone and Butler (2004), the ability to select short DNA oligonucleotide sequences capable of binding solely to their intended target was of great importance in developing nucleic acid based detection technologies. Freely available primer design software such as primer3 screens for potential hairpin and primer-dimer interactions while selecting a set of primer pairs.

2.3.3.2 General Rules for PCR Primer Design

According to Innis and Gelfand (1990), the rules for primer design are:

1. Primers should be 17-28 bases in length.
2. The (G+C) content of the primers should be 50-60 per cent.
3. The 3′ of the primers should end in G or C or CG or GC so as to prevent "breathing" of ends and to increase the efficiency of priming.
4. T_m values between 55-80 °C are preferred.
5. Avoid primers with 3′ complementarity, as otherwise primer dimers will be synthesized preferentially to any other product.
6. Primer self-complementarity (ability to form secondary structures such as hairpins) should be avoided.

7. Runs of three or more Cs or Gs at the 3'-ends of primers may promote mispriming at G or C-rich sequences because of stability of annealing and should be avoided.

According to Roux (1995) PCR primers should have 40-60 per cent GC content and a 3'-terminal *"G/C* clamp" (at least one or two 3' G's and/or C's); be similar in size (18-25 bases) Tm values, and nucleotide ratios; and be free of repetitive motifs, palindromes, excessive degeneracy, and long stretches of polypurines or polypyrimidines.

Lexa *et al.* (2001) described Virtual PCR as a promising programme, which accepts user-given primers and then conducts a similarity search using NCBI Blast to identify sequences in public databases that were complementary to any two primers. Input was limited to only two primers at a time, making batch analysis a difficult job.

2.3.3.3 Design of Degenerate Primers

A PCR primer sequence is called degenerate if some of its positions have several possible bases. The degeneracy of the primer is the number of unique sequence combinations it contains. Degenerate primers are mixtures of similar, but not identical, primers. These mixed primers are used to find new genes or gene families, amplify same gene from different organisms, as the genes themselves are similar but not identical.

These primers can also be used to amplify conserved sequences of a gene or genes from the genome of an organism or to get the nucleotide sequence after having sequenced some amino acids from a protein of interest. The most common use of degenerate primers is when the amino acid sequence of a protein is known. One can reverse translate this sequence to determine all of the possible nucleotide sequences that could encode that amino acid sequence.

Rose *et al.* (2003) developed a new primer design strategy for PCR amplification of distantly related gene sequences based on consensus-degenerate hybrid oligonucleotide primers (CODEHOPs). Programme has been written to design CODEHOP PCR primers from conserved blocks of amino acids within multiple-aligned protein sequences. Each CODEHOP consisted of a pool of related primers containing all possible nucleotide sequences encoding 3–4 highly conserved amino acids within a 3' degenerate core. The CODEHOP designer (http://blocks.fhcrc.org/codehop.html) was linked to BlockMaker and the Multiple Alignment Processor within the Blocks Database World Wide Web (http://blocks.fhcrc.org).

Linhart and Shamir (2005) implemented a programme called HYDEN for designing highly degenerate primers for a set of genomic sequences. Also they reported the success of the programme in several applications, one of which was an experimental scheme for identifying all human olfactory receptor (OR) genes. In that project, HYDEN was used to design primers with degeneracies up to 10^{10} that amplified with high specificity many novel genes of that family, tripling the number of OR genes known at that time.

According to Lamprecht *et al.* (2008) GeneFisher was a sophisticated interactive programme for designing degenerate primers. The procedure leads to isolation of genes in a target organism using multiple alignments of related genes from different organisms. The term "gene fishing" was referred to the technique where PCR was used to isolate a postulated but unknown target sequence from a pool of DNA.

2.3.4 Cloning of the PCR Products

The ease of adding nucleotide sequences to the termini of PCR products has led to the development of varieties of strategies for their cloning, because the cloning of PCR products is typically the first step in achieving an experimental goal. The efficiency of the cloning procedure is usually the primary consideration

and cloning strategies should be simple in design and execution, requiring a minimum of enzymatic steps.

Cloning methods are separated into two groups, namely, ligation-dependent cloning and ligation-independent cloning. The ligation-dependent cloning methods for PCR products can be further classified into three types based on the ends of the DNA as blunt-end cloning, sticky end cloning and T-A cloning (Clark, 1988).

Oliner *et al.* (1993) reported an efficient method to clone PCR products, exploiting endogenous *E. coli* enzymatic activities. Polymerase chain reaction products were engineered to contain terminal sequences identical to sequences at the two ends of a linearized vector.

Tillet and Neilan (1999) described a simple method for the cloning of PCR products without the need for post amplification enzymatic treatment. Tailed PCR primer sets were used to create complimentary staggered overhangs on both insert and vector by post-PCR denaturation- hybridization reaction. The avoidance of post-amplification enzymatic procedures made the technique rapid and reliable, avoiding need for multiple sub-cloning steps.

2.3.4.1 T-A Cloning

T-A cloning is one of the most popular methods of cloning PCR products amplified using Taq and other DNA polymerases. This enzyme lacks 5'-3' proofreading activity and adds a single 3'dA overhang to each end of the PCR product. As a result the PCR product can be directly cloned into a linearized cloning vector that have single base 3'dT overhangs on each ends. Such vectors called T-vectors. The PCR product with an overhang is mixed with this vector in high proportion. The complimentary overhangs of a T-vector and the PCR products hybridize. The result is a recombinant DNA, the recombination being bought about by DNA ligase.

Taq DNA polymerase is the most widely used thermostable DNA polymerase and it has terminal transferase activity. In the presence of four deoxyribonucleotide triphosphates (dNTPs), dA was preferentially added to the 3′ termini of PCR amplified duplex DNA molecules leaving a single 3′-dA overhang (Clark, 1988).

Aslanidis and Jong (1990) documented an efficient method of cloning complex PCR mixtures, resulting in libraries exclusively consisting of recombinant clones. The 5′ ends of the primers used to generate clonable PCR fragments contained an additional 12 nucleotides sequence lacking dCMP. As a result, the amplification products included a 12-nucleotides sequence lacking dGMP at their 3′ ends.

Holton and Graham (1991) reported a novel approach in the design of vectors that could be used for direct cloning of PCR products that possessed 3′-overhangs. They exploited the fact that a ddT-tailed vector can ligate directly to a PCR product that contained a 3′- overhanging dA, by forming a phosphodiester bond between the vector's 5′-PO_4 and the PCR product's 3′-OH group from the overhanging dA.

Kovalic *et al.* (1991) described a method for direct cloning of DNA fragments generated by primed enzymatic amplification that was based on digestibility of the cloning vector with *XcmI* restriction endonuclease to yield a unit linear molecule with 3′ unpaired dT residues at both ends. Such vectors were capable of efficient ligation with products that were obtained by primed enzymatic amplification with Taq DNA polymerase bearing 3′ unpaired dA residues at both ends.

Marchuk *et al.* (1991) developed a procedure that was at least one hundred fold more efficient than attempting to clone PCR products into a blunt-ended cut vector. In this procedure, PCR products were gel purified to avoid cloning spurious bands and ligated to the vector at 14 °C. The vector and PCR

products had complimentary single base 3'-overhangs. Vector self-ligation events were prohibited by the 3'-dT overhang, and concatamerization of the insert was prohibited by the unphosphorylated 5' end, contributed by the oligonucleotide primer, as well as the 3'-dA overhang added by Taq polymerase during the PCR-reaction.

Mead *et al.* (1991) devised a simple, universal cloning strategy that permited the direct ligation of PCR amplified nucleic acid to a compatible vector preparation. This approach took advantage of single 3'-dA extension that *Thermus aquaticus*, *Thermus flavus*, and *Thermococcus litoralis* DNA polymerases added to the termini of amplified nucleic acid.

According to Novy *et al.* (1996) the factor known to affect the T-vector cloning efficiency was the number of PCR cycles performed to produce the insert. The amount of active Taq DNA polymerase decreased with each PCR cycle, whereas the amount of template increased. Performing excessive PCR cycles in the absence of sufficient Taq DNA polymerase could lead to conditions in which ragged ends were produced due to incomplete elongation.

Zhou and Gomez-Sanchez (2000) stated that T-A cloning was one of the simplest and most efficient methods for the cloning of PCR products. The procedure exploited the terminal transferase activity of certain thermophilic DNA polymerases, including *Thermus aquaticus* (Taq) polymerase.

Lee *et al.* (2007) tested the ability of Taq DNA polymerase to amplify long DNA fragments and showed that, if the conditions were set properly, Taq could successfully perform the "long PCR" up to 24 kb. The conditions included (1) longer primers (2) a two step cycling and (3) a long buffer. He proposed that the most important requirements were the survival rate of Taq DNA polymerase at high temperatures and that of the primers against the 5′ to 3′ exonuclease activity of Taq DNA polymerase.

2.3.4.2 E. coli as a Host Organism

Cohen *et al.* (1972) reported the transformation of *E. coli* cells treated with $CaCl_2$ to multiple antibiotic resistances by purified R-factor DNA. Transformed bacteria acquired a closed, circular, transferable DNA species having resistance, fertility and sedimentation characteristics of the parent R-factor.

Oliner *et al.* (1993) described an efficient method to clone PCR products exploiting endogenous *E. coli* enzymatic activities. PCR products were engineered to contain terminal sequences identical to sequences at the two ends of a linearized vector. Polymerase chain reaction product and vector DNA were then simply co-transfected in to *E. coli* strain JC8679, obviating the requirement for enzymatic treatment of the PCR product or *in vitro* ligation.

Lee (1996) stated that *E. coli* was the most widely used prokaryotic system for the synthesis of heterologous proteins. Once an optimal expression system had been constructed, protein production could be enhanced by increasing the production of protein per cell per unit time (specific productivity) or by increasing the cell concentration per unit time (cell productivity).

Topcu (2000) compared a number of parameters that were important in the ligation of the PCR amplified DNA inserts into plasmid vectors and their efficient transformation into bacterial cells. The results showed that the T4 polynucleotide kinase-treated group, without further enzymatic manipulation at an insert to vector ratio of 3:1 gave the highest recombination efficiency when 10 μg/ml DNA and 20 units T4 DNA ligase were applied for ligation for 12 hr at 4 °C.

2.3.4.2.1 Preparation of Competent E. coli for Transformation

Most species of bacteria, including *E. coli*, take up only limited amount of DNA under normal circumstances. In order to transform this species efficiently,

the bacteria have to undergo some form of physical and/or chemical treatment that enhances their ability to take up DNA. Cells that have undergone this treatment are said to be competent. Cells are made competent by a process that uses calcium chloride and heat shock.

Morrison (1977) described that all the methods of rendering cells competent for uptake of DNA involved, briefly: (i) several successive washes in cold NaCl, $MgCl_2$ and/or $CaCl_2$ solutions (ii) incubation in $CaCl_2$ at 0°C with DNA (iii) a short incubation at some higher temperatures before dilution into growth medium and (iv) selection of transformants. He showed that calcium treated cells frozen and stored at -82°C displayed good competence after thawing.

Chung *et al.* (1989) developed a simple, one step procedure for the preparation of competent *E. coli* that used a transformation and storage solution (LB broth containing 10 per cent (w/v) polyethylene glycol, 5 per cent (v/v) dimethyl sulphoxide, and 50M Mg^{2+} at pH 6.5).

Nakata *et al.* (1997) stated that transformation of *E. coli* was first described by Mandel and Higa, who reported that *E. coli* cells after treatment with calcium chloride, could take up bacteriophage λ DNA and produce viable phage particles. Modifications that improved the efficiency of transformation included prolonged exposure of cells to $CaCl_2$, substitution of Ca ions by other cat ions such as Rb^+, Mn^{2+}, K^+ and addition of other compounds such as Dimethyl sulphoxide, dithiothreitol and cobalt hexamine chloride.

Tsen *et al.* (2002) reported that *E. coli* did have a natural transformation process. Strains of *E. coli* could incorporate extracellular plasmids into cytoplasm 'naturally' at low frequencies. Their results showed that natural transformation with plasmid existed in *E. coli*.

2.3.4.2.2 Lysogeny Broth (LB)

Lysogeny Broth (LB), a nutritionally rich medium, is primarily used for the growth of bacteria. It is also known as Luria Broth or Luria-Bertani broth. It continues to be one of the most common media used for maintaining and cultivating recombinant strains of *E. coli*.

According to Miller (1987), LB media were nutritionally rich suitable for the growth of pure culture like recombinant strains *E. coli* and derived strains were deficient in vitamin B synthesis and they were not able to grow on nutritionally poor media. Tryptone and Yeast extract served as a source of nitrogen, sulfur and carbon while yeast extract also contained vitamin B complex. Sodium chloride provided sodium ions for the membrane transport and maintained osmotic equilibrium of the medium.

2.4 DNA SEQUENCING

DNA sequencing is the biochemical method for determining the order of nucleotide bases adenine, guanine, cytosine and thymine in a DNA oligonucleotide. In 1975, Frederich Sanger and coworkers developed chain termination method for DNA sequencing.

Maxam and Gilbert (1977) reported that DNA could be sequenced by a chemical procedure that broke a terminally labeled DNA molecule partially at each repetition of a base. The lengths of the labeled fragments then identify the positions of that base. They described reactions that cleaved DNA preferentially at guanines, adenines, cytosines and thymines equally and at cytosines alone. When the products of these four reactions were resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequence could be read from the pattern of radioactive bands.

Sanger *et al.* (1977) explained a new method of DNA sequencing, which was similar but more rapid and accurate than the "plus and minus" method. This

technique made use of the 2', 3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which acted as specific chain terminating inhibitors of DNA polymerase.

Chan (2005) documented that faster sequencing methods would undoubtedly lead to faster single nucleotide polymorphism discovery. With the completion of the human genome sequence, there is now a focus on developing new sequencing methodologies that would enable "personal genomics" or the routine study of our individual genomes.

Eid *et al.* (2009) presented single-molecule, real-time sequencing data obtained from a DNA polymerase performing uninterrupted template-directed synthesis using four distinguishable fluorescently labeled deoxyribonucleoside triphosphates (dNTPs). They detected the temporal order of their enzymatic incorporation into a growing DNA strand with zero-mode waveguide nanostructure arrays, which provided optical observation volume confinement and enabled parallel, simultaneous detection of thousands of single-molecule sequencing reactions.

2.4.1 Basic Local Alignment Search Tool (BLAST)

The Basic Local Alignment Search Tool (BLAST) finds region of local similarity between sequences. The programme compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationship between sequences as well as helps to identify members of gene families.

The National Centre for Biotechnology Information (NCBI) maintains a BLAST server with a home page at http:// www.ncbi.nlm.nih.gov/BLAST/. Blast is a heuristic programme that finds short matches between two sequences and attempts to start alignments from these 'hot spots'. In addition to performing alignments, BLAST provides statistical information to help decipher the biological significance of the alignment.

Karlin and Altschul (1990) presented a theory that provided precise numerical formulae for assessing the statistical significance of any region with high aggregate score. A second class of results described the composition of high-scoring segments. In certain contexts, these permitted the choice of scoring system, which was "optimal" for distinguishing biologically relevant patterns.

Tatusova and Madden (1999) described 'BLAST 2 sequences' as a new BLAST based tool for aligning two protein or nucleotide sequences. BLAST 2 sequences utilized the BLAST algorithm for pairwise DNA-DNA or protein-protein comparison. A World Wide Web version of the programme can be used interactively at the NCBI site (http://www.ncbi.nlm.nih.gov/gorf/bl2.+++html).

Deng *et al.* (2007) reported that ViroBLAST was a stand-alone BLAST web interface for nucleotide and amino acid sequence similarity searches. It extended the utility of BLAST to query against multiple sequence databases and user sequence datasets, and provided friendly output to easily parse and navigate BLAST results. ViroBLAST was readily useful for all research areas that require BLAST functions and was also available online.

2.4.2 Clustal W

Clustal W is a multiple sequence alignment programme for nucleotide and protein sequences. Clustal uses a method called pair wise progressive sequence alignment. This heuristic method first does a pairwise sequence alignment for all the sequence pairs that can be clustered from the sequence set. A dendrograph of the sequence is then done according to the pair wise similarity of the sequences. Finally multiple sequence alignment is constructed by aligning sequences in the order, defined by the guide tree. It is available at http://www2.ebi.ac.uk/clustalw/

Higgins and Sharp (1988) described an approach for performing multiple alignments of large numbers of amino acids or nucleotide sequences. The method was based on first deriving a phylogenetic tree from a matrix of all pair wise sequence similarity scores, obtained using a fast pair wise alignment algorithm.

Then the multiple alignments were achieved from a series of pair wise alignments of clusters of sequences, following the order of branching in the tree.

Thompson *et al.* (1994) stated that the sensitivity of the commonly used progressive multiple sequence alignment method had been greatly improved for the alignment of divergent protein sequences through sequence weighting, position specific gap penalties and weight matrix choice. These modifications are incorporated in to a new programme CLUSTAL W which was freely available.

Chenna *et al.* (2003) mentioned that the Clustal series of programmes were widely used in molecular biology for the multiple alignments of both nucleic acid and protein sequences and for preparing phylogenetic trees.

2.4.3 Software for Three Dimensional Structure Predictions of Proteins

Lambert *et al.* (2002) presented EsyPred3D, a new automated homology modeling programme. The method got benefit of the increased alignment performances of a new alignment strategy. Alignments were obtained by combining, weighing and screening the results of several multiple alignment programmes. The final three dimensional structure was built using the modeling package MODELLER.

SWISS MODEL workspace is a web-based integrated service dedicated to protein structure homology modeling. It assists and guides the user in building protein homology models at different levels of complexity (Arnold *et al.*, 2006).

According to Raman *et al.* (2008), protein structure prediction and refinement were carried out using the open source software Rosetta developed in David Baker's research group at the University of Washington. Rosetta was a software package and a protein structure prediction algorithm. Core of the Rosetta algorithm are potential energy functions for computing the interactions within and between micro molecules, as well as optimizing methods for determining the lowest energy structure for an amino acid sequence.

Materials and Methods

3. MATERIALS AND METHODS

3.1 EXPERIMENTAL ANIMALS

The study was carried out on Indian Elephants (*Elephas maximus indicus*) maintained in captivity in and around the state of Kerala.

3.2 COLLECTION OF BLOOD SAMPLES

Four blood samples (5 ml each) were collected, from the ear vein aseptically using sterile disposable syringes and transferred immediately into sterile centrifuge tubes containing 0.5 *M* EDTA (1 mg/ml) as anticoagulant. After collection of blood, tubes were capped and shaken gently to facilitate thorough mixing of blood with anticoagulant. The blood samples were brought in ice to the laboratory without delay and stored at 4°C till processed.

3.3 ISOLATION OF DNA FROM WHOLE BLOOD

Genomic DNA was extracted from whole blood using the standard phenol-chloroform extraction procedure (Sambrook *et al.*, 1989) with modifications. The procedure* was as follows:

1. Five ml of blood collected in a 15 ml centrifuge tube was centrifuged at 4000 rpm for 10 min and the plasma was discarded, leaving erythrocytes and leucocytes.

2. Two to three volumes of ice-cold RBC lysis buffer were added and mixed well. The entire mixture was incubated in ice with occasional shaking for 10 min for the complete lysis of erythrocytes.

3. The tubes were centrifuged at 4500 rpm for 10 min to pellet the contents. The leucocytes were recovered as pellet and the red coloured supernatant, containing the lysed erythrocytes and haemoglobin was discarded.

**Composition and methods of preparation of reagents and buffers are provided in Annexure-I.*

4. Steps 2 & 3 were repeated, till the pellet was clear without any unlysed erythrocytes.

5. The white cell pellet was washed twice in 10 ml of Tris buffered saline by vigorous vortexing, followed by centrifugation at 4000 rpm for 10 min.

6. The washed white cell pellet from 5 ml of blood was completely resuspended by vortexing in 5 ml of saline EDTA buffer so that no clumps remained. To this cell suspension, 25 µl of proteinase-K (20 mg/ml in water) and 0.25 ml of 20 per cent SDS were added and mixed gently and incubated at 50°C in a water bath with occasional shaking for a period of at least three hours.

7. The digested samples were cooled to room temperature and 300 µl of 5 *M* sodium chloride solution was added and mixed well by vortexing. An equal volume of Tris saturated phenol (pH 7.8) was added and mixed gently by inversion of the tubes for 10 min and centrifuged at 4500 rpm for 15 min.

8. The aqueous phase containing DNA was collected in fresh tubes, to which an equal volume of saturated phenol:chloroform:isoamyl alcohol (25:24:1) was added. The contents were mixed thoroughly by inversion of the tube and centrifuged at 4500 rpm for 15 min.

9. The aqueous phase was collected in fresh tubes, to which an equal volume of chloroform:isoamyl alcohol (24:1) was added, mixed and centrifuged at 4500 rpm for 15 min and this step was repeated.

10. The upper aqueous phase was carefully transferred to a sterile 50 ml beaker and 1/10th volume of 3 *M* sodium acetate (pH 5.5) was added and mixed well.

11. An equal volume of chilled isopropyl alcohol was added and the precipitated DNA was spooled out using a fresh micropipette tip, washed in 70 per cent ethanol and air-dried.

12 Dried DNA was resuspended in 0.5 ml of Tris EDTA buffer and stored at -20°C.

3.4 DETERMINATION OF YIELD AND PURITY OF DNA

Twenty microlitres of the DNA stock solution was diluted to 2 ml with sterile distilled water, giving a dilution of 100 times. Optical densities (OD) were measured at 260 nm and 280 nm using a 2 ml cuvette in a UV spectrophotometer (Jenway, UK). Sterile distilled water was used as blank. Yield and purity of DNA samples were estimated as follows.

3.4.1 Yield of DNA Samples

An OD of one at 260 nm wavelength corresponds to approximately 50 µg/ml of double stranded DNA. Concentration of DNA stock solution was calculated as follows.

Concentration of DNA stock solution (µg/ml) = OD_{260} x Dilution factor x 50

3.4.2 Purity of DNA Samples

The purity of DNA samples was assessed by estimating the ratio between the readings at 260 and 280 nm. Pure DNA samples have OD_{260}/OD_{280} ratios of 1.8 and above.

3.4.3 Checking Quality and Quantity of DNA

The quality of DNA samples was first assessed electrophoretically on one per cent agarose gel in 1X TAE buffer in a horizontal submarine gel

electrophoresis unit. The quantity of DNA was estimated by multiplying the concentration and the volume of DNA stock solution.

3.5 PCR AMPLIFICATION OF ELEPHANT GROWTH HORMONE GENE

3.5.1 Template DNA Preparation for PCR

Working solutions of DNA samples were prepared from the DNA stock solution by diluting with sterile triple distilled water to get a final concentration of 50 ng/µl and were stored at -20 °C. Two microlitres of this working solution was used for every 50 µl PCR reaction.

3.5.2 Design of Primers

The DNA sequence information on growth hormone (GH) gene of elephants is not available in literature. Therefore different strategies were adapted to amplify the gene. In one of these approaches, different forward and reverse primers were designed based on DNA sequence information available in other species using the software Primer3 (v.0.4.0) available at the site *http://frodo.wi.mit.edu.* The primers designed on the basis of the bovine sequence (Accession # M57764; Gordon *et. al.*, 1983) downloaded from the National Centre for Biotechnology Information (NCBI) site (*http://www.ncbi.nlm.nih.gov*) are presented in the Table 3.1.

The PCR amplification of the gene was attempted by using different combinations of forward and reverse primers in order to get the amplification of the complete gene or two overlapping fragments.

Table 3.1 Primers designed on the basis of the bovine sequence (Accession #M57764)

Sl. No.	Name	Sequence	Starting Nucleotide Position
1	GHELP-F666	5'-CCAGTTCACCAGACGACTCA -3'	666
2	GHELP-F1726	5'-ACTGCTCCTCATCCAGTCGT -3'	1726
3	GHELP-R1745	5'-ACGACTGGATGAGGAGCAGT -3'	1745
4	GHELP-R2403	5'-TGGGAGTGGCACCTTCCAGGGTCA-3'	2403
5	GHELP-R2439	5'-GCGATGCAATTTCCTCATTT -3'	2439

3.5.2.1 Design of Degenerate Primers

Another approach was to design degenerate primers for the amplification of GH gene based on the known amino acid sequence of the gene of African elephants (Wallis, 2008). The GH of African savanna elephant is a protein hormone of about 190 amino acids. The amino acid sequence was reverse translated into the nucleotide sequence using Reverse Translate tool at the Sequence Manipulate Suit site (http://www.bioinformatics.org/sms2/) to obtain a 570 bp nucleotide sequence corresponding to the exonic region of the gene.

Reverse Translate accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned as output. Codon usage accepts one or more DNA sequences and returns the number and frequency of each codon type. Since the programme also compares the frequencies of codons that code for the same amino acid, it can be used to assess whether a sequence shows a preference for particular synonymous codons.

The forward and reverse primers were designed online from the 570 bp consensus sequence derived, using the software Primer3 (v. 0.4.0). Primers statistics were checked with the help of PCR Primer Stats available at Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/pcr_primer_stats.html).

The different degenerate primers designed on the basis of the above consensus sequence are presented in Table 3.2.

Table 3.2 Degenerate primers designed on the basis of known amino acid sequence of the gene of African elephant

Sl. No.	Name	Sequence	Starting Nucleotide Position
1	13ELPGH-F	5'- CCGCTGAGCAGCCTGTTT-3'	13
2	24ELPGH-F	5'-CCTGTTCGCCAACGCCGTGCTGA-3'	24
3	527ELPGH-R	5'-CTCAGGTAGGTCTCGGCCTTGTGCA-3'	527
4	539ELPGH-R	5'- CATTTCATCACGCGCAGATA -3'	539

3.5.2.2 Design of Primers Based on the Flanking Sequences of Goat Growth Hormone

Different forward and reverse primers were also designed separately based on the 5' and 3' flanking sequences of goat GH gene sequences described by Kioka *et. al.* (1989). The different primers designed using this strategy is presented in the Table 3.3.

Table 3.3 Primers designed based on the 5′ and 3′ flanking sequences of goat growth hormone gene

Sl. No.	Name	Sequence	Starting Nucleotide Position
1	FOR-11	5′-ATGACCCCAGAGAAGGAA-3′	-167
2	FOR-141	5′-GCCAGGGTATAAAAAGGG-3′	-37
3	REV-32	5′-GGGGAGGGGTAACAACAGATG-3′	1721
4	REV131	5′-AGAATGACACCTACTCAGACAAT-3′	1820
5	REV147	5′-CCCCACCCCCTAGAATAGAA-3′	1836

The PCR amplification was attempted by using different combinations of forward and reverse primers in order to get the amplification of the complete gene and part of the 5′ and 3′ flanking sequences.

3.5.3 Synthesis and Dilution of the Primers

The primers were custom synthesized commercially (Integrated DNA Technologies, Inc., USA) and obtained in lyophilized form. They were reconstituted in sterile triple distilled water to make a stock solution of 200 p*M*/µl concentration. The solutions were incubated at room temperature for one hour and then stored at -20°C. Working solutions of the primers were prepared by dilution from the stocks.

3.5.4 PCR Conditions

The PCR assay was performed in 200 µl reaction tubes with a total volume of 50 µl, containing 100 ng of DNA template, 1X PCR buffer, 1.75m*M* $MgCl_2$, 50p*M* of each primer, 200µ*M* of each dNTP and 1unit Taq DNA polymerase. A negative control, containing all reaction components except template DNA was also set up to check any contamination, if any.

The tubes were spun briefly and placed in the thermal cycler. The cycling conditions for DNA amplification included an initial denaturation at 94°C for 3 min and 10 cycles each of denaturation at 94°C for 1 min, annealing at 54.2°C for 1 min and extension at 72°C for 3 min, followed by another 20 cycles of denaturation at 94°C for 1 min, annealing at 54.2°C for 1 min and extension at 72°C for 3 min plus addition of 20 sec each per cycle, followed by final extension at 72°C for 10 min. The PCR products were kept at -20°C till further analysis.

3.5.5 Agarose Gel Electrophoresis of Amplified DNA

The PCR products were checked in one per cent agarose gel in 1 X TAE buffer in a horizontal submarine electrophoresis unit. Appropriately sized gel tray was cleaned and sealed. The tray was placed on a leveled surface and comb kept in proper position in the tray. One per cent agarose in TAE buffer was heated in a microwave oven until it was a clear solution, cooled to 60°C and 0.5 μg/ml of ethidium bromide was added and mixed well. The mixture was cooled to hand bearable temperature and carefully poured into the gel tray avoiding air bubbles. After polymerization, the comb and sealing were removed gently and the tray was immersed in the buffer tank (Amersham Pharmacia Biotech., USA) containing 1X TAE buffer. Two microlitres of the PCR product was mixed with a small drop of 6 X gel loading buffer and loaded into the wells carefully. As a molecular size standard, 500 bp DNA ladder (Bangalore Genei Pvt. Ltd.) was also mixed with gel loading buffer and loaded into one of the wells.

Electrophoresis was carried out at 2V/cm until the bromophenol blue dye migrated more than $2/3^{rd}$ of length of the gel. The gel, which was stained with ethidium bromide, was visualized under a UV-trans illuminator (Hoefer, USA) and checked for amplification of target DNA.

3.6 GEL PURIFICATION OF PCR PRODUCT

For gel purification of PCR products, a preparative gel electrophoresis was carried out in the same manner as for checking the PCR product. The adjacent wells of the comb were combined with the help of cello tape so that approximately 50 µl of the PCR products could be loaded into a single well. The DNA band was cut from the agarose gel using a fine scalpel under UV light. The PCR product was purified from the gel slice using the Ultra Clean™ GelSpin™ DNA purification Kit (Catalog # 12400-50, MO BIO Laboratories, Inc. USA) as per the manufacturer's instructions as follows:

1. The gel band weight was determined using a separate tube. The steps were carried out directly in the spin filter as the gel band weighed only 0.181 g.
2. Placed the gel in the spin filter basket and ensured that it was resting on the white filter membrane so that the gel would be completely melted in the spin filter.
3. Added 0.54 ml gel bind buffer (i.e. three volumes of the gel band, to the gel slice and ensured that the gel was submerged in the gel bind buffer and closed the lid.
4. Incubated for 2 min at 55°C, inverted once and incubated until the gel was melted.
5. Inverted once to mix and centrifuged the spin filter for 10 sec at 10,000 x *g* so that the melted gel and unwanted salts flew through the spin filter membrane as the DNA binds to the membrane.
6. Removed the spin filter and vortexed the collection tube for 5 sec to mix the flow through so as to make the salt concentration homogenous.
7. Reloaded all the liquid from the collection tube back on to the spin filter, allowing any unbound DNA bind to the spin filter by passing the DNA

through the spin filter second time, thus increasing the yield of DNA from the melted agarose.

8. Centrifuged for 10 sec at 10000 x *g* and discarded flow through liquid and replaced spin filter basket, thus allowing the entire available DNA being bound to the spin filter, the discarded liquid containing only melted agarose and salt.
9. Added 300 µl of Gelwash buffer composed of ethanol solution that keeps the DNA bound to the spin filter while washing away residual salt and melted agarose.
10. Centrifuged for 10 sec at 10000 x *g* and discarded flow through and centrifuged again for 30 sec at 10000 x *g* to dry out any residual gel wash because the ethanol contained in the buffer can inhibit downstream applications for the recovered DNA.
11. Carefully transferred filter basket to a clean collection tube provided and added 50 µl of elution buffer (10 m*M* Tris) or water directly on to the center of the white spin filter membrane. The elution buffer gets in contact with the entire surface on the spin filter membrane for efficient recovery of the DNA. The DNA is released from the spin filter membrane at this point because there is no salt to keep it bound.
12. Centrifuged for 30 sec at 10000 x *g* and discarded the filter basket. The DNA flows through the spin filter membrane and into the collection tube which was stored at -20 °C for further work.

3.7 CLONING GROWTH HORMONE GENE IN pGEM®-T EASY VECTOR

The gel purified PCR product was cloned using the pGEM®-T Easy Vector System I (Catalog # A1360, Promega Corporation, Madison, USA). The TA cloning is a simple and an efficient method for cloning of PCR products, which exploits the terminal transferase activity of certain thermophilic DNA polymerases including the Taq DNA polymerase. Taq DNA polymerase has non-

template dependent activity, which preferentially adds a single adenosine to the 3′-ends of a double stranded DNA molecule and thus most of the molecules amplified by Taq DNA polymerase possess single 3′-dA overhang. The use of a pGEM®-T EASY VECTOR which has single 3′-dT overhangs on both ends allows direct, high-efficiency cloning of PCR products carrying 3′–dA overhang, facilitated by complementarity between the PCR product's 3′-dA overhangs and vector's 3′-dT overhangs.

3.7.1 The pGEM®-T Easy Vector System

This vector system is convenient for the cloning of PCR products. The vectors are prepared by cutting the pGEM®-T Easy Vectors with *EcoRV* and adding a 3′ terminal thymidine to both ends. They offer all of the advantages of the pGEM®-T Vector Systems with the added convenience of recognition sites for *EcoRI* and *NotI* flanking the insertion site. Thus several options for removal of the desired insert DNA with a single restriction digestion are provided.

The high-copy-number pGEM®-T Easy Vector contain T7 and SP6 RNA polymerase promoters flanking a multiple cloning site (MCS) within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows recombinant clones to be directly identified by color screening on indicator plates. This vector contains multiple restriction sites within the MCS. It also contains the origin of replication of the filamentous phage f1 for the preparation of single-stranded DNA (ssDNA). The map details of the vector are shown in Fig. 3.1.

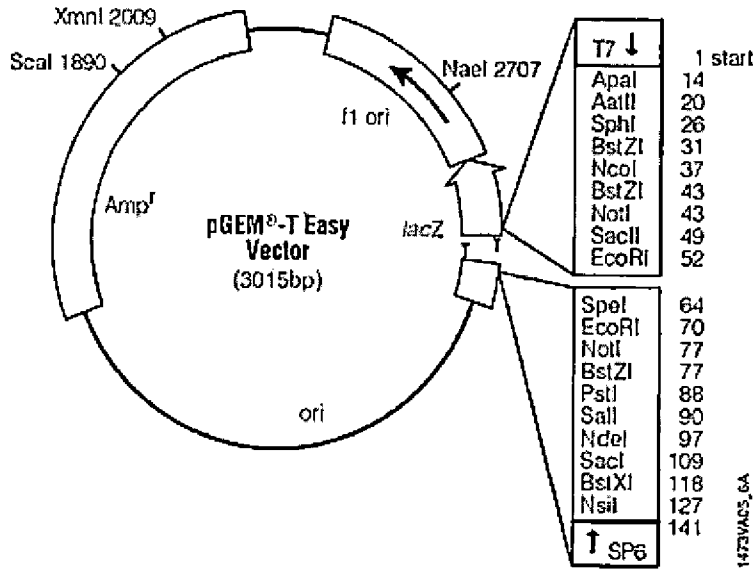


Fig. 3.1 The map details of pGEM®-T Easy Vector showing various sequence reference points

3.7.1.1 The Sequence Reference Points of pGEM®-T Easy Vector

T7 RNA Polymerase transcription initiation site	1
Multiple cloning region	10–128
SP6 RNA Polymerase promoter (–17 to +3)	139–158
SP6 RNA Polymerase transcription initiation site	141
pUC/M13 Reverse Sequencing Primer binding site	176–197
LacZ start codon	180
Lac operator	200–216
ß-lactamase coding region	1337–2197
Phage f1 region	2380–2835
Lac operon sequences	2836–2996, 166–395
pUC/M13 Forward Sequencing Primer binding site	2949–2972
T7 RNA Polymerase promoter (–17 to +3)	2999-3

3.7.1.2 Materials Provided in the Kit (pGEM®-T Easy Vector System I)

Material	Quantity	Store at
pGEM®-T Easy Vector (50ng/µl)	1.2µg	-20 °C
Control Insert DNA (4ng/µl)	12 µl	-20 °C
T4 DNA Ligase	100 µl	-20 °C
2X Rapid Ligation Buffer, T4 DNA Ligase	200 µl	-20 °C

3.7.2 Additional Materials Required for Cloning

1. Competent cells
2. Isopropyl β-D-1-thiogalactopyraranoside (IPTG)
3. 5-Bromo-4-Chloro-3-indolyl-β-D-galactopyranoside (X-gal) (20 mg/ml in Dimethyl formamide)
4. Ampicillin (100 mg/ml)
5. TE buffer (10mM Tris Cl, 1m*M* EDTA, pH 8.0)
6. LB medium and agar

3.7.3 Ligation of Growth Hormone Gene into the pGEM®-T Easy Vector

The pGEM®-T Easy Vector, the control insert and the gel purified PCR products were thawed and centrifuged briefly to collect the contents at the bottom of the vials.

3.7.3.1 Optimizing Insert: Vector Molar Ratios

The pGEM®-T Easy Vector System has been optimized using a 1:1 molar ratio of the control insert DNA to the vectors. The pGEM®-T Easy Vectors are approximately 3 kb and are supplied at 50 ng/µl. To calculate the appropriate amount of PCR product (insert) to include in the ligation reaction, the following equation was used.

$$\frac{\text{ng of vector(50)} \times \text{kb size of insert (1716)}}{\text{kb size of vector (3015)}} \times \text{insert:vector molar ratio (1:1)} = \text{ng of insert (28.5)}$$

For the 1712 bp amplicon, the amount of PCR product required for a 1:1 ratio was calculated as 28.2 ng. After calculating the amount required, the PCR product was diluted to the required concentration in such a way that the volume of PCR product did not exceed 30 per cent of the total volume.

3.7.4 Ligation Reaction

The ligation reactions were performed in three different tubes, one each for the PCR product, the positive control insert and the one to check the self-ligation of the linearised vector as follows.

Components	PCR product	Positive Control	Self - ligation Control
2X Rapid Ligation Buffer, T4 DNA Ligase	5µl	5µl	5µl
pGEM®-T Easy Vector (50ng)	1µl	1µl	1µl
PCR product	2 µl	-	-
Control Insert DNA	–	2µl	-
T4 DNA Ligase (3 Weiss units /µl)	1µl	1µl	1µl
Deionized water to a final volume of :	**10µl**	**10µl**	**10µl**

The reaction components were mixed and spun briefly to collect contents at the bottom. It was incubated at room temperature (25 °C) for one hour and stored at -20 °C till the time of transformation.

3.7.5 Transformation

The pGEM®-T Easy Vector ligated PCR products were transformed to *E. coli* using the Genei™ Transformation Kit (Catalog # KT07, Bangalore Genei Pvt. Ltd) as per the manufacturer's instructions as follows.

3.7.5.1 Materials Provided in the Kit

Materials	Quantity	Store at
Ampicillin	100 mg	4 °C
0.1 *M* $CaCl_2$ (Sterile)	25 ml	4 °C
Host (modified *E. coli*)	1 vial	4 °C
IPTG	1 ml	-20 °C
Plasmid DNA	25 µl	-20 °C
X-gal	1 ml	-20 °C
LB Broth	15 g	RT
Agar	5 g	RT

3.7.5.2 Additional Materials Required

1. Equipment: Spectrophotometer, Refrigerated centrifuge, Incubator shaker and water bath
2. Glassware: Conical flask, Petri plates, Pipettes, Spreader
3. Distilled water, Capped Centrifuge tubes, Crushed ice, Micropipette

3.7.5.3 Preparation of Competent Cells

1. The host cells were revived by adding 0.1 ml of the LB media in to the lyophilized vial and a loopful of suspension was streaked on to LB plates and incubated at 37°C overnight.
2. Inoculated a single colony into 5 ml of LB medium and incubated in a shaker at 37°C overnight.
3. Inoculated 1 ml of overnight culture into 100 ml LB medium in a 1 L conical flask and incubated at 37°C in a shaker for about three hours until the OD A_{600} reached 0.23-0.26.

4. Chilled the culture flask on ice for 10-20 min and transferred the culture aseptically into sterile centrifuge tubes and spun down at 6000 rpm for 8 min in a refrigerated centrifuge at 4°C.

5. Discarded the supernatant and to the cell pellet added 15 ml of cold 0.1 *M* $CaCl_2$ solution under aseptic conditions. The cell pellet was resuspended gently in the solution using a pre-chilled pipette, keeping the tubes in ice.

6. The tube was incubated in ice for 30 min and centrifuged at 6000 rpm for 8 min at 4°C.

7. The supernatant was discarded and the cell pellet was resuspended gently in 0.6 ml of cold 0.1*M* $CaCl_2$ solution.

8. The competent cells were aseptically split in to 100 µl aliquots into six prechilled vials, taking care not to remove the centrifuge tubes from ice.

9. The competent cells prepared were used for transformation experiment immediately, as the efficiency of transformation drops on storage at temperature higher than -70°C.

3.7.5.4 Transformation Protocol

1. The three different ligation mixtures, containing the PCR product, the control insert and the self-ligation control were added to three aliquots of 100 µl of competent cells, gently tapped and incubated on ice for 20 min.
2. Heat shocked the cells by placing the vials in 42 °C water bath for 2 min and then returned the vials to ice to chill for 5 min.
3. Added 1 ml of LB broth aseptically to the vials and incubated at 37 °C in a shaker for an hour so as to allow the bacteria to recover and express the antibiotic resistance.

4. Labeled three LB-Amp plates with X-gal and IPTG as a, b, c and pipetted 100 µl of LB broth onto each plate and added 25, 50 and 100 µl of the transformed cells to plates a, b and c respectively. Mixed well and spread using a spreader.
5. Repeated the step 4 for other two aliquots of competent cells transformed.
6. Another LB-Amp plate was plated with 100 µl of competent cells that has not been transformed, to check for any cell contamination and labeled as control plate.
7. Incubated the plates at 37 °C overnight.

3.7.5.5 Screening of Transformants

The pGEM®-T Easy Vector in the experiment carries ampicillin resistance marker that enables only transformed cells to grow on LB-Amp plates. The ampicillin-sensitive non-transformants do not produce colonies on the selective medium.

3.7.5.6 Screening of Recombinants

The recombinant clones among the transformed cells were identified by the insertional-inactivation of the β-galactosidase (LacZ) gene in the multicloning site of the pGEM®-T Easy Vector, referred to as the Blue–White Screening. The clones carrying non-recombinant plasmids were blue in colour in the presence of X-gal and IPTG whereas the recombinant clones appeared white in colour.

3.8 ISOLATION OF RECOMBINANT PLASMID FROM POSITIVE CLONES

The recombinant plasmid carrying the GH gene insert was isolated from the white clones by a modified SDS-alkaline lysis method using the AxyPrep Plasmid Miniprep Kit (Axygen Biosciences, USA) as per the manufacturer's instructions as follows.

3.8.1 Materials Provided in the Kit

Materials	Description
AxyPrep Column	Special DNA binding columns
2 ml microfuge tube	-
1.5 ml microfuge tube	-
RNase A	-
Buffer S1	Resuspension buffer, to which RNase A is added and stored at 4 °C
Buffer S2	Lysis buffer, Storage: RT
Buffer S3	Neutralisation buffer, Storage: RT
Buffer W2 concentrate	Desalting buffer, to which absolute ethanol is added and stored at RT
Eluent	2.5 mM Tris-Cl, Storage: RT

3.8.2 Protocol for Plasmid Isolation

1. Inoculated 3.5 ml of LB media with white clones and incubated at 37 °C in a shaker overnight. The culture was centrifuged at 12,000 X *g* for 1 min to pellet the bacteria and decanted the supernatant.

2. The bacterial pellet was resuspended in 250 µl of Buffer S1 by vortexing, ensuring complete resuspension of the bacteria.

3. To the cell suspension, 250 µl of Buffer S2 was added and mixed gently by inverting the tube for 4-6 times.

4. To the above suspension, 350 µl of Buffer S3 was added and mixed gently by inverting 6-8 times, followed by centrifugation at 12,000 X *g* for 10 min to clarify the lysate.

5. Placed the AxyPrep column into an uncapped 2 ml microfuge tube and transferred the clarified supernatant from step 4 to the AxyPrep column. It was then spun at 12,000 X *g* for 1 min.

6 Pipetted 700 µl of Buffer W2 into the AxyPrep column and centrifuged at 12,000 X *g* for 1 min.

7. The filtrate from the microfuge tube was discarded and AxyPrep column was put back into the microfuge tube and the step 6 was repeated.

8. The filtrate was discarded the microfuge tube and the AxyPrep column was put back into the microfuge tube and centrifuged at 12,000 X *g* for 1 min.

9. The AxyPrep column was transferred to a clean 1.5 ml microfuge tube and the purified plasmid DNA was eluted from the column by adding 70 µl of eluent pre-warmed to 60 °C, followed by centrifugation at 12,000 X *g* for 1 min, after keeping at room temperature for 1 min.

10. The filtrate containing pure plasmid DNA was stored at -20 °C till further use.

3.8.3 Agarose Gel Electrophoresis

The recombinant plasmid DNA was checked in one per cent agarose gel using 1 X TAE buffer in a horizontal submarine electrophoresis unit as explained earlier.

3.9 DNA SEQUENCING

Sequencing was carried out commercially (Bioserve Hyderabad, Pvt. Ltd.) by the dideoxynucleotide sequencing method using an automated DNA sequencer (Applied Biosystems, USA).

Primer walking was the method of choice as the DNA of interest was ~ 1.7 kb that cannot be covered with a single read. The initial sequencing was performed from each end using M13 universal forward and M13 universal reverse primers. New sequencing primers were custom designed from the initial reads and sequencing reactions were continued to obtain further sequence of the insert. This process was repeated until the entire length of the insert was covered. From the M13 universal forward end, the primers designed for sequencing were 96 INS 1F and 96 INS 2F. From the M13 universal reverse end, the only primer designed for primer walking was 96 INS 1R as the reverse direction sequence was having a strong C stretch, sequencing was dropping at this stretch of "C" and so it was difficult to proceed from this direction. The primers used for sequencing the insert completely are shown below.

1. M13 universal forward
2. 96 INS 1F- (5′-TAG GAT GGG GGC ACT AAC TCT G-3′)
3. 96 INS 2F- (5′- CCC GAG GGC CAG AGG TAC T -3′)
4. 96 INS 1R- (5′- GTA ATA GAT GCC CAG CAG CTA G -3′)
5. M13 universal reverse

The direction of primer walking and the sequence of the different fragments were in the following order.

M13 universal forward -----> 96 INS 1F-----> 96 INS 2F -----> 96 INS 1R -----> M13 universal reverse

The complete sequence of the PCR product was arrived at by combining together the sequences of all the three forward reactions with the reverse complement of the sequences of two reverse primers followed by deletion of the overlapping sequences.

3.10 SEQUENCE ANALYSIS

3.10.1 DNA Sequence Analysis

The 1712 bp sequence obtained through sequencing was analysed by BLAST search at NCBI site for homology using BLASTn programme (http://www.ncbi.nlm.nih.gov/BLAST). The raw sequence was annotated based on the porcine GH gene (Accession # M17704). The intron, exon and other structural and functional regions of the sequence were established. For interspecies nucleotide sequence homology between exonic regions of the GH gene in Indian elephant with other species, individual exons were compared by BLAST2 programme. The BLASTn programme compares a nucleotide query sequence against a nucleotide sequence database. The 1712 bp nucleotide sequence of the elephant GH gene was submitted as query sequence and selected the MegaBLAST programme for searching highly similar sequences from the DNA database.

3.10.2 Multiple Sequence Analysis of DNA

The multiple sequence alignment was done by the widely used computer programme EBI tool ClustalW (http://align.genome.jp/). This programme is available from European Bioinformatics Institute (EBI) ftp server. The input sequences were loaded in FASTA format. Using this programme the elephant sequence was compared simultaneously with sequences of other species to highlight the areas of homology and divergence.

3.10.3 Submission of the Sequence to GenBank

The sequence of GH gene of Indian elephant was submitted to the GenBank using BankIt option (http://www.ncbi.nlm.nih.gov/BankIt/) at the NCBI site and obtained the unique accession number for the new sequence.

3.10.4 Protein Sequence Analysis and Structure Prediction

The exonic regions were compiled together and using the programme ExPASy Translate tool (http://www.expasy.ch/tools/dna.html) the amino acid sequence of the signal peptide and the amino acid residues of the polypeptide chain were predicted. The protein sequence was analysed by using the BLASTp programme at the NCBI site. This programme compares an amino acid query sequence against a protein sequence database. The predicted amino acid sequence was entered as query sequence and selected the PROTEIN database for the search.

A multiple sequence alignment of the predicted protein sequence was carried out by ClustalW programme with other mammalian species. A phylogramme showing the evolutionary relationship was also constructed from the multiple alignment results.

The secondary structure of the elephant GH was predicted using the PAPIA system (http://mbs.cbrc.jp/papia/papia.html) of Computational Biology Research Center, Japan. The amino acid sequence in one letter code was submitted for the structure prediction.

The three dimensional structure of the sequenced protein was predicted using the automatic homology or comparative modeling server, the Swiss Model Server (http://swissmodel.expasy.org/). This server builds three-dimensional models for proteins based on homologues of known structure. The amino acid sequence in one letter code was made as input. The predicted structure was viewed with RasMol, which is molecular visualization software for proteins, DNA and macromolecules.

Results

4. RESULTS

4.1 ISOLATION OF GENOMIC DNA

DNA samples were isolated from whole blood obtained from four Indian elephants.

4.1.1 Yield and Quality of DNA

The concentration and ratio of optical densities at 260 and 280 nm of DNA obtained from the whole blood sample selected for the experiment was 215 µg/ml and 1.71, respectively. The concentration and purity of DNA obtained from five ml of whole blood from the four samples are given in Table 4.1. The samples, checked for quality by agarose gel electrophoresis, appeared as single band without sheared fragments indicating good quality of DNA.

4.2 PCR AMPLIFICATION OF GROWTH HORMONE GENE

The primers for amplification of the growth hormone (GH) gene were designed on the basis of the bovine sequence (Accession # M 57764; Gordon *et. al.*, 1983), known amino acid sequence of the gene of African elephant (Wallis, 2008) and also based on the 5′ and 3′ flanking sequences of goat GH gene sequence described by Kioka *et. al.* (1989). The primers were custom synthesized for PCR amplification of the samples. PCR amplification was attempted on all the samples using the different primer pair combinations and the products were checked by agarose gel electrophoresis.

4.2.1 Different Combinations of the Designed Primers

For successful amplification of the PCR product, different combinations of the designed primers were tried in order to get the amplification of the complete gene or two overlapping fragments. The list of the different combinations of the primers used is presented in Table 4.2.

Table 4.1 Concentration and purity of DNA obtained from blood samples

Sample No.	OD_{260}	OD_{280}	OD_{260} /OD_{280}	Dilution factor (DF)	Conc.(µg/ml) A_{260} x DF x 50
1	0.048	0.028	1.71	100	215
2	0.055	0.034	1.61	100	198
3	0.043	0.027	1.59	100	209
4	0.068	0.055	1.23	100	195

Table 4.2 Different combinations of the forward and reverse primers tested to amplify the GH gene of Indian elephant

Sl. No.	Combinations of the forward and reverse primers	Expected product size
1	GHELP-F666 and GHELP-R2403	1738
2	GHELP-F666 and GHELP-R2439	1774
3	GHELP-F666 and GHELP-R1745	1080
4	GHELP-F1726 and GHELP-R2403	678
5	GHELP-F1726 and GHELP-R2439	714
6	13ELPGH-F and 527ELPGH-R	Can't predict
7	13ELPGH-F and 539ELPGH-R	Can't predict
8	24ELPGH-F and 527ELPGH-R	Can't predict
9	24ELPGH-F and 539ELPGH-R	Can't predict
10	FOR-11 and REV-32	Can't predict
11	FOR-11 and REV-131	Can't predict
12	FOR-11 and REV-147	Can't predict
13	FOR-141 and REV-32	Can't predict
14	FOR-141 and REV-131	Can't predict
15	FOR-141 and REV-147	Can't predict

4.2.2 Optimization of the PCR Protocol

The composition of the PCR, temperature and time of the cycles were optimized for the efficient amplification of the product from each set of primers. To determine the correct primer combination, magnesium chloride concentration, annealing temperature, Taq DNA polymerase, it took around 130 different PCR assays. Amongst all of the above set of primers used, only FOR-141 and REV-131 combination was successful in amplifying a product of the expected size. The reaction components optimized for the above PCR are presented in Table 4.3. The annealing temperature was optimized by the gradient option in the thermal cycler and 54.2°C was found to be optimum for the assay. The standardized temperature and cycling parameters are presented in Table 4.4.

The $MgCl_2$ concentration was found to be an important parameter for the efficient amplification of the gene. A concentration of 1.75 m*M* in the reaction mixture resulted in robust amplification of the product without any non-specific products.

4.2.3 Agarose Gel Electrophoresis of PCR Products

The PCR products were checked for amplification by electrophoresis on one per cent agarose gels using 500 bp markers as the size standard. The amplified products were found to be approximately 1.7 kb from all the DNA samples (Plate 1). There was no amplification in the negative control.

4.3 PURIFICATION OF DNA FROM AGAROSE GEL

The PCR products were purified from the gel slice using the Ultra Clean™ GelSpin™ DNA purification Kit (Catalog # 12400-50, MO BIO Laboratories, Inc. USA). The gel purification resulted in more than 50 per cent recovery of the PCR product. The purified PCR products, free of salts, buffers, enzymes, ethidium bromide and dNTPs, were collected in 50 µl elution buffer and stored at -20°C for the cloning assay.

Table 4.3 Standardized conditions for the PCR assay to amplify the GH gene of Indian elephant

Sl. No.	Parameter	Value
1.	Template DNA	100 ng
2.	$MgCl_2$	1.75 m*M*
3.	10X Reaction Buffer µl	5
4.	dNTPs	200 µ*M*
5.	Forward Primer	50 *pM*
6.	Reverse Primer	50 *pM*
7.	Taq DNA Polymerase (XT20 Taq DNA polymerase)	1 U
8.	Reaction Volume	50 µl

Table 4.4 Standardized temperature and cycling conditions for the PCR

Sl. No.	Parameter	Temperature/Time
1.	Initial Denaturation	94.0 ^{0}C for 3 min
2.	Denaturation	94.0 ^{0}C for 1 min
3.	Annealing	54.2^{0}C for 1 min
4.	Extension	72.0^{0}C for 3 min
5.	No. of Cycles	10
6.	Denaturation	94.0 ^{0}C for 1 min
7.	Annealing	54.2 ^{0}C for 1 min
8.	Extension	72 °C for 3 min + addition of 20 sec /cycle
9.	No. of cycles	20
10.	Final extension	72.0 ^{0}C for 10 min
11.	Hold	4 ^{0}C forever

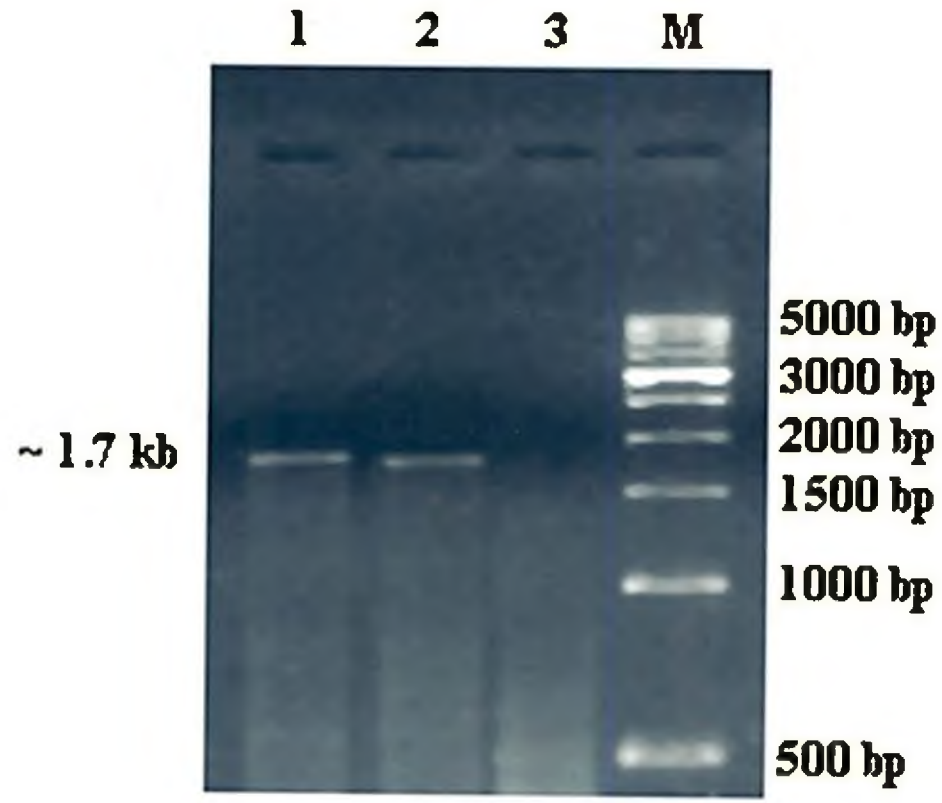


Plate 1 Gel showing amplified fragment of the Indian elephant GH gene
Lanes 1-2: ~ 1.7 kb amplified fragment of GH gene
Lane 3: Negative control
M : 500 bp ladder as molecular size marker

4.4 TA CLONING

The gel purified PCR product was cloned using the pGEM®-T Easy Vector System I (Catalog # A1360, Promega Corporation, Madison, USA). The ligation reactions were performed in three different tubes, one each for the PCR product, the positive control insert and to check the self-ligation of the linearised vector. After ligation, the recombinant vectors containing the PCR product and the control insert were checked by electrophoresis. The sizes were approximately 4.7 and 3.5 kb, respectively indicating successful ligation reaction (Plate 2).

4.4.1 Transformation

Competent cells for transformation of cloned vector were prepared by Genei™ Transformation kit (Catalog # KT 07, Bangalore Genei Pvt. Ltd.) as per the manufacturer's protocol. Transformation of the competent cells was carried out with three different ligation mixtures, namely the one with PCR product, the positive control insert and the self-ligation control. In each case, the transformants were plated onto three LB plates with 25, 50 and 100 µl of transformed cells. In the case of PCR product, although clones were grown in all the three plates, the one with 100 µl showed sufficient number uniformly spread colonies with more than 40 per cent of positive colonies (Plate 3). The control insert plate with 100 µl showed more than hundred colonies of which more than 60 per cent were white and the remaining being blue colonies. The number of colonies was less in the other control plates in proportion to the volume of the transformed cells added. The plates with self-ligation yielded only a few blue colonies.

4.5 ISOLATION OF PLASMID FROM POSITIVE CLONES

The recombinant plasmid carrying the GH gene insert was isolated from the genuine white colonies after propagating in LB-ampicillin broth for 12 hours. The isolated plasmid DNA was checked by agarose gel run to ensure that the

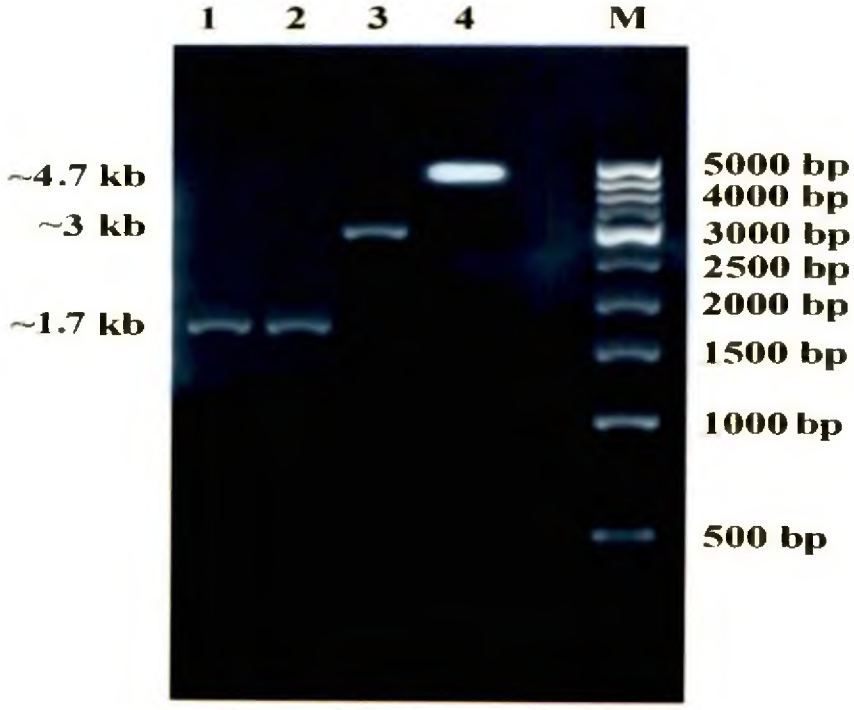


Plate 2 Gel showing purified PCR product, vector and vector with insert DNA

Lane 1-2: Gel purified PCR product (~1.7 kb)
Lane 3: pGEM-T Easy Vector (~3 kb)
Lane 4: Recombinant vector (~4.7 kb)
Lane M: 500 bp DNA ladder

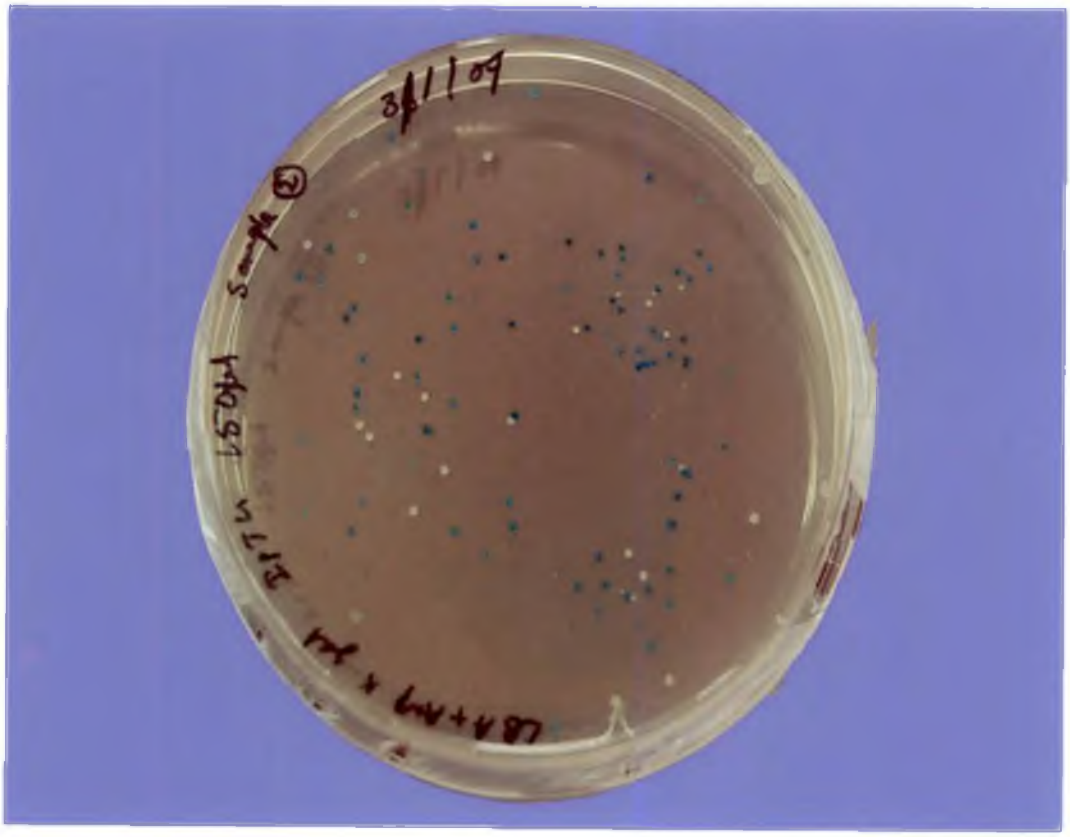


Plate 3 Result of transformation. Petri plate showing recombinant white clones and negative blue clones

recombinant plasmid contain the target gene insert. The plasmid band was found to be having a molecular weight of approximately 4.7 kb ensuring that the recombinant vector carries the DNA of interest. The purified plasmid DNA yield was found to be sufficient for subsequent sequencing steps.

4.6 DNA SEQUENCING

The T-vector cloned amplified product was sequenced by the dideoxy termination method. The sequencing results revealed 1712 bp fragment of the gene, containing an initiation codon 'ATG' for translation. For complete sequencing, primer walking sequencing method was performed using three additional primers in addition to the M13 universal forward and M13 universal reverse primers. The complete nucleotide sequence of the GH gene is presented in Fig. 4.1

4.7 SEQUENCE ANALYSIS

4.7.1 DNA Sequence Analysis

The BLAST analysis of the nucleotide sequence of the GH gene of Indian elephant obtained through sequencing revealed 77 to 95 per cent identity in three different regions with that of dolphin (*Delphinus delphis*; Accession # AJ492191.1). The sequence was also found to be having 75 to 96 per cent identity with that of pig (Accession # M17704.1) and 77 to 81 per cent with that of hippopotamus (*Hippopotamus amphibius*; Accession # AJ575422.1) GH genes, respectively (Table 4.5). The annotation of the GH gene sequence based on porcine GH gene sequence revealed the presence of 5 exons of 10, 161, 117, 162 and 201 bp separated by 4 introns of 269, 215, 123 and 249 bp (Fig. 4.1). A comparison of the lengths of exons and introns of the gene in different species is presented in Table 4.6. The homology of the exonic regions of elephant with different species ranged from 88 to 96 per cent as shown in Fig. 4.2. The results of ClustalW multiple nucleotide sequence alignment of the complete sequence showed a high degree of homology with a score of 78 with pig and dolphin, 76 with hippopotamus and Indian cattle while only 66 with buffalo.

Table 4.5 Results of BLASTn of the growth hormone (GH) gene of Indian elephant's nucleotide sequence showing the percentage of identity with other species

Accession	Description	Maximum score	Total score	Query coverage (per cent)	E value	Maximum identity (per cent)
# AJ492191.1	*Delphinus delphis* GH gene	442	1098	79	3e-120	95
# M17704.1	Porcine GH gene	383	1030	80	2e-102	96
# AJ575422.1	*Hippopotamus amphibius* gh1 GH gene	538	968	90	4e-149	81
# AJ831741.1	*Balaenoptera physalus* GH gene precursor	529	966	91	2e-146	81
# U25973.1	*Felis catus* GH mRNA	327	948	43	8e-86	95
# Z23067.1	C.familiaris mRNA for GH	313	797	36	2e-81	95
# EU376045.1	*Vulpes lagopus* GH gene	300	795	38	2e-77	91
# AJ575419.1	*Camelus dromedarius* GH gene	366	628	30	2e-97	90
# DQ530374.1	*Lama pacos* GH gene	555	555	43	4e-154	80
# EU939446.1	*Equus caballus* GH gene	542	542	42	3e-150	80
# X12546.1	Ovine GH gene	291	541	30	1e-74	91

Table 4.6 Comparison of the lengths of exons and introns of the Indian elephant GH gene in different species

Species	No. of nucleotide residues								
	Exon 1	**Intron 1**	**Exon 2**	**Intron 2**	**Exon 3**	**Intron 3**	**Exon 4**	**Intron 4**	**Exon 5**
Elephas maximus	10	269	161	215	117	123	162	249	201
Sus scrofa	10	242	161	210	117	197	162	278	201
Delphinus delphis	10	253	161	203	117	228	162	275	201
Oryctolagus cuniculus	10	261	161	188	117	114	162	261	201
Vulpes lagopus	10	222	161	219	117	167	162	459	201
Hippopotamus amphibius	10	242	161	205	117	272	162	264	201
Cervus elephas	13	247	161	227	117	229	162	280	201
Ovis aries	13	246	161	227	117	229	162	275	201
Capra hircus	13	247	161	227	117	229	162	277	201
Bubalus bubalis	13	201	161	227	117	229	162	273	201
Bos indicus	13	248	161	227	117	229	162	272	201

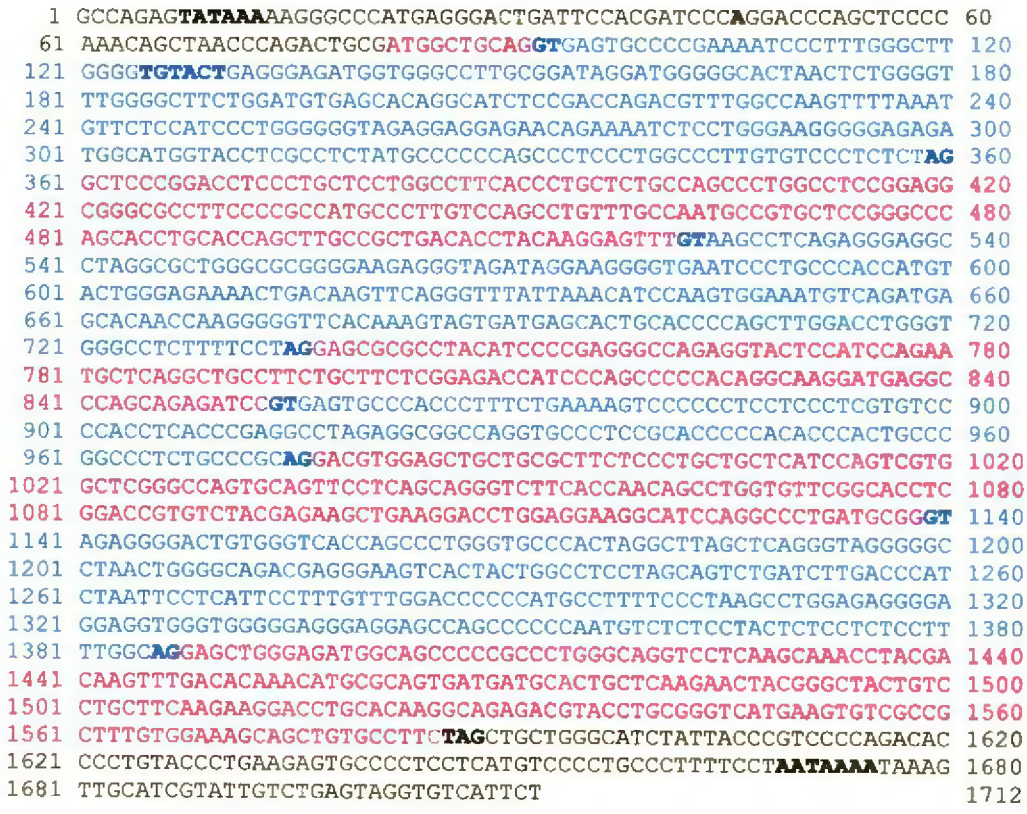


Fig 4.1 Indian elephant GH gene nucleotide sequence showing exons (red), introns (blue) and regulatory elements (Bold)

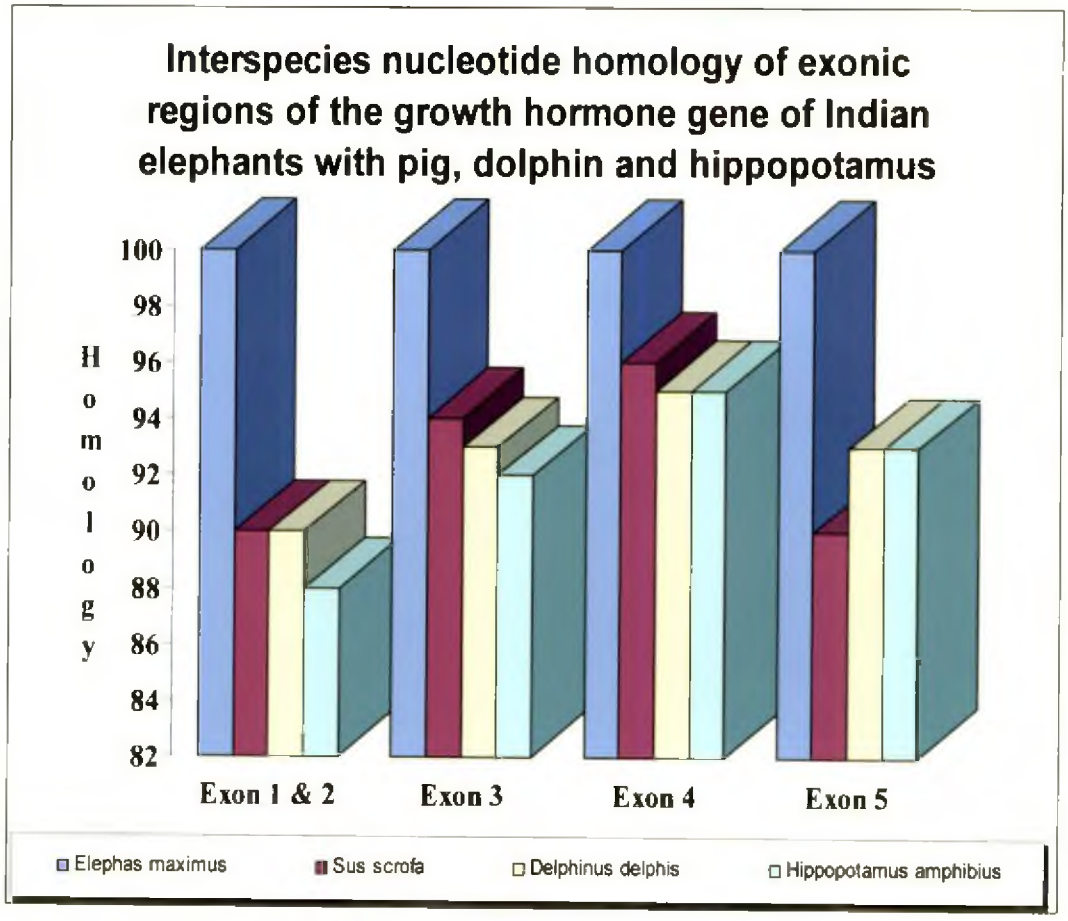


Fig 4.2 Nucleotide sequence homology of the exonic regions of the GH gene of Indian elephant with that of most closely related species.

The complete alignment of the corresponding sequences of Indian elephant and other livestock species is shown in Annexure-IV.

4.7.1.1 GenBank Accession Number

The new nucleotide sequence data of Indian elephant GH gene determined has been submitted to the GenBank DNA database and obtained the unique Accession Number: GQ141104.

4.7.2 Protein Sequence Analysis

The Indian elephant GH gene has an open reading frame (ORF) of 648 nucleotides encoding a signal peptide of 26 amino acid residues and a mature protein of 190 amino acid residues. The single letter code representing the amino acids and their triplet codons are given in the Annexure-V. The signal peptide was found to be 73 per cent homologous to its *Sus scrofa* counterpart showing variations at seven amino acid residues between the two proteins.

The 190 amino acid residues of the mature polypeptide chain was found to be 99 per cent homologous to that of African elephant (Accession # P20392) showing only one variant amino acid at position 127 while porcine showed 97 per cent homology with four variants. A comparison of the deduced amino acid sequence of the GH with the amino acid sequence of African elephant and porcine is presented in Fig. 4.3. Similar to African elephant GH, the Indian elephant also contains 190 amino acid residues with *Phe* at both NH_2- and COOH-terminal. The two proteins differ only at position 127 where *Gly* (GGA) was substituted by *Glu* in African elephant while with porcine protein, variations at two nucleotide bp were found 127 and 133 positions; *Gly* (GGA) was replaced by *Glu* (GAG) and *Pro* (CCT) was replaced by *Ala* (GCA), respectively.

The protein sequence analysis by BLASTp also revealed 95 per cent identity with the corresponding protein of camel (Accession # Q7YRR6.1) and 94

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	IEGH	216	2	AEGH	190	99
1	IEGH	216	3	PGH	216	94
2	AEGH	190	3	PGH	216	98

IEGH= Indian elephant growth hormone protein amino acids
AEGH= African elephant growth hormone protein amino acids
PGHP= Pig growth hormone protein amino acids

CLUSTAL 2.0.10 multiple sequence alignment

```
IEGH  MAAGSRTSLLLAFTLLCQPWPPEAGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEFERA 60
AEGH  ------------------------FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERA 34
PGH   MAAGPRTSALLAFALLCLPWTREVGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEFERA 60
                                ************************************

IEGH  YIPEGQRYSIQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQFLSR 120
AEGH  YIPEGQRYSIQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQFLSR 94
PGH   YIPEGQRYSIQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQFLSR 120
      ************************************************************

IEGH  VFTNSLVFGTSDRVYEKLKDLEEGIQALMRELGDGSPRPGQVLKQTYDKFDTNMRSDDAL 180
AEGH  VFTNSLVFGTSDRVYEKLKDLEEGIQALMRELEDGSPRPGQVLKQTYDKFDTNMRSDDAL 154
PGH   VFTNSLVFGTSDRVYEKLKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSDDAL 180
      ********************************* ***** .**:********** :******

IEGH  LKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
AEGH  LKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 190
PGH   LKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
      *************************************
```

Fig 4.3 Comparison of the deduced amino acid sequence of the GH gene of Indian elephant with that of African elephant and pig. The variant amino acids are shown in bold red letters.

per cent identity with both dolphin (Accession # Q8MI73.1) and hippopotamus (Accession # Q7YQB8.1), respectively (Table 4.7). The ClustalW results of amino acid sequences also produced a high alignment score with African elephant, pig, dolphin, hippopotamus and Indian cattle of about 99, 94, 94, 94 and 87, respectively. The ClustalW protein alignment score for GH of pig, dolphin and hippopotamus with that of Indian elephant was found to be equal. The ClustalW multiple protein sequence alignment of GH of Indian elephant and that of other mammalian species depicting the homology and divergence is presented in Annexure-VI and the constructed phylogramme is shown in the Fig. 4.4. The distance phylogramme of mammals placed Indian elephant along with African elephant and opossum while members of Bovidae family and pigs were in a separate cluster.

4.7.3 Protein Structure Prediction

The secondary structure of the GH of Indian elephant predicted using the PAPIA system (http://mbs.cbrc.jp/papia/papia.html) is presented in Plate 4. The predicted secondary structure showed the larger α-helical lobe formed by four sections of the polypeptide chain (residues 3-34, 47-80, 110-126 and 148-172) while the smaller lobe, which encompasses a small anti-parallel β-sheet and a small irregular structure formed the remaining structure of the polypeptide chain.

The three dimensional (3D) structure was predicted using automatic comparative modeling server, the Swiss Model Server. The 3D model was built using the 3D structure 1bp3A as template, which has 65.45 per cent sequence identity. The predicted 3D protein structure obtained from the above server and the 3D structure of human GH down loaded from the protein data bank (http://www.rcsb.org/pdb; PDB structure Id=1HGU) are presented in Plate 5. It reveals that the α-helices, β-sheets and random coils in the predicted protein structure are similar to that of the human GH structure, which is the only GH protein whose 3D structure is available in the protein data bank.

Table 4.7 Protein sequence analyses by BLASTp showing percentage of identity with GH gene of Indian elephant

Accession	Length	Organism	Identity (per cent)	Score (max)	E-Value (min)
# ABG67748.1	216	*Lama pacos*	95	428	2e-118
# Q7YRR6.1	216	*Camelus dromedarius*	95	428	2e-118
# CAA80601.1	216	*Canis familiaris*	95	427	3e-118
# Q8MI73.1	216	*Delphinus delphis*	94	425	1e-117
# AAS89356.1	216	*Sus scrofa*	94	424	2e-117
# Q7YQB8.1	216	*Hippopotamus amphibius*	94	424	2e-117
# BAF69023.2	216	*Equus caballus*	93	421	2e-116
# P20392.1	190	*Loxodanta africana*	99	392	1e-107
# ABC18165.1	217	*Bos taurus*	86	372	1e-101
# ABC18163.1	217	*Bubalus bubalis*	87	360	3e-98
# ABY61234.1	217	*Bos indicus*	87	357	3e-97

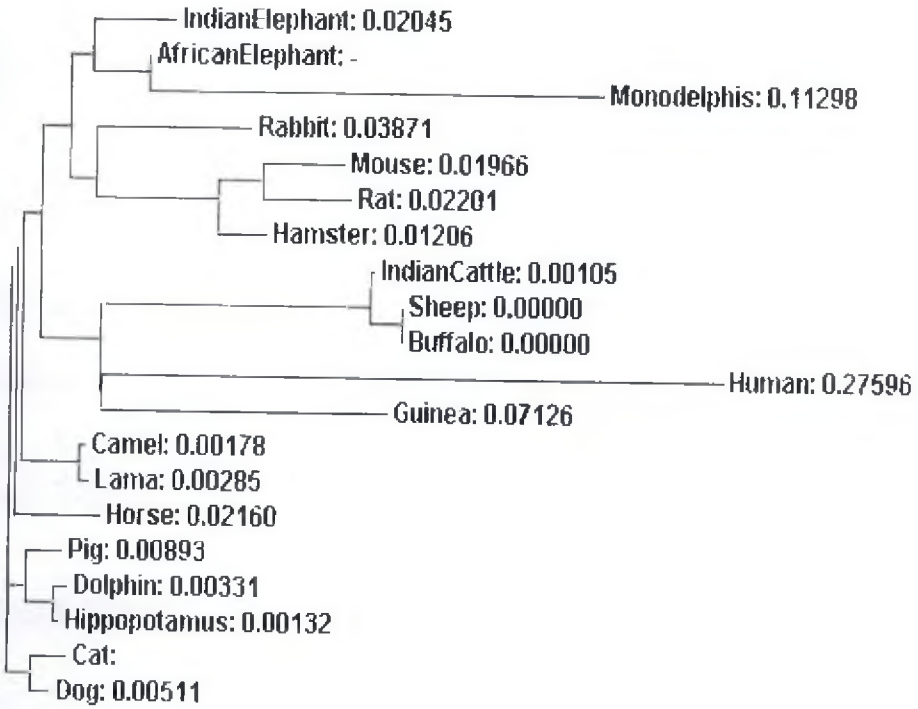


Fig. 4.4 Phylogramme constructed by ClustalW depicting the relationships between species and rate of evolution

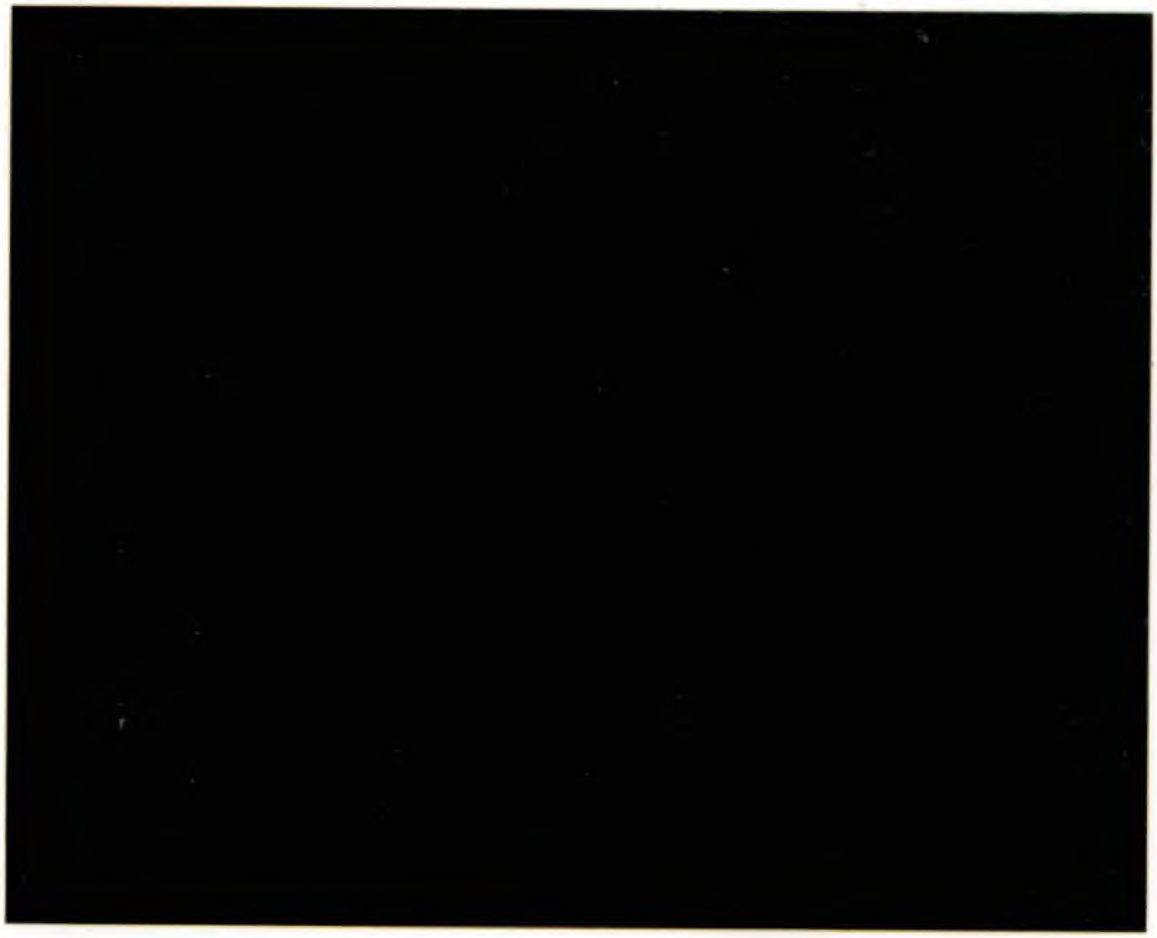
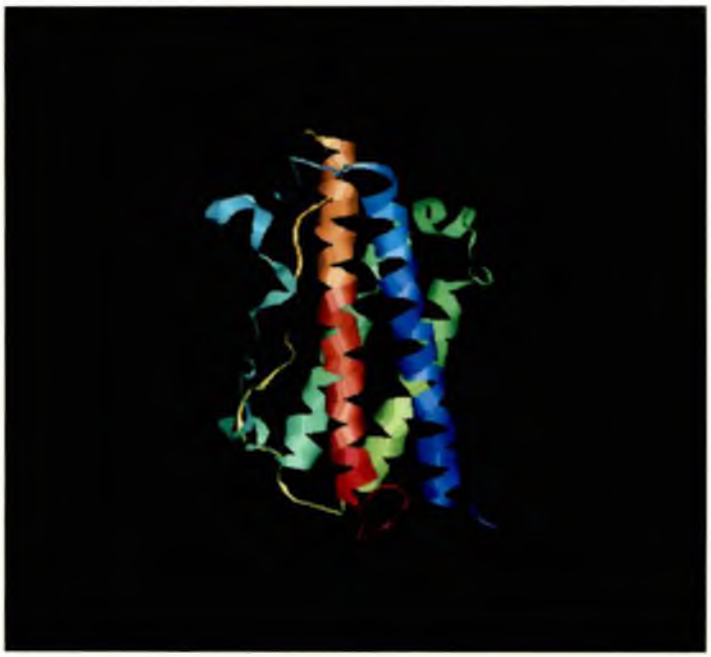


Plate 4 Predicted secondary structure of the GH gene of Indian elephant using PAPIA system showing sites of alpha helices, beta sheets and coils

H: alpha helix **E: beta sheet** **C: coil**

A)



B)

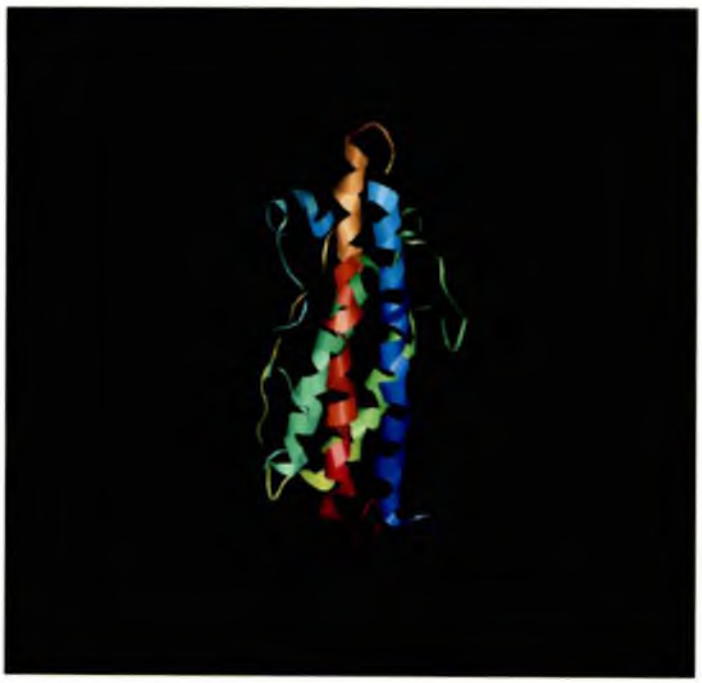


Plate 5 A) Predicted 3D model of Indian elephant GH protein (Swiss Model Server) B) Human GH protein 3D structure (PDB structure Id=1HGU)

Discussion

5. DISCUSSION

The elephants are an integral part of the daily life in Kerala. They are loved, revered, groomed and given a prestigious place in the state's culture. Elephants in Kerala are often referred to as the 'sons of the *sahya*.' The elephant is the state animal of Kerala and is featured on the emblem of the Government of Kerala. Almost all of the festivals in Kerala include at least one richly caparisoned elephant. Elephants carry the deity during annual festival processions and ceremonial circumambulations in the temples. A study on nucleotide and amino acid sequence analysis would throw light on the elephant growth hormone (GH) gene architecture and the extent of functional and structural similarity with that of other species. A wider knowledge of both the structure and function of elephant GH gene will provide the potential for genetic modification.

5.1 ISOLATION OF GENOMIC DNA

DNA isolation from whole blood was carried out using the standard phenol chloroform extraction procedure (Sambrook *et al.*, 1989), with the modification of overnight incubation of the WBC suspension with SDS and proteinase-K. Phenol chloroform extraction procedure is a common and efficient technique for DNA isolation from blood (Tantia *et al.*, 2004; Araujo *et al.*, 2006; Hirbo *et al.*, 2006).

5.2 PCR ANALYSIS

The polymerase chain reaction (PCR) is an *in vitro* method of nucleic acid synthesis that enables the specific replication of a targeted segment of DNA, providing a rapid, highly sensitive and specific means of nucleic acid detection and isolation. In molecular biology, PCR is a technique to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating millions or more copies of a particular DNA sequence. A standard reaction mixture contains the sample DNA, two oligonucleotide primers, thermostable Taq DNA

polymerase and four deoxyribonucleotide triphosphates (dNTPs) in a buffered solution. Through repeated cycles of heat denaturation of the DNA, annealing of the primers to their complimentary sequences, and extension of annealed primers, an exponential accumulation of the specific target fragment is achieved (Saiki *et al.*, 1988).

5.2.1 Optimization of PCR

The polymerase chain reaction (PCR) is a relatively simple technique that amplifies a DNA template to produce specific DNA fragments *in vitro*. PCR has a several qualitative characteristics such as specificity, efficiency (yield of product), sensitivity and fidelity (error rate). PCR conditions were optimized for the primers selected to obtain specific products and avoiding multiple bands on the gel. If the PCR conditions are well optimized, most of the added primers and dNTPs will be used up during the PCR amplifications, thus there will be only very few interfering factors during cloning and sequencing. For cloning the PCR fragment into a vector it is necessary to ensure that there is only one major PCR band visible on the gel.

Magnesium chloride concentration and annealing temperature of the primer were critical in obtaining specific amplifications. The annealing temperature for the primer pair was optimized using gradient thermal cycler in the range of 50 to 60°C. An annealing temperature of 54.2°C for 1 min was found optimal for the reaction. According to Don *et al.* (1991), the occurrence of spurious bands can be reduced by increasing the annealing temperature of PCR. Specific product was obtained at 1.75 m*M* $MgCl_2$ concentration.

Other conditions affecting the specificity of the PCR include concentration of enzyme and primers as well as the annealing time, extension time and number of cycles (Cha and Thilly, 1993). These conditions were kept constant for all the primers used. A concentration of 1U/reaction of Taq DNA

polymerase and 50 p*M* of primer per reaction gave satisfactory results. The time for denaturation was 1 min and for extension 3 min for the first 10 cycles and 3 min plus 20 sec/cycle for the remaining 20 cycles, followed by a final extension time of 10 min.

5.3 GEL PURIFICATION OF DNA

Purification of PCR products after amplification is usually recommended to obtain accurate sequencing data. Agarose contains various impurities which may inhibit downstream reactions if not efficiently removed from the DNA. Modern spin-column kits are very good at removing impurities. When cutting out the DNA band from the gel, we must visualize the band, in an ethidium bromide stained gel, in a dark-room on a UV trans-illuminator. The trans-illuminator was set to a long-wavelength UV and minimized the amount of time the DNA is exposed. This is because the UV light causes mutation in DNA at a measurable rate.

The unused primers, if contaminate the PCR product, can act as extension primers during sequencing, resulting in the generation of an additional set of dye-labeled sequencing fragments which can make sequencing data interpretation difficult. Post-purified PCR products were subjected to an agarose gel run along with an appropriate DNA ladder to examine the quality of the purified product. If smearing or multiple bands are present, it is not possible to obtain a high quality sequence data. The PCR product purified using the Ultra Clean™ GelSpin™ DNA purification Kit (Catalog # 12400-50, MO BIO Laboratories, Inc. USA) was found to be of high quality suitable for sequencing.

5.4 CLONING

The cloning of genes, gene fragments and other DNA sequences is a fundamental part of molecular biology. This enables us to manipulate and study the function of a particular sequence. The PCR products generated using a non-

proofreading DNA polymerase, which lacks, 3′→5′ exonuclease activity, have a single template-independent nucleotide at the 3′ end of the DNA strand (Clark, 1988; Newton and Graham, 1994). This single-nucleotide, which is most commonly a 'dA' residue, allows hybridization with and cloning into T-vectors, which have a complementary 3′ single dT overhang.

5.4.1 The pGEM®-T Easy Cloning Vector

The pGEM®-T Easy Vector System was a convenient system for the cloning of PCR products. These vectors take advantage of the dA overhangs on PCR products after amplification with Taq DNA polymerase by providing compatible dT ends for ligation (Mezei and Storts, 1994; Robles and Doers, 1994). The single 3′-dT overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases.

5.4.2 Ligation

The amount of insert DNA for ligation was estimated by comparing the staining intensity with that of DNA molecular weight standards of known quantity on an ethidium bromide-stained agarose gel. In most cases, either a 1:1 or a 1:3 molar ratio of vector:insert works well. In the present study a 1:1 ratio gave excellent results as evidenced by large proportion of positive clones following transformation.

5.4.3 Transformation of Competent Cells

5.4.3.1 Competent Cells

Transforming a newly constructed plasmid into competent *E. coli* cells is the primary method to propagate and select for the clone or clones of interest. Competent bacterial cells are receptive to importing foreign DNA and replicating

it. High-quality competent *E. coli* is an integral part of a successful cloning protocol. In the present study JM109 competent cells were prepared according to the manufacturer's instructions. These competent cells when transformed with the ligated plasmid DNA via the heat-shock method resulted in sufficient numbers of transformant clones. These host cells are reported to be an ideal host for many molecular biology applications (Yanisch-Perron *et al.*, 1985) and can be used for α-complementation of β-galactosidase for blue/white screening.

5.4.3.2 Transformation and Screening

The ligated product was transformed into competent cells of *E. coli* JM109 cells prepared by $CaCl_2$ treatment as per the manufacture's instructions. Recombinants were selected through blue-white screening on LB-amp plates. Successful cloning of an insert into the T-Vector disrupts the coding sequence of the β-galactosidase α-peptide leads to failure of α-complementation resulting in formation of white colonies on X-gal/IPTG indicator plates following transformation of competent cells where as simple vector reclosure without insert results in blue colonies.

5.5 DNA SEQUENCING

Presence of the insert in single white colonies was confirmed by PCR with the same primer combination and purified plasmids from positive clones were sequenced at the DNA Sequencing Facility, Bioserve, Biotechnologies (India), Pvt. Ltd. Hyderabad. A nucleotide sequence was determined by the dideoxy-chain termination method. For complete sequencing primer walking method was performed.

5.5.1 Structure of the Growth Hormone Gene

The sequence data of 1712 bp obtained from sequencing of the cloned PCR product provide the complete structural organization of the transcription unit of the gene. The transcription unit with a size of 1507 bp begins at position

82 as evidenced by the presence of the start codon 'ATG'. Most of the novel GH gene sequences include a substantial stretch of sequence upstream of the initiation codon. This region is known to include a number of regulatory elements, one of such regulatory element, TATA box was seen at position 8 of the sequenced gene. The other further upstream regulatory elements seen in GH gene such as binding sites for transcription factor Pit-1 (Theill and Karin, 1993), thyroid hormone receptor (TRE, Glass *et al.,* 1987), a cyclic AMP regulated element (Shepard *et al.*, 1994) and a negative regulatory element (NRE3), which probably represents a binding site for transcription factor YY1 (Park & Roe, 1996) could not be examined as these regions were not available in the sequenced data.

The sequence included the putative transcription start point (tsp) at position 46. Growth hormone gene of Indian elephant is found to be composed of five exons and four introns. A glucocorticoid response element (ACANNNTGTNCT) is present in the first intron of the human GH gene (Slater *et al.,* 1985). A part of this consensus sequence (TGTCCT), which was found in the first intron of the goat GH gene (Wallis, 2008), could also be seen (TGTACT) in the Indian elephant. The four intervening sequences begin and end with the consensus sequences GT at 5' and AG at 3' boundaries, respectively and these sequences involve in the splicing process in which the sequences corresponding to the introns are removed from the primary transcript of the gene as reported by Breathnach and Chambon (1981). The TAG stop codon was formed by nucleotides 1586–1588, followed by untranslated sequence containing a potential polyadenylation signal (AATAAAA) at nucleotides 1669–1675.

5.5.1.1 Comparison of Growth Hormone with Porcine and African Elephant

The nucleotide sequences obtained for Indian elephant GH gene was found to be having 75 to 96 per cent identity with that of pig. The 190 amino acid composition of the GH protein determined from the sequence was found to be in good agreement with the corresponding sequence of African elephant with only

one variant at position 127, where, *Gly* (GGA) in Indian elephant was substituted by *Glu* in African elephant (Accession # P20392.1). The derived amino acid sequence contained the 26 amino acid residues of the signal peptide which was found to be 90 per cent homologous to that of both *Bos taurus* (Accession # ABC18166.1) and *Ovis aries* (Accession # AAB64121.1) GH genes.

Comparison of the Indian elephant GH gene with porcine showed identical structural organization and extensive homology within the transcription unit. The exon-intron boundaries in the porcine gene occurred at the codons of the amino acid residues, Gly-4 (intron 1), Phe-57 (intron 2), Ser-96 (intron 3) and Arg-150 (intron 4). All these four residues were also conserved in Indian and African elephants as well as in pigs. This strict homology in the sites of insertion of introns suggests that the exon-intron organization of these genes was established before the divergence of these species. The positions of the exon-intron boundaries are also conserved as evidenced from similar sizes of the exons but the introns sizes vary in different species.

Examination of the nucleotide sequences for conserved regions thought to be involved in gene transcription and RNA processing reveals a number of features in common with other mammalian single-copy genes. The dinucleotides GT and AG occurring at the 5′ and 3′ ends of the introns, respectively is a feature in common with other eukaryotic genes (Mount, 1982).

Characterization of sequence homology between the Indian elephant and other animal species by using BLAST2 sequence alignment programme demonstrated a high degree of homology within the exons. All the exons except the first showed equal numbers of nucleotide residues in all these species. In elephant, the first exon was found to be having 10 nucleotide residues similar to that reported in pig, dolphin, rabbit and hippopotamus as against 13 residues found in species such as cattle, buffalo, sheep and goat. Evidence for some homology was also seen in intron 1, which showed maximum 84 per cent

similarity with that of giraffe. In contrast, intron 2, 3 and 4 showed no significant similarity both in length and in sequence with other animal species.

A molecule of GH gene comprised 190 amino acid residues in addition to the 26 amino acid residues as signal peptide. The alignment of the 190 amino acids sequence of the Indian elephant and African elephant showed a high degree of homology (99 per cent) with only one amino acid variant residues and with porcine it showed 97 per cent homology with four variants. This data denoted the slow basal rate of evolution shown by mammalian GHs and also showed that the GH gene is strongly conserved among species. Signal peptide sequences seems to be evolved more rapidly as there was less homology among them as compared to mature GH sequences, and there was no correlation between the rates of evolution of these two sequences.

The predicted secondary structure showed that four sections of the polypeptide chain formed the larger α-helical lobe (residues 3-34, 47-80, 110-126 and 148-172), while the smaller lobe, which encompasses a small anti-parallel beta-sheet and a small irregular structure formed the remaining structure of the polypeptide chain. This predicted structure was found to be similar to that of human GH, which was reported to be 45 per cent helical consisting of four α chains, 8 helices and 87 residues (Kabsch and Sander, 1983). Predicted tertiary structure of the Indian elephant GH was highly homologous with that of human GH structure, which is the only GH protein whose 3D structure is available in the protein data bank.

The review of available literature did not reveal any reference on DNA sequence information of any of the functional genes in elephants. Though a few microsatellite loci have been cloned from Indian as well as African elephants, DNA sequence information on functional genes is not yet reported in both species of elephants. This work seems to be the first of its kind in establishing the DNA and protein sequence data and structural organization of growth hormone gene in Indian elephants. The analysis of sequence data revealed that though the

DNA sequence identity with other species is low (only 75 to 96 per cent identity with porcine) there is high similarity of the protein at the amino acid sequence level, the difference being a single residue in African elephant and a couple of residues only in pig. This observation revealed a significant variation in the codon usage between Indian and African elephants and other species of animals. The signal peptide and intron sequences seem to be evolved more rapidly as there was less homology as compared to exonic sequences.

Overall, the structures of Indian elephant GH gene was found to be similar to that of African elephant and porcine as evidenced from the high degree of amino acid sequence identity (99 – 97 per cent) and it may be possible to clone the GH gene from African elephant using primers designed based on the Indian elephant GH gene sequence reported in this study.

Summary

6. SUMMARY

The awareness of the value of elephants in this world, especially in Kerala was the stimulus for this study. Despite being the largest terrestrial mammal, little is known about its genome, its organization and structure. Very few genes or DNA markers have been identified or studied in elephants. A study on nucleotide and amino acid sequence analysis would throw light on the elephant growth hormone (GH) gene architecture and the interspecies comparison of the GH protein sequence would also give an idea about the evolutionary status of the animal.

Genomic DNA was isolated from Indian elephant using the phenol-chloroform extraction procedure. A 1712 bp fragment of GH gene which covers the complete transcriptional unit of the gene and flanking sequences was amplified by PCR using specific primers designed on the basis of the 5′ and 3′ flanking sequences of goat GH gene. The PCR assay was performed in 200 µl reaction tubes with a total volume of 50 µl, containing 100 ng of DNA template, 1 X PCR buffer, 1.75 m*M* $MgCl_2$, 50 p*M* of each primer, 200 µ*M* of each dNTP and one unit *Taq* DNA polymerase.

The PCR product was purified by a preparative gel electrophoresis using a gel elution kit (Mo Bio laboratories, USA) following manufacturer's instructions. The purified DNA was analyzed on agarose gel, checked for the presence of any smearing, primer-dimer or additional bands, and quantified before using it for cloning.

The gel purified PCR product was cloned into pGEM®-T Easy cloning vector prior to sequencing. The ligation reaction was carried out in 10 µl volume and contained 2 µl of PCR product, 5 µl of 2X rapid ligation buffer, 1 µl T-Vector and 1 µl Instant T4 DNA ligase. The ligated product was examined in one per cent agarose gel in 1 X TAE buffer in a horizontal submarine electrophoresis unit and kept at -20 °C till transformation.

Transformation was induced by treating host *E. coli* cells with ice-cold calcium chloride followed by a brief exposure to high temperature using the Genei Transformation Kit. The recombinant plasmid and the competent bacterial cells were mixed and incubated at 42°C in a water bath for 2 min and then returned to ice for 5 min. The mixture was then mixed with 1 ml of LB broth and incubated in a shaker at 37°C for one hour. These transformed cells at different volumes were added on to LB-Amp plates with X-gal and IPTG and incubated for 37°C overnight. The cells carrying recombinant plasmid produced white colonies.

The recombinant plasmid carrying GH gene insert was isolated from the white clones by a modified SDS-alkaline lysis method using the AxyPrep Plasmid Miniprep Kit. The purified plasmid gene insert was examined by a one per cent agarose gel run and the band showed a molecular size of approximately 4.7 kb. The purified recombinant plasmid was sequenced by the dideoxynucleotide sequencing method using an automated DNA sequencer (Applied Biosystems, USA) at Bioserve, Biotechnologies (India), Pvt. Ltd. Hyderabad.

The 1712 bp nucleotide sequence of the GH gene and the predicted protein structure was compared with corresponding sequences from pig, African elephant and other species through BLAST analysis. The nucleotide sequence showed 75 to 96 per cent homology with pig and 77 to 95 per cent with that of dolphin GH genes, respectively. Multiple sequence analysis was done with the corresponding nucleotide and amino acid sequences of livestock species and a phylogramme was constructed using the ClustalW programme. The distance phylogramme of mammals placed Indian elephant along with African elephant and opossum while members of Bovidae family and pigs were in a separate cluster. The results of ClustalW multiple nucleotide sequence alignment showed homologies with a scores of 78 with pig and dolphin, 76 with hippopotamus and Indian cattle and 66 with buffalo. The ClustalW results of amino acid sequences

also produced a high alignment score with African elephant, pig, dolphin, hippopotamus and Indian cattle of 99, 94, 94, 94 and 87, respectively.

The Indian elephant GH contains 190 amino acid residues with both amino- and carboxyl-terminal *Phe* and alignment of the amino acid sequence with African counterpart showed high degree of homology with a single variant amino acid at position 127 where *Gly* (GGA) was substituted by *Glu* in African elephants. Whereas, in the case of pig variations at two nucleotide bp were found 127 and 133 positions; *Gly* (GGA) was replaced by *Glu* (GAG) and *Pro* (CCT) was replaced by *Ala* (GCA), respectively.

The secondary structure of the elephant GH was predicted using the PAPIA system (http://mbs.cbrc.jp/papia/papia.html) of Computational Biology Research Center, Japan. The predicted secondary structure showed that the larger α-helical lobe is formed by four sections of the polypeptide chain (residues 3-34, 47-80, 110-126 and 148-172) while the smaller lobe, which encompassed a small anti-parallel beta-sheet and a small irregular structure formed the remaining structure of the polypeptide chain.

The three dimensional (3D) structure of the protein was predicted using automatic comparative modeling server, the Swiss Model Server. The 3D model was built using the 3D structure 1bp3A as template which has 65.45 per cent sequence identity. The predicted 3D protein structure obtained from the above server and the 3D structure of human GH revealed that the α-helices, β-sheets and random coils in the predicted protein are similar to that of the human GH structure.

The review of available literature did not reveal any reference on DNA sequence information of any of the functional genes in elephants. This work seems to be the first of its kind in establishing the DNA and protein sequence data and structural organization of growth hormone gene in Indian elephant. The analysis of sequence data revealed that though the DNA sequence identity with

other species is low (only 75 to 96 per cent identity with porcine) there is high similarity of the protein at the amino acid sequence level, the difference being a single residue in African elephant and a couple of residues only in pig. This observation revealed a significant variation in the codon usage between Indian elephant and African elephants and other species of animals.

Overall, the structures of Indian elephant GH gene was found to be similar to that of African elephant and porcine as evidenced from the high degree of amino acid sequence identity (99 – 97 per cent) and it may be possible to clone the GH gene from African elephant using primers designed based on the Indian elephant GH gene sequence reported in this study.

References

REFERENCES

Ahmad, N.N., Cu-Unjieng, A.B. and Donoso, L.A. 1995. Modification of standard proteinase K/phenol method for DNA isolation to improve yield and purity from frozen blood. *J. Med. Genet.* 32(2): 129-130.

Araujo, A.M., Guimaraes, S.E.F., Machado, T.M.M., Lopes, P.S., Pereira, C.S., Silva, F.L.R., Rodrigues, M.T., Columbiano, V.S. and Fonseca, C.G. 2006. Genetic diversity between herds of Alpine and Saanen dairy goats and the naturalized Brazilian Moxoto breed. *Genet. Mol. Biol.* 29: 67-74.

Aravindakshan, T.V., Nainar, A.M. and Nachimuthu, K. 1998. Extraction of DNA from cattle blood cells. *Cheiron* 27: 5-6.

Arnold, K., Bordoli, L., Kopp, J. and Schwed, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics.* 22(2): 195-201.

Aslanidis, C. and Jong, P.J.D. 1990. Ligation indépendant cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18 (20): 6069-6074.

Barsh, G.S., Seeburg, P.H. and Gelinas, R.E. 1983. The human growth hormone gene family: structure and evolution of the chromosomal locus. *Nucleic Acids Res.* 11 (12): 3939-3958.

Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. 1981. Primary structure and evolution of rat growth hormone gene. *Proc. Natl. Sci.* 78(8): 4867-4871.

Bellair, J.T. 1972. Ovine growth hormone sequence of the C-terminal 68 amino acids. *Biochem. Biophys. Res. Commun.* 46 (3): 1128-1134.

Bewely, T.A. and Li, C.H. 1970. Primary structures of human pituitary growth hormone and sheep pituitary lactogenic hormone compared. *Science* 168 (3937): 1361-1362.

Blin, N. and Stafford, D.W. 1976. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* 3(9): 2303-2308.

Breathnach, R. and Chambon, P. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50: 349–383.

Cha, R.S. and Thilly, W.G. 1993. Specificity, efficiency, and fidelity of PCR. *Genome Res.* 3: 18-29.

Chan, E.Y. 2005. Advances in sequencing technology. *Mutat. Res.* 573: 13-40.

Chen, H.C., Wilhelmi, A.L. and Howard, S.C. 1970. Purification and characteristics of porcine growth hormone. *J. Biol. Chem.*. 245(13): 3402-3406.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. 2003. Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res.* 31(13): 3497-3500.

Chung, C.T., Niemela, S.L. and Miller, R.H. 1989. One step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA.* 86 (7): 2172-2175.

Clark, J.M.1988. Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Res.* 16(20): 9677-9686.

Clegg, C.D., Ritz, K. and Griffiths, B.S. 1997. Direct extraction of microbial community DNA from humified upland soils. *Lett. Appl. Microbiol.* 25(1): 30-33.

Cohen, S.N., Chang, A.C.Y. and Hsu, L. 1972. Nonchromosomal antibiotic resistance in bacteria; genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Nat. Acad. Sci. USA.* 69 (8): 2110-2114.

Cole, H.H. 1938. The endocrine control of growth. *J. Anim. Sci.1938*: 20-26. Available: http://jas.fass.org/cgi/reprint/1938/1/20.pdf

Conde, R.D., Paladini, A.C., Santome, J.A. and Dellacha, J.M. 1973. Isolation purification and caractérisation of equine growth hormone. *Eur. J. Biochem.* 32 (3): 563-568.

Das, P., Meyer, L., Seyfert, H.M., Brockmann, G. and Schwerin, M. 1996. Structure of the growth hormone-encoding gene and its promoter in mice. *Genetics* 169(2): 209-213.

Deane, E.E. and Woo, N.Y.S. 2006. Molecular cloning of growth hormone from silver sea bream: effects of abiotic and biotic stress on transcriptional and translation expression. *Biochem. Biophys. Res. Commun.* 342 (4): 1077-1082.

Deng, W., Nickle, D.C., Learn, G.H., Maust, B. and Mullins, J.I. 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple data bases and users datasets. *Bioinformatics.* 23 (17): 2334-2336.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* 19: 4008.

Douglas, A.M., Georgalis, A.M., Benton, L.R., Canavan, K. L. and Atchison, B.A. 1992. Purification of human leukocyte DNA: Proteinase K is not necessary. *Anal. Biochem.* 201(2): 362-365.

Edgerton, F. 1931. *Elephant-lore of the Hindus.* Reprint 1985. Motilal Banarasidass Publication, Delhi. 129 P.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P. and Rank, D. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910): 133-138.

Einarsdottir, I.E., Sakata, S. and Bjornsson, B.T. 2002. Atlantic halibut growth hormone: structure and plasma levels of sexually mature males and females during photoperiod-regulated annual cycles. *Gen. Comp. Endocrinol.* 127(1): 94-104.

Fernando, P. and Lande, R. 2000. Molecular genetic and behavioural analysis of social organization in the Asian elephant (*Elephas maximum*). *Behav. Ecol. Sociobiol.* 48(1): 84-91.

Fernando, P., Pfrender, M.E., Encalada, S.E. and Lande, R. 2000. Mitochondrial DNA variation, phylogeography and population structure of Asian elephants. *Heredity* 84(3): 362-372.

Fernando, P., Vidya, T.N.C., Payne, J., Stuewe, M., Davison, G., Alfred, R.J., Andav, P., Bosi, E., Killbourn, A. and Melnick, D.J. 2003. DNA analysis indicates that Asian elephants are native to Borneo and are therefore a high priority for conservation. *PLOS Biol.* 1(1): 110-115.

Fleischer, R.C., Perry, E.A., Muralidharan, K., Stevens, E.E. and Wemmer, C.M. 2001. Phylogeography of the Asian elephant (*Elephas maximum*) based on mitochondrial DNA. *Evolution* 55(9): 1882-1892.

Forsyth, I.A. and Wallis, M. 2002. Growth hormone and prolactin- molecular and functional evolution. *J. Mammary Gland Biol.* 7(3): 291-312.

Glass, C.K., Franco, R., Weinberger, C., Albert, V.R., Evans, R.M. and Rosenfeld, M.G. 1987. A *c-erb-A* binding site in rat growth hormone gene mediates *trans*-activation by thyroid hormone. *Nature* 329: 738-741.

Gordon, D.F., Quick, D.P., Erwin, C.R., Donelson, J.E. and Maurer, R.A. 1983. Nucleotide sequence of the bovine growth hormone chromosomal gene. *Mol. Cell Endocrinol.* 33(1): 81-95.

Gupta, S.K., Thangaraj, K. and Singh, L. 2006. A simple and inexpensive molecular method for sexing and identification of the forensic samples of elephant origin. *J. Forensic Sci.* 51(4): 805-807.

Hartree, A.S. 1966. Separation and partial purification of the protein hormones from the human pituitary glands. *Biochem. J.* 100 (3): 754-761.

Higgins, D.G. and Sharp, P.M. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Genetics* 73(1): 237-244.

Hirbo, J., Muigai, A., Naqvi, A.N., Rege, E.D. and Hanotte, O. 2006. The genetic diversity of trans-caucasian native sheep breeds. *Asian-Aust. J. Anim. Sci.* 19: 943-952.

Holton, T.A. and Graham, M.W. 1991. A simple and efficient method for direct cloning of PCR products using ddT tailed vectors. *Nucleic Acids Res.* 19 (5): 1156.

Hongbao, M. 2005. Development application of polymerase chain reaction (PCR). *J. Am. Sci.* 1(3): 1-47.

Horikawa, R., Gaylinn, B.D., Lyons, C.E., Jr. and Thorner, M.O. 2001. Molecular cloning of ovine and bovine growth hormone-releasing hormone receptors: The ovine receptor is C- terminally truncated. *Endocrinology* 142 (6): 2660-2668.

Hu, Z., Levievre,V., Tam, J., Chang, J.W., Fuerizalida, G., Zhou, X. and Waschek, J.A. 2000. Molecular cloning of growth hormone releasing hormone/ pituitary adenylyl cyclase- activating polypeptide in the frog *Xenopus laevis*: Brain distribution and regulation after castration. *Endocrinology* 141 (9): 3366-3376.

Hungerford, D.A., Chandra, H.S., Snyder, R.L., Ulmer, F.A. and Jr. 1966. Chromosomes of three elephants, two Asian (*Elephas maximum*) and one African (*Loxodonta Africana*). *Cytogenetics* 5 (3-4): 243-246.

Innis, M.A. and Gelfand, D.H. 1990. Optimization of PCRs. *PCR protocols*. (eds. Innis, M.A., Gelfand ,D.H., Snisky, J.J and White, T.J). Academic press, New York, pp.3-12.

Jeong, K.J. and Lee, S.Y. 2003. Enhanced production of recombinant proteins in *Escherichia coli* by filamentation suppression. *Appl. Environ. Microbiol.* 69 (2): 1295-1298.

Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12): 2577-2637.

Kalia, A., Rattan, A. and Chopra, P. 1999. A method for extraction of high quality and high quantity genomic DNA generally applicable to pathogenic bacteria. *Anal. Biochem.* 275 (1): 1-5.

Karlin, S. and Altschul, S.F. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA*. 87(6): 2264-2268.

Kioka, N., Manabe, E., Abe, M., Hashi, H., Yato, M., Okuno, M., Yamano, Y., Sakai, H., Komano, T., Utsumi, K. and Iritani, A. 1989. Cloning and sequencing of goat growth hormone gene. *Agic. Biol. Chem.* 53(6): 1583-1587.

Kovalic, D., Kwak, J.H. and Weisblum, B. 1991. General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Res.* 19(16): 4560.

Lagziel, A., Lipkin, E. and Soller, M. 1996. Association between SSCP haplotypes at the bovine growth hormone gene and milk protein percentage. *Genetics* 142 (3): 945-951.

Lambert, C., Leonard, N., Bolle, X.D. and Depiereux, E. 2002. Esy Pred 3D: prediction of proteins 3D structures. *Bioinformatics* 18(9): 1250-1256.

Lamprecht, A., Margaria, T., Steffen, B., Sczyrba, A., Hartmeier, S. and Giegerich, R. 2008. GeneFisher-P: variations of GeneFisher as processes in Bio-jETI. *BMC Bioinformatics* 9(4): S13.

Lee, S. Y. 1996. High cell density cultures of *Escherichia coli*. *Trends Biotechnol.* 14(3): 98-105.

Lee, H., Kim, K.N. and Chae, Y.K. 2007. Reevaluating the capability of Taq DNA polymerase: Long PCR amplification. *Protein Peptide Lett.* 14: 321-323.

Lexa, M., Horak, J., and Brzobohaty, B. 2001. Virtual PCR. *Bioinformatics.* 17: 192-193.

Li, C.H., Evas, H.M. and Simpson, M.E. 1945. Isolation and properties of the anterior hypophyseal growth hormone. *J. Biol. Chem.* 159(2): 353-366. Available: http://www.jbc.org/cgi/reprint/159/2/353.pdf

Li, C.H., Bewley, T.A., Chung, D. and Oosthuizen, M.M. 1987. Elephant growth hormone, isolation and characterization. *Int. J. Pept. Protein Res.* 29(1): 62-67.

Linhart, C. and Shamir, R. 2005. The degenerate primer design problem: theory and applications. *J. Comput. Biol.* 12(4): 431-456.

Lioupis, A., Nevo, E. and Wallis, M. 1999. Cloning and characterization of gene encoding mole rat (spa lax ehrenbergi) growth hormone. *J. Mol. Endocrinol.* 22 (1): 29-36.

Lioupis, A., Wallis, O.C. and Wallis, M. 1997. Cloning and characterization of the gene encoding red deer (cervus elaphus) growth hormone: Implication for the molecular evolution of growth hormone in arteriodactyls. *J. Mol. Endocrinol.* 19 (3): 259-266.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. 1991. Constructing of T vectors, rapid and general method for direct cloning of unmodified PCR products. *Nucleic Acids Res.* 19 (5): 1154.

Maxam, A.M. and Gilbert, W. 1977. A New method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 74(2): 560-564.

Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. and Smith, L.M. 1991. A universal method for the direct cloning of PCR amplified nucleic Acid. *Bio. Technol.* 9 (7): 657-663.

Mezei, L.M. and Storts, D.R. 1994. Purification of PCR products. *PCR Technology: Current Innovations.* (eds. Griffin, H.G. and Griffin, A.M.). CRC Press, Boca Raton. pp. 21-27.

Miller, H. 1987. Practical aspects of preparing phage and plasmid DNA: Growth, maintenance and storage of bacteria and bacteriophage. *Methods Enzymol.* 152: 145-170.

Miller, W.L., Martial, J.A. and Baxter, J.D. 1980. Molecular cloning of DNA complimentary to bovine growth hormone mRNA. *J. Biol. Chem.* 255 (16): 7521-7524.

Morrison, D.A. 1977. Transformation in *Escherichia coli*: Cryogenic preservation of competent cells. *J. Bacteriol.* 132(1): 349-351.

Mount, S.M. 1982. A catalogue of splice junction sequences. *Nucl. Acids Res.* 10: 459-472.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. And Erlich, H.A. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold springer Harbor Symposia on Quantitative Biology.* 51: 263-273.

Nakata, V., Tang, X. and Yokoyama, K.K. 1997. Preparation of competent cells for high efficiency plasmid transformation of *Escherichia coli. Methods in Molecular Biology.* 69: 129-137.

Newton, C.R. and Graham, A. 1994. *PCR.* BIOS Scientific Publishers, Ltd., Oxford, UK. 13p.

Novy, R.E., Yaeger, K.W. and Kolb, K.M. 1996. Perfectly blunt cloning: A superior method for cloning PCR products or any DNA. *Innovations* 6 (1): 7-11.

Oliner, J.D., Kinzler, K.W. and Vogelstein, B. 1993. In vivo cloning of PCR products in E. coli. *Nucleic Acids Res.* 21(22): 5192-5197.

Orian, J.M., Mahoney, J.V. and Brandon, M.R. 1988. Cloning and sequencing of the ovine growth hormone gene. *Nucleic Acids Res.* 16(18): 9046.

Park, K.Y. and Roe, J.H. 1996. Identification of a negative regulatory site in the upstream region of bovine growth hormone gene. *Biochem. Biophys. Res. Commun.* 219: 354–358.

Perzer, S. and Mannhatter, C. 1991. A rapid method for the isolation of genomic DNA from citrated whole blood. *Biochem. J.* 273(1): 229-231.

Raman, S., Qian, B., Baker, D. and Walker, R.C. 2008. Advances in rosetta protein structure prediction on massively parallel systems. *IBM J. Res. Dev.* 52 (1): 7-17.

Robles, J. and Doers, M. 1994. pGEM®-T vector systems troubleshooting guide. *Promega Notes.* 45: 19-20.

Rose, T.M., Henikoff, J.G. and Henikoff, S. 2003. CODEHOP (Consensus-Degenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acid Res.* 31(13): 3763-3766.

Roskam, W.G. and Rougeon, F. 1979. Molecular cloning and nucleotide sequence of human growth hormone structural gene. *Nucleic Acids Res.* 7(2): 305-320.

Roux, K.H. 1995. Optimization and troubleshooting in PCR. *PCR methods Appl.* 4 (5): 185-194.

Rybicki, E. 2001. PCR primer design and reaction optimization. In: Coyne, V.E., James, M.D., Reid, S.J. and Rybichi, E.P. (ed.), *Molecular Biology Techniques Manual* (3rd ed.). Available: http://www.mcb.vct.ac.za/pcroptim.htm (23rd Oct. 2008)

Rychlik, W. 1995. Selection of primers for polymerase chain reaction. *Mol. Biotechnol.* 3 (2): 129-134.

Rychlik, W., Spencer, W.J. and Rouds, R.E. 1990. Optimization of the annealing temperature for DNA amplification in vitro. *Nucl. Acids. Res.* 18 (21): 6409-6412.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239 (4839): 487-491.

Sakthikumar, A., Mukundan, G. and Raghunandanan, K.V. 1992. Chromosome profile of Indian elephants. *The Asian elephant ecology, Biology, diseases, conservation and management* (eds. Silas, E.G., Nair, M.K. and Nirmalan, G.). KAU, Lumiere printing works, Trichur. PP. 41-42.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual.* Second edition. Cold spring Harbor Laboratory Press. New York, P.1886.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad .Sci. USA.* 74(12): 5463-5467.

Santome, J.A., Dellacha, J.M., Paladini, A.C., Wolfenstein, C.E.M., Pena, C., Poskus, E., Daurat, S.T., Biscoglio, M.J., Desese, Z.M.M. and Sanguesa, A.V.F. 1971. The amino acid sequence of bovine growth hormone. *FEBS Lett.* 16(3): 198-200.

Sekine, S., Mizukami, T., Nishi, T., Kuwana, Y., Saito, A., Sato, M., Itoh, S. and Kawauchi, H. 1985. Cloning and expression of cDNA for salmon growth in *Escherichia coli. Proc. Natl. Acad. Sci. USA.* 82 (13): 4306-4310.

Shepard, A.R., Zhang, W., Eberhardt, N.L. 1994. Two CGTCA motifs and a GHF1/Pit1 binding site mediate cAMP- dependent protein kinase A regulation of human growth hormone gene expression in rat anterior pituitary GC cells. *J. Biol. Chem.* 269:1804-1814.

Slater, E.P., Rabenau, O., Karin, M., Baxter, J.D. and Beato, M. 1985. Glucocorticoid receptor binding and activation of a heterologous promoter by dexamethasone by the first intron of the human growth hormone gene. *Mol. Cell. Biol.* 5: 2984–2992.

Stephen, C.Y.Ip., Zhang, X. and Leung, F.C. 2001. Genomic growth hormone gene polymorphisms in native Chinese chickens. *Exp. Biol. Med.* 226 (5): 458-462.

Surendran, P.N. 1992. Status of elephants in Kerala stata, India. *The Asian elephant ecology, Biology, diseases, conservation and management* (eds. Silas, E.G., Nair, M.K. and Nirmalan, G.). KAU, Lumiere printing works, Trichur. PP. 6-8.

Tantia, M.S., Behl, R., Sheoran, N., Singh, R. and Vijh, R.K. 2004. Microsatellite data analysis for conservation of two goat breeds. *Indian J. Anim. Sci.* 74: 761-767.

Tatusova, T.A. and Madden, T.L. 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174(2): 247-250.

Theill, L.E. & Karin, M. 1993. Transcriptional control of GH expression and anterior pituitary development. *Endocr. Rev.* 14:670–689.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22): 4673-4680.

Tillet, D. and Neilan, B.A. 1999. Enzyme free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Res.* 27 (19): e26.

Toogood, A.A., Harvey, S., Thorner, M.O. and Gaylinn, B.D. 2006. Cloning of the chicken pituitary receptor for growth hormone releasing hormone. *Endocrinology.* 147 (4): 1838-1846.

Topcu, Z. 2000. An optimized recipe for cloning of the polymerase chain reaction amplified DNA inserts into plasmid vectors. *Acta Biochem. Pol.* 47(3): 841-846.

Tsen, S.D., Fang, S.S., Chen, M.J., Chien, J.Y., Lee, C.C., Tsen, D.H.L. 2002. Natural plasmid Transformation in *Escherichia coli. J. Biomed. Sci.* 9(3): 246-252.

Vallone, P.M. and Butler, J.M. 2004. Autodimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 37 (2): 226-231.

Varma, S., Ghorpade, A., Tiwari, G., Das, P. and Gary, L.C. 1990. cDNA cloning & sequence analysis of bubaline growth hormone. *Mitochondrial DNA.* 10(2): 101-103

Venugopal, T., Mathavan, S. and Pandian, T.J. 2002. Molecular cloning of growth hormone encoding cDNA of Indian major carps by a modified rapid amplification of cDNA ends strategy. *J. Biosci.* 27 (3): 261-272.

Vidya, T.N.C., Fernando, P., Melnick, D.J. and Sreekumar, R. 2005. Population genetics structure and conservation of Asian elephants (*Elephas maximum*) across India. *Anim. Conserve.* 8(4): 377-388.

Vidya, T.N.C., Kumar, V.R., Arivazhagan, C. and Sukumar, R. 2003. Application of molecular sexing to free-ranging Asian elephant (*Elephas maximum*) populations in Southern India. *Curr. Sci.* 85(7): 1074-1077.

Wallis, M. 1969. The N- terminus of ox growth hormone. *FEBS Lett.* 3(2): 118-120.

Wallis, M. 2008. Mammalian genome projects reveal new growth hormone (GH) sequences characterization of the GH-encoding genes of armadillo (*Dasypus novemcinctus*), hedgehog (*Erinaceus europaeus*), bat (*Myotis lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel (*Spermophilus tridecemlineatus*), elephant (*Loxodanta africana*), cat (*Felis catus*) and opossum (*Monodelphis domestica*). *Gen. Comp. Endocr.* 155(2): 271-279.

Wallis, O.C. and Wallis, M. 1995. Cloning and characterization of the rabbit growth hormone-encoding gene. *Genetics* 163(2): 253-256.

White, B.A. 2005. Polymerase chain reaction (PCR): Design and optimization of reactions. *Enzyclopedia of Life Sciences*. 60: 25-30.

Wilhelmi, A.E. 1974. Chemistry of growth hormone. *Endocrinology: 4. The pituitary gland and its neuroendocrine control, Part 2* (eds. Greep, R.O. and Astwood, E.B.). American Physiological Society, Washington, D.C. 59-78.

Woychik, R.P., Camper, S.A., Lyons, R.H., Horovitz, S., Goodwin, E.C. and Rottman, F.M. 1982. Cloning and nucleotide sequencing of the bovine growth hormone gene. *Nucleic Acids Res.* 10 (22): 7197-7210.

Wu, D., Chen, M., Buchwald, M. and Phillips, R.A. 1995. A simple, rapid method for isolation of high quality genomic DNA from animal tissues. *Nucleic Acid Res.* 23 (24): 5087-5088.

Wu, D.Y., Ugozzoli, L., Pal, B.K., Qian, J. and Wallace. R.B. 1991. The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA Cell Biol.* 10(3): 233-238.

Xu, B., Moriyama, S., Zhang, P., Miao, H., Li, D. and Kawauchi, H. 2001. The complete amino acid sequence of growth hormone and partial amino acid sequence of prolactin and somatolactin from sea perch (*Lateolabrax japonicus*). *Aquaculture* 201(1-2): 117-136.

Yamano, Y., Oyabayashi, K., Okuno, M., Yato, M., Kioka, N., Manabe, E., Hashi, H., Sakai, H., Komano, T., Utsumi, K. and Zritani, A. 1988. Cloning and sequencing of cDNA that encodes goat growth hormone. *FEBS lett.* 228(2): 301-304.

Yanisch-Perron, C., Vieira, J. and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33: 103–119.

Yap, E.P.H. and McGee J.O.D. 1991. Short PCR product yields improved by lower denaturation temperatures. *Nucleic Acids Res.* 19 (7): 1713.

Zhoa, X., Li, J., Li, R., Zhang, S., Zhang, Y. and Yang, Z. 1996. Amplification, cloning and sequence comparison of the growth hormone gene for carp (*Cyprinus carpio*) by the polymerase chain reaction. *Chin, J. Biotechnol.* 12 (3): 161-167.

Zhou, M.Y. and Gomez-Sanchez, C.E. 2000. Universal TA cloning. *Curr. Issues Mol. Biol.* 2(1): 1-7.

Annexures

ANNEXURE – I

COMPOSITION OF REAGENTS AND BUFFERS USED IN THE STUDY

Agarose (1 %)

Weighed 0.6 g of agarose powder and mixed with 60 ml of 1 X TAE buffer in a conical flask. Solution was heated in a microwave oven until boiling and cooled slowly.

EDTA (0.5*M*, pH 8.3)

Dissolved 18.61 g of EDTA (disodium, dihydrate) in 80 ml of distilled water by bringing the pH to 8.3 with NaOH solution and volume made upto 100 ml. Stored at room temperature after filtration and autoclaving.

Ethidium Bromide (10 mg/ml)

Dissolved 100 mg of Ethidium bromide in 10 ml of distilled water. Solution stored at 4 °C in a dark coloured bottle.

Formamide dye / Stop buffer

Deionised formamide	98 %
Xylene cyanol	0.025 %
Bromophenol blue	0.025 %
0.5 *M* EDTA	10m*M*

Gel loading buffer

Bromophenol blue	0.25 %	50 mg
Xylene cyanol	0.25 %	50 mg

Sucrose	40 %	8 g

Components were mixed and stirred well in 20 ml distilled water and stored at 4 °C.

LB agar/broth

Ingredients	Grams/Litre
Tryptone	10.0
Yeast extract	5.0
Sodium chloride	5.0
Agar	10.0

Final pH 7.2 at 37 °C. Sterilized by autoclaving. Stored prepared media below 8 °C, protected from direct light. Stored dehydrated powder, in a dry place, in tightly-sealed containers at 2-25 °C.
For LB agar, added 1.5 % agar and autoclaved.

LB Ampicillin media:

Dissolved 100 mg of ampicillin in 1 ml sterile water to get Ampicillin stock concentration of 100 mg/ml. Stored at 4 °C.

Added ampicillin to LB broth or agar at a final concentration of 100 µg/ml, when the temperature of the media is around 40-45 °C.

LB-Amp. Plates with X-Gal. and IPTG:

Added 40 µl each of X-Gal and IPTG for every 20 ml of LB-Amp agar at 40-45 °C. Mixed well and poured media into required number of plates.

Phenol (Saturated, pH 7.8)

Commercially available crystalline phenol was melted at 65 °C in a water bath. Hydroxyquinolone was added to a final concentration of 0.1 percent. To the molten phenol, 0.5 *M* Tris HCl (pH 8.0) was added in equal volume and stirred for 30 min in a magnetic stirrer. The contents were transferred into a separating funnel. Lower phenolic phase was collected and mixed with equal volume of 0.1 *M* Tris HCl (pH 8.0) and stirred again for 30 min. The phenolic phase was collected and extraction repeated with 0.1*M* Tris HCl (pH 8.0) until the pH of phenolic phase was more than 7.8. Finally 0.1 volume of 0.01 *M* Tris HCl (pH 8.0) added and stored in a dark bottle at 4 °C.

RBC lysis buffer

Ammonium chloride	150 mM	8.0235 g
Potassium chloride	10 mM	0.7455 g
EDTA	0.1 mM	0.0372 g

Dissolved the contents in distilled water and volume made up to 1000 ml. Stored at 4 °C after filtration and autoclaving.

Sodium acetate

Dissolved 40.824 g of Sodium acetate in 70 ml of distilled water and pH adjusted to 5.5 with glacial acetic acid. Volume was made up to 100 ml, autoclaved and stored at room temperature.

Sodium chloride (5 M)

Dissolved 29.22 g of sodium chloride in 80 ml distilled water and volume made up to 100 ml. Solution filtered and stored at room temperature.

Sodium chloride- EDTA (SE) buffer (pH, 8.0)

Sodium chloride	75 mM	4.383 g
EDTA	35 mM	9.306 g

Dissolved the contents in 900 ml distilled water and pH adjusted to 8.0. Made up the volume to 1000 ml, filtered, autoclaved and stored at 4°C.

Tris Acetate EDTA (TAE) buffer (50X)

Tris base	48.4 g
Glacial acetic acid	11.42 ml
0.5 M EDTA (pH 8.0)	20 ml
Distilled water up to	1000 ml

Autoclaved and stored at room temperature.

Tris-Borate (TBE) buffer pH 8.3) 10X

Tris base	108.0 g
Boric acid	55.0 g
EDTA	9.3 g

Dissolved in 700 ml of distilled water and pH adjusted to 8.3. Volume was made up to 1000 ml, autoclaved and stored at room temperature.

Tris Buffered Saline (TBS) pH 7.4

Sodium chloride	140 mM	8.18 g
Potassium chloride	0.5 mM	0.0373 g
Tris base	0.25 mM	0.0303 g

Dissolved in 900 ml of distilled water and pH adjusted to 7.4. Made up the volume to 1000 ml, filtered, autoclaved and stored at 4 °C.

Tris EDTA (TE) buffer (pH 8.0)

Tris base	10 mM	1.2114 g
EDTA	0.1 mM	0.3722 g

Dissolved in 900 ml of distilled water and adjusted the pH to 8.0. Made up the volume to 1000 ml, filtered, autoclaved and stored at 4 °C.

Tris 1*M* (pH 8.0)

Tris base	121.14 g

Added distilled water up to 1000 ml, pH adjusted to 8.0, filtered and stored at room temperature.

ANNEXURE – II

SOURCES OF IMPORTANT CHEMICALS AND ENZYMES USED IN THIS STUDY

(A) CHEMICALS

Agar	- Himedia Laboratories,. Mumbai
Agarose (Low EED)	- Bangalore Genei Pvt. Ltd.
Ammonium chloride	- SRL, Bombay
Boric acid	- SRL, Bombay
Chloroform	- Merck
Crystalline phenol	- Merck
Di-sodium hydrogen orthophosphate	- SRL, Bombay
dNTPs	- Finn Enzymes
EDTA	- SRL, Bombay
Ethanol	- Merck
Ethidium bromide	- BDH lab, England
6 X gel loading buffer	- Bangalore Genei Pvt. Ltd.
Glacial acetic acid	- BDH-E, Merck (India) Ltd.
Isoamyl alcohol	- Merck
Isopropyl alcohol	- SRL, Bombay
Potassium chloride	- SRL, Bombay
Sodium acetate	- SRL, Bombay
Sodium chloride	- SRL, Bombay

Tris base	- SRL, Bombay
Tryptone	- Himedia Laboratories, Mumbai
Yeast extract	- Himedia Laboratories, Mumbai

(B) PRIMERS

Integrated DNA Technologies, Inc.

1710 Commercial Park, Coralville, IA (www.idtdna.com)

(C) MOLECULAR MARKERS

500 bp DNA Ladder	- Bangalore Genei Pvt. Ltd.

(D) ENZYMES

Taq DNA polymerase	- Bangalore Genei Pvt. Ltd.
Proteinase-K	- Bangalore Genei Pvt. Ltd.

(E) KITS

Transformation kit	- Bangalore Genei Pvt. Ltd.
Cloning kit	- Promega Corporation, USA
Plasmid Miniprep kit	- Axygen Biosciences, USA

ANNEXURE – III

ABBREVIATIONS

A	Absorbance
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide BLAST
BLASTp	Protein BLAST
bp	base pair
$CaCl_2$	Calcium Chloride
cDNA	complementary Deoxyribo Nucleic acid
CDS	Coding Sequence
cm	centimeter
Da	Dalton
DNA	Deoxyribo Nucleic Acid
dNTP.	Deoxyribo Nucleotide Triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylene Diamine Tetraacetic Acid
E-value	Expect value
ExPASy	Expert Protein Analysis System
FAO	Food and Agricultural Organization
g	gram
g	Relative centrifugal force
i.e.	that is
IPTG	Isopropyl β-D-1-thiogalactopyranoside.
kb	Kilo basepair
kDa	Kilo Dalton
kg	kilogram
L	litre
LB-Amp	Luria Bertani-Ampicillin
M	moles per litre

MCS	Multi Cloning Site
mg	milligram
$MgCl_2$	Magnesium chloride
m*M.*	millimolar
mRNA	messenger Ribo Nucleic Acid
MYR	Million year
NaCl	Sodium Chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
NH_4Cl	Ammonium Chloride
ng	nano gram
nm	nanometer
OD	Optical Density
PAPIA	Parallel Protein Information Analysis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
p*M*	Pico moles
RFLP	Restriction Fragment Length Polymorphism
rpm	Revolutions per minute
RT	Room temperature
s	second
SDS	Sodium Dodecyl Sulphate
ss	single stranded
UV	ultra violet
WBC	white blood cell
X-gal	5-bromo-4-chloro-3-indolyl- beta-D- galactopyranoside
µg	microgram
µl	microlitre
µm	micrometer
°C	degree Celsius
3D	Three Dimensional

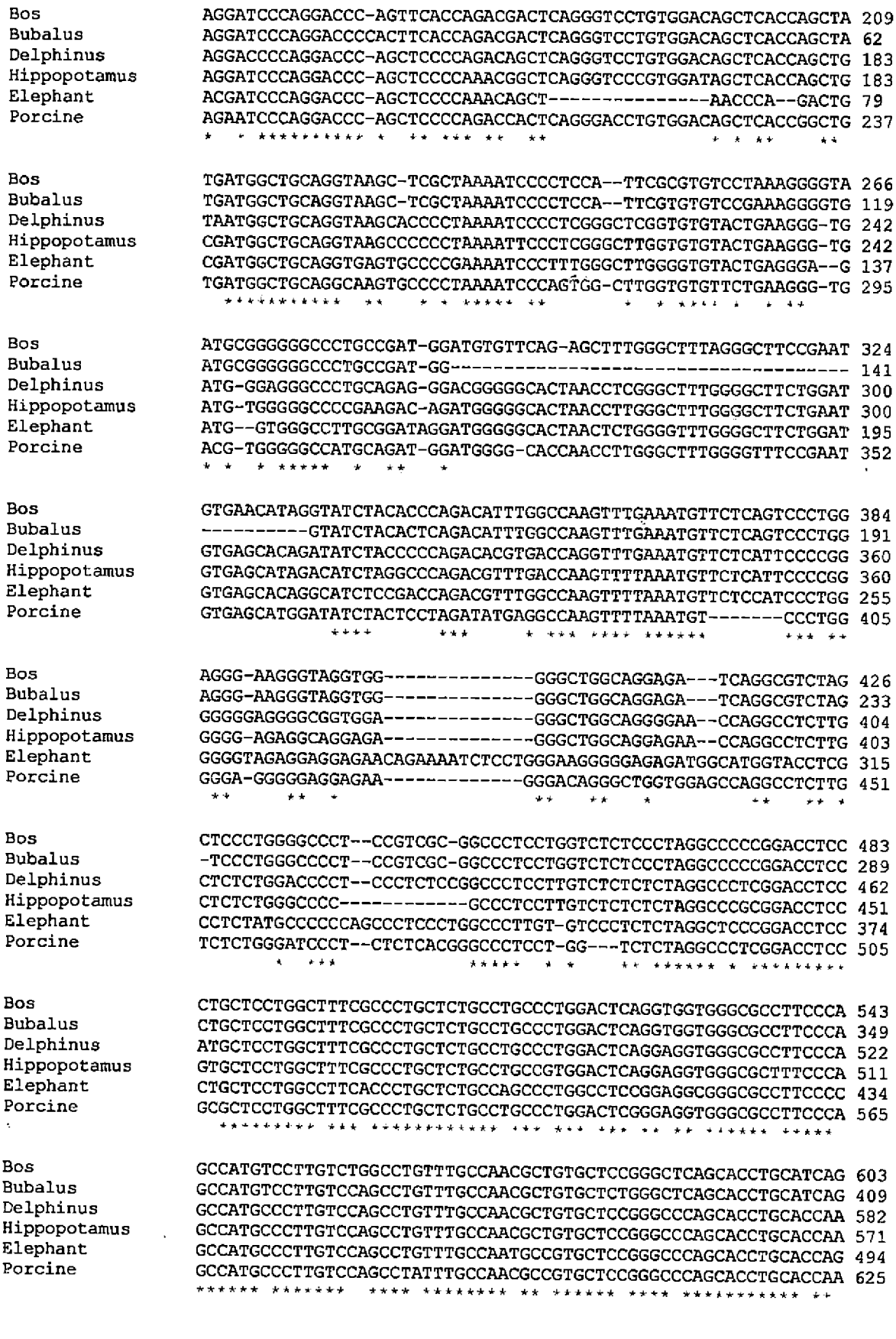
ANNEXURE – IV

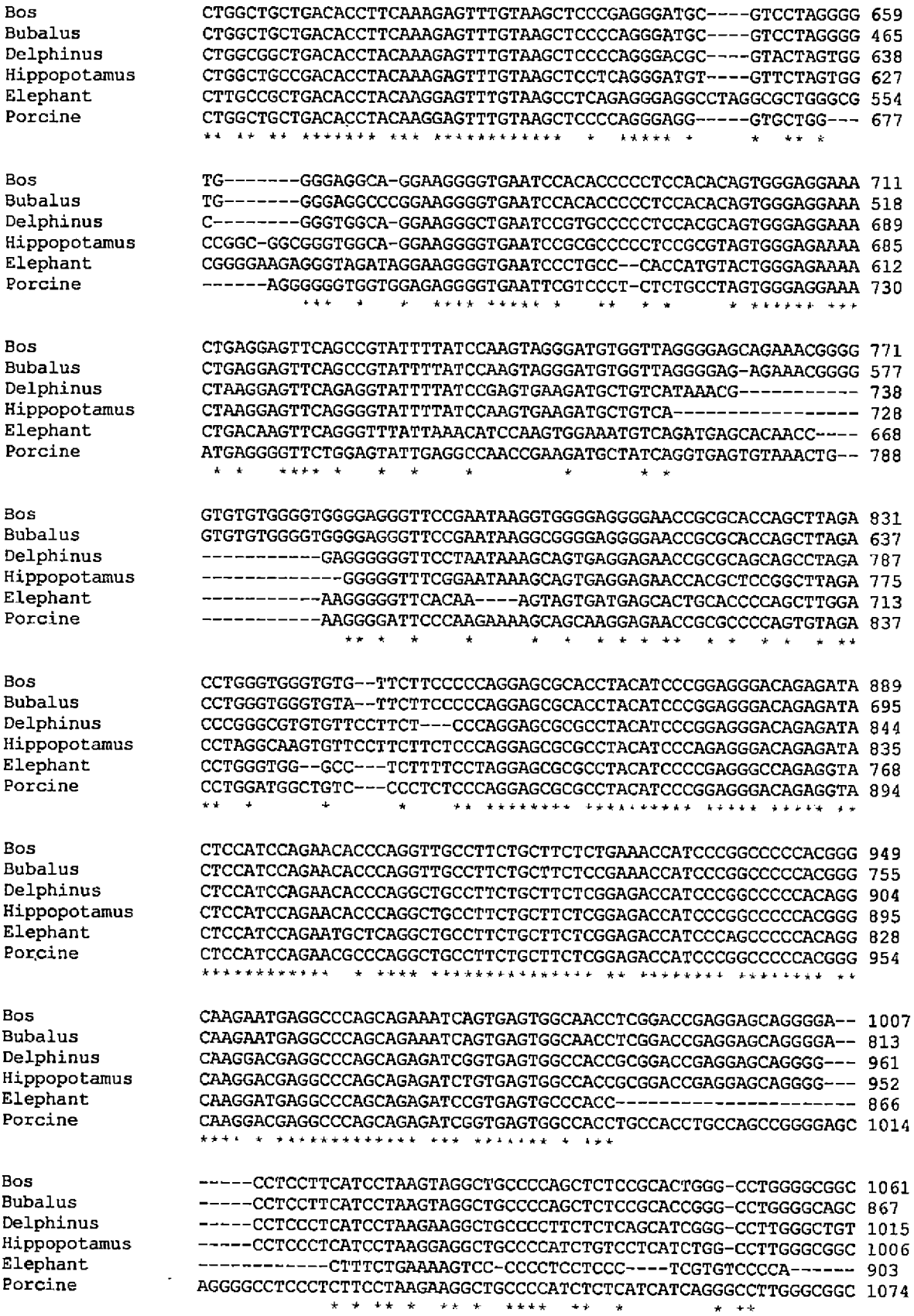
CLUSTALW MULTIPLE NUCLEOTIDE SEQUENCE ALIGNMENT OF INDIAN ELEPHANT GROWTH HORMONE GENE WITH CORRESPONDING REGIONS OF OTHER LIVESTOCK SPECIES

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Elephant	1712	2	Porcine	2231	78
1	Elephant	1712	3	Delphinus	1900	78
1	Elephant	1712	4	Hippopotamus	1887	76
1	Elephant	1712	5	Bos	2061	76
1	Elephant	1712	6	Bubalus	1798	66
2	Porcine	2231	3	Delphinus	1900	85
2	Porcine	2231	4	Hippopotamus	1887	82
2	Porcine	2231	5	Bos	2061	78
2	Porcine	2231	6	Bubalus	1798	79
3	Delphinus	1900	4	Hippopotamus	1887	88
3	Delphinus	1900	5	Bos	2061	84
3	Delphinus	1900	6	Bubalus	1798	81
4	Hippopotamus	1887	5	Bos	2061	83
4	Hippopotamus	1887	6	Bubalus	1798	79
5	Bos	2061	6	Bubalus	1798	97

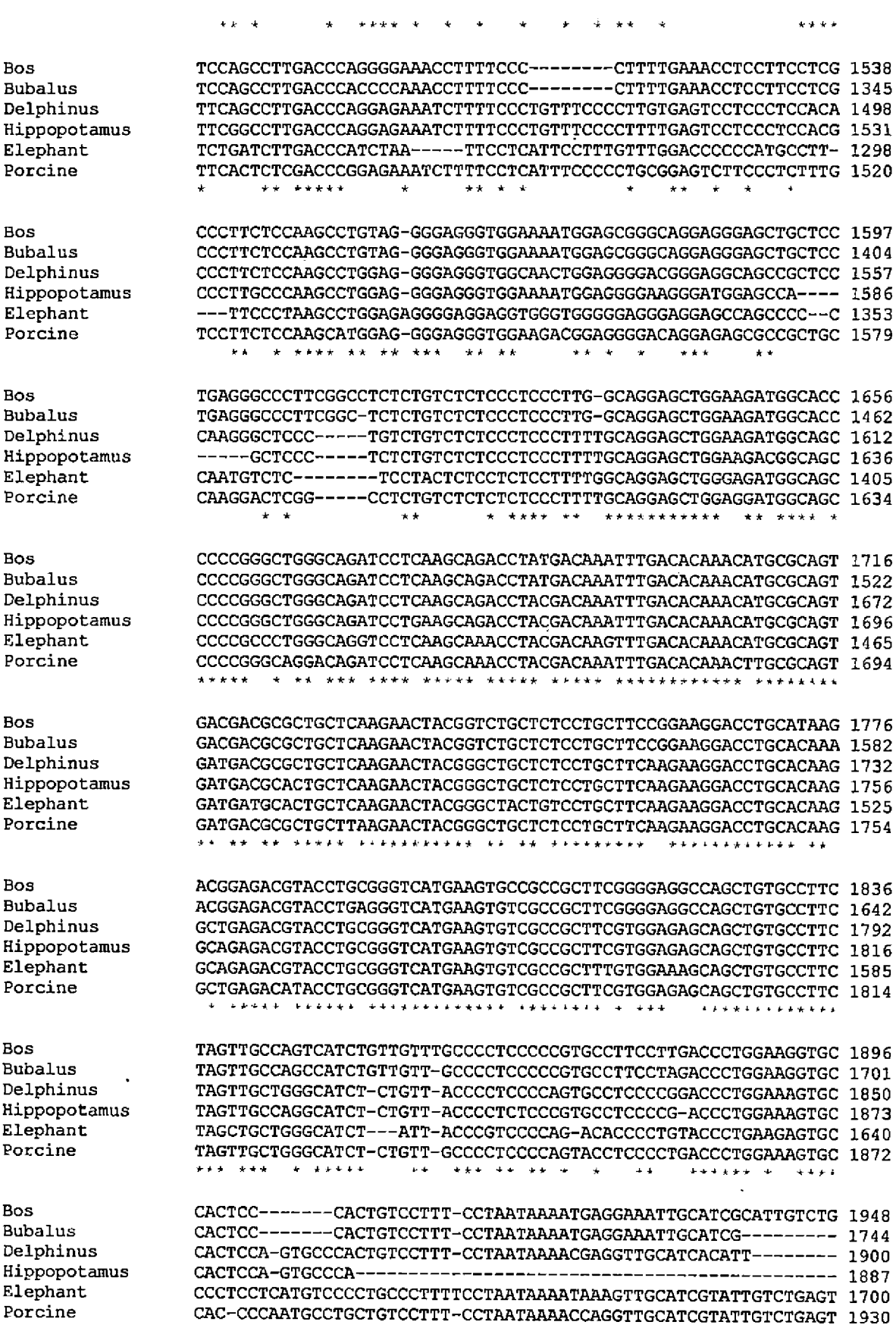
Elephant = *Elephas maximus*
Porcine= *Sus scrofa*
Delphinus= *Delphinus delphin*
Hippopotamus= *Hippopotamus amphibius*
Bos= *Bos indicus*
Bubalus= *Bubalus bubalis*

```
CLUSTAL 2.0.10 multiple sequence alignment

Bos             ----------------------------GGGAACAGGATGAGTGAGAGGAGGTTCTAAA 31
Bubalus         ------------------------------------------------------------
Delphinus       ------------------------------------------------------TAAA 4
Hippopotamus    ------------------------------------------------------TAAA 4
Elephant        ------------------------------------------------------------
Porcine         CCCGGGGGACATGACCCCAGAGGAGGAGCGGGAACAGGATGAGTGGGAGGAGGTTCTAAA 60

Bos             TTATCCATTAGCACAGGCT-GCCAGTGGTCCTTGCATAAATGTATAGAGCACACAGGTGG 90
Bubalus         ------------------------------------------------------------
Delphinus       TTATCCATTAGCCCAGGCCCGCCGGTGGCCCGTGCATAAATGTATAGAGAAAATAGGTGG 64
Hippopotamus    TTATCCATTAGCAAAGGCCCGCCAGTGGCCCATGCATAAATGTATAGAGAAAATAGGTGG 64
Elephant        ------------------------------------------------------------
Porcine         TTATCCATTAGCACATGCCTGCCAGTGGGCCATGCATAAATGTATAGAGAAAATAGGTGG 120

Bos             GGGGAAAGGGAGAGAGAGAAGAAGCCAGGGTATAAAAATGGCCCAGCAGGGACCAATTCC 150
Bubalus         ----------------------------------------------------------CC 2
Delphinus       GGGCAGAGGGAGAGAGAGAAGAGGCCAGGGTATAAAAACGGCCCACAAGGGACCAATTCC 124
Hippopotamus    GGGCAGAGGGAAAGAGAGAAGAGGCCAGGGTATAAGAAGGGCCCACGAAGGACCAATTCC 124
Elephant        ----------------------GCCAGAGTATAAAAAGGGCCCATGAGGGACTGATTCC 37
Porcine         GGGCAGAGGGA--GAGAGAAGAGGCCAGGGTATAAAAAGGGCCCAAAAGGGACCAATTCC 178
                                                                        **
```


Bos AGGATCCCAGGACCC-AGTTCACCAGACGACTCAGGGTCCTGTGGACAGCTCACCAGCTA 209
Bubalus AGGATCCCAGGACCCCACTTCACCAGACGACTCAGGGTCCTGTGGACAGCTCACCAGCTA 62
Delphinus AGGACCCCAGGACCC-AGCTCCCCAGACAGCTCAGGGTCCTGTGGACAGCTCACCAGCTG 183
Hippopotamus AGGATCCCAGGACCC-AGCTCCCCAAACGGCTCAGGGTCCCGTGGATAGCTCACCAGCTG 183
Elephant ACGATCCCAGGACCC-AGCTCCCCAAACAGCT----------AACCCA--GACTG 79
Porcine AGAATCCCAGGACCC-AGCTCCCCAGACCACTCAGGGACCTGTGGACAGCTCACCGGCTG 237
Bos TGATGGCTGCAGGTAAGC-TCGCTAAAATCCCCTCCA--TTCGCGTGTCCTAAAGGGGTA 266
Bubalus TGATGGCTGCAGGTAAGC-TCGCTAAAATCCCCTCCA--TTCGTGTGTCCGAAAGGGGTG 119
Delphinus TAATGGCTGCAGGTAAGCACCCCTAAAATCCCCTCGGGCTCGGTGTGTACTGAAGGG-TG 242
Hippopotamus CGATGGCTGCAGGTAAGCCCCCCTAAAATTCCCTCGGGCTTGGTGTGTACTGAAGGG-TG 242
Elephant CGATGGCTGCAGGTGAGTGCCCCGAAAATCCCTTTGGGCTTGGGGTGTACTGAGGGA--G 137
Porcine TGATGGCTGCAGGCAAGTGCCCCTAAAATCCCAGTGG-CTTGGTGTGTTCTGAAGGG-TG 295
Bos ATGCGGGGGGCCCTGCCGAT-GGATGTGTTCAG-AGCTTTGGGCTTTAGGGCTTCCGAAT 324
Bubalus ATGCGGGGGGCCCTGCCGAT-GG---------- 141
Delphinus ATG-GGAGGGCCCTGCAGAG-GGACGGGGGCACTAACCTCGGGCTTTGGGGCTTCTGGAT 300
Hippopotamus ATG-TGGGGGCCCCGAAGAC-AGATGGGGGCACTAACCTTGGGCTTTGGGGCTTCTGAAT 300
Elephant ATG--GTGGGCCTTGCGGATAGGATGGGGGCACTAACTCTGGGGTTTGGGGCTTCTGGAT 195
Porcine ACG-TGGGGGCCATGCAGAT-GGATGGGG-CACCAACCTTGGGCTTTGGGGTTTCCGAAT 352
Bos GTGAACATAGGTATCTACACCCAGACATTTGGCCAAGTTTGAAATGTTCTCAGTCCCTGG 384
Bubalus ----------GTATCTACACTCAGACATTTGGCCAAGTTTGAAATGTTCTCAGTCCCTGG 191
Delphinus GTGAGCACAGATATCTACCCCCAGACACGTGACCAGGTTTGAAATGTTCTCATTCCCCGG 360
Hippopotamus GTGAGCATAGACATCTAGGCCCAGACGTTTGACCAAGTTTTAAATGTTCTCATTCCCCGG 360
Elephant GTGAGCACAGGCATCTCCGACCAGACGTTTGGCCAAGTTTTAAATGTTCTCCATCCCTGG 255
Porcine GTGAGCATGGATATCTACTCCTAGATATGAGGCCAAGTTTTAAATGT-------CCCTGG 405
Bos AGGG-AAGGGTAGGTGG----------GGGCTGGCAGGAGA---TCAGGCGTCTAG 426
Bubalus AGGG-AAGGGTAGGTGG----------GGGCTGGCAGGAGA---TCAGGCGTCTAG 233
Delphinus GGGGGAGGGGCGGTGGA----------GGGCTGGCAGGGGAA--CCAGGCCTCTTG 404
Hippopotamus GGGG-AGAGGCAGGAGA----------GGGCTGGCAGGAGAA--CCAGGCCTCTTG 403
Elephant GGGGTAGAGGAGGAGAACAGAAAATCTCCTGGGAAGGGGGAGAGATGGCATGGTACCTCG 315
Porcine GGGA-GGGGGAGGAGAA----------GGGACAGGGCTGGTGGAGCCAGGCCTCTTG 451
Bos CTCCCTGGGGCCCT--CCGTCGC-GGCCCTCCTGGTCTCTCCCTAGGCCCCCGGACCTCC 483
Bubalus -TCCCTGGGCCCCT--CCGTCGC-GGCCCTCCTGGTCTCTCCCTAGGCCCCCGGACCTCC 289
Delphinus CTCTCTGGACCCCT--CCCTCTCCGGCCCTCCTTGTCTCTCTCTAGGCCCTCGGACCTCC 462
Hippopotamus CTCTCTGGGCCCC----------GCCCTCCTTGTCTCTCTCTAGGCCCGCGGACCTCC 451
Elephant CCTCTATGCCCCCCAGCCCTCCCTGGCCCTTGT-GTCCCTCTCTAGGCTCCCGGACCTCC 374
Porcine TCTCTGGGATCCCT--CTCTCACGGGCCCTCCT-GG---TCTCTAGGCCCTCGGACCTCC 505
Bos CTGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCCTGGACTCAGGTGGTGGGCGCCTTCCCA 543
Bubalus CTGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCCTGGACTCAGGTGGTGGGCGCCTTCCCA 349
Delphinus ATGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCCTGGACTCAGGAGGTGGGCGCCTTCCCA 522
Hippopotamus GTGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCGTGGACTCAGGAGGTGGGCGCTTTCCCA 511
Elephant CTGCTCCTGGCCTTCACCCTGCTCTGCCAGCCCTGGCCTCCGGAGGCGGGCGCCTTCCCC 434
Porcine GCGCTCCTGGCTTTCGCCCTGCTCTGCCTGCCCTGGACTCGGGAGGTGGGCGCCTTCCCA 565
Bos GCCATGTCCTTGTCTGGCCTGTTTGCCAACGCTGTGCTCCGGGCTCAGCACCTGCATCAG 603
Bubalus GCCATGTCCTTGTCCAGCCTGTTTGCCAACGCTGTGCTCTGGGCTCAGCACCTGCATCAG 409
Delphinus GCCATGCCCTTGTCCAGCCTGTTTGCCAACGCTGTGCTCCGGGCCCAGCACCTGCACCAA 582
Hippopotamus GCCATGCCCTTGTCCAGCCTGTTTGCCAACGCTGTGCTCCGGGCCCAGCACCTGCACCAA 571
Elephant GCCATGCCCTTGTCCAGCCTGTTTGCCAATGCCGTGCTCCGGGCCCAGCACCTGCACCAG 494
Porcine GCCATGCCCTTGTCCAGCCTATTTGCCAACGCCGTGCTCCGGGCCCAGCACCTGCACCAA 625


Bos CTGGCTGCTGACACCTTCAAAGAGTTTGTAAGCTCCCGAGGGATGC----GTCCTAGGGG 659
Bubalus CTGGCTGCTGACACCTTCAAAGAGTTTGTAAGCTCCCCAGGGATGC----GTCCTAGGGG 465
Delphinus CTGGCGGCTGACACCTACAAAGAGTTTGTAAGCTCCCCAGGGACGC----GTACTAGTGG 638
Hippopotamus CTGGCTGCCGACACCTACAAAGAGTTTGTAAGCTCCTCAGGGATGT----GTTCTAGTGG 627
Elephant CTTGCCGCTGACACCTACAAGGAGTTTGTAAGCCTCAGAGGGAGGCCTAGGCGCTGGGCG 554
Porcine CTGGCTGCTGACACCTACAAGGAGTTTGTAAGCTCCCCAGGGAGG-----GTGCTGG--- 677
Bos TG--------GGGAGGCA-GGAAGGGGTGAATCCACACCCCCTCCACACAGTGGGAGGAAA 711
Bubalus TG--------GGGAGGCCCGGAAGGGGTGAATCCACACCCCCTCCACACAGTGGGAGGAAA 518
Delphinus C---------GGGTGGCA-GGAAGGGCTGAATCCGTGCCCCCTCCACGCAGTGGGAGGAAA 689
Hippopotamus CCGGC-GGCGGGTGGCA-GGAAGGGGTGAATCCGCGCCCCCTCCGCGTAGTGGGAGAAAA 685
Elephant CGGGGAAGAGGGTAGATAGGAAGGGGTGAATCCCTGCC--CACCATGTACTGGGAGAAAA 612
Porcine ------AGGGGGGTGGTGGAGAGGGGTGAATTCGTCCCT-CTCTGCCTAGTGGGAGGAAA 730
Bos CTGAGGAGTTCAGCCGTATTTTATCCAAGTAGGGATGTGGTTAGGGGAGCAGAAACGGGG 771
Bubalus CTGAGGAGTTCAGCCGTATTTTATCCAAGTAGGGATGTGGTTAGGGGAG-AGAAACGGGG 577
Delphinus CTAAGGAGTTCAGAGGTATTTTATCCGAGTGAAGATGCTGTCATAAACG---------- 738
Hippopotamus CTAAGGAGTTCAGGGGTATTTTATCCAAGTGAAGATGCTGTCA---------- 728
Elephant CTGACAAGTTCAGGGTTTATTAAACATCCAAGTGGAAATGTCAGATGAGCACAACC----- 668
Porcine ATGAGGGGTTCTGGAGTATTGAGGCCAACCGAAGATGCTATCAGGTGAGTGTAAACTG-- 788
Bos GTGTGTGGGGTGGGGAGGGTTCCGAATAAGGTGGGGAGGGGAACCGCGCACCAGCTTAGA 831
Bubalus GTGTGTGGGGTGGGGAGGGTTCCGAATAAGGCGGGGAGGGGAACCGCGCACCAGCTTAGA 637
Delphinus ----------GAGGGGGGTTCCTAATAAAGCAGTGAGGAGAACCGCGCAGCAGCCTAGA 787
Hippopotamus ----------GGGGGTTTCGGAATAAAGCAGTGAGGAGAACCACGCTCCGGCTTAGA 775
Elephant ----------AAGGGGGTTCACAA----AGTAGTGATGAGCACTGCACCCCAGCTTGGA 713
Porcine ----------AAGGGGATTCCCAAGAAAAGCAGCAAGGAGAACCGCGCCCCAGTGTAGA 837
Bos CCTGGGTGGGTGTG--TTCTTCCCCCAGGAGCGCACCTACATCCCGGAGGGACAGAGATA 889
Bubalus CCTGGGTGGGTGTA--TTCTTCCCCCAGGAGCGCACCTACATCCCGGAGGGACAGAGATA 695
Delphinus CCCGGGCGTGTGTTCCTTCT---CCCAGGAGCGCGCCTACATCCCGGAGGGACAGAGATA 844
Hippopotamus CCTAGGCAAGTGTTCCTTCTTCTCCCAGGAGCGCGCCTACATCCCAGAGGGACAGAGATA 835
Elephant CCTGGGTGG--GCC---TCTTTTCCTAGGAGCGCGCCTACATCCCCGAGGGCCAGAGGTA 768
Porcine CCTGGATGGCTGTC---CCCTCTCCCAGGAGCGCGCCTACATCCCGGAGGGACAGAGGTA 894
Bos CTCCATCCAGAACACCCAGGTTGCCTTCTGCTTCTCTGAAACCATCCCGGCCCCCACGGG 949
Bubalus CTCCATCCAGAACACCCAGGTTGCCTTCTGCTTCTCCGAAACCATCCCGGCCCCCACGGG 755
Delphinus CTCCATCCAGAACACCCAGGCTGCCTTCTGCTTCTCGGAGACCATCCCGGCCCCCACAGG 904
Hippopotamus CTCCATCCAGAACACCCAGGCTGCCTTCTGCTTCTCGGAGACCATCCCGGCCCCCACGGG 895
Elephant CTCCATCCAGAATGCTCAGGCTGCCTTCTGCTTCTCGGAGACCATCCCAGCCCCCACAGG 828
Porcine CTCCATCCAGAACGCCCAGGCTGCCTTCTGCTTCTCGGAGACCATCCCGGCCCCCACGGG 954
Bos CAAGAATGAGGCCCAGCAGAAATCAGTGAGTGGCAACCTCGGACCGAGGAGCAGGGGA-- 1007
Bubalus CAAGAATGAGGCCCAGCAGAAATCAGTGAGTGGCAACCTCGGACCGAGGAGCAGGGGA-- 813
Delphinus CAAGGACGAGGCCCAGCAGAGATCGGTGAGTGGCCACCGCGGACCGAGGAGCAGGGG--- 961
Hippopotamus CAAGGACGAGGCCCAGCAGAGATCTGTGAGTGGCCACCGCGGACCGAGGAGCAGGGG--- 952
Elephant CAAGGATGAGGCCCAGCAGAGATCCGTGAGTGCCCACC---------- 866
Porcine CAAGGACGAGGCCCAGCAGAGATCGGTGAGTGGCCACCTGCCACCTGCCAGCCGGGGAGC 1014
Bos -----CCTCCTTCATCCTAAGTAGGCTGCCCCAGCTCTCCGCACTGGG-CCTGGGGCGGC 1061
Bubalus -----CCTCCTTCATCCTAAGTAGGCTGCCCCAGCTCTCCGCACCGGG-CCTGGGGCAGC 867
Delphinus -----CCTCCCTCATCCTAAGAAGGCTGCCCCTTCTCTCAGCATCGGG-CCTTGGGCTGT 1015
Hippopotamus -----CCTCCCTCATCCTAAGGAGGCTGCCCCATCTGTCCTCATCTGG-CCTTGGGCGGC 1006
Elephant ----------CTTTCTGAAAAGTCC-CCCCTCCTCCC----TCGTGTCCCCA------ 903
Porcine AGGGGCCTCCCTCTTCCTAAGAAGGCTGCCCCATCTCTCATCATCAGGGCCTTGGGCGGC 1074

```
Bos              CTTCTCCCCGAGGTGGCG-GAGGTTGTTGG----------------ATGGCAGTGGAGG 1103
Bubalus          CTTCTCCCCGAGGTGGCCCGAGAATGTTGG----------------ATGGCAGTGGAGG 910
Delphinus        CTTCTCTCCGAGGAGGTG-CAGGGTGGTGG----------------CCGGTGGCGGAAG 1057
Hippopotamus     CTTCTCCCCGAGGTGGCG-GAGGGTGGTGGTTGGCGGAGGGTGGTGGCTGGCGGCGGAAG 1065
Elephant         CCTCACCCGAGGCCTA-------------------------------------------G 920
Porcine          CTTCTCCCCGAGCTGGTG--GGGGTGATGG----------------TGGCAGAGGGCG 1114
                 * ** * *   *                                              *

Bos              ATGATGG-----TGGGCGGT--GGTGGC-------------------AGGAGGTCCTCG 1136
Bubalus          ATCATGG-----TGGGCGGT--GGTGGC-------------------AGGAGGTCCTCG 943
Delphinus        GTGACGGATGG-CGGAGGGA-GGACCGC-------------------GGCTGCCCCCCT 1095
Hippopotamus     GTGATGGACAGACGGACGGACGGACGGCAGGAGGTGGTGGTTGGAGGTGGGAGGCCCACA 1125
Elephant         A-------------GGCGGCCAGG----------------------------TGCCCTCC 939
Porcine          G-------------GGTGGTGAGGGGGA------------------------CGCCCACC 1137
                              *  **   *                                 ** *

Bos              GGCAGAGGCCGACCTTGCAGGGCTGCCCCA-AGCCCGCGGCACCCACCGACCACCCATCT 1195
Bubalus          GGCAGAGGCCGACCTTGCAGGGCTGCCCCA-A-CCCGCGGCACCCACCGACCACCCATCT 1001
Delphinus        CCC-----CCGCCCCCGC-CACCGACCTCGCGGCGCCCACCGCCCACCACCCACCCATCT 1149
Hippopotamus     GGCAGAGGCTGGCCTTGCACAGCAGCCCCG-AGCCCTCGGCGCCCAGCACCCACCCATCT 1184
Elephant         GC----------------------ACCCCC---------ACACCCACTGCCCGGCCCTCT 968
Porcine          GGC---------------------GGAGGCA---------GCGCCCCCCATCCACGCATCT 1168
                                         *           * ***     **   * ***

Bos              GCCAGCAGGACTTGGAGCTGCTTCGCATCTCACTGCTCCTCATCCAGTCGTGGCTTGGGC 1255
Bubalus          GCCAGCAGGACTTGGAGCTGCTTCGCATCTCACTGCTCCTTATCCAGTCGTGGCTTGGGC 1061
Delphinus        GCCCGCAGGACGTGGAGCTGCTCCGCTTCTCACTGCTGCTCATCCAGTCGTGGCTCGGGC 1209
Hippopotamus     GTCCGCAGGACGTGGAGCTGCTCCGCTTCTCGCTGCTGCTCATCCAGTCGTGGCTTGGGC 1244
Elephant         GCCCGCAGGACGTGGAGCTGCTGCGCTTCTCCCTGCTGCTCATCCAGTCGTGGCTCGGGC 1028
Porcine          GCCCGCAGGACGTGGAGCTGCTGCGCTTCTCGCTGCTGCTCATCCAGTCGTGGCTCGGGC 1228
                 * * ******* ********** *** **** ***** ** ************** ****

Bos              CCCTGCAGTTCCTCAGCAGAGTCTTCACCAACAGCTTGGTGTTTGGCACCTCGGACCGTG 1315
Bubalus          CCCTGCAGTTCCTCAGCAGAGTCTTCACCAACAGCTTGGTGTTTGGCACCTCGGACCGTG 1121
Delphinus        CCGTGCAGTTCCTCAGCAGGGTCTTCACCAACAGCCTGGTGTTTGGCACCTCGGACCGCG 1269
Hippopotamus     CGGTGCAGTTCCTCAGCAGGGTCTTCACCAACAGCCTGGTGTTTGGCACCTCGGACCGCG 1304
Elephant         CAGTGCAGTTCCTCAGCAGGGTCTTCACCAACAGCCTGGTGTTCGGCACCTCGGACCGTG 1088
Porcine          CCGTGCAGTTCCTCAGCAGGGTCTTCACCAACAGCCTGGTGTTTGGCACCTCAGACCGCG 1288
                 *  ***************** ************** ******* ******** ***** *

Bos              TCTATGAGAAGCTGAAGGACCTGGAGGAAGGCATCCTGGCCCTGATGCGGGTGGGGATGG 1375
Bubalus          TCTATGAGAAGCTGAAGGACCTGGAGGAAGGCATCCTGGCCCTGATGCGGGTGGGGATGG 1181
Delphinus        TCTATGAGAAGCTGAAGGACCTGGAGGAGGGCATCCAGGCCCTGATGCGGGTGGGGATGG 1329
Hippopotamus     TCTATGAGAAGCTGAAGGACCTGGAGGAAGGCATCCAGGCCCTGATGCGGGTGGGGATGG 1364
Elephant         TCTACGAGAAGCTGAAGGACCTGGAGGAAGGCATCCAGGCCCTGATGCGGGTAGAGGGGA 1148
Porcine          TCTACGAGAAGCTGAAGGACCTGGAGGAGGGCATCCAGGCCCTGATGCGGGTGGGGAGGC 1348
                 **** ************************** ******** ************** * *  *

Bos              CGTTGTGGGTCCCTTCCATG-TGGGGGCCATGCCCGCCCTCTCCTGGCTTAGCCA----- 1429
Bubalus          CGTTGTGGGTCCCTTCCATGCTGGGGGCCATGCCCGCCCTCTCCTGGCTTAGCCA----- 1236
Delphinus        CGTAGTGGGTCC--CCCAGACTGGGA-CCACGCCCGCCCTCTCCTGGCTTAGCCGAG-GG 1385
Hippopotamus     CGTTGCGGGTCC--TCCACCCTGGGC-CCATGCCCGCCCCCTCCTGGCTTAGCCGAG-GG 1420
Elephant         CTGT--GGGTCA--CCAGCCCTGGG-----TGCCCACTA------GGCTTAGCTCAG-GG 1192
Porcine          GCGCTCGGGTCC--CGCACACTGGGGCCCATGCCGGCTCTCTCCCGGCTGAGCGGAGCGG 1406
                       *****          ****      ***  *        **** ***

Bos              --GGAGAATGCACGTGGGCTGGGGGAGACAGATCCC-TGCTCTCTCCCTCTTTCTAGCAG 1486
Bubalus          --GGAGAACGCACGTGGGCTGGGGGAGAGAGATCCC-TGCTCTCTCTCTCTTCCTAGCAG 1293
Delphinus        ATGGGGGGCTCCCGTGGGCTGGGGGAGAGAGGTGCC-TGCTCTCTCTGT------GGCAG 1438
Hippopotamus     GTGGGGGGCGCACGTGGGCTGGGGGAGAGAGGTACT-TGCTCTCTCTA--------GCAG 1471
Elephant         TAGGGGGCCTAACTGGGGCAGACGAGGGAAGTCAC--TACTGGCCTCCT------AGCAG 1244
Porcine          TGGGGGGACGCACGTGGGCTGGGGGAGAGGGTCCCGATGCTCTCTCTGT------AGCAG 1460
```


Bos TCCAGCCTTGACCCAGGGGAAACCTTTTCCC--------CTTTTGAAACCTCCTTCCTCG 1538
Bubalus TCCAGCCTTGACCCACCCCAAACCTTTTCCC--------CTTTTGAAACCTCCTTCCTCG 1345
Delphinus TTCAGCCTTGACCCAGGAGAAATCTTTTCCCTGTTTCCCCTTGTGAGTCCTCCCTCCACA 1498
Hippopotamus TTCGGCCTTGACCCAGGAGAAATCTTTTCCCTGTTTCCCCTTTTGAGTCCTCCCTCCACG 1531
Elephant TCTGATCTTGACCCATCTAA-----TTCCTCATTCCTTTGTTTGGACCCCCCATGCCTT- 1298
Porcine TTCACTCTCGACCCGGAGAAATCTTTTCCTCATTTCCCCCTGCGGAGTCTTCCCTCTTTG 1520
Bos CCCTTCTCCAAGCCTGTAG-GGGAGGGTGGAAAATGGAGCGGGCAGGAGGGAGCTGCTCC 1597
Bubalus CCCTTCTCCAAGCCTGTAG-GGGAGGGTGGAAAATGGAGCGGGCAGGAGGGAGCTGCTCC 1404
Delphinus CCCTTCTCCAAGCCTGGAG-GGGAGGGTGGCAACTGGAGGGGACGGGAGGCAGCCGCTCC 1557
Hippopotamus CCCTTGCCCAAGCCTGGAG-GGGAGGGTGGAAAATGGAGGGGAAGGGATGGAGCCA---- 1586
Elephant ---TTCCCTAAGCCTGGAGAGGGGAGGAGGTGGGTGGGGGAGGGAGGAGCCAGCCCC--C 1353
Porcine TCCTTCTCCAAGCATGGAG-GGGAGGGTGGAAGACGGAGGGGACAGGAGAGCGCCGCTGC 1579
Bos TGAGGGCCCTTCGGCCTCTCTGTCTCTCCCTCCCTTG-GCAGGAGCTGGAAGATGGCACC 1656
Bubalus TGAGGGCCCTTCGGC-TCTCTGTCTCTCCCTCCCTTG-GCAGGAGCTGGAAGATGGCACC 1462
Delphinus CAAGGGCTCCC-----TGTCTGTCTCTCCCTCCCTTTTGCAGGAGCTGGAAGATGGCAGC 1612
Hippopotamus -----GCTCCC-----TCTCTGTCTCTCCCTCCCTTTTGCAGGAGCTGGAAGACGGCAGC 1636
Elephant CAATGTCTC--------TCCTACTCTCCTCTCCTTTTGGCAGGAGCTGGGAGATGGCAGC 1405
Porcine CAAGGACTCGG-----CCTCTGTCTCTCTCTCCCTTTTGCAGGAGCTGGAGGATGGCAGC 1634
Bos CCCCGGGCTGGGCAGATCCTCAAGCAGACCTATGACAAATTTGACACAAACATGCGCAGT 1716
Bubalus CCCCGGGCTGGGCAGATCCTCAAGCAGACCTATGACAAATTTGACACAAACATGCGCAGT 1522
Delphinus CCCCGGGCTGGGCAGATCCTCAAGCAGACCTACGACAAATTTGACACAAACATGCGCAGT 1672
Hippopotamus CCCCGGGCTGGGCAGATCCTGAAGCAGACCTACGACAAATTTGACACAAACATGCGCAGT 1696
Elephant CCCCGCCCTGGGCAGGTCCTCAAGCAAACCTACGACAAGTTTGACACAAACATGCGCAGT 1465
Porcine CCCCGGGCAGGACAGATCCTCAAGCAAACCTACGACAAATTTGACACAAACTTGCGCAGT 1694
Bos GACGACGCGCTGCTCAAGAACTACGGTCTGCTCTCCTGCTTCCGGAAGGACCTGCATAAG 1776
Bubalus GACGACGCGCTGCTCAAGAACTACGGTCTGCTCTCCTGCTTCCGGAAGGACCTGCACAAA 1582
Delphinus GATGACGCGCTGCTCAAGAACTACGGGCTGCTCTCCTGCTTCAAGAAGGACCTGCACAAG 1732
Hippopotamus GATGACGCACTGCTCAAGAACTACGGGCTGCTCTCCTGCTTCAAGAAGGACCTGCACAAG 1756
Elephant GATGATGCACTGCTCAAGAACTACGGGCTACTGTCCTGCTTCAAGAAGGACCTGCACAAG 1525
Porcine GATGACGCGCTGCTTAAGAACTACGGGCTGCTCTCCTGCTTCAAGAAGGACCTGCACAAG 1754
Bos ACGGAGACGTACCTGCGGGTCATGAAGTGCCGCCGCTTCGGGGAGGCCAGCTGTGCCTTC 1836
Bubalus ACGGAGACGTACCTGAGGGTCATGAAGTGTCGCCGCTTCGGGGAGGCCAGCTGTGCCTTC 1642
Delphinus GCTGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTCGTGGAGAGCAGCTGTGCCTTC 1792
Hippopotamus GCAGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTCGTGGAGAGCAGCTGTGCCTTC 1816
Elephant GCAGAGACGTACCTGCGGGTCATGAAGTGTCGCCGCTTTGTGGAAAGCAGCTGTGCCTTC 1585
Porcine GCTGAGACATACCTGCGGGTCATGAAGTGTCGCCGCTTCGTGGAGAGCAGCTGTGCCTTC 1814
Bos TAGTTGCCAGTCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGC 1896
Bubalus TAGTTGCCAGCCATCTGTTGTT-GCCCCTCCCCCGTGCCTTCCTAGACCCTGGAAGGTGC 1701
Delphinus TAGTTGCTGGGCATCT-CTGTT-ACCCCTCCCCAGTGCCTCCCCGGACCCTGGAAAGTGC 1850
Hippopotamus TAGTTGCCAGGCATCT-CTGTT-ACCCCTCTCCCGTGCCTCCCCG-ACCCTGGAAAGTGC 1873
Elephant TAGCTGCTGGGCATCT---ATT-ACCCGTCCCCAG-ACACCCCTGTACCCTGAAGAGTGC 1640
Porcine TAGTTGCTGGGCATCT-CTGTT-GCCCCTCCCCAGTACCTCCCCTGACCCTGGAAAGTGC 1872
Bos CACTCC-------CACTGTCCTTT-CCTAATAAAATGAGGAAATTGCATCGCATTGTCTG 1948
Bubalus CACTCC-------CACTGTCCTTT-CCTAATAAAATGAGGAAATTGCATCG--------- 1744
Delphinus CACTCCA-GTGCCCACTGTCCTTT-CCTAATAAAACGAGGTTGCATCACATT-------- 1900
Hippopotamus CACTCCA-GTGCCCA---------- 1887
Elephant CCCTCCTCATGTCCCCTGCCCTTTTCCTAATAAAATAAAGTTGCATCGTATTGTCTGAGT 1700
Porcine CAC-CCCAATGCCTGCTGTCCTTT-CCTAATAAAACCAGGTTGCATCGTATTGTCTGAGT 1930

```
                    * *  **

Bos                 AGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG 2008
Bubalus             ----------------------GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG 1783
Delphinus           ------------------------------------------------------------
Hippopotamus        ------------------------------------------------------------
Elephant            AGGTGTCATTCT--------------------------------------------------- 1712
Porcine             AGGTGTCACTCTGCGATGGAGGGAGGTGGGGCAGTAGGGCAAGGGGTGGGGGTGGGAAGA 1990

Bos                 GAAGACAATAGCAGGCATGCTGGGGATGCTGTGGGCTCTATGGGTACCCAGGT------- 2061
Bubalus             GAAGACAATAGCAGG------------------------------------------- 1798
Delphinus           ------------------------------------------------------------
Hippopotamus        ------------------------------------------------------------
Elephant            ------------------------------------------------------------
Porcine             CAACCTGCAGGCATCCTTGGGGGTCTCCTGGGGACCTAGACACTGAATGATGGTTGACCC 2050

Bos                 ------------------------------------------------------------
Bubalus             ------------------------------------------------------------
Delphinus           ------------------------------------------------------------
Hippopotamus        ------------------------------------------------------------
Elephant            ------------------------------------------------------------
Porcine             GGCTTCTTCCTGGGCTTGAAAGAGCAGGCACATTACCTTCTCTCTGTTACACACCCACTG 2110

Bos                 ------------------------------------------------------------
Bubalus             ------------------------------------------------------------
Delphinus           ------------------------------------------------------------
Hippopotamus        ------------------------------------------------------------
Elephant            ------------------------------------------------------------
Porcine             CACCCACTGCTCAGGTCTGCAGTCCAGCTTGCTGGGCACTCATAGGTCAGGACCACCCCC 2170

Bos                 ------------------------------------------------------------
Bubalus             ------------------------------------------------------------
Delphinus           ------------------------------------------------------------
Hippopotamus        ------------------------------------------------------------
Elephant            ------------------------------------------------------------
Porcine             CATCCTGCTACACCCCCCGCCTCCCATAAAGTACCCAAGAATGGAAAGAGATGAAAGCAA 2230

Bos                 -
Bubalus             -
Delphinus           -
Hippopotamus        -
Elephant            -
Porcine             G 2231
```

ANNEXURE -V

ABBREVIATIONS AND TRIPLET CODONS OF AMINO ACIDS

Amino Acid	Abbreviation		Codons
	3 letter	Single Letter	
Isoleucine	Ile	I	ATT, ATC, ATA
Leucine	Leu	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	Val	V	GTT, GTC, GTA, GTG
Phenylalanine	Phe	F	TTT, TTC
Methionine	Met	M	ATG
Cysteine	Cys	C	TGT, TGC
Alanine	Ala	A	GCT, GCC, GCA, GCG
Glycine	Gly	G	GGT, GGC, GGA, GGG
Proline	Pro	P	CCT, CCC, CCA, CCG
Threonine	Thr	T	ACT, ACC, ACA, ACG
Serine	Ser	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Tyr	Y	TAT, TAC
Tryptophan	Trp	W	TGG
Glutamine	Gln	Q	CAA, CAG
Asparagine	Asn	N	AAT, AAC
Histidine	His	H	CAT, CAC
Glutamic acid	Glu	E	GAA, GAG
Aspartic acid	Asp	D	GAT, GAC
Lysine	Lys	K	AAA, AAG
Arginine	Arg	R	CGT, CGC, CGA, CGG, AGA, AGG
Termination	Ter	X	TAA, TAG, TGA

ANNEXURE –VI

CLUSTALW MULTIPLE PROTEIN SEQUENCE ALIGNMENT OF INDIAN ELEPHANT GH WITH OTHER SPECIES

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	Indian Elephant	216	2	African Elephant	190	99
1	Indian Elephant	216	3	Porcine	216	94
1	Indian Elephant	216	4	Delphinus	216	94
1	Indian Elephant	216	5	Hippopotamus	216	94
1	Indian Elephant	216	6	Bos	217	87
2	African Elephant	190	3	Porcine	216	98
2	African Elephant	190	4	Delphinus	216	98
2	African Elephant	190	5	Hippopotamus	216	98
2	African Elephant	190	6	Bos	217	90
3	Porcine	216	4	Delphinus	216	98
3	Porcine	216	5	Hippopotamus	216	98
3	Porcine	216	6	Bos	217	90
4	Delphinus	216	5	Hippopotamus	216	99
4	Delphinus	216	6	Bos	217	91
5	Hippopotamus	216	6	Bos	217	91

CLUSTAL 2.0.10 multiple sequence alignment

```
Sheep              MMAAGPRTS--LLLAFTLLCLPWTQVVGAFPAMSLSGLFANAVLRAQHLHQLAADTFKEF 58
Buffalo            MMAAGPRTS--LLLAFTLLCLPWTQVVGAFPAMSLSGLFANAVLRAQHLHQLAADTFKEF 58
Bos                MMAAGPRTS--LLLAFALLCLPWTQVVGAFPAMSLSGLFANAVLRAQHLHQLAADTFKEF 58
African Elephant   ----------------------------FPAMPLSSLFANAVLRAQHLHQLAADTYKEF 31
Monodelphis        -MAPGMRVCLLLLIAFTLLG---PQRAAAFPAMPLSSLFANAVLRAQHLHQLVADTYKEF 56
Indian Elephant    -MAAGSRTS--LLLAFTLLCQPWPPEAGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Mouse              -MATDSRTS--WLLTVSLLCLLWPQEASAFPAMPLSSLFSNAVLRAQHLHQLAADTYKEF 57
Rat                -MAADSQTP--WLLTFSLLCLLWPQEAGALPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Rabbit             -MAAGSWTA--GLLAFALLCLPWPQEASAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Camel              -MAAGPRTS--VLLAFTLLCLPWPQEAGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Lama               -MAAGPRTS--MLLAFTLLCLPWPQEAGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Delphinus          -MAAGPRTS--MLLAFALLCLPWTQEVGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Hippopotamus       -MAAGPRTS--VLLAFALLCLPWTQEVGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Porcine            -MAAGPRTS--ALLAFALLCLPWTREVGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Cat                -MAAGPRNS--VLLAFALLCLPWPQEVGTFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Equus              -MAAGPRTS--VLLAFGLLCLPWPQDVGAFPAMPLSSLFANAVLRAQHLHQLAADTYKEF 57
Human              -MATGSRTS--LLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEF 57
                                                 :*::.** ** **:***::****. **::**
```

```
Sheep              ERTYIPEGQRYS-IQNTQVAFCFSETIPAPTGKNEAQQKSDLELLRISLLLIQSWLGPLQ 117
Buffalo            ERTYIPEGQRYS-IQNTQVAFCFSETIPAPTGKNEAQQKSDLELLRISLLLIQSWLGPLQ 117
Bos                ERTYIPEGQRYS-IQNTQVAFCFSETIPAPTGKNEAQQKSDLELLRISLLLIQSWLGPLQ 117
African Elephant   ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 90
Monodelphis        ERTYIPEAQRHS-IQSTQTAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLSPVQ 115
Indian Elephant    ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Mouse              ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKEEAQQRTDMELLRFSLLLIQSWLGPVQ 116
Rat                ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKEEAQQRTDMELLRFSLLLIQSWLGPVQ 116
Rabbit             ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDMELLRFSLLLIQSWLGPVQ 116
Camel              ERTYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Lama               ERTYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Delphinus          ERAYIPEGQRYS-IQNTQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Hippopotamus       ERAYIPEGQRYS-IQNTQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Porcine            ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Cat                ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDVELLRFSLLLIQSWLGPVQ 116
Equus              ERAYIPEGQRYS-IQNAQAAFCFSETIPAPTGKDEAQQRSDMELLRFSLLLIQSWLGPVQ 116
Human              EEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQ 117
                   *.:***: *::* :*..*.::****:**:*:.::*:**::::****:********* *:*

Sheep              FLSRVFTNSLVFGTSDR-VYEKLKDLEEGILALMRELEDVTPRAGQILKQTYDKFDTNMR 176
Buffalo            FLSRVFTNSLVFGTSDR-VYEKLKDLEEGILALMRELEDVTPRAGQILKQTYDKFDTNMR 176
Bos                FLSRVFTNSLVFGTSDR-VYEKLKDLEEGILALMRELEDGTPRAGQILKQTYDKFDTNMR 176
African Elephant   FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRPGQVLKQTYDKFDTNMR 149
Monodelphis        FLSRVFTNSLVFGTSDR-VYEKLRDLEEGIQALMQELEDGSSRGGLVLKTTYDKFDTNLR 174
Indian Elephant    FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELGDGSPRPGQVLKQTYDKFDTNMR 175
Mouse              FLSRIFTNSLMFGTSDR-VYEKLKDLEEGIQALMQELEDGSPRVGQILKQTYDKFDANMR 175
Rat                FLSRIFTNSLMFGTSDR-VYEKLKDLEEGIQALMQELEDGSPRIGQILKQTYDKFDANMR 175
Rabbit             FLSRAFTNTLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRVGQLLKQTYDKFDTNLR 175
Camel              FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRAGQILRQTYDKFDTNLR 175
Lama               FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRAGQILRQTYDKFDTNLR 175
Delphinus          FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNMR 175
Hippopotamus       FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNMR 175
Porcine            FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLR 175
Cat                FLSRVFTNSLVFGTSDR-VYEKLKDLEEGIQALMRELEDGSPRGGQILKQTYDKFDTNLR 175
Equus              LLSRVFTNSLVFGTSDR-VYEKLRDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLR 175
Human              FLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSH 177
                   :*   *:*:*::*:**  **: *:****** :** .* * :.* * ::: **.***:* :

Sheep              SDDALLKNYGLLSCFRKDLHKTETYLRVMKCRRFGEASCAF 217
Buffalo            SDDALLKNYGLLSCFRKDLHKTETYLRVMKCRRFGEASCAF 217
Bos                SDDALLKNYGLLSCFRKDLHKTETYLRVMKCRRFGEASCAF 217
African Elephant   SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 190
Monodelphis        SDEALLKNYGLLSCFKKDLHKAETYLRVMECRRFVESSCAF 215
Indian Elephant    SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Mouse              SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Rat                SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFAESSCAF 216
Rabbit             GDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCVF 216
Camel              SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Lama               SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Delphinus          SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Hippopotamus       SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Porcine            SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Cat                SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Equus              SDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF 216
Human              NDDALLKNYGLLYCFRKDMDKVETFLRIVQCR-SVEGSCGF 217
                   .*:********* **:**:.*.**:**:::**   *.** *
```

CLONING AND SEQUENCE ANALYSIS OF THE GROWTH HORMONE GENE IN INDIAN ELEPHANTS

BHOSALE R.A.

Abstract of the thesis submitted in partial fulfilment of the requirement for the degree of

Master of Veterinary Science

Faculty of Veterinary and Animal Sciences
Kerala Agricultural University, Thrissur

2009

Department of Animal Breeding, Genetics & Biostatistics
COLLEGE OF VETERINARY AND ANIMAL SCIENCES
MANNUTHY, THRISSUR-680651
KERALA, INDIA

ABSTRACT

The study was undertaken with the objectives of cloning and sequence analysis of the growth hormone (GH) gene of the Indian elephant. The GH is a peptide hormone of about 190 amino acids, synthesized and secreted by cells called somatotrophs. Growth hormone is a major participant in the control of several complex physiologic processes, including growth and metabolism.

The genomic DNA was isolated from blood samples and a 1712 bp fragment of the entire transcriptional unit of the GH was amplified by PCR using synthetic oligonucleotide primer pairs designed based on the 5′ and 3′ flanking sequences of goat GH gene. The gel purified PCR product was ligated into the pGEM®-T Easy cloning vector and was transformed by giving heat shock to competent *E. coli* cells prepared by $CaCl_2$ treatment. The recombinant clones among the transformed cells were identified by *Blue–White Screening* and the recombinant plasmid carrying the insert gene was isolated from the white clones by a modified SDS-alkaline lysis method. The 1.7 kb GH gene insert in the vector was sequenced by the dideoxynucleotide sequencing method with primer walking using an automated DNA sequencer.

The nucleotide sequence showed 75 to 96 per cent homology with pig and 77 to 95 per cent with that of dolphin GH genes, respectively. The exon-intron boundaries in the porcine gene occurred at the codons of the amino acid residues, Gly-4 (intron 1), Phe-57 (intron 2), Ser-96 (intron 3) and Arg-150 (intron 4). All these four residues were conserved in both species of elephants as well as in pig. This strict homology in the sites of insertion of introns suggests that the exon-intron organization of these genes was established before the divergence of these species. The positions of the exon-intron boundaries are also conserved as evidenced from similar sizes of the exons. Evidence for some homology was also seen in intron 1, which showed maximum 84 per cent similarity with that of giraffe. In contrast, intron 2, 3 and 4 showed no significant similarity both in length and in sequence with other animal species.

The Indian elephant GH gene has an open reading frame (ORF) of 648 nucleotides encoding a signal peptide of 26 amino acid residues and a mature protein of 190 amino acid residues with both NH_2- and COOH- terminal *Phe*. Alignment of this sequence with African elephant counterpart showed that 189 amino acid residues are identical with only one variant while, with pig sequence it showed 186 identical residues with four variants.

The predicted secondary structure showed that the larger α-helical lobe was formed by four sections of the polypeptide chain (residues 3-34, 47-80, 110-126 and 148-172) while the smaller lobe, which encompassed a small anti-parallel beta-sheet and a small irregular structure formed the remaining structure of the polypeptide chain. The predicted 3D protein structure obtained from the Swiss Model Server and the 3D structure of human GH protein revealed that the α-helices, β-sheets and random coils in the predicted protein are similar to that of the human GH structure.

This work seems to be the first of its kind in establishing the DNA and protein sequence data and structural organization of growth hormone gene in Indian elephants. The analysis of sequence data revealed that there was high similarity of the protein at the amino acid sequence level, the difference being a single residue in African elephant and four variants only in pig. This observation revealed a significant variation in the codon usage between Indian and African elephants and other species of animals. The signal peptide and intron sequences seem to be evolved more rapidly as there was less homology as compared to exonic sequences.

Overall, the structures of Indian elephant GH gene was found to be similar to that of African elephant and porcine as evidenced from the high degree of amino acid sequence identity (99 – 97 per cent) and it may be possible to clone the GH gene from African elephant using primers designed based on the Indian elephant GH gene sequence reported in this study.