Chapter-I

Introduction

1.1 Background of the Study:

The Chelonians are the oldest living reptiles that evolved 200 million years ago, extending valuable service as scavengers in the seas, rivers and other water bodies. Turtle refers to all species included in the order Testudines, covering turtles, tortoises and terrapins. They bear scales, lay eggs and ectothermic in nature. The common names generally signify the habitat of the species and living pattern. Turtles bear webbed feet for swimming and spend most of their life in water. Freshwater turtles live in fresh water like rivers, ponds and lakes. In addition to swimming, they also climb to the banks, logs or rocks to bask in the sunshine. Tortoises are land-dweller and they do not possess webbed feet and their feet are round and stumpy as suitable for walking on land. Terrapins are often found in brackish and swampy areas, where they spend most of the time both in land and water. Turtles play a major role in cleaning the river and tanks and thus significantly controlling pollution in the freshwater system with its wide range of diversity (Moll and Moll, 2004).

Turtles are among the world’s most endangered vertebrates and almost half of their 328 recognized species (452 taxa) facing threatened with extinction (Turtle conservation coalition, 2011). Presently, eight species (10 taxa) of tortoise and freshwater turtles have become extinct since 1500 AD and out of this 320 species (442 taxa) of living turtles and tortoise, 7 species are marine and 313 species (435 taxa) are living freshwater turtles and tortoises (Turtle Taxonomy Working Group, 2010). According to the draft Red List of IUCN (Tortoise and Freshwater Turtle Specialist Group) status, 156 turtle species (47.6 % of total species) are recognized as globally threatened while 90 (27.4%) are Critically Endangered or Endangered species (Turtle Taxonomy Working Group, 2010).

The turtle population in India is declining due to over exploitation either for food or trade and habitat destruction (Rao, 1985; Choudhury et al., 1993). Their biological needs are affected by changes in the natural environment, being altered by
developmental projects, cultivation, removal of sand from their nesting ground and many others. The other reasons for decline of some of the species are due to illegal slaughtering for meat, egg predation, water pollution and human disturbances (Gupta, 2000).

The Indian subcontinent with 33 species have been enriched with the most diverse chelonian fauna of the world with five (5) marine, twenty four (24) freshwater and four (4) land species (Das, 1985, 1995, 2002; Das and Andrews, 1997; Das and Bhupathy, 2009). Out of 28 species of freshwater turtles and tortoise occur in India, 39.3% (11 of 28 taxa) are listed as either Endangered or Critically Endangered in the IUCN Red List (IUCN, 2007). The Indo-Gangetic plain and the Terai region has diverse chelonian fauna with the occurrence of 20 species (Rao, 1990; 2009; Das, 2002). Endowed not only with hardshell, softshell and tortoises group of turtle, but also this region stands for wide range of biological diversity (Shrestha, 2001). Once considered as extinct, *P. sylhetensis* has been rediscovered in new location in the Northeastern states and occupies distinct position in the evolutionary tree being natural scavangers. This group occupies a wide range of habitats and the most threatened clade of turtles (Van Dijk *et al.*, 2000). The freshwater turtles and tortoises of this region are facing extinction due to excessive hunting for meat and rapid loss of habitat. The scientific community declared seven species of turtles of this region as endangered in the Schedules of Indian Wildlife (Protection) Act, 1972 and in the Red Data List (Rao, 1989; IUCN, 2010a).

The turtle family Geoemydidae is the most diverse family of living turtles, encompassing about 70 species and 23 genera, distributed globally in Asia, Europe, North Africa as well as Central and South America (Iverson, 1992; Van Dijk *et al.*, 2000; IUCN, 2011). This group also occupies a wide range of habitat, from highly aquatic (*Batagur* and *Malayemys*) to highly terrestrial (*Geoemyda*). It is also the most threatened clade of turtles which is due to the over exploitation for supply in international wildlife trade in Asia (Van Dijk *et al.*, 2000; IUCN, 2010b).

The turtles under the genus Pangshura are small-sized species, viz. *P.tecta, P. sylhetensis, P. tentoria* and *P. smithii* (Das, 2001). The genus Pangshura, having maximum shell length of 20–26.5 cm (Ernst *et al.,* 2000; Sarma, 2007), comparatively smaller than *Batagur* (maximum shell lengths 48–58 cm; Ernst *et al.,* 2000) is characterized by sexual dimorphism (Ernst *et al.,* 2000; Sarma, 2007), now placed them into two distinct genera (Das, 2001).
The name *Pangshura* refers to small-bodied *Kachuga* (Das, 2001; Schleich and Käste, 2002). However, these species are fragmented to sub species namely *P. tentoria tentoria*, *P. tentoria circumdata*, *P. tentoria flaviventer*, *P. smithii* and *P. smithii pallidipes*. Morphological characters have long been used to unite turtles as monophyletic group and to resolve the phylogenetic position of species (Gaffney et al., 1991). Molecular characterization have been introduced in turtle research, which in turn provides stronger support for the arrangements identified from morphology alone (Shaffer et al., 1997; Fujita et al., 2004; Baruah, 2010).

**Table 1.1**: Species and sub-species of *Pangshura* analyzed in the present study (Fritz and Havas, 2007).

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Taxon</th>
<th>CITES</th>
<th>IUCN</th>
<th>WLPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Roofed Turtle</td>
<td><em>Pangshura tecta</em> (Gray, 1831)</td>
<td>Appendix I</td>
<td>Low Risk, Near Threatened</td>
<td>Schedule I</td>
</tr>
<tr>
<td>Assam Roofed Turtle</td>
<td><em>Pangshura sylhetensis</em> (Jerdon, 1870)</td>
<td>Appendix II</td>
<td>Endangered</td>
<td>Schedule I</td>
</tr>
<tr>
<td>Brown Roofed Turtle</td>
<td><em>Pangshura smithii smithii</em> (Gray, 1863)</td>
<td>Appendix II</td>
<td>Low Risk, Near Threatened</td>
<td>Not Listed</td>
</tr>
<tr>
<td></td>
<td><em>Pangshura smithii pallidipes</em> (Moll, 1987)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indian Tent Turtle</td>
<td><em>Pangshura tentoria tentoria</em> (Gray, 1834)</td>
<td>Appendix II</td>
<td>Low Risk, Least Concern</td>
<td>Not listed</td>
</tr>
<tr>
<td></td>
<td><em>Pangshura tentoria flaviventer</em> (Günther, 1864)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Pangshura tentoria circumdata</em> (Mertens, 1969)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The understanding of genetic variation of *Pangshura* is listed as top priority research in the ‘Conservation Action Plan for Endangered Turtles and Tortoises of India’ (CFH/MCBT, 2006). Therefore, it is quite relevant that molecular understanding at species and sub-species level of the *Pangshura* group in particular and the turtle group in general will help to formulate conservation strategies for many of the endangered turtle group. Thus, the genetic differentiation based on well defined markers has contributed considerably of how species evolved, diverged and revealed unexpected genetic structuring within many species complex (Bickford et al., 2007). A phylogeny based on DNA data (Spinks et al., 2004) showed that *Kachuga* was paraphyletic and so removed *flaviventer*, *smithii*, *sylhetensis*, *tecta*, and *tentoria* into the genus *Pangshura* (Table 1.1). Schleich and Käste (2002) elevated *P. flaviventer* to a full species status.
Molecular characterization is now a standard tool in taxonomic and phylogenetic studies, focusing mostly on genes in the mitochondrial genome or DNA loci in the nuclear genome. In particular, the application of the polymerase chain reaction (PCR) has increased the availability of molecular technologies while decreasing the cost (Engstrom et al., 2007). Recent advances in molecular biology give rise to the development of in silico methods for sequence analysis, protein structure prediction, functional annotation and evolutionary studies, which have been used in this study. In the present investigation, molecular characterization based on mitochondrial and nuclear genes along with protein sequence analysis from the Indian freshwater turtles of the genus Pangshura has been performed. The study represents the first hand information generation that includes field data collection, sample collection and analyses including DNA isolation, PCR, microsatellite based ISSR marker development, DNA sequencing followed by application of in silico tools for sequence comparison, multi-gene phylogeny, RNA secondary structure prediction and tertiary structure prediction of various proteins to predict the molecular genetic variation within the chelonian genus Pangshura as well as to identify their homology and evolutionary profile. The Phylogeny and Phylogeography studies were considered to avoid complexity in morphometric analysis and sequence similarity searching for tracing the orthologs.

Knowledge of genetics is increasingly recognized as a critical element of conservation biology (Moritz, 1994; Soltis and Gitzendanner, 1999). Molecular techniques and methods of statistical analysis derived from evolutionary theory can be used to estimate how genetic diversity is apportioned spatially, how rapidly diversity will be lost over time, to identify crucial forces contributing to present and future loss of diversity, to gain insight into fundamental aspects of an organism’s biology, and to provide information guidance for conservation and management (Moritz, 1999; Reed and Frankham, 2003; DeYoung and Honeycutt, 2005; Whiteley et al., 2006). Despite the clear importance of genetics as a foundation for understanding turtle biology and directing turtle conservation action, there is a paucity of freshwater turtle conservation genetics studies in India.

The present study has been made for the first time to address the question of intra and interspecific differentiation of Pangshura genus using genomic inter-simple-sequence
repeats polymerase chain reaction (ISSR-PCR) along with multi-gene sequence based phylogeny. In the study, a close look has also been given on the geographical lineage of genetic variability of the freshwater turtle group in the Northern and Northeastern parts of India and in the bordering areas of Bangladesh, Bhutan, Nepal and Myanmar.

1.2 Relevance of the Study:
The tributaries of the Ganges and the Brahmaputra rivers hold different species of turtles. The steady decline in population of different species of freshwater turtles in the river systems in India has prompted research and conservation programmes on freshwater turtles in different parts of India (Moll, 1984; Rao, 1990; Choudhury and Bhupathy, 1993). The Asian turtle crisis has brought to light the devastating decline in turtle abundance in Asia. Asian turtle trade numbers are astounding- 15,500 metric tons annually, 10.3 million turtles annually, about 28,000 turtles per day (Van Dijk et al., 2000; Turtle Conservation Fund, 2002). These turtles are used as pets, for religious purposes, as food, and in traditional Chinese medicine. Schlaffer et al. (2005) documented the impact of trade in amphibians and reptiles in and out of the United States and expressed explicit concern for 19 species of turtles. The researchers are using molecular genetic techniques to answer resource management questions involving the maintenance of genetic diversity, delineation of appropriate units for management, mating systems, and the genetic history of individuals.

The present study has been conducted in the northeastern and northern (Brahmaputra-Ganges region) parts of India. The Brahmaputra-Ganges region of India and Bangladesh exhibits the greatest range overlap for turtles in a single area in the world, with 19 species known from each of 4 connected hydrologic unit compartments [(HUCs) (Figure 1.1) (Iverson, 1992)]. The richness of turtles in the Brahmaputra-Ganga Basin is in part because of the overlap in the 2 faunas. The lower Gangetic Plains is Moist Deciduous Forest Ecoregion, representing the third-most species-rich turtle Ecoregion in the world (18 species), accounted for 12 species. The Upper Gangetic Plains Moist Deciduous Forests Ecoregion, although clearly needed for conservation of the Ganges River system and its fauna, does not gain any more turtle species, just more range of the same species. Other Ecoregions that include the same high turtle species richness (but are within the Indo-Burma and Himalaya biodiversity hotspots) include Meghalaya Subtropical Forests, Brahmaputra Valley Evergreen
Forests, Mizoram–Manipur–Kachin Rainforests, and Terai–Duar Savannas and Grasslands. In addition, climate change studies suggest that glacial melt in the Himalaya will affect water flow in the Ganges-Brahmaputra Basin (Xu et al., 2009), thus adding another conservation concern to this priority area. In the global analysis of Tortoise and freshwater turtles (Buhlmann et al., 2009), distribution range of 4 Pangshura species have been addressed (Table 1.2).

![Figure 1.1](image.jpg)

**Figure 1.1:** The study area belongs to world’s greatest turtle richness area, based on the co-occurrence of species in hydrologic unit compartment in the Ganges-Brahmaputra river basin drainages of India and Bangladesh in South Asia (Buhlmann et al., 2009).

About 42% of freshwater turtles and tortoises are considered to be facing a high risk of extinction, and are in need of urgent conservation action (Turtle Conservation Coalition, 2011). This includes a need to (i) assess our current state of knowledge regarding the application of genetics for studies of freshwater turtles and tortoises and (ii) determine future research directions. Since the advent of techniques to view condensed chromosomes in the early 1900s, researchers began studying turtle genomes (Oguma, 1936, 1937; Risley, 1936) to quantify genetic variation among and within species and to reconstruct the relationships among taxa. In the decades that
followed, an array of more powerful and finely discriminating techniques and genetic markers were developed that are now used to investigate (i) genetic relationships among taxa (phylogeny), (ii) genetic diversity and structure and patterns of gene flow among populations, (iii) mating systems and the extent of multiple paternity, (iv) histories of captive lineages, (v) origins of forensic specimens, and (vi) aspects of molecular evolution within taxa (Bull et al., 1974; Carr and Bickham, 1981; Bickham et al., 1985; Avise et al., 1992).

Table 1.2: Pangshura species addressed in the analysis of Buhlmann et al. (2009) and their primary region of occurrence, range in sq. km, and their percentage of occurrence in existing Global Conservation Strategies (GCS): Biodiversity Hotspots (BH) and Other Wilderness Area (OWA).

<table>
<thead>
<tr>
<th>Species</th>
<th>Range (Sq km)</th>
<th>Geogr Reg.</th>
<th>% BH Identification</th>
<th>% BH</th>
<th>% OWA</th>
<th>First-priority Ecoregion</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. smithii</em></td>
<td>1151781</td>
<td>AS</td>
<td>29.5</td>
<td>HIMA,INBU</td>
<td>1</td>
<td>Northwestern thorn scrub Forest</td>
</tr>
<tr>
<td><em>P. sylhetensis</em></td>
<td>288548</td>
<td>AS</td>
<td>72.4</td>
<td>HIMA,INBU</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td><em>P. tecta</em></td>
<td>1775971</td>
<td>AS</td>
<td>12.1</td>
<td>HIMA</td>
<td>2</td>
<td>Upper Gangetic Plains moist deciduous forests</td>
</tr>
<tr>
<td><em>P. tentoria</em></td>
<td>1288375</td>
<td>AS</td>
<td>22.7</td>
<td>HIMA,INBU</td>
<td>1</td>
<td>Upper Gangetic Plains moist deciduous forests</td>
</tr>
</tbody>
</table>

AS- Asia; HIMA- Himalaya; INBU- India-Bangladesh

Despite of a good member of molecular phylogenetic interaction (Honda et al., 2002a; Barth et al., 2004; Spinks et al., 2004; Sasaki et al., 2006; Praschag et al., 2007a) the genus Pangshura has poorly been addressed. Praschag et al. (2007) could not establish any distinctness of the subspecies of *P. smithii* and *P. tentoria* with the help of mtDNA (cyt b gene) data. Therefore, a detailed study on molecular characterization of the genus Pangshura with interdisciplinary approaches has been designed for proper indentification of species and sub-species of Pangshura. Understanding differentiation of these turtles at molecular level would be beneficial for taxonomic understanding as well as formulating effective conservation strategy.
1.3 Patterns of distribution and diversification:

In the present investigation, different molecular tools (DNA, RNA and protein sequences) along with morphology and morphometry of the Pangshura, are used to trace the relationship among various populations belonging to different region associated with past geological and ecological events. The past geological and ecological events are not only responsible for the distribution and diversification of species; but also equally responsible for intraspecific morphological variation in each region. The morphological variation is a well known natural phenomenon. Due to the past historical events, certain species are isolated in different region and adapted accordingly, showing the intraspecific morphological variation. Therefore, the present study would be helpful in correlating the genetic and morphology together to understand the presence of different ecomorphs of Pangshura in northeastern and northern region of India.

Multi-gene phylogeny and phylogeography have been addressed in the present investigation to determine appropriate conservation units as well as identifying priority areas for effective conservation management of freshwater turtles within the Indian subcontinent. The chelonian conservation effort will be more successful, if accompanied by genetic investigations. Data generated by conservation genetics study will provide improve management plan. Apart from these, the present study will also be helpful for in-situ conservation of endangered chelonians.

Recent advances in molecular biology give rise to the development of in silico methods for sequence analysis, protein structure and function prediction and evolutionary studies, which have been utilized in the present investigation. The Bioinformatics based methods are additional methods for identification and characterization of samples protected under Convention International Trade in Endangered Species (CITES) and will allow to improve the work for the conservation of the endangered species.

1.4 Aims and Objectives:

The present study seeks to determine appropriate genetic diversity within the freshwater turtle genus Pangshura and to identify priority areas for effective
conservation management of freshwater turtles in northeast India and parts of northern India. Based on complete taxon sampling, the study quantifies the genetic variation within and among the population of Pangshura genus in two Indian River systems (Ganga and Brahmaputra) along with the comparison. The present study will also add a genetic dimension to the ongoing efforts to understand the ecology, demography and natural history of the endangered and endemic Assam Roofed turtle Pangshura sylhetensis (Jerdon, 1870).

The present study took up the following aims and objectives:

I. Identification of Pangshura species by comparative sequence analysis of selected genes.

II. Analysis of genetic diversity among the Pangshura species using microsatellite based molecular marker.

III. Study of multi-gene phylogeny of the genus Pangshura with reference to phylogeography of P. sylhetensis:

(a) To see the concordance of morphological data with genetic data along with spatial pattern of distribution.

(b) To test whether the currently recognized sub-species of P. tentoria correspond well with the multi-gene sequence data.

(c) To investigate the genetic relationship between the taxa belong to Ganges and the Brahmaputra river system.

(d) To address the issue of morphological and genetic determinants of the Pangshura species.

IV. In silico molecular characterization of genus Pangshura through evolutionary analysis and homology modelling of selected proteins.

The null hypothesis for this study is that genetic and morphological variations are not correlated to geographical variation.
1.5 Review of literature:

Study of reptilian fauna has been started long back both nationally and internationally in different countries of the world. In the past some notable contributions have been made towards the understanding of taxonomy, ecology and biogeography of the turtle in world. Some of the earlier studies are Gray (1863), Jerdon (1870), Bramble (1971), Bour (1975), Bhaskar (1978), Auffenberg and Iverson (1979), Grumly (1984), Honegger (1980), Ahmed et al. (1986), Das (1986), Dimention et al. (1987), Bartlett (1994), Das (1997), Das and Bhupathy (2001), Gupta and Guha (2002). Few studies conducted on Indian freshwater turtles have mainly dealt with taxonomy and their broad distribution ranges (Smith, 1933; Pritchard, 1979; Daniel, 1983; Das, 1985; Tikader and Sharma, 1985).

The beginning of the studies of Testudines in India date back as the middle of the nineteenth century and the first consolidated work was published by Gunther (1846) in his book "The Reptiles of British India". Some of the outstanding contribution to the study of Indian testudines were (i) Descriptive catalogue of the Reptiles of British India (Theobald, 1876), (ii) The fauna of British India Reptilia and Batrachia (Boulenger, 1889), (iii) Notes on various Indian Testudines (Anderson, 1872), (iv) A new tortoise from Travancore (Ferguson, 1903), (v) Numerous papers on Indian testudines (Annandale, 1914), (vi) Cataologue of the Shielded Reptiles and numerous papers on Indian Testudines (Gray, 1872), (vii) Many papers of living and fossil Testudines of India (Lydekker, 1885, 1889) and (viii) The fauna of British India Loricata, Testudines- The latest and most consolidated account of all the Indian species and subspecies (Smith, 1931).

Genetic data have been used in turtle conservation to evaluate the genetic variability within and among population (Janzen et al., 1997; Souza et al., 2002; Schwartz and Karl, 2005), to recognize the existence of cryptic taxa (Russello et al., 2006) and to reveal migratory patterns (Bowen and Avise, 1996). Similarly microsatellites have recently been used to evaluate the genetic consequences of recent population bottlenecks (Kuo and Janzen, 2004; Waldick et al., 2002), estimate population size and migration rates (Nichols and Freeman, 2004), assess natal dispersal (Berry et al., 2004), detect hybridization (Burns et al., 2003) and to provide identification in wildlife forensics (Avise, 2004).
Taxonomy and phylogenetic position of the turtle family Geoemydidae is still entangled despite numerous studies derived from morphological, chromosomal, and molecular analysis (McDowell, 1964; Bramble, 1971; Hirayama, 1984; Gaffney and Meylan, 1988; Yasukawa et al., 2001; Honda et al., 2002b; Spinks et al., 2004) and the clade is currently considered to be the sister-group to land tortoises (Joyce, 2007). Specially, within the Geoemydidae, relationships among Pyxidea, Cuora, Pangshura (=Kachuga), Batagur and Cyclemys are always controversial. Further studies on these species are in high priority. Based on the patchy taxon sampling, Praschag et al. (2007a) established a hypothesis for Batagur, Callagur, Hardella, Kachuga and Pangshura. However, very little is known about the freshwater ehelonians of the Northeastern region of India. This is evident that even new species of turtles were described from the region recently (Fritz et al., 2008, Praschag et al., 2009) and also species added to Indian ehelonians not known earlier (Pawar and Choudhury, 2000). Though, little efforts have been made to investigate the ehelonians of the region, yet some of the notable works have already been carried out in turtle family from other parts of the world (Das, 1991; Choudhury, 1995; Pawar and Choudhury, 2000; Sengupta et al., 2000; Praschag and Gemel, 2002; Fritz et al., 2008; Das et al., 2010; Praschag et al., 2011).

DNA-based markers have gained popularity in recent years in assessment of genetic relationship among species. Of these markers, Inter Simple Sequence Repeat (ISSR) markers (Gupta et al., 1994; Zietkiewicz et al., 1994; Wu et al., 1999; Marmi et al., 2006; Fritz et al., 2008; Guicking et al., 2009; Bu et al., 2011) are often used in phenetic studies. Being polymorphic (Bomet and Branchard, 2001) and ubiquitous in the genome (Tautz and Renz, 1984), ISSR markers have the advantages of SSR (Simple Sequence Repeat) markers, while bypassing the major obstacle to the development of SSR marker, as SSR requires knowledge on flanking sequences. Fritz et al. (2005) used ISSR-PCR genomic data to generate nuclear fingerprints and examined the evolutionary relationship of 5 species of the genus Testudo. Hence, ISSR markers are suitable for use in species distinction, where extensive information on DNA sequences is not yet available (Ratnaparkhe et al., 1998; Meloni et al., 2006).
RNA structures are essential in many biological processes and are often conserved in evolution. Such conserved structures are found in tRNA (Sprinzl et al., 1998), rRNA (Wuyts et al., 2001; 2002), tmRNA (Zwieb et al., 2003), RNase P RNA (Brown, 1999) and SRP RNA (Rosenblad et al., 2003). Many computational methods have been developed for predicting RNA structures (Bachellerie et al., 2002; Hofacker et al., 2002; Perriquet et al., 2003; Hofacker et al., 2004; Knight et al., 2004). Recent advances in bioinformatics and *in silico* methods bring some sorts of efficient procedure to predict tertiary structure of proteins from their sequence. The most successful protein tertiary structure prediction method is the homology modelling (*CASP2, http://prediction.llnl.gov*) which focuses on the use of structural templates derived from known structures to build an all-atom model of a target protein (Deane and Blundell, 2000).

Typically the early work with a new genetic approach was mostly theoretical, with the assessment of the utility of each technique and tool followed. For example, a debate over the utility of molecular markers versus morphological data in determining phylogeny was raised regarding turtles (Seidel and Lucchino, 1981; Shaffer et al., 1997; Iverson, 1998; McLuckie et al., 1999), which instigated analytical approaches for dealing with diverse data sets (Pupko et al., 2002; Wortley and Scotland, 2006). In contrast, the application of molecular markers in studies of mating systems was rapidly adopted since early genetic studies demonstrated that actual paternity not necessarily correspond to expectations of observed matings (Galbraith, 1993; Galbraith et al., 1993; Valenzuela, 2000; Pearse and Avise, 2001; Pearse et al., 2002; Roques et al., 2004; Johnston et al., 2006). In recent years, new techniques that allow faster through output, which have expanded the range of species that be studied with genetic techniques as well as the types of questions that should be addressed using this technology (Zhang and Hewitt, 2003; Parham et al., 2006a; McGaugh et al., 2007). Declining costs, greater global communication, and web-based availability of DNA sequences and primers (e.g., GenBank) have facilitated collaborations among researchers and increased sample size, have convincingly made the genetic studies more amenable to rigorous statistical analysis (Holmes, 2003). In addition to these developments, increasingly sophisticated software packages to process and analyze data are becoming available to an ever-expanding user community (Felsenstein, 2004; Pearse and Crandall, 2004). These conceptual and technological advances have
increased the role of genetics in studies of wildlife including turtles and contend that they can play a larger role in advancing turtle conservation (Alacs et al., 2007; McGaugh et al., 2007).

Because of multiple copies of the mitochondrion exist in each cell, mtDNA analysis would be useful in identifying the taxonomic or geographic origin of unidentifiable or poor quality samples (Roman and Bowen, 2000; Hsieh et al., 2006). Mitochondrial DNA can be used in conjunction with other datasets, including either morphological or nuclear molecular markers, to identify hybrid individuals (Hsieh et al., 2006). Molecular techniques can also be used for assessing origins of individuals (Gaur et al., 2006). Contemporary efforts are being made in the same research area for detailed molecular understanding about polymorphism and phylogenetic studies on other Indian freshwater turtle genera viz. Nilssonia (Prachag et al., 2007b), Cyclemys (Fritz et al., 2008; Prachag et al., 2009), Lissemys (Prachag et al., 2011).

1.5.1 Selection of Molecular markers:

A. Mitochondrial DNA:

The vertebrate mitochondrial DNAs (mtDNAs) characterized to date are double-stranded circular molecules of approximately 15–17 kb in length that contains 37 genes (Boore, 1999) and codes 22 tRNAs, 2 rRNAs and 13 proteins plus one noncoding region that contains signals for the mitochondrial replication and transcription (Figure 1.2) (Wolstenholme, 1992). Because of maternal transmission and lack of recombination, mitochondrial genome is inherited as a single locus (Avise, 2004). These features along with a relatively high mutation rate, make sequences from the mtDNA locus ideal for many kinds of evolutionary studies. Bowen et al. (1989) and Lamb et al. (1989) were the first workers to apply mtDNA data to chelonian questions, using variation in mtDNA to assess phylogeographic structure in Chelonia mydas and Gopherus agassizii, respectively. The first complete mitochondrial genome sequenced from a turtle (Pelomedusa subrufa) was used to assess the phylogenetic position of turtles relative to other amniotes (Zardoya and Meyer, 1998), while the first study to use mitochondrial genome data exclusively in turtles examined the phylogenetic relationships of a small group of Old World tortoises including Testudo, Indotestudo, and Malacochersus (Parham et al., 2006b).
Such large datasets are often necessary to resolve uncertain or incorrect relationships recovered from smaller DNA fragment data (Cummings et al., 1995; Zardoya and Meyer, 1996). Because of its compact size, multiple copy status in a cell, maternal inheritance and lack of recombination, mitochondrial genome has been widely used as a marker for evolutionary and population genetic studies (Dutton et al., 1996; Curole and Kocher, 1999; Wu et al., 1998, 1999; Parham et al., 2001; Stuart and Parham, 2004; Spinks et al., 2009a).

**Figure 1.2:** Complete mitochondrial genome organization of turtles. The tRNAs are identified by the single-letter amino acid code.

Turtle mtDNAs have been proposed to evolve at slower rates than those of other vertebrates (Avise et al., 1992; Serb et al., 2001; Feldman and Parham, 2002). Mitochondrial DNA (mtDNA) sequence data have been informative in both phylogeography and in systematics (Hillis et al., 1996). The most frequently used genes in deep phylogenetic studies are the slowly evolving 12s rRNA (Shaffer et al., 1997) and moderately evolving cytochrome b (cyt b) (Shaffer et al., 1997; Spinks et al., 2004). Cyt b, ND4 and other protein coding genes have been most useful for studies among closely related species (Caccone et al., 1999a,b; Engstrom et al., 2002; Feldman and Parham, 2002) and for phylogeographic studies within species (Starkey
et al., 2003; Spinks and Shaffer, 2005). The control region is widely used in population and intraspecific level studies because of its high rate of mutation (Starkey et al., 2003; Pearse et al., 2006a); yet, some studies have noted equal or greater levels of variation in protein coding genes (Spinks and Shaffer, 2005).

In contrast to the mitochondrial genome, the nuclear genome contains a huge number of coding and non-coding regions (introns and intergenic spacers) that are subject to different mutation mechanisms and rates (Li, 1997). Thus the nuclear genome offers a virtually unlimited sets of potential markers that are informative across the entire range of phylogenetic divergence and could be applied to a wider array of questions relative to mtDNA data, including studies of adaptive radiation, life histories, hybridization, species delimitation, and phylogenetic inference including estimates of divergence times (Zhang and Hewitt, 2003; Avise, 2004; Near et al., 2005). Karl et al. (1992) for the first time used nuclear markers (restriction digests of anonymous loci) to estimate global population structure of the marine turtle Chelonia mydas. The inheritance of the mitochondrial genome as a single linkage group, the frequent demonstration of hybridization and natural selection, and lineage sorting often require additional nuclear data to test mtDNA-based phylogenetic hypothesis. In addition, a strong base compositional bias and rapid rate of evolution may hinder mtDNA-based phylogenetic analysis as a result of homoplasy (Naylor and Brown, 1998; Garcia-Machado et al., 1999; Wiens and Hollingsworth, 2000). Thus, while the mtDNA has provided and will continue to provide invaluable utility to phylogenetics, it is important to identify independent markers that can complement those of the mitochondrial genome.

The mitochondrial DNA fragments (in total 2286 bp) have been well established to reveal differences and phylogenetic relationships among chelonian terminal taxa (Engstrom et al., 2002; Spinks et al., 2004; Praschag et al., 2007a,b; Vargas-Ramírez et al., 2008, 2010; Fritz et al., 2010, 2011). The nuclear R35 intron 1 has been used for inter and intraspecific analysis of turtles including Geoemydids (Engstrom et al., 2004; Spinks et al., 2004; Spinks and Shaffer, 2005; Near et al., 2005, Praschag et al., 2009).

Vargas-Ramirez et al. (2008) used mitochondrial and nuclear DNA sequences in order to determine the phylogenetic relationships between the three genera
(Erymnochelys, Peltocephalus and Podocnemis) and the eight species of Podocnemididae. They have used cyt b gene and part of the tRNA serine gene (tRNA-ser), D-loop, cytochrome oxidase subunit I (COI) and 12S rRNA gene. With respect to nuclear genes, they produced partial sequences of the recombination-activating gene 2 (RAG2), the intron 1 of the RNA fingerprint protein 35 (R35) and the neurotrophin-3 (NT3) gene. More recently, Vargas-Ramirez et al. (2010) investigated the phylogeographic differentiation of the widely distributed African helmeted terrapin Pelomedusa subrufa based on 1503 base pairs of mitochondrial DNA (partial cyt b and ND4 genes with adjacent tRNAs) and 1937 bp of nuclear DNA (partial RAG1, RAG2, R35 genes).

In the present study, two mitochondrial (16S rRNA and ND4) and two nuclear genes (RAG2 and R35 Intron-1) were included (Bickham et al., 1996; Dutton et al., 1996; Starkey, 1997; Georges et al., 1998; Noonan 2000; Feldman and Parham, 2001; Engstrom and McCord, 2002; Honda et al., 2002b; Cunningham, 2002; Engstrom et al., 2004; Krenz et al., 2005; Le et al., 2006; Praschag et al., 2011). The nuclear RAG2 gene and the R35 nuclear intron were successful in resolving deeper nodes of chelonian phylogeny (Engstrom et al., 2004; Parham et al., 2004; Spinks et al., 2004; Near et al., 2005; Noonan and Chippindale, 2006; Fritz and Bininda-Emonds, 2007; Xia et al., 2011). The combined approach of using nuclear and mitochondrial genes has been verified to be useful in addressing the relationships of turtles and tortoises (Engstrom et al., 2004; Le et al., 2006).

Combining nuclear gene regions can resolve the Testudines phylogenetic history, which has long terminal branches that may result in ambiguous placement of some taxa (Bergsten, 2005). In fact, multiple nuclear genes have been informative about the placement of turtles within Amniota (Hedges and Poling, 1999; Iwabe et al., 2004), single loci have been useful for resolving relationships within Testudines (Fujita et al., 2004) and combination of nuclear and mtDNA indicated the separation between Platysternidae and Chelydridae (Krenz et al., 2005).

B. Nuclear Loci:

Because of the recognized limitations of mtDNA, increased attention is being paid to the nuclear genome as an additional and independent source of data for phylogenetic,
phylogeographic and population genetic analysis (Bruford and Wayne, 1993; Groth and Barrowclough, 1999; Hare, 2001). The three sources of nuclear data, most commonly used, include size polymorphisms at microsatellite loci, and sequence data from nuclear protein-coding genes and introns. In contrast to mtDNA, nuclear protein-coding genes and introns tend to evolve more slowly (Prychitko and Moore, 1997, 2000; Groth and Barrowclough, 1999; Birks and Edwards, 2002; Caccone et al., 2004; Engstrom et al., 2004), making them less prone to excessive homoplasy-a common problem among mitochondrial genes over deeper divergences. Nuclear introns have the further advantage of being free from many of the evolutionary constraints imposed on protein-coding sequences, resulting in little base compositional bias, relatively low transition/transversion ratio, and little among-site rate heterogeneity (Armstrong et al., 2001; Prychitko and Moore, 2003; Fujita et al., 2004). One disadvantage of nuclear DNA is that the slow rate of evolution, which minimizes homoplasy on long timescales, can also reduce variation on shorter timescales (Birks and Edwards, 2002). This characteristic can limit its utility in phylogeographic and population genetic studies of turtles (Spinks and Shaffer, 2005).

Intron sequence has shown great utility in interspecific phylogenetics (Engstrom et al., 2004), but due to lack of functional constraint they can be difficult to align across deep phylogenies (Fujita et al., 2004; Loytynoja and Goldman, 2005). Protein-coding genes have proven usefulness in interspecific phylogenies at many levels (Georges et al., 1998) and will be crucial in testing the location of the root of the turtle tree (Krenz et al., 2005; Near et al., 2005) along with understanding the placement of turtles relative to other amniotes (Hedges and Poling, 1999). The nuclear introns and protein coding genes are biparentally inherited, detection of heterozygotes is a useful tool in the identification of interspecific hybrids (Stuart and Parham, 2004; Spinks et al., 2009a; Xia et al., 2011). Another less-explored source of nuclear gene data is the rapidly growing field of developmental genetics. Many genes have been cloned from cDNAs for studies of sex determination (Valenzuela et al., 2006), morphological development and gene expression (Chien et al., 2005) and chromosome evolution (Kuraku et al., 2005, 2006; Matsuda et al., 2005). Many of these genes have proven application in large scale phylogenetic analysis (Chien et al., 2006; Kuraku et al., 2005, 2006; Matsuda et al., 2005).
The R35 gene-RNA fingerprint protein 35:

The RNA fingerprint protein 35 (R35) belongs to a very large and extremely diverse superfamily of 7-transmembrane proteins (Friedel et al., 2001). R35 is a polypeptide consisting of 723 amino acids (Fujita et al., 2004). Sequence comparisons indicate that R35 may act as a G-protein-coupled receptor (GPCR) that is developmentally regulated and highly expressed in the cerebellum, spinal cord and ganglion cells of chick embryos (Friedel et al., 2001). Friedel et al. (2001) reported an intron of approximately 1000–2000 bp resided in the transmembrane domain 6 in several birds (Meleagris gallopavo, Anser anser, Struthio camelus, and Sitta europea), a tortoise (Geochelone denticulata) and snake (Elaphe guttata) (Fujita et al., 2004).

The nuclear genome provides an incredible diversity of markers available to evolutionary biologists. There are coding and non-coding sequences that evolve at different rates, allowing investigators to make a broad range of inferences from intrapopulation dynamics to the tree of life based on markers from independent linkage groups (Barns et al., 1996; Sunnucks, 2000). Despite these advantages of employing nuclear markers in phylogenetic studies, their use remains sparse when compared to mtDNA. Nuclear introns provide an alternate to use coding sequence as genetic markers. As they are relatively free from many of the functional constraints of coding regions, introns tend to have an elevated rate of evolution when compared to coding sequence (Graur and Li, 2000). In addition, introns tend to have uniform and even base composition and unbiased substitution patterns, resulting in minimal among-site rate heterogeneity and a relatively low transition/transversion ratio (Armstrong et al., 2001; Birks and Edwards, 2002; Engstrom et al., 2004). The recent popularity of introns in phylogenetic studies has demonstrated their utility in a variety of phylogenetic analysis of plants and animals (Prychitko and Moore, 1997; Johnson and Clayton, 2000; Armstrong et al., 2001; Baker et al., 2001; DeBry and Seshadri, 2001; Birks and Edwards, 2002; Howarth and Baum, 2002; Weibel and Moore, 2002).

C. Microsatellite:

Short Tandem Repeat (STR) Loci Microsatellites are the popular genetic markers for determining population structure and revealing differentiation among population and individuals (Bruford and Wayne, 1993). Microsatellites or STRs are non-coding repetitive DNA sequences composed of a variable number of tandemly repeating
motifs. On average, STRs have mutation rates between $10^{-2}$ and $10^{-5}$ per gamete per generation (Page and Holmes, 1998) and thus can provide the resolution to differentiate individuals and populations, even within small geographic areas. Microsatellites are bi-parentally inherited (unless associated with a sex chromosome) and co-dominant, thereby allowing both alleles at a locus to be identified in heterozygotes. Microsatellites are generally considered selectively neutral (McGaugh et al., 2007) and their simple Mendelian transmission makes them useful for assessing genetic diversity. In freshwater turtles and tortoises microsatellites have been used in population genetics (Ciofi et al., 2002; Kuo and Janzen, 2004), conservation genetics (Sites et al., 1999; Cunningham et al., 2002; Pearse et al., 2006b; Bu et al., 2011) and in paternity and mating systems (Valenzuela, 2000; Pearse et al., 2004; Roques et al., 2006). In addition, STRs are well-suited to address future concerns in turtle biology such as interspecies hybridization (Roy et al., 1994, 1996; Williams et al., 2005; Xia et al., 2011) and forensic detection of wildlife poaching (Manel et al., 2002). The process of finding microsatellite markers can unfortunately be very time-consuming and expensive.

Molecular markers are particularly promising for conservation genetics research, since they are less affected by selection pressures than morphology, thus providing an independent assessment of relationships unaffected by convergent selection for similar natural history traits (Avise, 1994, 2000).

The concept of using nuclear introns as phylogenetic markers was first introduced approximately 20 years ago (Lessa, 1992; Slade et al., 1993, 1994). Since then, intron data partitions have been used as stand-alone independent markers and as nuclear corroborative support for mtDNA partitions in numerous studies (Prychitko and Moore, 1997; Oakley and Phillips, 1999; Pitra et al., 2000; Johnson and Clayton, 2000; Rockman et al., 2001; Rowe and Honeycutt, 2002; Birks and Edwards, 2002; Creer et al., 2003). Although introns have been shown to be involved in alternative splicing (Leicht et al., 1995) and gene regulation mechanisms (Gasch et al., 1989; Alder et al., 1992; Kirby et al., 1995; Prychitko and Moore, 2003), empirical approaches have shown that they can be considered as neutral markers that possess a number of traits that are desirable for molecular phylogenetics (Friesen et al., 1997). The overt lack of functional constraints and the fact that most bases have the potential
to yield phylogenetically informative sites mean that introns generally exhibit lower levels of homoplasy than mtDNA (Slade et al., 1994; Prychitko and Moore, 2000, 2003).

1.5.2 Genetic Studies of Freshwater Turtles and Tortoises:

Powerful molecular techniques have been developed over many decades for resolving genetic relationships, population genetic structure, patterns of gene flow, mating systems and the amount of genetic diversity. Genetic studies of turtles were among the earliest and the rapid application of new genetic tools and analytical techniques is still apparent in the literature on turtles.

Cytochrome $b$ (cyt $b$) was first used in turtle studies in 1994 to address the systematics of *Graptemys* (Lamb et al., 1994) and became the predominant mtDNA marker. Second to cyt $b$ was the control region, initially used in a phylogeographic study (Walker et al., 1995). Sequencing of the complete mtDNA was first used to study molecular evolution and turtle affinities in 1998 (Zardoya and Meyer, 1998a) and later sequencing of the complete mtDNA control region was used to construct a phylogeny of *Kinosternum flavescens* (Serb et al., 2001). Compared to the mtDNA markers used, a broader diversity of nuclear markers was found in the studies, with a major shift in the type of markers used over the years of the study. Allozyme/protein variants and karyotypes predominated in the earlier studies.

Phylogenetic studies rely most strongly on cyt $b$, 12S rRNA, ND4, tRNAs, 16S rRNA, and a variety of allozyme/protein variants. Data relating to molecular evolution have mostly focus on nuclear DNA, particularly at the level of chromosomes, using detailed techniques to determine karyotypes and investigate chromosome damage including the use of flow cytometry. Earlier studies on Population Genetic studies relied heavily on nuclear markers, initially using allozyme/protein variants (33.3%), which have been largely supplanted by microsatellites (38.5%). Phylogeography studies preferred sequencing the control region (48.3%) and cyt $b$ (34.5%), but began with the use of mtDNA-restriction fragment length polymorphisms (RFLPs) in 1989 (Lamb et al., 1989). Increasingly, phylogenetic research on turtles are providing analysis of both mtDNA and nuclear
data sets, particularly as awareness increases about the limitations of mtDNA data (Bazin et al., 2006; Mulligan et al., 2006).

Although mitochondrial DNA (mtDNA) trees are predicted to be good estimators of species trees, gene tree topologies may not reflect species tree due to the risk of historical lineage sorting and/or horizontal gene transfer (Pamilo and Nei 1988; Avise 1989; Moore, 1995). Therefore, nuclear DNA gene trees can be used to support mtDNA gene trees in the search of organismal, as opposed to gene, phylogenies (Moore, 1995, 1997). Nuclear introns have the potential to fill this niche in the dynamic field of molecular markers. The mtDNA normally evolves rapidly at the sequence level (Vawter and Brown, 1986), due in part to an elevated mutation rate resulting from the mitochondrial inefficiency in repairing replication errors and DNA damage (Brown et al., 1979; Clayton, 1984). Therefore, noncoding nuclear introns would be expected to have a greatly reduced rate of molecular evolution. However, in protein coding mtDNA sequences, there is a bias for transitions over transversions (Kimura, 1980; Moritz et al., 1987) and only the third and sometimes first position synonymous nucleotide sites are free from functional constraints. Thus two thirds of all bases in protein-coding sequences are less likely to yield phylogenetic signal than the remaining third base, which is inherently prone to homoplasy (Allard et al., 1999; Bjorkland, 1999). Conversely, introns are apparently free from functional constraints and all bases have the potential to yield phylogenetically informative sites with lower levels of homoplasy and lower transition/tranversion ratios (Slade et al., 1993; Palumbi and Baker, 1994; Slade et al., 1994; Prychitko and Moore, 1997, 2000).

1.6 Future Conservation Efforts:

The northeastern region of India is at the conjunction of the Himalaya and Indo-Burma biodiversity hotspots, enriched with a diversity of 23 turtle species has earned the destination of being included in the priority of turtle conservation area (Choudhury, 1995a; 1995b; 1998; Das, 2001; Buhlmann et al., 2009; Ahmed et al., 2009). Despite the fact that freshwater turtles and tortoises have been the focus of recent and intense conservation attention (Van Dijk et al., 2000), there are still a little work carried out for most of the Critically Endangered or data deficient turtle taxa. The necessary role that conservation genetic studies have in turtle conservation is apparent and suggests areas demanding future research. Since 1950, several studies
used genetic techniques to confirm the presence of natural hybrids in wild population as well as in captive population (Zweig and Crenshaw, 1957; Seidel and Atkins, 1987; Georges et al., 2002; Spinks et al., 2004; Spinks and Shaffer, 2007; Xia et al., 2011).

*P. sylhetensis* is one of the endangered species in the IUCN Red list (IUCN, 2010), while the *P. tecta* is listed in CITES Appendix-I and the other three species namely *P. tentoria* and *P. smithii* are in the CITES Appendix-II. Two of the species namely *P. sylhetensis* and *P. tecta* are listed in the Schedule I of Indian Wildlife (Protection) Act, 1972. As such the species appears to be rare within its limited range (Das, 1991; Rashid and Khan, 2000) although considered as abundant in Kaziranga National Park (Das et al., 2010) as well as some other localities within Assam (Sarma, 2007); Das et al., (2010) emphasized that intensive survey and genetics studies are needed to document population genetics, status and threats of *P. sylhetensis*.

Genetic study can help with issues as diverse as maximizing breeding strategies in captive assurance colonies, to identifying cryptic diversity, to clarify the phylogenetic prioritization of key taxa (Shaffer et al., 2007, 2008). It is expected that the findings of the present study will be a contribution to the world turtle genetics research and the Turtle of the World (TOTW) Project i.e. an ongoing effort to sequence every species of turtle on earth for ~20 independent genes and to build a phylogeny for all 320 or so species of living turtles and tortoises (Barley et al., 2010; Thompson and Shaffer, 2010 a, b). Considering all the above points into account, an attempt has been made in the present investigation to identify the *Pangshura* species through modern molecular tools and techniques.