Molecular evolution study in China: progress and future promise
Ya-ping Zhang1,2,* and Song Ge3

1 Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 JiaoChangDongLu, Kunming, Yunnan 650223, People’s Republic of China
2 Laboratory for Conservation and Utilization of Bioresources, Yunnan University, Kunming, Yunnan 650091, People’s Republic of China
3 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, People’s Republic of China

China has a large land area with highly diverse topography, climate and vegetation, and animal resources and is ranked eighth in the world and first in the Northern Hemisphere on richness of biodiversity. Even though little work on molecular evolution had been reported a decade ago, studies on both the evolution of macromolecules and the molecular phylogeny have become active in China in recent years. This review highlights some of the interesting and important developments in molecular evolution study in China. Chinese scientists have made significant contribution on the methods inferring phylogeny and biogeography of animals and plants in East Asia using molecular data. Studies on population and conservation genetics of animals and plants, such as Golden monkey and Chinese sturgeon, provided useful information for conserving the endangered species. East and South Asia has been demonstrated to be one of the centres of domestication. Origin and evolution of genes and gene families have been explored, which shed new insight on the genetic mechanism of adaptation. In the genomic era, Chinese researchers also made a transition from single-gene to a genomic investigation approach. Considering the fact that amazing progress has been made in the past few years, and more and more talented young scientists are entering field, the future of molecular evolution study in China holds much promise.

Keywords: molecular evolution; China; phylogeny; biogeography; conservation genetics; gene

1. INTRODUCTION
China has a large land area with highly diverse topography, climate and vegetation, and a wealth of animal and plant resources. China is ranked eighth in the world and first in the Northern Hemisphere on richness of biodiversity (Zhang 1997). China is the home for approximately 10% of the world’s biodiversity (Wu et al. 2004), which provides rich resources for studying molecular evolution. Even though there has been a strong research community in China in documenting the flora and the fauna, and remarkable achievements have been made over the past decades, little was done on the study of molecular evolution in China until a decade ago. Fortunately, the situation has been totally changed in recent years. The molecular evolution studies on the evolution of macromolecules and the molecular phylogeny became active and drew much international attention. For example, Chinese scientists contributed 26 papers in two important molecular evolution journals: Molecular Biology and Evolution and Molecular Phylogenetics and Evolution in 2004. We shall mention four key senior scientists, Professors Jiiazheng Tan (C. C. Tan), Li-ming Shi, Yu-yi Chen and De-yuan Hong, who have made much effort to promote molecular evolution studies in China. In 1990, the Key Laboratory of Cellular and Molecular Evolution was founded by Professor Li-ming Shi at Kunming Institute of Zoology, the Chinese Academy of Sciences. Since then, Kunming Institute of Zoology became the research centre in the area of molecular evolution in China. Such a key laboratory system is valuable to promote new disciplines in China. Here, we have made no attempt to cover all progress in molecular evolution, rather to consider interesting and important developments in China during the recent decade.

2. MOLECULAR PHYLOGENY
During the past 20 years, molecular phylogenetics has dramatically reshaped our views of organism relationships and evolution at all taxonomic levels of the hierarchy of life (Soltis & Soltis 2000). This trend has been manifested from recent studies on many groups in China.

Chinese scientists have made useful contributions on the methods inferring phylogeny. Both maximum-likelihood and maximum-parsimony (MP) methods play key roles for inferring phylogeny from DNA sequences. Felsenstein's (1981) maximum-likelihood method assumes that the rate of nucleotide substitution is constant over different nucleotide sites. This assumption is sometimes unrealistic. To solve this problem, Yang (1993) extended Felsenstein's method...
to the case where substitution rates over sites are described by the gamma distribution. He presented a numerical example to show that the method fits the data better than did previous models (Yang 1993). Yang (1996) investigated the assumptions underlying the MP method by studying the way the method works. He found that parsimony involves very stringent assumptions concerning the process of sequence evolution, such as constancy of substitution rates between nucleotides, constancy of rates across nucleotide sites and equal branch lengths in the tree. When a simple model of nucleotide substitution was assumed to generate data, the probability that MP recovers the true topology could be as high as, or even higher than, that of the likelihood method. When the assumed model became more complex and realistic, the probability that MP recovers the true topology, and especially its performance relative to that of the likelihood method, generally deteriorates. The likelihood method appears preferable to parsimony (Yang 1996).

Restriction fragment length polymorphism of mitochondrial DNA (mtDNA) was used to study phylogenetic relationships of the Old World monkeys (OWMs) and two panda at the early stage (Zhang & Shi 1989, 1991). DNA sequencing became the dominant technique in recent years, as DNA sequences can provide most abundant information for phylogenetic analysis. Even though mtDNA is still the most popular genetic marker and has been widely used for inferring phylogeny of animals, new nuclear markers have also been explored in recent years (Yu et al. 2004; Yu & Zhang 2005). More nuclear genes are expected to be used in the near future. In addition, combined datasets of multiple genes can be useful to resolve some ambiguous phylogeny (Murphy et al. 2001; Yu & Zhang 2005, 2006). A similar trend has been observed in the studies on plants. Knowledge of the relationships among animal and plant groups is essential to understand the history of life, but also relevant to many studies of molecular evolution and developmental genetics.

3. MOLECULAR PHYLOGENY AND BIOGEOGRAPHY OF ANIMALS

As a biodiversity hot spot, China is an important region to understand origin and dispersal of many animals. There are over 3500 species of vertebrates and vascular plants in China. Phylogenetic analysis at the molecular level has shed some light on understanding the pattern of such biodiversity and the process involved.

The mammalian order Carnivora includes 11 families and classically has been divided into two monophyletic superfamilies, Caniformia and Feliformia. Despite numerous efforts, however, evolutionary relationships within and among the diverse families of living carnivores remain controversial (Wozencraft 1989; Zhang & Ryder 1993; Pecon Slattery & O’Brien 1995; Flynn & Nedbal 1998). Yu et al. (2004b) investigated the phylogenetic relationships among 37 living species of order Carnivora using nuclear sequence data from exon 1 of the interphotoreceptor retinoid-binding protein gene and first intron of the transthyretin gene. Their results strongly supported the red panda as the closest lineage to procyonid–mustelid (i.e. Musteloidea) clade followed by pinnipeds (Otaridae and Phocidae), Ursidae (including the giant panda) and Canidae, the most recently diverged clade in the superfamily Caniformia. Four feliform families, namely the monophyletic Herpestidae, Hyaenidae and Felidae, and paraphyletic Viverridae, were consistently recovered convincingly. Yu et al. (2004a) analysed the phylogenetic relationships of all the seven bear species with seven genes. Their results supported basal position of the spectacled bear and sister-taxa association of the brown and polar bear, of Asiatic and American black bears, and of the sloth bear and sun bear, respectively. However, additional DNA sequences from more nuclear genes were needed to fully resolve the phylogeny of this family. In a further analysis of Caniformia using one mitochondrial and four nuclear genes (4417 bp), Yu & Zhang (2006) provided additional evidence to support the phylogenetic relationships recovered previously. To assess the evolutionary history of pantherine lineage of the cat family Felidae, six mitochondrial genes and three nuclear genes (≈6500 bp) were analysed. The monophyletic feature of the Panthera genus in pantherine lineage was confirmed and interspecific affinities within this genus revealed a novel branching pattern, with Panthera tigris diverging first in the Panthera genus, followed by P. onca, P. leo and the last two sister species P. parda and P. uncia. In addition, the close association of Neofelis nebulosa to Panthera, the phylogenetic redefinition of Otocolobus manul within the domestic cat group and the relatedness of Acinonyx jubatus and Puma concolor were all important findings in the resulting phylogenies (Yu & Zhang 2005). The evolutionary relationships in the genus Macaca, the most widely distributed non-human primate, are still under debate. Li & Zhang (2005) sequenced five mitochondrial gene fragments from 40 individuals of 8 species: Macaca mulatta; M. cyclopis; M. fascicularis; M. arctoides; M. assamensis; M. thibetana; M. silenus and M. leonina. Combining with the data from M. sylvanus, they constructed phylogenetic trees using MP and Bayesian methods. Their results agree with earlier studies, suggesting that the mitochondrial lineages of M. arctoides share a close evolutionary relationship with the mitochondrial lineages of fascicularis group macaques (and M. fascicularis, specifically), M. mulatta (with respect to M. cyclopis), M. assamensis assamensis (with respect to M. thibetana) and M. leonina (with respect to M. silenus) are paraphyletic. Those studies demonstrated the power of an approach using multiple genes.

Molecular phylogenetic analysis is not only powerful to resolve the phylogeny relationships, but also useful to assess biogeographic scenarios underlying the geographical structure of lineages, to explore the factors shaping the geographical structure and to understand adaptation of animals.

Yu et al. (2000) investigated the molecular phylogenetic relationships among worldwide species of Ochotona. Their results strongly indicated three major clusters: the shrub–steppe group; the northern group; and the mountain group. They suggested that the differentiation of this genus in the Palearctic Region was closely related to the gradual uplifting of the Tibet
(Qinghai–Xizang) Plateau, and that vicariance might have played a major role in the differentiation of this genus on the Plateau. On the other hand, the North American species, *Ochotona princeps*, is most probably a dispersal event, which might have happened during the Pliocene through the opening of the Bering Strait. These findings support the hypothesis that pikes have entered the steppe environment several times and that morphological similarities within steppe dwellers were due to convergent evolution (Yu et al. 2000). The phylogenetic analysis of musk deer based on the complete mitochondrial cytochrome *b* gene sequence supported the historic dispersion of musk deer in China from north to south, and Himalayan and the Hengduan Mountain areas in southwestern China as a divergence centre for musk deer (Su et al. 1999).

Oriental voles of the genus *Ochotona* are predominantly distributed along the southeastern shoulder of the Qinghai–Tibetan Plateau. Luo et al. (2004) investigated phylogeny of this genus based on mitochondrial gene sequences to test whether vicariance could explain the observed high species diversity in this area by correlating the estimated divergence times to species distribution patterns and the corresponding palaeogeographic events. Their results suggest that: (i) the eight species of Oriental voles form a monophyletic group with two distinct clades and these two clades should be considered as valid subgenera—*Eothenomys* and *Antelomys*, (ii) Japanese red-backed voles are more closely related to the genus *Clethrionomys* than to continental Asian *Eothenomys* taxa, (iii) the genus *Clethrionomys*, as presently defined, is paraphyletic, and (iv) the process of speciation of Oriental voles appears to be related to the Trans-Himalayan formation via three recent uplift events of the Qinghai–Tibetan Plateau within the last 3.6 Myr, as well as to the effects of the Mid-Quaternary ice age.

The genus *Lepus* is monophyletic with three unique species groups: North American; Eurasian; and African. Ancestral area analysis indicated that ancestral *Lepus* arose in North America, and then dispersed into Eurasia via the Bering Land Bridge eventually extending to Africa. Brooks parsimony analysis showed that dispersal events, followed by subsequent speciation, have occurred in other geographical areas as well and resulted in the rapid radiation and speciation of *Lepus*. The genus appears to have arisen ca 10.76 Myr ago (±0.86 Myr ago), with most speciation events occurring during the Pliocene epoch (5.65 ± 1.15–1.12 ± 0.47 Myr ago) (Wu et al. 2005).

The phylogenetic analysis of the subfamily Xenocyprinae from China suggested that species of *Xenocypris* and *Plagiognathops* form a monophyletic group that is sister to the genera *Distochoodon* and *Pseudobrama* (Xiao et al. 2001). Xiao et al. found that the introgressive hybridization might occur among the populations of *Xenocypris argentea* and *X. davidi*. The spatial distributions of mtDNA lineages among populations of *Xenocypris* were compatible with the major geographical region, which indicated that the relationship between Hupeh+Hunan and Fujian is closer than that between Hupeh+Hunan and Sichuan. From a perspective of parasite investigation, their data suggested that the fauna of *Hexamita* in Xenocyprinae could be used to infer the phylogeny of their hosts (Xiao et al. 2001). The family Sisoridae is one of the largest and most diverse Asiatic catfish families, most species occurring in the water systems of the Qinghai–Tibetan Plateau and East Himalayas. Gao et al. (2005) used mitochondrial cytochrome *b* and 16S rRNA gene sequences to clarify existing gaps in phylogenetics and to test conflicting vicariant and dispersal biogeographic hypotheses of Chinese sisorids using dispersal–vicariance analysis and weighted ancestral area analysis in combination with palaeogeographic data as well as molecular clock calibration. Their results suggest that: (i) Chinese sisorid catfishes form a monophyletic group with two distinct clades; (ii) the glyptosternoid is a monophyletic group and *Glyptosternum, Glaridoglanis* and *Exostoma* are three basal species having a primitive position among them; (iii) a hypothesis referring to *Pseudechenes* as the sister group of the glyptosternoid was supported; (iv) the genus *Pareuchiloglanis* is not monophyletic; (v) the uplift of Qinghai–Tibetan Plateau played a primary role in the speciation and radiation of the Chinese sisorids; and (vi) an evolutionary scenario combining the aspects of both vicariance and dispersal theory is necessary to explain the distribution pattern of the glyptosternoids. They tentatively estimated that the glyptosternoids most possibly originated in the Oligocene–Miocene boundary (19–24 Myr ago) and radiated from the Miocene to Pleistocene, along with a centre of origin in the Irrawaddy–Tsango drainages and several rapid speciations in a relatively short time.

More than 10 species within the freshwater fish genus *Sinocyclocheilus* adapt to caves and show different degrees of degeneration of eyes and pigmentation. Therefore, this genus can be useful to study evolutionary developmental mechanisms and the role of natural selection and adaptation in cave animals. To better understand these processes, Xiao et al. (2005) investigated the phylogenetic relationships of 31 recognized species based on the complete mitochondrial cytochrome *b* gene (1140 bp) and partial NADH dehydrogenase subunit 4 (*ND4*) gene (1032 bp). Their phylogenetic results showed that all species except for two surface species, *Sinocyclocheilus jii* and *S. macrolepis*, clustered as five major monophyletic clades (I, II, III, IV and V) with strong supports. *S. jii* was the most basal species in all the analyses, but the position of *S. macrolepis* was not resolved. The cave species were polyphyletic and occurred in these five major clades. The results indicate that adaptation to cave environments has occurred multiple times during the evolutionary history of *Sinocyclocheilus*. The branching orders among the clades I, II, III and IV were not resolved, and this might be due to early rapid radiation in *Sinocyclocheilus*. All species distributed in Yunnan except for *S. rhinoceros* and *S. hyalinus* formed a strongly supported monophyletic group (clade V), probably reflecting their common origin. Such results suggested that the diversification of *Sinocyclocheilus* in Yunnan may correlate with the uplifting of Yunnan Plateau.

The sand lizards of the genus *Phrynocephalus* (Family Agamidae; Kaup 1825) distributed from northern-western China to Turkey, and are one of the major
components of the Central Asian desert fauna. Pang et al. (2003) investigated the phylogenetic relationships among most Chinese species of lizards in the genus Phrynocephalus using four mitochondrial gene fragments (12S rRNA, 16S rRNA, cytochrome b and ND4-trRNALEU). Their results showed that there are two major clades representing Chinese Phrynocephalus species: the viviparous group and the oviparous group. All the analyses left the nodes for the oviparous group, the most basal clade within the oviparous group and Phrynocephalus mystaceus unresolved. The phylogenies further suggest that the monophyly of the viviparous species might have resulted from vicariance, while recent dispersal might have been important in generating the pattern of variation among the oviparous species.

Chen et al. (2004) reconstructed the phylogeny of the family Eurytomidae based on nuclear 28S and 18S rRNA genes, and mitochondrial 16S rRNA and cytochrome oxidase I genes. Their analysis revealed a significant incongruence between the mitochondrial and the nuclear genes. Their choice of results from the nuclear genes as their preferred hypothesis suggested that the family Eurytomidae was not a monophyletic group; neither were the genera Plutarchia and Bruchophagus and Eurytoma. The monophyly of genera Sycophila and Phytarcha was well supported, as was the close association of the genera Aiolomorphus, Temeniptilus, Bebrateloides and Phylloxeroxenus. Their phylogeny also revealed an anticipated pattern in which species groups from the genera Eurytoma and Bruchophagus were often more closely related to other small genera than to other species groups of the same genus. However, more sequence data from both nuclear and mitochondrial genes are required to test their hypothesis. Furthermore, this family can be useful to explore the factors resulting in the incongruence between the mitochondrial and the nuclear genes.

4. MOLECULAR PHYLOGENY OF PLANTS
Recent advances in phylogenetic methodologies, together with the availability of new types of data (in particular, the DNA sequences), have significantly increased the accuracy of phylogenetic reconstruction of plants.

Pinaceae is the largest extant family of gymnosperms comprising 11 genera and more than 200 species (Farjon 1990). Many species of the pine family constitute the major forest elements in the northern temperate region. Owing to morphological convergence within the family, Pinaceae has been a phylogenetically complex group (Farjon 1990). Using sequences of the chloroplast matK gene, the mitochondrial nad5 gene and the low-copy nuclear gene 4CL, Wang et al. (2000) studied interfamilial relationships of all the extant genera in Pinaceae. A combined analysis of the three gene sequences generated a well-resolved and strongly supported phylogeny that agrees to a certain extent with the previous phylogenetic hypotheses based on morphological, anatomical and immunological data. Disagreement between the previous hypotheses and the three-genome phylogeny suggested that morphology of both vegetative and reproductive organs has undergone convergent evolution within the pine family. The strongly supported monophyly of Nothotsuga longibracteata, Tsuga mertensiana and Tsuga canadensis on all the three gene phylogenies provided evidence against the previous hypotheses of interfamilial hybrid origins of N. longibracteata and T. mertensiana. This study with a complete generic sampling of the family provided important insights into the phylogenetic relationships within the family by examining congruence and incongruence of gene phylogenies among the three genomes.

Chen et al. (1999) presented a comprehensive analysis of the phylogenetics and the systematics of the family Betulaceae from very diverse perspectives, including molecular, morphological and fossil evidence. They used rbcL, internal transcribed spacer (ITS) and morphological data to assess the phylogeny of Betulaceae, which was found to be monophyletic, with Casuarinaceae as its sister group. Within Betulaceae, two sister clades were evident, corresponding to the two traditional subfamilies Betuloideae (Alnus and Betula) and Coryloideae (Corylus, Ostryopsis, Carpinus and Ostrya). The molecular phylogenetic relationships among the extant genera were compatible with inferences from ecological evolution and the extensive fossil record (Chen et al. 1999).

As an important group of the subclass Hamamelidaceae of angiosperms, Fagales comprise eight families: Betulaceae, Casuarinaceae, Fagaceae, Juglandaceae, Myricaceae, Nothofagaceae, Rhoipteleaceae and Ticondendraceae. Although the concept of the broadly defined order has received strong support from molecular systematic studies and its monophyly has been well established (see the reviews by Li et al. 2004), interfamilial relationships have not been resolved with certainty. Li et al. (2004) sequenced six regions from three plant genomes (the plastid trnL-F, matK, rbcL and atpB; the mitochondrial matR; and the nuclear 18S rDNA) for all 31 extant genera representing eight families of the order. At the familial level, the same phylogenetic relationships were inferred from five different analyses of these data. Nothofagus, followed by Fagaceae, were subsequent sisters to the rest of the order. Fagaceae are then sister to the core ‘higher’ hamamelids, which consist of two main subclades. The combined datasets provide the best-supported estimate of evolutionary relationships within Fagales. In addition, this study suggested that the combination of different sequences from several species within the same genus representing a terminal taxon had little influence on phylogenetic accuracy, and inclusion of taxa with some missing data in combined datasets did not have a major impact on the topology (Li et al. 2004).

The Lythraceae sensu lato (s.l.) comprise small to large trees, shrubs, and perennial and annual herbs adapted to a wide variety of vegetation types, including mangrove swamps, rainforests, seasonally dry savannas, coastal dunes, freshwater marshes and shallow waters of ponds and rivers (Huang & Shi 2002). All members of the family share numerous specialized features with the rest of the order Myrtales. The phylogenetic relationships of Lythraceae s.l. were investigated by parsimony and likelihood analyses of 83 accessions representing 23 species; 16 genera have been assigned to the family at various times.
Phylogenetic analyses based on the three datasets (chloroplast rbcL gene, psaA-psyc spacer and the nuclear ITS) strongly support the monophyly of the Laevistemonaleae s.l., in which the satellite genera Dypsis, Pinanga, Sonneratia and Trapia were included. Paraphyly of subfamily Lythroidae (= Laevistemonaleae sensu stricto) is proposed with the other four monotypic subfamilies nested within (Huang & Shi 2002).

There are many other examples of phylogenetic inferences at the intergeneric and higher levels and the application of molecular systematic approaches have resolved many debates involving a variety of vascular plant groups (Shi et al. 2000; Ge et al. 2002; Kong et al. 2002; Meng et al. 2003; Zhang et al. 2003a,b; Huang & Shi 2002; Guo & Ge 2005).

One of the important tasks of phylogenetic reconstruction is to reveal the closest relatives of model species and economically important crops. An obvious example is the study of the Asian cultivated rice (Oryza sativa) and its relatives. Rice is one of the world’s most important crops and is becoming an excellent model for various biological studies, in particular after the completion of the rice genome projects (Yu et al. 2002; Guo & Ge 2005). Although considerable effort has been devoted to studying rice genetics and breeding, the phylogeny of the rice genus (Oryza) has been less explored than that of the other major crops and model plants such as arabidopsis, maize, wheat, soybean and cotton (Solitis & Solitis 2000). The recent decade has seen a series of molecular phylogenetic investigations undertaken in China to examine the relationships of rice and its relatives at different taxonomic levels. Based on the sequence data of two nuclear genes (Adh1 and Adh2) and a chloroplast gene (matK), Ge et al. (1999) identified a new or the tenth genome type (HK) in the genus and reconstructed the phylogeny of all the 10 genomes and 23 Oryza species. The phylogeny indicated clearly that the G-genome species is the earliest divergent lineage, whereas the A-genome group, which contains the cultivated rice, is a recently diverged and rapidly radiated lineage within the rice genus. Further multiple-gene phylogenetic studies on the rice tribe based on chloroplast, mitochondridal and nuclear genes revealed that the rice tribe formed two monophyletic groups, and Leersia is the most closely related genus to Oryza (Ge et al. 2002; Guo & Ge 2005). To better understand the phylogenetic relationships of the Asian cultivated rice and its close relatives, which have been controversial for a long time, Zhu & Ge (2005) sequenced the fast evolving introns of four nuclear genes for eight A-genome species and reconstructed their phylogeny. The resulting phylogeny demonstrated that the Australian endemic, O. meridionalis, is the earliest divergent lineage and supported the previous opinion to treat O. rufipogon and O. nivara as a single species. More importantly, they indicated that two subspecies of O. sativa (ssp. indica and ssp. japonica) formed two separate monophyletic groups, suggestive of their independent origins from two ancestor lineages that separated from each other ca 400 000 years ago (Zhu & Ge 2005).

To put the divergence of the rice genus and tribe into a temporal framework, Guo & Ge (2005) and Zhu & Ge (2005) recently dated the divergence times of the main lineages in the tribe and genus, respectively, based on the molecular clock approach. The estimates suggest that the rice tribe originated roughly 35 Myr ago, Oryza and Leersia separated from the rest of the tribe ca 20 Myr ago, and from each other 14 Myr ago. Within Oryza, the age of the deepest split between the most basal G genome and the remaining genomes was estimated at ca 9 Myr, and the diversification of the A-genome group occurred ca 2 Myr ago. Despite a number of limitations to the use of clocks based on sequence data, the divergence estimates for the main lineages in Oryza and its related genera provide a rough timeframe for understanding the evolutionary tempo of this important plant group and establish a framework for studying various biological questions using rice as a model system.

As the closely related group to rice, bamboos are a plant group of great ecological and economic importance in China. Based on partial sequences of the nuclear GBSSI gene and ribosomal ITS spacer, Guo & Li (2004) constructed the phylogeny of eight genera of the Thamnocalamus group and its allies that include the main food bamboos for the giant pandas and other rare fauna of the Himalayas and the adjacent areas. The results supported the monophyly of the Thamnocalamus group and its allies, but Thamnocalamus group per se was resolved as polyphyletic. In addition, the current limitation of Thamnocalamus, Fargesia (including Borinda) and Yushania may not reflect the true phylogenetic relationships of the complex. Kiwifruit (Actinidia delicosa) is a popular fruit worldwide; it was domesticated from a single fruit collected in Central China and has been cultivated in New Zealand since 1904 (Li et al. 2002). China is both the distributional centre and the genetic diversity centre of Actinidia species, because 62 out of a total 66 species in this genus are found in China and high diversity exists within species. Based on molecular phylogenetic analyses of 23 Actinidia species, Li et al. (2002) found that neither ITS nor chloroplast matK phylogeny supported monophyly of any of the four sections of the genus, suggestive of the necessity of revision of infrageneric classification of Actinidia. In addition, polymorphic ITS sequences within accessions of some species implied historical events from hybridization or introgression. These, together with other examples, have highlighted the value of a phylogenetic framework in the effective utility and conservation management of the valuable genetic diversity.

5. MOLECULAR BIOGEOGRAPHY AND SPECIATION IN PLANTS

While China has the most diverse flora of any country in the North Temperate with very high levels of endemism, the diversity of plant genera and species is not uniformly distributed, but concentrated in the south central part of China (Ying 2001). Recent advance in phylogenetic reconstruction using DNA sequences provides an excellent opportunity for addressing biogeographic questions concerning plant biogeography and endemism.

The Qinghai–Tibet Plateau is the highest plateau in the world. Its short period of formation, since the
Pliocene, has considerably influenced the structure and evolution of its component flora (Shi et al. 1998). Although glaciations during the Quaternary period covered almost the entire plateau, and thus led to large-scale recession and extinction of many species, the plateau has an exceptionally diverse flora with approximately 4385 species of 1174 genera in 189 families. In contrast to the fact that over 25% of the total number of species are endemic to the plateau, less than 2% of the total number of genera have been considered endemic (Wu 1987), leading to a hypothesis that the Quaternary ice sheet had wiped out the flora of the plateau and the current flora migrated or originated from adjacent areas (Wulff 1944; Wu 1987).

Nannoglottis (Asteraceae) is a genus with about eight species endemic to the Qinghai–Tibet Plateau. Previous taxonomic studies have suggested its relationships with four different tribes of Asteraceae. Based on sequence data from chloroplast ndhF and trnL-F, and nuclear ITS, Liu et al. (2002) found that all sampled species of Nannoglottis formed a well-defined monophyly and were most closely related to the tribe Astereae. Despite the very early divergence of Nannoglottis in the Astereae, the tribe must be regarded to have its origin in the Southern Hemisphere rather than in Asia, because Asteraceae and its major lineages (tribes) are supposed to have originated in the Southern Hemisphere based on different lines of evidence (Liu et al. 2002). Since the 23–32 Myr divergence time between Nannoglottis and other Astereae predated the formation of the plateau, Nannoglottis seems to have reached the Qinghai–Tibet area in the Oligocene–Eocene and then re-diversified with the uplift of the plateau. Liu et al. (2002) also estimated that the ‘alpine shrub’ versus ‘coniferous forest’ divergence within Nannoglottis was ca 3.4 Myr ago, when the plateau began its first large-scale uplifting and the coniferous vegetation began to appear.

Most of the current species at the ‘coniferous forest’ clade of the genus are estimated to have originated from 1.02 to 1.94 Myr ago, when the second and third uprisings of the plateau occurred. The diversification and radiation of Nannoglottis in the Qinghai–Tibet Plateau suggested by the molecular phylogenetic study agreed well with the known geological and palaeobotanical histories of the Qinghai–Tibet Plateau (Liu et al. 2002). Similarly, Yang et al. (2003) also detected a high level of ITS divergence for 42 Pedicularis species sampled from this region and suggested a relatively ancient origin and diversification of the genus followed by migration of different floristic components into this region.

The Hengduan Mountain region is located in the southeast part of the Qinghai–Tibet Plateau and is one of the three Chinese hotspots of plant diversity (Ying 2001). The uplift of the Qinghai–Tibet Plateau in the past 40 Myr has caused the regional climate to cool down markedly and changed the regional topography. This has led to a sharp change in the vegetation and speciation and evolution (Ying 2001). An excellent example is Pinus densata, which is distributed in the Hengduan Mountain region and the adjacent area.

Previous morphological, allozyme and chloroplast DNA data indicated that this species is a hybrid pine originated from hybridization between P. tabulaeformis and P. yunnanensis and the different populations of P. densata have different evolutionary histories (Wang & Szmidt 1994). Further analyses of allozyme and restriction sites of chloroplast and mitochondrial fragments based on population samples revealed different evolutionary histories among P densata populations and that P. tabulaeformis and P. yunnanensis had acted as both mother and father donors (Wang et al. 2001; Song et al. 2003). In addition, P. densata populations have a stabilized hybrid nature and maintained successful sexual reproduction with normal fertility and high fecundity despite significant founder effects; the backcross happened during the population establishment of the hybrid pine (Wang et al. 2001; Song et al. 2003). P. densata represents an example of diploid hybrid speciation in an extreme ecological habitat that is both spatially and ecologically separated from that of its parents.

Polyploidization followed by hybridization between species is another force driving plant speciation. Low-copy nuclear genes, which contain fast-evolving introns and are less susceptible to concerted evolution, are powerful markers for reconstructing allotetraploidization. Using sequences of biparental inherited nuclear Adh1 and Adh2 genes as well as the maternally inherited matK gene, Ge et al. (1999) reconstructed the origin of the allotetraploid species in Oryza. In contrast to a single origin of three CD-genome species, the BC-genome species had different origins because their maternal parents had either a B or a C genome (Ge et al. 1999). A study with additional genes and samples further demonstrated that the CD-genome species originated from the hybridization between the C- and E-genome species, and that the C-genome species served as the maternal parent (Bao & Ge 2004). Zhang et al. (2002a-c) also studied the origin and evolution of some tetraploid wheat using nuclear ITS sequences.

6. POPULATION AND CONSERVATION GENETICS OF ANIMALS

Knowledge of population history and genetic structure is not only necessary to understand the evolutionary process of the species, but also useful for conserving the endangered species.

Liu et al. (2006a) studied the demographic history and genetic structure of Lateolabrax maculatus and L. japonicus in the northwestern Pacific using the cytochrome b gene and control region sequences. The demographic history of the two species was examined using neutrality tests and mismatch distribution analyses, and results indicated Pleistocene population expansion in both the species. Estimates of population expansion time suggested earlier population expansion in L. japonicus than in L. maculatus. Molecular variance analyses showed differential genetic structuring for these two closely related species. Their results indicated that L. japonicus is panmictic throughout its range. In contrast, populations of L. maculatus showed statistically significant levels of genetic structuring. Pattern of isolation by distance was observed in...
L. maculatus, suggesting that L. maculatus is in genetic equilibrium. In contrast, L. japonicus did not exhibit isolation by distance (Liu et al. 2006a). Such kind of study can provide rich information on the population history and the genetic structure of animal species.

Climatic oscillations during the Pleistocene ice ages produced great changes in species’ geographical distribution and abundance, which could be expected to have genetic consequences. To investigate the effects of Pleistocene climatic changes on the evolution in Japanese anchovy, Liu et al. (2006b) determined mtDNA control region sequences for 241 individuals from 13 localities and 37 individuals of Australian anchovy. Japanese anchovy and Australian anchovy are reciprocally monophyletic and a Late Pleistocene transequatorial divergence between the two species was indicated. Analyses of molecular variance and the conventional population statistic $F_{ST}$ revealed no significant genetic structure throughout the range of Japanese anchovy. Both mismatch distribution analyses and neutrality tests suggested a Late Pleistocene population expansion for both Japanese anchovy (79 000–317 000 years ago) and Australian anchovy (45 000–178 000 years ago; Liu et al. 2006b).

Combining microsatellites with mitochondrial sequences is useful for studying the genetic structure of animal species. Shao et al. (2004) investigated seven mainland and island Asian populations of Bombus ignitus using nine microsatellite markers and the mitochondrial cytochrome $b$ (cytb) gene sequences. While microsatellite markers showed high genetic variability, no sequence variation was found in the cytb gene fragment analysed. Gene diversities per locus per population ranged from 0.378 to 0.992. AMOVA test and most pairwise $F_{ST}$ values showed significantly genetic differentiation between mainland and island populations. The cytb sequence data and microsatellite bottleneck tests indicated that almost all populations were subjected to recent bottleneck. Their results suggest that B. ignitus populations diverged due to recent bottleneck and geographical isolation (Shao et al. 2004).

It is generally believed that there is a negative correlation between the level of genetic diversity and the endangerment of a species. However, what much genetic diversity really reflects is the demographic history of the species rather than its fitness. A recent study of Li et al. (2003) provided an example of a proper interpretation of the genetic diversity. The Golden monkey (Rhinopithecus roxellana) is a well-known primate, which is distributed in the central part of mainland China, with a population size around 10 000–20 000. Forty-four allozyme loci were surveyed in 32 individuals, none of which were found to be polymorphic. The void of polymorphism compared with that of other non-human primates is surprising, particularly considering that the current population size is many times larger than some other endangered species. Coalescent approaches were used to explore various scenarios of population bottleneck, and it was concluded that the most recent bottleneck could have happened within the last 15 000 years. Moreover, the proposed simulation approach could be useful for researchers who need to analyse zero or low polymorphism data (Li et al. 2003).

The red panda (Ailurus fulgens) is an endangered species and its present distribution is restricted to the isolated mountain ranges in western China (Sichuan, Yunnan, and Tibet provinces) and the Himalayan Mountain chain of Nepal, India, Bhutan and Burma. Li et al. (2005a) did not observe significant differentiation between subspecies. They suggest that the present population structure has resulted from habitat fragmentation and expansion from glacial refugia. Owing to its habitat requirements, it is probable that the red panda has undergone bottlenecks and population expansions several times in the recent past. The present population may exhibit a pattern reminiscent of a relatively recent population expansion. Similar to the red panda, significant genetic differentiation has not been detected among the giant panda populations, and a population bottleneck was inferred (Zhang et al. 2002a–c).

The Chinese sturgeon, Acipenser sinensis, is an endangered anadromous fish, mainly endemic to the Yangtze River, China. Its spawning migratory route was blocked in 1981 by the Gezhouba Dam, which caused a drastic decline of the natural population. Zhang et al. (2003a,b) investigated mtDNA sequence variation in this species. Their results indicate that the Chinese sturgeon underwent a population expansion in the past, but there is no indication of a historic genetic bottleneck. Strikingly, the ratios of effective female population size ($N_e$) to the census female population size ($N_0$) are unusually high (0.77–0.93). This may be interpreted as the result of a current or recent bottleneck in the population, which is more probably caused by human intervention, rather than evolutionary forces (Zhang et al. 2003a,b).

These studies demonstrate how population genetic analysis can help to understand an endangered species. A species with low genetic diversity could have the ability to expand if the habitat is conserved.

7. ORIGIN OF DOMESTIC ANIMALS
The domestic animals were extremely important in the process of human civilization. Understanding the origin and dispersal of domestic animals is not only important by itself, but can also provide valuable information to trace the recent human evolution as well. Archaeological finds from Mesolithic sites around the world indicate that the dog was the first domestic animal. The origin of the domestic dog from wolves was established, but the number of founding events, as well as where and when these occurred, is not known. To address these questions, Savolainen et al. (2002) examined the mtDNA sequence variation among domestic dogs representing all major populations worldwide. They found that more than 95% of all the sequences belonged to three universally represented phylogenetic groups, showing a common origin from a single gene pool for all dog populations. A larger genetic variation in East Asia compared with Southwest Asia and Europe strongly suggests an East Asian origin ca 15 000 years ago for the domestic dog (Savolainen et al. 2002).

Domestic chickens have long been important to human societies for food, religious, entertainment and decorative uses, yet the origin and diffusion of chickens...
in the world remain unclear. Liu et al. (2006c) assessed the origin of chicken by analysing the mtDNA sequences from 834 domestic chickens across Eurasia as well as 66 red jungle fowls (Gallus gallus) from Southeast Asia and China. They revealed nine highly divergent mtDNA clades (A–I), of which seven clades contained both the red jungle fowls and domestic chickens. The clades A, B and E are distributed ubiquitously in Eurasia, while the other clades were restricted to South and Southeast Asia. Clade C was mainly distributed in Japan and Southeast China, while clades F and G were exclusive to Yunnan, China. The geographical distribution of clade D was closely related to the distribution of the pastime of cock fighting. Statistical tests detect population expansion within each subclade. These distinct distribution patterns and expansion signatures suggest that different clades may originate from different regions, such as Yunnan, South and Southwest China and/or surrounding areas, and the Indian subcontinent, respectively, which support the theory of multiple origins in South and Southeast Asia (Liu et al. 2005).

China has numerous native domestic goat breeds, but so far there has been no extensive study on genetic diversity, population demographic history and origin of Chinese goats. Chen et al. (2005) examined the genetic diversity and phylogeographic structure of Chinese domestic goats. They revealed that there were four mtDNA lineages (A, B, C and D) identified in Chinese goats, in which lineage A was predominant, lineage B was moderate and lineages C and D were at low frequency. These results further support the multiple maternal origins of domestic goats. They detected two subclades in lineage B, in which one was unique to eastern Asia, another was shared between eastern and southern Asia. A larger genetic variation in eastern Asia than southern Asia and the pattern of phylogeographic variation in lineage B suggest that at least one subclade of lineage B originated from eastern Asia. There was no significantly geographical structuring in Chinese goat populations, which suggested that there existed strong gene flow among goat populations caused by extensive transportation of goats in history (Chen et al. 2005).

To investigate the origin and genetic diversity of Chinese domestic sheep, Chen et al. (2006) analysed mtDNA control region of 449 Chinese autochthonous sheep from 19 breeds/populations from 13 geographical regions. Phylogenetic analysis showed that all the three previously defined lineages A, B and C were found in all sampled Chinese sheep populations, except for the absence of lineage C in four populations. The pattern of genetic variation in lineage A, together with the divergence time between the two central founder haplotypes, suggested that two independent domestication events have occurred in sheep lineage A. The high levels of intrapopulation diversity in Chinese sheep and the weak phylogeographic structuring indicated that three geographically independent domestication events have occurred and the domestication place was not only confined to the Near East, but also occurred in other regions (Chen et al. 2006).

These and other studies suggest that East and South Asia is one of the centres for domestication, which provides a good opportunity for Chinese scientists. Further studies using more genetic markers, and on other domestic animals in this region, will shed new light on the origin of domestic animals.

8. ORIGIN OF GENES

Origin of genes is a fundamental process during the evolution of organisms (Long et al. 2003). The mechanism of origin of genes and the evolutionary forces behind the functional divergence of newly originated genes are not yet fully understood. Genome-wide analysis and phylogenetic inference of genes and gene families have shed some light on this issue.

Recently, Wang et al. (2004) discovered a unique young gene family of Drosophila. The new gene family was named monkey-king (mkg) following an ancient Chinese legend in which the monkey king could instantly produce many offspring transformed from its hairs, to account for several new genes that were created in this gene family in a very short time. This mkg family could be used to answer a number of important questions in the study of new gene origination. First, they directly showed how a eukaryotic gene fissions for the first time. The process by which mkg originated demonstrates that gene duplication followed by complementary partial degeneration is a simple, but efficient mechanism for gene fission. This also provides a general mechanism to generate new introns in a previously intronless gene. In addition, this gene family also showed that regulatory sequences could be created quickly (Wang et al. 2004).

Complete genome sequences demonstrated that duplication is a major source for the origin of genes. It is widely accepted that there is often an acceleration of the rate of evolution following gene duplication, which could be driven either by relaxation of selective constraints or by positive selection. However, the relative contribution of these two factors is still controversial. A natural question is the role of gene duplication in organisinal adaptation and biodiversity (Zhang et al. 2002a–c).

Growth hormone (GH) is a classic molecule in the study of the molecular clock hypothesis, as it exhibits a relatively constant rate of evolution in most mammalian orders except primates and artiodactyls, where dramatically enhanced rates of evolution (25- to 50-fold) have been reported. Ye et al. (2005) determined 21 different GH-like sequences from four species of OWM and Hominoids. Their analysis demonstrated that multiple gene duplications and several gene conversion events both occurred in the evolutionary history of this gene family in OWM/Hominoids.

GH-N genes in hominoids and OWM are under strong purifying selection. In contrast, CSH genes in both lineages are probably not. GH-V genes in OWM and hominoids evolved at different evolutionary rates and underwent different selective constraints. Recent phylogenetic analysis of Li et al. (2005c) revealed monophyly for New World monkey (NWM) GH-like genes with respect to those of OWM/hominoids, which indicates that independent gene duplications have occurred in NWM GH-like genes. Their results disclosed the complex history of the primate GH gene

Phil. Trans. R. Soc. B (2007)
family and raised intriguing questions on the consequences of these evolutionary events.

Wang et al. (2005a,b) identified and characterized three Spindlin (Spin) genes from medaka (Oryzias latipes). Their phylogenetic analysis demonstrated that those three genes resulted from gene duplication. Furthermore, Western blot analysis revealed significant expression differences of the three Spins among different tissues and during embryogenesis in medaka, and suggested that sequence and functional divergence might have occurred during evolution.

The enzyme chalcone synthase (CHS) is a key enzyme in the biosynthesis of flavonoids, which are important for the pigmentation of flowers and act as attractants to the pollinators. The CHS gene family is probably one of the most studied gene families in plants. Growing evidence shows that genes encoding CHS constitute a multigene family in which the copy number varies among plant species and functional divergence appears to have occurred repeatedly (Yang et al. 2002). Plants in the genus Dendranthema (Asteraceae) have white, yellow and pink flowers, exhibiting considerable variation in flower colour. Based on the phylogenetic analyses of the 18 CHS genes cloned and sequenced from six Dendranthema species and those available for other genera of Asteraceae in GenBank, Yang et al. (2002) found that except for two pseudogenes, the functional Dendranthema CHS genes formed three well-supported subfamilies: SF1, SF2 and SF3. The inferred phylogeny of the CHS genes of Dendranthema and Gerbera suggests that these genes originated as a result of duplications before divergence of these two genera, and the function of Dendranthema CHS genes have diverged in a similar fashion to the Gerbera CHS genes, i.e. the genes of SF1 and SF3 code for typical CHS enzymes expressed during different stages of development, whereas the genes of SF2 code for another enzyme that is different from CHS in substrate specificity and reaction. More importantly, much higher non-synonymous–synonymous rate ratio for the lineage ancestral to SF2 was detected, suggesting that positive selection appeared to have driven the divergence of SF2 from SF1 and SF3 (Yang et al. 2002).

Gene duplication may arise from region-specific duplication or genome-wide polyploidization, and ranks among the most important evolutionary mechanisms affecting plant genome evolution. Since gymnosperms are characterized by a large nuclear genome and highly complex gene families, it is of interest to investigate the frequency of gene duplication and the evolutionary fate and the consequences of duplicate genes. Wang et al. (2000) studied intergeneric relationships of Pinaceae using a low-copy nuclear 4CL gene and found that duplication and deletion of the 4CL gene occurred at a tempo such that paralogous loci were maintained within but not between genera. They also found that exons of the 4CL gene have diverged approximately twice as fast as the matK gene and five times more rapidly than the nad5 gene. Further investigation on the evolution of 4CL gene in Larix, another genus in Pinaceae, demonstrated that frequent duplication/deletion appears to be a common evolutionary phenomenon. This gene family and paralogous genes differ greatly in their evolution rate (Wei & Wang 2004).

To examine whether the amino acid substitution rates are the same between duplicate genes of a ubiquitous NF-Y transcription factor, Yang et al. (2005a,b) used likelihood-ratio tests to evaluate the influence of selection on evolution of duplicated NF-Y genes in the Arabidopsis and rice genomes by comparing the conservative and radical amino acid substitution rates. The results indicated that some NF-YB and NF-YC duplicates showed significant evidence of asymmetric evolution, but not the NF-YA duplicates, and most amino acid replacements in the NF-YB and NF-YC duplicates resulted in changes in hydropathy, polar requirement and polarity. They suggested that relaxed selective constraints following gene duplication are most probably responsible for the unequal evolutionary rates and distinct divergence patterns of duplicate NF-Y genes, and positive selection might have promoted amino acid hydropathy changes in the NF-YN duplicates (Yang et al. 2005a,b).

These studies support the idea that relaxation of selective constraints is common following gene duplication, but positive selection may operate to drive the function divergence among duplicates.

9. EVOLUTION OF GENES

SRY is a Y-chromosomal gene that is pivotal in initiating the development of testis and a determinant of male sex in mammals. A higher rate of non-synonymous nucleotide substitution than that of synonymous substitution in the terminal regions of SRY has been reported in primates, with suggestions that these regions may be subject to positive Darwinian selection (Whitfield et al. 1993). Wang et al. (2002) sequenced the SRY genes of nine OWMs and revealed reduced rate of non-synonymous substitution and action of purifying selection in the terminal regions of SRY in OWMs. While confirming earlier results of high non-synonymous substitution rates of SRY in hominoids (Whitfield et al. 1993), they were not able to reject the neutral evolution hypothesis. As such, whether the rapid evolution of SRY in hominoids is due to positive selection or relaxed functional constraints remains an open question. Thus, SRY shows a complex pattern of erratic evolution among different groups of primates, raising an intriguing possibility of varied selective pressure on this fundamentally important gene in evolution (Wang et al. 2002).

Yang & Gui (2004) studied transferrin polymorphism in the three subspecies of polyploid Carassius auratus. DNA polymorphism of extremely high extent was shown for the transferrin gene by the 248 segregation sites among coding region sequences of its alleles. The deduced amino acid sequences of the transferrin alleles showed variable theoretical physicochemical parameters; positive selection was observed. Furthermore, the corresponding sites to these selected codons were collectively located at two planes in the crystallographic structure of rabbit transferrin, which suggested that the rapid evolution of C. auratus transferrin might correlate to its adaptation to variable environmental elements such as oxygen pressure. The minimal 26 recombination
events were detected among coding sequences of *C. auratus* transferrin, with partial mosaic sequences and breakpoints. Phylogenetic analyses revealed multiple antique allelic lineages of transferrin, which was estimated to diverge 15–20 Myr ago. All these features strongly suggested the role of balancing selection in long persistence of high transferrin polymorphism in *C. auratus* (Yang & Gui 2004).

Protein tyrosine phosphatase (PTP) consists of two superfamilies, the phosphatase I with a single low-molecular-weight PTP (lmwPTP) family and the phosphatase II including both the higher-molecular-weight PTP (hmwPTP) and the dual specificity phosphatase (DSP) families, but the phosphatases I and II are usually considered to be the result of convergent evolution (Stone & Dixon 1994; Barford et al. 1995). However, Huang (2003) demonstrated that lmwPTPs and hmwPTPs/DSPs are remotely related in evolution, and all these three PTP families might have resulted from a common ancestral gene by a series of duplications, fusions and circular permutations. The circular permutation in PTPs is caused by a reading frame difference, which is similar to that in DNA methyltransferases (Jeltsch 1999). Nevertheless, the evolutionary mechanism of circular permutation in *PTP* genes seems to be more complicated than that in DNA methyltransferase genes. Both mechanisms in PTPs and DNA methyltransferases can be used to explain how some protein families and superfamilies came to be formed by circular permutations during molecular evolution.

Glycyl-tRNA synthetases (GlyRSs) with z2/z2 tetramer are distributed in most bacteria, but those with z2 dimer are in archaebacteria, eukaryotes and a few bacteria. Tan & Huang (2005) found that the anticodon-binding domains (ABDs) of dimeric and tetrameric GlyRSs are non-homologous, but their catalytic central domains are homologous. The ABD structures of dimeric and tetrameric GlyRSs are clearly different, but the functions of both GlyRSs in dimer and tetramer forms are the same. Thus, the mechanisms of GlyRS recognizing tRNA_Gly in dimeric and tetrameric GlyRSs should be different. The results suggest that the same function in some proteins can be realized by mechanisms changing during evolution from bacteria to eukaryotes, which also implies a special evolutionary way in GlyRS recognizing tRNA_Gly.

Divergence of proteins in signalling pathways requires ligand and receptor coevolution to maintain or improve binding affinity and/or specificity. Li et al. (2005b) provided a clear case of coevolution between the prolactin (*PRL*) gene and its receptor (prolactin receptor, PRLR) in mammals. First, they observed episodic evolution of the extracellular domain (ECD) and the intracellular domain (ICD) of the PRLR, which is closely consistent with that seen in *PRL*. Correlated evolution was demonstrated both between PRL and its receptor and between the two domains of the PRLR using Pearson’s correlation coefficient. On comparing the ratio of the non-synonymous substitution rate to synonymous substitution rate (ω) for each branch of the star phylogeny of mammalian PRLR, separately for the ECD and the transmembrane domain/intracellular domain (TMD/ICD), they observed a lower ω ratio for ECD than TMD/ICD along those branches leading to pig, dog and rabbit, but a higher ratio for ECD than TMD/ICD on the branches leading to primates, rodents and ruminants, on which bursts of rapid evolution were observed. These observations can be best explained by coevolution between PRL and its receptor and between the two domains of the PRLR (Li et al. 2005b).

To better understand the evolution of genes, it is necessary to take network interaction of genes into account in future studies.

### 10. FROM SINGLE GENE TO GENOME APPROACH

In the genomic era, it is evident that there is a transition from single-gene investigation to a genomic approach.

The explosion of genomic studies in *Arabidopsis*, rice, maize and other model plants is providing unprecedented opportunities to address many biological questions, including evolution of gene and genome. For example, members of the *GRAS* gene family encode transcriptional regulators that have diverse functions in plant growth and development such as gibberellin signal transduction, root radial patterning, axillary meristem formation, phytochrome A signal transduction and gametogenesis (Tian et al. 2004). To take full use of the *Arabidopsis* and rice genome sequences and bioinformatic approaches, Tian et al. (2004) identified 57 and 32 *GRAS* genes in rice and *Arabidopsis*, respectively, and the *GRAS* gene family could be divided into eight subfamilies, with distinct conserved domains and functions. They also found that both genomic/segmental duplication and tandem duplication contributed to the expansion of the *GRAS* gene family in the rice and *Arabidopsis* genomes. The existence of *GRAS*-like genes in bryophytes suggests that GRAS is an ancient family of transcription factors, which arose before the appearance of land plants over 400 Myr ago (Tian et al. 2004).

S-phase kinase-associated protein 1 (*SKP1*) is a core component of SCF ubiquitin ligases and mediates protein degradation, thereby regulating eukaryotic fundamental processes. Despite a single *SKP1* gene in protists, fungi and some vertebrates, many animal and plant species possess multiple *SKP1* homologues. By comparing the entire sets of *SKP1* homologues of rice and *Arabidopsis* as well as fungi and animals, Kong et al. (2004) demonstrated that multiple *SKP1* homologues from the same species have evolved at highly heterogeneous rates and that the differences in evolutionary rate were so large that true phylogenies were not recoverable from the full dataset. They indicated that only when the original dataset was partitioned into sets of genes with slow, medium and rapid rates of evolution and analysed separately, better-resolved relationships were observed. They divided the *SKP1* genes into three categories. The first are the slowly evolving *SKP1* homologues that are relatively highly conserved in sequence and expressed widely and/or at high levels, suggesting that they have evolved under functional constraint and served the most fundamental function(s). The second category consists of the rapidly evolving members that are structurally more diverse and usually have limited expression patterns and higher
dN/dS values, suggesting that they might have evolved under relaxed or altered constraint, or even under positive selection. The final category includes some rapidly evolving members that might have lost their original function(s) and/or acquired new function(s) or become pseudogenes, as suggested by their expression patterns, dN/dS values, and amino acid changes at key positions (Kong et al. 2004).

In the rice genome, large-scale duplication events have been recently uncovered, but different interpretations (aneuploid or palaeopolyploid) were proposed regarding the extent of the duplications based on the draft sequence of the genome of the subspecies japonica (Paterson et al. 2005). Using independent dating approaches and the genome sequence data of the other subspecies of rice (indica), Wang et al. (2005a,b) detected 10 duplicated blocks on all 12 chromosomes and thus supported whole-genome duplication. In addition, they inferred that the whole-genome duplication occurred ca 70 Myr ago and a segmental duplication at ca 5 Myr ago, involving chromosomes 11 and 12. Yu et al. (2005) published an improved, near-complete genome analysis of two subspecies of rice and also reported that there was evidence in the rice DNA sequences for a whole-genome duplication ca 55–70 Myr ago.

Shi et al. (2003) identified 10 and 30 putative bitter taste receptor T2R genes from the draft human and mouse genome sequences, respectively. Their phylogenetic analysis of the T2R genes suggests that these genes can be classified into three main groups, i.e. A, B and C. They demonstrated that tandem gene duplication is the primary source of new T2Rs. For closely related paralogous genes, a significantly higher rate of synonymous nucleotide substitution than that of synonymous substitution was observed in the extra-cellular regions of T2Rs, which are presumably involved in tastant-binding. This suggests the role of positive selection in the diversification of newly duplicated T2R genes. Since many natural poisonous substances are bitter, they conjectured that the mammalian T2R genes are under diversifying selection for the ability to recognize a diverse array of poisons that the organisms may encounter in exploring new habitats and diets.

In mammals, the vomeronasal organ (VNO) of the olfactory system is a chemosensory organ specialized in the detection of pheromones, chemical signals that induce innate reproductive and social behaviours between the members of the same species. The pheromone receptors are encoded by two distinct and complex superfamilies, named V1R and V2R.

Shi et al. (2005) used computational methods to identify 95 and 62 new putative V1R genes from the draft rat and mouse genome sequence, respectively. The rat V1R repertoire consists of 11 subfamilies, 10 of which are shared with the mouse while the rat appears to lack the H and I subfamilies found in mouse and possesses one unique subfamily (M). Their analysis reveals that many subfamilies underwent an expansion very close to the split of mouse and rat. The non-synonymous and synonymous rate ratios for most of these clusters were much higher than one, suggesting the role of positive selection in the diversification of these duplicated V1R genes. They speculate that the V1R genes have evolved under positive Darwinian selection to maintain the ability to discriminate between large and complex pheromonal mixtures. Grus et al. (2005) describe the V1R repertoires of dog, cow and opossum, based on their draft genome sequences. Phylogenetic analysis of placental V1R genes suggests multiple losses of ancestral genes in carnivores and artiodactyls and gains of many new genes by gene duplication in rodents, manifesting massive gene births and deaths. Their results show a concordance between the V1R repertoire size and the complexity of VNO morphology, suggesting that the latter could indicate the sophistication of pheromone communications within species (Grus et al. 2005).

Based on a computational analysis of the mouse and rat genome sequences, Yang et al. (2005a,b) report the first global draft of the V2R vomeronasal receptor gene repertoire, composed of approximately 200 genes and pseudogenes. Rodent V2Rs are subject to rapid gene births/deaths and accelerated amino acid substitutions, probably reflecting the species-specific nature of pheromones. Vertebrate V2Rs appear to have originated twice prior to the emergence of the VNO in ancestral tetrapods.

Those studies demonstrate that the genomic approach is very powerful and necessary for understanding the evolution of genes and gene families.

We thank Professor Zhu Chen for valuable comments and Dr Yun Gao for editing the manuscript. This work was supported by the State Key Basic Research and Development Plan (2003CB415105), the Program for Key International S and T Cooperation Project of China (2001CB711103), the Natural Science Foundation of Yunnan Province and the National Natural Science Foundation of China.

REFERENCES


Phil. Trans. R. Soc. B (2007)