Species complex, species concepts and characterization of cryptic diversity: vignettes from Indian systems

K. Praveen Karanth

Species are one of the fundamental units of evolution and to this date, are largely described based on morphological characters. However, with the advent of molecular tools it has become apparent that many widely distributed species harbour cryptic diversity. Nevertheless, delimiting the members of such complexes is a challenge given these species exhibit very little or no morphological differences among them. In this regard, approaches that use multiple lines of evidence for species delimitation have gained ascendency. Here, I review a few recent studies from India, where application of an integrative approach has provided us with new insights into the systematics and evolution of these groups. Further, I discuss the implications of this approach on taxonomy in particular, and on ecological and evolutionary research in general.

Much of the biological diversity has been characterized and classified using morphological characters. Description of a new species involves finding diagnostic character(s) that separates it from other related entities. Furthermore, morphological characters are also used to place the new species in the broader classification (genus, family, order, etc.). This system of naming and classification of organisms, also referred to as Linnaean taxonomy, has been in place since the 1750s, and has contributed immensely in the description and characterization of the planet's biological wealth. In turn, taxonomy has been the bedrock on which reside other streams of biology. However, with the advent of molecular tools, morphology-based taxonomy has been undergoing a period of flux.

Molecular data have influenced taxonomy in two principal ways. First, molecular phylogenies have challenged the classification of many groups at multiple levels of the taxonomic hierarchy. Secondly, with the advent of molecular data, taxonomists have been compelled to revisit the definition of species and the tools used to delimit them. To illustrate these points, I briefly discuss the molecular studies undertaken on the langurs of the Indian subcontinent over the last two decades. Among Indian langurs, the Hanuman langur (*Semnopithecus entellus*), a widespread Indian primate was long considered a distinct taxon and assigned to the monotypic genus *Semnopithecus*, while the remaining langurs were placed in the genus *Trachypithecus* along with other Southeast Asian langurs. However, molecular studies suggest that Nilgiri and purple-faced langurs, which have restricted distributions in southwest India and Sri Lanka respectively, are more closely related to Hanuman langurs than to the rest of the langurs in the genus *Trachypithecus*[1,2] (Figure 1). Accordingly, these two langurs are currently placed in *Semnopithecus*, such that the taxonomy reflects their evolutionary relationships. Thus molecular data have helped us revise the classification of these langurs.

These molecular studies also highlight another problematic area in taxonomy, namely identification and diagnosis of closely related and morphologically similar species. For example, the widely distributed Hanuman langur was long considered as a single species. However, molecular studies suggest that this taxon constitutes a complex with multiple species[3,4]. In the mitochondrial tree shown in Figure 1, the so-called Hanuman langur is polyphyletic with respect to Nilgiri and purple-faced langurs. Thus even though the Hanuman langurs from across their range exhibit overall morphological similarity, they are not monophyletic in the molecular tree. In recent times numerous molecular studies have indicated that many widespread species consist of multiple species[5–11], henceforth referred to as species complex, and often these species exhibit very little or no morphological differences among them (cryptic species). Given that species are described largely based on morphological characters, delimiting members of such complexes is often challenging due to lack of morphological differences among them. This problem brings to fore another related issue of what constitutes a species, i.e. what is a species?

The species problem

Species are one of the principal units of evolution[12]. Therefore, it is disconcerting that there is much confusion regarding what constitutes a species[13]. This is apparent from the plethora of species concepts

Figure 1. Molecular systematics of Indian langurs. The tree shown here is based on mitochondrial cytochrome *b* gene (Karanth[40]), *Semnopithecus* and *Trachypithecus* are the two genera of langurs and leaf monkeys of Asia respectively (see text for details). NI, North India; SI, South India and SL, Sri Lanka.

that have been proposed[14], and the numerous reviews written on this topic in recent times[15]. The earliest and perhaps the most influential and widely accepted species concept is Ernst Mayr's biological species concept (BSC). The BSC states that 'species are groups of interbreeding natural populations that are reproductively isolated from other such groups'[16]. Interestingly, when new species are described, reproductive isolation is almost never tested. Implicit in the morphology-based species delimitation is the assumption that morphological divergence can be taken as a surrogate for reproductive isolation. This assumption is reasonable in cases where characters used to diagnose species are directly or indirectly associated with reproduction, such as genital morphology. However, for a vast majority of species the characters used to describe them have no relationship with reproductive isolation. Additionally, the use of morphological characters in species delimitation is difficult in the case of cryptic species discussed above. By far the major issue with BSC is the problem of hybridization between 'good species'[17]. For example, among the langurs of the Indian subcontinent, *Semnopithecus priam* and *Semnopithecus johnii* hybridize in areas where their ranges overlap[3]. These two species look very different, occupy different habitats and have long been accepted as distinct species ('good species'). Thus BSC has been problematic as there are always exceptions to reproductive isolation and in most cases reproductive isolation is difficult to test. Similarly, other species concepts also have limitations ranging from lack of universality to difficulty in implementing species delimiting criteria[13,14,18]. Nevertheless, in recent years, a semblance of consensus seems to be emerging with respect to species concept and operational criteria used to delimit them. In this regard, the general concept of species (GCS) proposed by Kevin de Queiroz[13,15,19] provides the conceptual framework for what constitutes a species, and integrative taxonomy provides the operational framework to delimit species[20].

According to GCS, species are 'segments of separately evolving metapopulation lineages'[19]. GCS attempts to unify all previous species concepts by suggesting that these concepts refer to the plethora of subprocesses that occur during the process of speciation. Importantly, all species concepts consider species as separately evolving lineages, as in GCS, but differ only on how these lineages are identified (delimited). Thus during speciation the daughter lineages undergo divergence along multiple axes such as morphology, behaviour, ecology, molecular, reproductive traits, etc. (Table 1 and Figure 2 *a*). Divergence along each of these axes pertains to variations of the different species concepts (Table 1). For example, if the two lineages occupy different niches (i.e. ecological divergence), then they are considered as separate species according to the ecological species concept. Similarly, if the lineages are reproductively isolated (divergence in reproductive traits), then they are considered as separate species according to BSC. Ideally a 'good species' pair would exhibit divergence along all these axes (Figure 2 *a*); however, according to de Queiroz[19], divergence along any one of these axes is evidence for the existence of two species. Additionally, divergence along multiple axes would increase our confidence in assigning them to separate species.

It must be noted that these subprocesses (divergence along multiple axes) that occur during speciation do not follow a particular order[13]. For some species pairs reproductive isolation might occur first followed by molecular and morphological divergence (species pair 1, Figure 2 *a*), while for others morphological divergence might occur first followed by divergence along other axes (species pair 2, Figure 2 *a*). How and when these subprocesses occur would depend on the selection regime experienced by the diverging lineages. An important implication of the above observation is that one cannot use the same set of species delimitation criteria for different species pairs[15]. Additionally, the number of axes along which species pairs exhibit divergence would depend on time since speciation, i.e. where in the speciation continuum the species pair lies. Younger pairs would show divergence along fewer axes, whereas older pairs would exhibit divergence along more number of axes.

Among some species pairs certain subprocesses might not have occurred as the

Table 1. Some of the subprocesses that occur during speciation (see refs 16, 19 for details)

| Subprocess	Delimitation criteria	Species concept
Divergence in reproductive traits (μ)	Reproductive isolation	Biological species concept
Molecular divergence (β)	Reciprocal monophyly	Phylogenetic species concept
Morphological divergence (π)	Diagnostic characters	Typological species concept
Behavioural divergence (Ω)	Mate recognition mechanism	Recognition species concept
Ecological divergence (α)	Occupy different niches	Ecological species concept
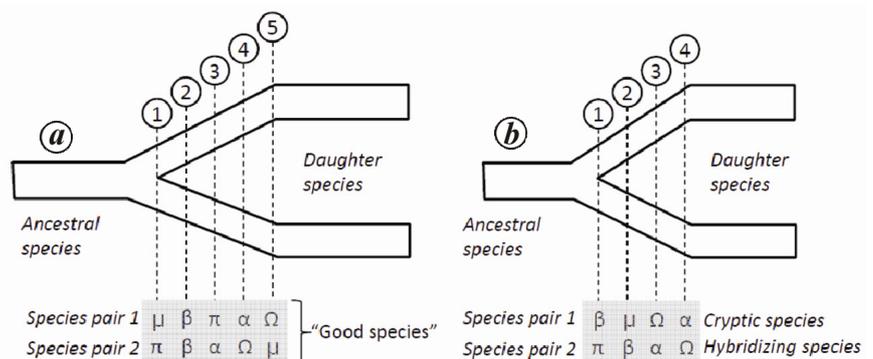
Figure 2. Various subprocesses that occur during speciation. Symbols represent the subprocesses: μ, Divergence in reproductive traits; β, Molecular divergence; π, Morphological divergence; Ω, Behavioural divergence; α, Ecological divergence (see text and Table 1 for details). *a*, Good species pairs that exhibit all subprocesses, however, the sequence in which these subprocesses occur is different for species pairs 1 and 2. *b*, Species pairs where one of the subprocesses has not occurred.

pairs may be in early stages of speciation[19], or due to lack of selection for divergence along that axis. For example, if there is no selective advantage for daughter species to diverge in morphology from the ancestral form, then the daughter species will look identical, a condition we would call cryptic species (species pair 1, Figure 2 *b*). Similarly, reproductive isolation as a character might not evolve if there is no selection for it; for example, in species that have evolved in allopatry. Such species might hybridize when their distributions overlap due to subsequent range expansion (species pair 2, Figure 2 *b*). Clearly use of a single, uniform criterion or method for species delimitation can be problematic[14]. This is because the chosen criterion might not be under selection for divergence between daughter species. Thus taxonomy needs to be pluralistic and incorporate multiple approaches for species delimitation[14,20–26]. These developments have led to the birth of 'integrative taxonomy', which is defined as 'the science that aims to delimit the units of life's diversity from multiple and complementary perspectives (phylogeography, comparative morphology, population genetics, ecology, development, behaviour, etc.)'[20]. Integrative taxonomy also provides the operational basis for GCS, as it espouses the use of multiple lines of evidence to delineate species boundaries[27].

Integrative taxonomy of Indian taxa

The Indian subcontinent is a vast landmass bounded by the Himalaya in the north and the Indian Ocean in the south. This landmass is topologically very heterogeneous with many mountain ranges and rivers. These physiographic features could serve as potential barriers to species dispersal and have been implicated in inter- and intraspecific diversification[28–31]. Additionally, there is much variation in climatic conditions (rainfall and temperature) across this landmass, which has in turn influenced local vegetation. Given this setting, it is conceivable that many widely distributed species might have undergone diversification across these varied climatic regimes and barriers. In the case of the Hanuman langurs, multiple lines of evidence (molecular, morphological and ecological data)

suggest that this widely distributed taxon consists of at least three distinct species[3,32,33]. These species occupy different ecological zones[33] and are also separated by river barriers[34]. Similarly, phylogenetic approaches in conjunction with morphological data suggest that the widely distributed gecko species, *Hemidactylus brookii* is a complex with at least five species[35,36]. This *brookii* radiation consists of species that have adapted to different substrata, in that some are commensal species while others are rock- and ground-dwelling. Thus, widely distributed taxa often tend to harbour cryptic diversity that is detected when multiple lines of evidence are used to delimit species. Interestingly, among certain invertebrates and small vertebrates, cryptic species are being detected at a smaller spatial scale. For example, the centipede species complex in the genus *Digitipes* is endemic to the Western Ghats, a mountain chain along the west coast of India[27]. Similarly, many species complexes have also being reported from frog groups that are confined to the Western Ghats[37–39].

Table 2 lists some of the species complexes that have been reported from India in recent years. In most of these cases, the species numbers have doubled when molecular data were used to characterize diversity in these groups. There are numerous widely distributed species

in India, many of which might be species complexes. Furthermore, very little is known about cryptic diversity among invertebrates. Taken together, these observations suggest that a large part of India's cryptic diversity is yet to be characterized. Thus, the current approximation of India's biodiversity is an underestimate of the true diversity even for well-studied groups. The identification and characterization of this cryptic diversity is perhaps one of the major challenges of taxonomy today and an important frontier in biodiversity research.

Characterizing cryptic diversity has important conservation implications. One obvious reason that was alluded to in the previous section is that knowledge of biodiversity of an area is an important first step in conservation. When many species remain undetected, we fail to fully appreciate the biotic wealth of that area. All subsequent conservation measures or decisions might be flawed as they are targeted towards a subset of species. Understanding cryptic diversity is also important for species-centric conservation measure. For example, Hanuman langur was accorded low conservation priority as it was considered as a widespread and common species[40]. However, some of the newly erected species in this complex have restricted distribution and habitat requirement[3]. Thus, populations

Table 2. Some species complexes reported from India

Species	TT	IT	Reference
Primates			
Semnopithecus entellus	1	>3	3
Reptiles			
Sitana	2	>5	48
Geckoella	7	14*	49
Hemidactylus brookii	1	5	35
Cyrtopodion aravallense	1	7*	50
Cyrtodactylus (I)	5	22*	51
Amphibians			
Raorchestes	43	52	39
Micrixalus	12	26	38
Hylarana (WG–SL)	7	14	52
Nyctibatrachus	15	27	37
Invertebrates			
Digitepes	3	6	27, 53
Itaropsis	1	>3	47
Birds			
Brachypteryx major	1	>2	30

TT, Traditional taxonomy; IT, Integrative taxonomy. *Based only on molecular data, *Cyrtodactylus* and *Hylarana* are also distributed outside the Indian subregion; the cited papers have looked at diversification in India (I) and the Western Ghats–Sri Lanka (WG–SL) hotspot.

of these species might now be threatened. Additionally, many widely distributed species are also used as a model system for comparative studies in ecology and evolution. Typically in these studies multiple populations from different parts of the species' range are compared to better understand the evolution of a certain trait. These studies assume that the same species is being evaluated under different ecological conditions. For example, Newton[41] reviewed infanticidal behaviour in Hanuman langur based on data collected from different populations of this species from across its range. However, the revised taxonomy suggests that the comparative framework used in this study is actually interspecific rather than intraspecific. Clearly in the light of the revised taxonomy, the results of comparative studies done on Hanuman langurs from across the range need to be reinterpreted[4]. Therefore, the taxonomic status of the 'widely distributed' species should be ascertained before they could be used in comparative studies.

Detection and characterization of cryptic diversity

How do we detect cryptic diversity? The logical first step is to assume that a widely distributed species might constitute a species complex (for reasons discussed above). Typically, molecular tools are used to determine spatial distribution of genetic variation in such a taxon and this information is in turn used to identify cryptic species. To this end, DNA samples are collected from across the range of the taxon (Figure 3 a). These samples are used to generate phylogenetic trees based on rapidly evolving nuclear and mitochondrial markers (or chloroplast markers in case of plants). If the phylogenetic analysis reveals multiple clades that are confined to different geographical areas (geographical clades), then one possible conclusion is that the chosen taxon might be a species complex (Figure 3 b). These geographical clades are considered as putative species and additional lines of evidence are then used to test if these putative species have diverged along other axes. For example, after identifying putative species based on molecular data, morphology can be revisited to identify new characters or combinations of previously described
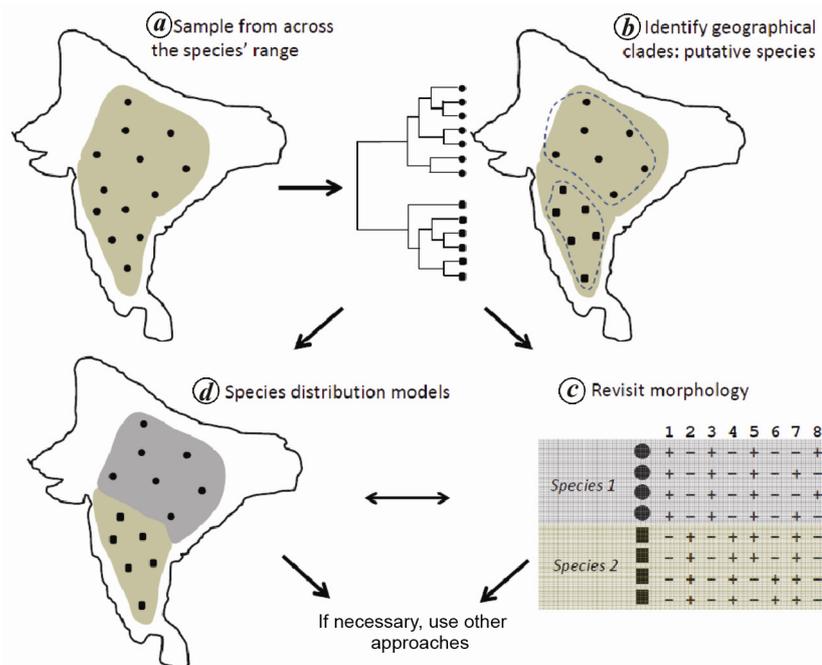

Figure 3. Identification and characterization of cryptic diversity (see text for details). *a*, Shaded area represents distribution of a species suspected to be a complex and dots indicate sampling location. *b*, Squares and circles represent the two putative species based on molecular data, while areas enclosed by dashed line show the distributions of the two geographical clades. *c*, Numbers 1–8 are morphological characters typed in each putative species; + and − indicate presence and absence of a character respectively. *d*, Shaded areas are the predicted distributions of the two putative species; these areas do not overlap, suggesting significant divergence in the ecological axis.

characters to diagnose members of each clade[35] (Figure 3 c). In the absence of discernable morphological differences between putative species, as in the case of cryptic species, other non-morphology-based approaches are explored to test the multi-species hypothesis. For example, species distribution models can be used to determine if these putative species exhibit significant divergence along the ecological axis[27] (Figure 3 d). In turn, cryptic species could be diagnosed and described based on characters other than morphology, such as behavioural, acoustic and molecular data.

However, the presence of multiple geographical clades does not always imply existence of multiple species. Such pattern can also be generated by population-level processes within species. Thus, it is important to distinguish population-level differentiation from species-level diversification[42]. In recent years, many coalescent-based species delimitation models have been proposed to address this issue[43–45]. Nevertheless, these methods are still in their infancy and often tend to overestimate true diversity (see Hedin[46], and the references therein).

Molecular data need not be the primary source for characterization of cryptic diversity. Any other kind of data (acoustic, behavioural, ecological, chemical) can be potentially used to identify putative species. For example, in case of crickets, a widely distributed species might exhibit different call characteristics across its range. Each call type could be considered as a putative species whose species identity can be further tested using other approaches (see Jaiswara et al.[47] for details).

Conclusion

It is apparent from this review and other studies[47] that there is much cryptic diversity in nature. It goes unnoticed because we classify entities based on characters we can perceive, i.e. morphological characters, but ignore characters that we cannot see, smell, hear or feel. However, evolution is occurring along these multiple axes, and a member of a species is able to distinguish conspecifics from heterospecifics based on a combination of cues. We humans are unable to detect

COMMENTARY

most of these cues due to our limited sensory ability. Thus, it is imperative that we use characters other than those based on 'typical' external morphology to distinguish cryptic species. Cryptic diversity needs to be characterized through the use of multidimensional integrative approach rather than methods that are unidimensional or rely on only one kind of data. To this end there is a need for change in the mindset of taxonomists practicing traditional morphology-based taxonomy[20]. Additionally, taxonomy journals should be open to the idea of using non-morphological characters for species descriptions.

1. Karanth, P. K., Singh, L., Collura, R. and Stewart, C.-B., *Mol. Phylogenet. Evol.*, 2008, **46**, 683–694.
2. Osterholz, M., Walter, L. and Roos, C., *BMC Evol. Biol.*, 2008, **8**, 58.
3. Ashalakshmi, C. N., Nag, C. and Karanth, K. P., *Conserv. Gene.*, 2015, **16**, 43–58.
4. Karanth, P. K., Singh, L. and Stewart, C.-B., *Mol. Phylogenet. Evol.*, 2010, **54**, 627–663.
5. Zamudio, K. R., Jones, K. B. and Ward, R. H., *Syst. Biol.*, 1997, **46**, 284–305.
6. Martin, A. P. and Bermingham, E., *Proc. R. Soc. London, Ser. B*, 2000, **267**, 1135–1141.
7. Wynn, A. and Heyer, W. R., *Trop. Zool.*, 2001, **14**, 255–285.
8. Leaché, A. and Reeder, T. D., *Syst. Biol.*, 2002, **51**, 44–68.
9. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. and Hallwachs, W., *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14812–14817.
10. Lohman, D. J. *et al.*, *Biol. Conserv.*, 2010, **143**, 1885–1890.
11. Sousa-Neves, T., Aleixo, A. and Sequeira, F., *Mol. Phylogenet. Evol.*, 2013, **68**, 410–424.
12. Mayr, E., *Philos. Sci.*, 1996, **63**, 262–277.
13. de Queiroz, K., In *Endless Forms: Species and Speciation* (eds Howard, D. J. and Berlocher, S. H.), Oxford University Press, New York, 1998, pp. 57–75.
14. Sites, J. W. and Marshall, J. C., *Annu. Rev. Ecol. Evol. Syst.*, 2004, **35**, 199–227.
15. de Queiroz, K., *Proc. Natl. Acad. Sci. USA (Suppl. 1)*, 2005, **102**, 6600–6607.
16. Mayr, E., *Animal Species and Evolution*, Harvard University Press, Cambridge, MA, USA, 1963.
17. Zink, R. M. and McKitrick, M. C., *Auk*, 1995, **112**, 701–719.
18. Saikia, U., Sharma, N. and Das, A., *Resonance*, 2008, 1049–1064.
19. de Queiroz, K., *Syst. Biol.*, 2007, **56**, 879–886.
20. Dayrat, B., *Biol. J. Linn. Soc.*, 2005, **85**, 407–415.
21. Balakrishnan, R., *Syst. Biol.*, 2005, **54**, 689–693.
22. Will, K. W., Mishler, B. D. and Wheeler, Q. D., *Syst. Biol.*, 2005, **54**, 844–851.
23. Yoder, A. D. *et al.*, *Proc. Natl. Acad. Sci. USA (Suppl. 1)*, 2005, **102**, 6587–6594.
24. Bond, J. E. and Stockman, A. K., *Syst. Biol.*, 2008, **57**, 628–646.
25. Padial, J. M., Miralles, A., De la Riva, I. and Vences, M., *Front. Zool.*, 2010, **7**, 16.
26. Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E. and Crozier, R. H., *Annu. Rev. Entomol.*, 2010, **55**, 421–38.
27. Joshi, J. and Karanth, K. P., *PLoS ONE*, 2012, **7**(8), e42225; doi:10.1371.
28. Ripley, S. D. and Beehler, B. M., *J. Biogeogr.*, 1990, **17**, 639–648.
29. Vidya, T., Fernando, P., Melnick, D. and Sukumar, R., *Heredity*, 2005, **94**, 71–80.
30. Robin, V. V., Sinha, A. and Ramakrishnan, U., *PLoS ONE*, 2010, **5**(10), e13321; doi:10.1371.
31. Robin, V. V., Vishnudas, C. K., Gupta, P. and Ramakrishnan, U., *Proc. R. Soc. London, Ser. B*, 2015, **282**, 20150861.
32. Nag, C., Karanth, P. K. and Vasudeva, G. K., *PLoS ONE*, 2014, **9**(2), e87804; doi:10.1371.
33. Nag, C., Pramod, P. and Karanth, K. P., *Int. J. Primatol.*, 2011, **32**, 830–848.
34. Roonwal, M. L., In *Current Primate Research* (eds Roonwal, M. L., Mohnot, S. M. and Rathore, N. S.), Jodhpur University Press, Jodhpur, 1984, pp. 93–151.
35. Lajmi, A., Giri, V. B. and Karanth, K. P., *Organisms Diver. Evol.*; doi:10.1007/s13127-016-0271-9,
36. Karanth, P. K., *Curr. Sci.*, 2015, **108**, 1847–1851.
37. Biju, S. D. *et al.*, *Zootaxa*, 2011, **3029**, 1–96.
38. Biju, S. D., Garg, S., Gururaja, K. V., Shouche, Y. and Walujkar, S. A., *Ceylon J. Sci. (Biol. Sci.)*, 2014, **43**, 37–123.
39. Vijayakumar, S. P., Dinesh, K. P., Prabhu, M. V. and Shanker, K., *Zootaxa*, 2014, **3893**, 451–488.
40. Karanth, P. K., *J. Gene.*, 2010, **89**, 393–399.
41. Newton, P., *Int. J. Primatol.*, 1988, **9**, 59–77.
42. Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H. and Tsai, P., *Philos. Trans. R. Soc. London, Ser. B*, 2008, **363**, 3893–3902.
43. Leaché, A. D. and Fujita, M. K., *Proc. R. Soc. London, Ser. B*, 2010, **277**, 3071–3077.
44. Zhang, J., Kapli, P., Pavlidis, P. and Stamatakis, A., *Bioinformatics*, 2013, **29**, 2869–2876.
45. Fujisawa, T. and Barraclough, T., *Syst. Biol.*, 2013, **62**, 707–724.
46. Hedin, M., *Mol. Ecol.*, 2015, **24**, 346–361.
47. Jaiswara, R., Balakrishnan, R., Robillard, T., Rao, K., Cruaud, C. and Desutter-Grandcolas, L., *Zool. J. Linn. Soc.*, 2012, **164**, 285–303.
48. Deepak, V. *et al.*, *Contrib. Zool.*, **85**(1), 67–111.
49. Agarwal, I. and Karanth, P. K., *Mol. Phylogenet. Evol.*, 2015, **82**, 193–199
50. Agarwal, I., Bauer, A. M., Jackman, T. R. and Karanth, P. K., *Zool. Scripta*, 2014, **43**, 455–471.
51. Agarwal, I., Bauer, A. M., Jackman, T. R. and Karanth, P. K., *Mol. Phylogenet. Evol.*, 2014, **80**, 145–155.
52. Biju, S. D., Garg, S., Mahony, S., Wijayathilaka, N., Senevirathne, G. and Meegaskumbura, M., *Contrib. Zool.*, 2014, **83**, 4.
53. Joshi, J. and Edgecombe, G. D., *Zootaxa*, 2013, **3626**, 099–145.
54. Bickford, D. *et al.*, *Trends Ecol. Evol.*, 2006, **22**, 148–156.

K. Praveen Karanth is in the Centre for Ecological Sciences, Indian Institute of Science, Bengaluru 560 012, India. e-mail: karanth@ces.iisc.ernet.in