Spider Research in the 21st Century

trends & perspectives

Edited by David Penney

Foreword by Norman I. Platnick

Sirri Scientific Press
Spider Research in the 21st Century

trends & perspectives

Edited by David Penney
Faculty of Life Sciences
The University of Manchester, UK

Foreword by Norman I. Platnick
American Museum of Natural History, New York, USA
Systematics

Progress in the study of spider diversity and evolution

INGI AGNARSSON, JONATHAN A. CODDINGTON, MATJAŽ KUNTER

Introduction

The field of systematics involves at least three major elements: biodiversity exploration (inventory); taxonomic discovery and description (taxonomy); and the estimation of phylogenetic relationships among these species (phylogeny). All elements experienced significant progress in the 20th century and progress continues in the early 21st. Here we review spider systematics in this broad sense encompassing these three fields.

At the beginning of the last century, biodiversity inventories continued as simple taxonomic checklists but were increasingly challenged by the growing, powerfully explanatory field of ecology. These influences led to improved study design (Coddington et al., 1991; Basset et al., 1997; Longino et al., 2002; Sorensen et al., 2002; Jimenez-Valverde & Lobo, 2006; Cardoso et al., 2008; Cardoso, 2009), as well as progress in empirical and analytical methodology such as estimation of species richness from sample data (Colwell & Coddington, 1994; Novotný & Basset, 2000; Gotelli & Colwell, 2003; Coddington et al., 2009; Colwell, 2011; Colwell et al., 2012). Classical exploratory collecting also transformed arachnology by discovering taxa, especially from austral and tropical regions, that did not fit easily into the schemas of 19th century taxonomists (Hickman, 1931; Forster, 1949, 1955; Lehtinen, 1967).

Taxonomic practice developed more slowly. Although current work in many aspects resembles that of a century ago, it does show signs of adopting novel tools. The use of molecular data for species identification (Hebert et al., 2003), and for circumscription and diagnosis (e.g. Macias-Hernandez et al., 2010; Hamilton et al., 2011; Hedin & Carlson, 2011; Satler et al., 2011; Keith & Hedin, 2012; Richardson & Gunter, 2012) are notable. As are improved methods of imaging and referencing images (Ramirez et al., 2007), computer

databasing and georeferencing of specimens (e.g. Goblin Spider PBI, http://research.amnh.org/oonopidae/), cyber-informatics (Miller et al., 2009, 2012a; Penev et al., 2009, 2010), collection digitization and linking of distribution data to public databases such as the Global Biodiversity Information Facility (GBIF http://www.gbif.org/). Another recent progress is the development of less subjective criteria for species delimitation (Bond & Stockman, 2008). Further, as noted by Coddington & Levi (1991), the invention of the humble ‘loan’ should not be underestimated. Historically, taxonomic quality was proportional to the number of major museums visited. Now museums routinely risk (spider) holotypes to any qualified researcher via the global postal system. The result has been a worldwide explosion in the quality of revisionary work, well worth the risk of occasional specimen loss.

Of the three elements mentioned, phylogenetics transformed the most in the 20th century. It evolved from effectively Aristotelian, to authoritarian, then numerical, and eventually to cladistic argumentation (Hennig, 1966). It expanded from an exclusively morphological domain to one embracing all evidence of heritable variation, currently dominated especially by DNA sequence data from independent genes, ‘loci’, or markers. In the near future, focus will increasingly be on a wide range of genomic data and genomic architecture, but that revolution in (spider) systematics is just beginning. Concurrent with these empirical advances, quantitative systematic methodology began as a branch of multivariate statistics, and soon surrendered to the philosophical first principle of grouping by synapomorphy (Farris, 1983). Now, use of more explicitly model-based methods dominates, in part because the mathematics of only four nucleotides (or tri-nucleotide codes for amino acids) is tractable, and also because such methods facilitate use of phylogenies to statistically test hypotheses. Powerful as this reductionist view of ‘genes’ may seem, it will change dramatically in the years ahead, and potentially away from the notion that aligning DNA sequences, *per se*, is the heart of the phylogenetic problem.

Systematics of the diverse group of spiders (Figure 1) followed this general pattern. Early on spider systematics relied mostly on binary – and idiosyncratic – evaluations of morphology, failed to distinguish apomorphy from plesiomorphy, and therefore proposed classifications in which about half of all groups were not natural. The 20th century transformation only began in earnest in the 1970s, and occurred in many steps.

The first major step was incorporating cladistic theory – relying only on shared derived homology, synapomorphy – in the establishment of a classification system. Early studies using phylogenetic logic by Platnick and colleagues (Platnick & Gertsch, 1976; Platnick, 1977) led to quantitative tests of spider phylogeny that rejected many classical spider groups founded on symplesiomorphy or convergence and thus began to abandon para- and polyphyletic classifications.
The incorporation of web-building behavioural data in phylogenetic analyses was a second and highly productive step, in spider phylogenetics (Eberhard, 1982a, 1987, 1990ac; Coddington, 1986ac). The construction of spider webs represents one of the most elegant and intricate examples of complex animal behaviour, yet one that is highly stereotypic. William Eberhard (Eberhard, 1982a) was the first to demonstrate clearly the power of web-building behavioural data in spider phylogenetics. Orb building behaviour offered solid support for many higher level clades, many of which remain as strongholds of phylogenetic knowledge of spiders, such as Araneoidea and Araneidae. Subsequently, Coddington (1986ab) rejuvenated and articulated in detail a provocative hypothesis based on web-building behaviour: that the orb-web arose only once in the group Orbiculariae. This claim was controversial because the types of silk used, and the organs involved in the production of the ‘sticky spiral’ of orb webs in the ‘classical orb-weavers’ (Araneoidea) and the cribellate orb-weavers (Deinopoidea) differed so dramatically that contemporary arachnologists generally agreed that the two web architectures were convergent. However, behaviour strongly supported the ‘single origin hypothesis’
and since then support has been increasing for Orbicularia from various data sources (Blackledge et al., 2009a, 2011; Dimitrov et al., 2012; Agnarsson et al., 2013). Nevertheless, no molecular analysis has robustly supported Orbicularia monophyly, either receiving very weak support from molecules alone (Dimitrov et al., 2012), or supported in a total evidence analysis excluding the orb behavioural data (Blackledge et al., 2009a). The hypothesis, therefore, remains controversial. The obvious next step, given current trends, will be testing Orbicularia monophyly via analysis of not one or a few genes, but hundreds or thousands of loci obtained from transcriptomes.

The third major step arose from the previous; the importance of silk manipulation in the lives of spiders suggested that spinneret morphology would reflect evolutionary history. Coddington (1989) used SEM microscopy to map the external morphology of spinnerets and spigots to histological work (reviewed in Kovoor, 1987). This rich character system has been influential in spider systematics ever since.

The fourth and last major step in 20th century spider systematics was the integration of DNA evidence, as in phylogenetics in general. Spider systematists were rather slow in adopting molecular evidence, starting with a few studies in the 1990s (Croom et al., 1991; Huber et al., 1993; Gillespie et al., 1994; Hedin, 1997; Zehethofer & Sturmbauer, 1998; Tan et al., 1999), but then taking off at the turn of the millennium. Molecular data continue to offer an independent test of phylogenies based on morphology and behaviour, and are refining our understanding of spider systematics.

Nevertheless, molecular phylogenetics in spiders has not yet lived up to its promise. Early on, primers were developed for a few promising markers, but since then few new markers applicable to all spiders have emerged. Nearly all molecular phylogenetics in spiders relies on relatively few markers most notably Cytochrome oxidase 1 (COI), Nitrogen dehydrogenase 1 (ND1), Histone 3 (H3) and the ribosomal genes 16S, 18S, and 28S, most of which do not consistently or clearly resolve old, deep, or ‘higher level’ nodes (see below). In this respect spider systematics lags behind some other groups of organisms, including certain arthropod groups, such as Lepidoptera, Drosophila, and ants. However, a few recent papers and conference presentations suggest new developments and use of novel loci in many laboratories. Thus, we are likely to see rapid and vast expansion in the scope of molecular data used for spider systematics.

Furthermore, molecular technology is currently advancing in huge leaps with next generation sequencing (NGS). Using techniques of transcriptomics it is now possible to amplify mRNA directly from tissue to construct cDNA libraries, and in turn, use those to cheaply employ random sequencing strategies that may result in many thousands of phylogenetically informative genetic markers (Mattila et al., 2012b). Such approaches, commonly referred
to as phylogenomics, will vastly enhance phylogenetic understanding and will clearly dominate the next phase of spider phylogenetics. The drawback is that cDNA libraries require tissue preserved especially for RNA extraction, either specimens preserved in appropriate solutions (e.g. RINalater), live specimens, or specimens frozen at extreme temperatures. Most existing collections, stored at room temperature in aqueous solutions of ethanol, cannot be used for mRNA-based methods. Instead, phylogenetically key taxa will have to be collected again using specialized preservation protocols to provide whole genome transcriptomes.

The hope is that whole-body transcriptomes from relatively few but carefully chosen lineages will provide primers for new, single-copy markers that can be amplified from most, if not all spider lineages. Collections of high quality DNA are the best source of material for such studies, but are only beginning to be assembled. Gene representation in transcriptomes depends on what mRNA is actually present. Ideally, balanced sampling requires mRNA from major organ systems such as muscles, nerves, gonads, and digestive, venom, and silk glands, both sexes, and several life stages. Over the next decade spider phylogenies will no doubt be based on dozens or hundreds of genes rather than six or less, and we expect that within the next decade, phylogenomic studies will become the norm, especially for relatively taxon poor studies.

Next generation sequencing (NGS) technologies such as Illumina or 454 sequencing can use amplification of whole genome DNA (not mRNA), and typically produce millions of short sequences (several hundred nucleotides) but ‘cover’ all parts of the genome multiple times. Assembling such short sequences correctly into accurate genes and gene orders (chromosomes), however, is difficult. Repetitive DNA sequences in exons, introns and intergenic sequences pose particular problems. Transcriptomes avoid such problems because cell machinery edits out non-expressed sequences, leaving essentially functional gene sequences. Given the difficulty of de novo genome assembly and annotation, it is quite likely that the next phase of phylogenetic inference in spiders will use genes detected via transcriptomes.

Molecular geneticists believe that sequencing whole genomes will soon be relatively common. Assuming appropriate reference genomes, phylogenomic studies based on whole genomes may soon appear for spiders, as they have for some model organisms (Drosophila, yeasts, mammals). Mere quantity of sequence data, therefore, may soon no longer be a problem. Instead, the challenges of the 21st century will be different, including creating and fine-tuning pipelines to process and analyze massive datasets.

It is further possible that old, rapid radiations in spiders may not yield readily to sequence data if informative point mutations are few, and have been overwritten by subsequent changes. In this regard, it is interesting that
in the phylogenomic study of Regier et al. (2010), robust support was found for clades throughout Arthropoda, with the notable exception of relationships within Chelicerata and Arachnida, most of which were poorly supported. This suggests that ancient splits in arachnid phylogeny, potentially including basal nodes within spiders, may require exceptional data to resolve unambiguously. Nevertheless, the clear, immediate challenge in spider systematics is to apply NGS technology widely to identify robustly supported phylogenies.

A persistent problem in phylogeny is taxon sampling, a problem that will grow as fresh/live specimens are needed for transcriptomic approaches. The global biodiversity crisis poses additional challenges in this respect as taxa are disappearing before being sampled. With well over 43,000 species of spider described to date (Platnick, 2013), all experts agree that we have not yet discovered even half of extant spider species (see below). With habitat loss and global change, the challenge will be to discover and describe species before they disappear. Further, most of the spider species described to date are known only from a few specimens preserved for morphology rather than molecules. In some cases the types have been destroyed or lost. Hence, solid advances in spider phylogeny require the acquisition of fresh specimens of thousands of species, before they go extinct.

Local and global dimensions of spider biodiversity

As in other poorly known and hyperdiverse groups, the most significant factor limiting growth in systematic knowledge of spiders is the number of araneologists discovering and studying them (Coddington et al., 1990; Platnick, 1991, 1999; Cassis et al., 2007; Foord et al., 2011).

Unlike some other animal groups (e.g. geophilomorph centipedes), spiders are not intrinsically difficult to discover or study. For understudied groups, it can be important to distinguish discovery – the first time a taxon is found in nature, from study – the systematic and phylogenetic placement of the taxon. Generally speaking, systematists use material already in museums, supplemented by what they personally collect. Material in museums depends, in turn, on the intensity and scope of global biodiversity exploration – the extent to which all species are represented in museum collections. For example, estimates of global spider species richness based on the ratio of new to previously known species in revisions, will be low if most species are ‘still out there.’ New to known species ratios will seem plausible to most araneologists because museums do contain an immense backlog of undescribed spider species, and thus, although based only on museum holdings, yield impressive global estimates. In revisions of tropical or austral taxa, about half of the species are new (Coddington & Levi, 1991), thus leading to estimates that about half of all species are known (Platnick, 1991).
All such estimates are fundamentally ‘back of the envelope’ educated guesses that depend strongly on which facts are used, and what assumptions are made, even though several use sophisticated statistics. Many appear in the context of arguing for particular science policy goals, such as the feasibility of global inventories (Platnick, 1991, 1999), greater than realized progress in taxonomy (Costello et al., 2011), the limited scope of global biodiversity (Stork, 1988) or less gloomy predictions of future extinction.

Figure 2 depicts the history of discovery over the last 210 years of currently recognized species in well-known (birds, mammals), moderately known (bees) and poorly known groups (spiders). Although the richness of all of these groups will likely increase as molecular data uncover ‘cryptic’ species among previously described forms, the time period covered was dominated by morphological species criteria, and so is probably not biased across groups. Bird richness has been studied more and seems clearly asymptotic; few birds remain to be discovered. Mammals and bees are less well known, their discovery seems less sigmoidal; the discovery of spider species has been accelerating for the last 60 years. For spiders, this acceleration is due both to discovery of new forms in nature and an increase in the productivity of active systematic araneologists. These graphs show that species discovery does eventually asymptote, even if only for relatively species-poor, intensively studied groups.

![Graph showing species discovery history](image)

**Figure 2.** History of species discovery in birds, mammals, bees, and spiders since Linnaeus’ 10th version of *Systema Naturae* (Linné, 1758). Numbers in parentheses indicate the approximate number of species currently recognized for each group.
If discovery of spider species shows no sign of slowing down, what about more inclusive groups, such as families? (By ‘family’ we mean a ‘large’ monophyletic lineage, presumed heuristically to correlate strongly with the Linnean rank of family.) After all, only about 8000–9000 families of life on Earth exist (Parker, 1982; Sayers et al., 2009; http://www.catalogueoflife.org/, 2012), and the de novo discovery in nature of new families of life (as distinct from nomenclatural acts driven by study, e.g. taxonomy or phylogenetic research) is generally less than ten per year, at least in eukaryotes. Figure 3 plots the earliest discovery (year of description) of a species placed in one of the currently recognized spider families. Only three families of extant spiders have literally been discovered since 1950: Gradungulidae (Gradungula sorenseini, Forster, 1955), Chummidae (Chumma inquieta, Jocqué, 2001), and most recently Trogloraptoridae (Trogloraptor marchingtoni, Griswold, Audisio & Ledford, 2012). In contrast, one of the most recently recognized families, Penestomidae (Miller et al., 2010b), was first ‘discovered’ (though not delimited or named) in 1902 (Penestomus planus Simon, 1902 (Simon, 1902), and the family Euctenizidae was first elevated to that rank in 2012 (Bond et al., 2012b), but existed as a subfamily-level clade before that (Raven, 1985). It seems that the discovery of spider families is about as asymptotic as the discovery of bird species. Even if the number of recognized spider families increases (as it surely will), the dates of discovery of that diversity remain constant. If the asymptote at this higher cladistic level holds true, therefore, we have some justification to claim that discovery of higher spider lineage diversity is nearly complete.

![Spider family discovery 1740–2010](image)

**Figure 3.** The earliest discovery (year of description) of spider species placed in one of the currently recognized spider families (110 at the time of the analysis).
Patterns at the generic level are also interesting. Roughly one third of spider genera are monotypic, and serve no grouping function. Although most of these are linyphiids and salticids, in which species can be so distinctive that genera are obscure, a significant fraction were described so inadequately (~250 before 1900) that only examination of the types themselves will show if generic status is justified and what the name means. All such names functionally are *nomina dubia*. About ten years ago one of us surveyed the databases – essentially shelf lists – available at that time of major spider collections and discovered that only about 1500–2000 of the then roughly 3500 generic names are used in daily practice by curators (Coddington, unpublished data). The reason is that the remainder were so poorly described and illustrated (if at all) that the name was unusable at ‘birth’. When examined, too many of the roughly half of all pragmatically unused names of spider genera will turn out to be senior synonyms of younger, better described clades. Current rules of nomenclature will consequently impose significant ‘churn’ and instability in global naming systems, because the International Commission on Zoological Nomenclature (ICZN) privileges old taxonomy, even bad taxonomy, over newer, better taxonomy. The burden of dealing with old, inadequate taxonomy is a significant drag on progress.

Patterns in spider diversity at local levels are quite different. First of all, species richness (alpha diversity) is clearly highest in the wet tropics, and relatively sparser towards the poles (Figure 4) (Silva, 1996; Silva & Coddington, 1996; Höfer & Brescovit, 2001; Scharff *et al*., 2003; Floren & Deeleman-Reinhold, 2005; Raizer *et al*., 2005; Finch *et al*., 2008). The overall local pattern in the neotropics, within and between sites, is consistent and fairly clear (Nogueira *et al*., 2006; Ricetti & Bonaldo, 2008). The scant data from the old world tropics are similar (Russell Smith & Stork, 1994, 1995; Sorensen *et al*., 2002; Sorensen, 2004; Floren & Deeleman-Reinhold, 2005).

On the other hand, higher taxon richness, e.g. families, and above, is not primarily tropical but rather austral, probably due to fundamental patterns in earth history (Platnick, 1991). Species ranges in north temperate regions are large (e.g. holarctic taxa), whereas tropical ranges, especially on altitudinal gradients, can be tiny. If species from three tropical sites from Figure 4 are compared along an altitudinal gradient (El Trapiche at 100 m, Rio Tigre at 500 m, and Cerro Uchumachi at 1900 m, Figure 5), observed faunal overlap never exceeds 3%. Between areas at roughly the same elevation (Pakita and Tambopata, Peru, roughly 100 m), the overlap is 5–20% depending on the family (Figure 5). Moreover, for each site, many more species were predicted to be present than were actually observed (Figure 4).

Estimating species richness and overlap among habitats and regions becomes exponentially difficult to do as more comparisons are made. The Peruvian work involved checking 1200 morphospecies across two sites
Figure 4. Observed and estimated spider diversity in one hectare plots at 11 tropical and one temperate site in the Americas and Africa.

Figure 5. A glimpse into beta diversity comparing observed and estimated (Chao1) species richness of 1ha plots in Peru and Bolivia. Observed species overlap between two sites in Peru at similar altitudes ranged from 5–20%, depending on family, and that among three sites at different altitudes in Bolivia ranged from 0.8–2.8%.
for ‘synonymies,’ and the Bolivian work almost 700 morphospecies across three sites (Coddington & Silva, unpublished data). Most species could not be named, either because they were new to science or because existing descriptions of known species were inadequate to identify specimens. Given that species are best described in the context of revisions, formally naming and describing rich tropical faunas as a byproduct of biodiversity research is effectively impossible. On the other hand, local inventories can publish images of the habitus and some aspects of diagnostic sexual morphology of observed ‘morphospecies’ on websites. Perhaps these semi-diagnosed informal taxa will help to accelerate alpha taxonomy (e.g. http://aracnologia.macn.gov.ar/ThaiPlot/Taxonomy/index.htm).

Local richness in reasonably intact moist tropical ecosystems is rarely less than several hundred observed species per hectare and estimates commonly approach 1000 depending on the estimator used (Figure 4, Coddington et al., 2009). In Guyana, a more or less average closed canopy tropical moist forest, per hectare abundances were estimated to approach 1–2 million, with the average size of an adult spider 2.8 mm (Coddington et al., 2009). The faunas of adjacent regions, especially if at different elevations, seem to have relatively few species in common. Overall, the local pattern, viewed through the lens of biodiversity research, implies phenomenal global richness in spider species.

Revisionary work, however, suggests fewer species. Casual examination of revisions that attempt to include all known specimens of species treated do not, overwhelmingly, feature species known from single sites or even single specimens. Thus a paradox emerges. The view from the museum suggests that species, if not widely distributed, are at least not overwhelmingly narrow endemics, but the view from the field suggests extremely high alpha and beta diversity that is very challenging to sample adequately, thus most of it remaining undersampled.

What explains the paradox? It is unlikely that spider taxonomists, who generally do not hesitate to describe species from single individuals, are self-censoring their work. One potential explanation clearly is undersampling (Coddington et al., 2009) (Figure 6). Simply put, traditional intensities of collecting are everywhere inadequate, but especially so where spider diversity is greatest. Widespread quantification of undersampling bias is a relatively recent phenomenon. The discipline best suited to its study was statistical ecology, but after a few early attempts (Fisher et al., 1943; Preston, 1948), the attention of the field in the latter half of the 20th century turned towards characterizing the relative abundance distribution as inferred from a small set of quantitative sampling methods chosen for statistical tractability, not accurately estimating richness (Krebs, 1999). At the same time, growing recognition of the extinction crisis and the need for quantitatively justified plans to save as much biodiversity as possible (May, 1988; Vane-Wright et al., 1991), made knowing
Figure 6. Sampling intensity versus % singletons in tropical biodiversity inventories of spiders, data from Coddington et al. (2009). Singletons are clearly inflated when inventories are underresourced, resulting in undersampling. Scales are logarithmic.

how much biodiversity existed and where it occurred a high priority. Colwell & Coddington (1994) reviewed the set of species richness estimators available at that time, and Colwell (2011) implemented them in the enormously influential statistical package EstimateS. For the first time, biodiversity research had an easy, intelligible way to estimate what it was missing. The implications of these techniques are still unfolding. At local levels, richness estimators inform sampling design and, in general, rationally justify enormously larger collecting efforts (Cardoso et al., 2004, 2008; Cardoso, 2009). Because richness estimators depend mostly on the frequency of ‘rare’ species, and because local natural history as well as population biology rejects the image of a viable species with all individuals so dispersed that breeding is unlikely, the two fields agree that ‘singletons’ are anomalous (Coddington et al., 2009). Exactly what explains inordinately high singleton frequency in inventories is hotly debated (Novotný & Basset, 2000; Bruno, 2002; Hubbell, 2005; Coddington et al., 2009).

As generalist predators, spider patterns will probably not be explained by patterns of prey, or prey host plants. These arguments, theoretical and empirical, are directly relevant to discussions or estimates of global spider species richness.
How many spiders are there?

Another way to estimate total species richness is to seek expert opinion on diverse subclades within a group. We requested opinions from taxonomic experts on the 15 most diverse spider families, together accounting for about 70% of total known spider diversity. Estimates ranged from expected being roughly equal to ‘known’ diversity (Theraphosidae), to estimates suggesting we know little more than 10% of diversity (Sparassidae). On average, the ratio of expected to known species according with expert opinion was 3.18, thus suggesting that roughly 2/3 of the diversity of the most diverse spider groups awaits discovery. Hamilton et al. (2010), similarly concluded that around 70% of arthropods are as yet undescribed. Note, however, the portion of ‘known’ species richness that is invalid (species described under more than one name – synonyms, and species described wrongly – invalid names), may in some cases be well above the average 10%, which would inflate the estimated diversity. This may especially hold for those families with common large animals, such as theraphosids, nephilids and certain araneids. The exact percentage may only be detected through comprehensive taxonomic revisions. To extrapolate this result to all spiders (Table 1), we used the following formula:

\[
Es = (Ks-S) \times Us
\]

where \( Es \) = estimated total species richness, \( Ks \) = known species, \( S \) = estimated rate of synonymy plus invalid names at 10%, and \( Us \) = the expected ratio of unknown to known species, hence \( Es = (Ks-Ks/10) \times 3.18 \). We note that generating a single parameter to estimate richness for very different clades, is grossly over simplistic. Thus we caution against using the numbers in the ‘Estimated’ column as an explicit prediction of diversity per family, it is rather simply a reflection on our lack of knowledge. Nevertheless, taken at face value, this metric predicts total spider richness to be at least 120,000. This value is inside the range estimated by Adis & Harvey (2000), an opinion-based estimate of 76,000–170,000. Of course, this is simply a compilation of educated guesses, and probably cautious at that.

First, researchers base such guesses, at least partially, on the ratio of new to known species they have encountered in the field, especially in the tropics. However, as the data from biodiversity inventories discussed above demonstrate, beta diversity in tropical spiders is typically high, in the few cases it has been studied. Species shared between sites are more likely to be widespread, and thus more likely to have been described.

Second, nearly all spider species are diagnosed morphologically, but molecular studies increasingly reveal ‘cryptic species’ that are morphologically difficult to distinguish. Although we should also anticipate molecularly diagnosed synonymies, the overall trend seems to predict multiples, not frac-
tions of known diversity (Waugh, 2007; Robinson et al., 2009). In spiders multiples are particularly likely in sedentary, poorly dispersing, morphologically uniform groups, such as mygalomorphs. Indeed, Bond & Stockman (2008) went to some length not to recognize plainly diagnosable evolutionary lineages as species. Our estimate highlights that the experiences of expert taxonomists with fieldwork and rate of taxonomic discovery, combined with point estimates of alpha and beta diversities, leave little doubt that much more remains to be discovered than has been discovered so far, and presaged by tropical biodiversity inventory data and molecular work, perhaps orders of magnitude more. We would not be surprised if true spider diversity, in the early 21st century, described and undescribed, exceeded 200,000 species.

Spiders are diverse morphologically (Figure 1) and most spider systematists focus on one or more subclades of spiders, typically a family or set of related families. Progress is thus unevenly spread through the spider tree of life, depending on the interest of active systematists at any given time. Spider taxonomy is quite an active field with on average over 500 new species described annually over the last 10 years (Figure 7A). Numbers of genera and species are rising consistently and linearly (Figure 7A and B ex Platnick 2001–2012). Assuming that the trend holds, 50,000 species will be known by 2025, 60,000 by 2045, and 100,000 species roughly a century from now. Of course extinction is also rising, so that some, presumably large, fraction of species will disappear before (and after) discovery. If one accepts that discovery and description have actually been accelerating over the last 50 years (Figures 2, 7B), 50,000 species may be known by 2020, and 100,000 species as early as 2060 (Figure 8). Evidence for an increased rate (Figure 7B) is less consistent, however, and thus the future is harder to predict. Nevertheless, a nearly complete knowledge of spider species diversity might be witnessed by current spider systematics students. The apparent increase in number of species described per year goes hand in hand with a sharp increase in the number of taxonomic papers published annually since the new millennium (Figure 9).

At present, spider taxonomy is active and training in systematics seems to be successfully countering the ‘taxonomic impediment’ (Rodman & Cody, 2003). While taxonomy is often perceived as low-budget science, taxonomic progress is nevertheless proportional to funding. Large grants targeting particular spider groups can have impressive effects, such as the Goblin spider Planetary Inventory, supported by the National Science Foundation. Soon after the ~2 million dollar grant was awarded, in 2006, knowledge of Oonopidae began to grow exponentially (Figure 9). In the first 250 years of taxonomy 1757–2006, taxonomists had described nearly 500 goblin spiders. Six years hence the total of known species has more than doubled, demonstrating that taxonomy, modernized through use of online tools, can be accelerated remarkably through additional funding.
<table>
<thead>
<tr>
<th>Family</th>
<th>Genera</th>
<th>Species</th>
<th>Estimated</th>
<th>Expert opinion</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salticidae</td>
<td>575</td>
<td>5423</td>
<td>15,507</td>
<td>15–20,000</td>
<td>Maddison</td>
</tr>
<tr>
<td>Linyphiidae</td>
<td>587</td>
<td>4412</td>
<td>12,616</td>
<td>7–11,000</td>
<td>Frick</td>
</tr>
<tr>
<td>Araneidae</td>
<td>168</td>
<td>3029</td>
<td>8661</td>
<td>6000</td>
<td>Kuntner</td>
</tr>
<tr>
<td>Lycosidae</td>
<td>120</td>
<td>2387</td>
<td>6825</td>
<td>4130</td>
<td>Santos</td>
</tr>
<tr>
<td>Theridiidae</td>
<td>119</td>
<td>2324</td>
<td>6645</td>
<td>11,600</td>
<td>Agnarsson</td>
</tr>
<tr>
<td>Thomisidae</td>
<td>174</td>
<td>2151</td>
<td>6151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gnaophosidae</td>
<td>117</td>
<td>2128</td>
<td>6085</td>
<td>3000</td>
<td>Platnick</td>
</tr>
<tr>
<td>Ageelenidae</td>
<td>68</td>
<td>1152</td>
<td>3294</td>
<td>1530</td>
<td>Spagna</td>
</tr>
<tr>
<td>Pholcidae</td>
<td>84</td>
<td>1270</td>
<td>3631</td>
<td>6770</td>
<td>Huber</td>
</tr>
<tr>
<td>Sparassidae</td>
<td>85</td>
<td>1114</td>
<td>3185</td>
<td>10,200</td>
<td>Jäger</td>
</tr>
<tr>
<td>Corinnidae</td>
<td>87</td>
<td>1016</td>
<td>2905</td>
<td>2200</td>
<td>Haddad</td>
</tr>
<tr>
<td>Zodariidae</td>
<td>77</td>
<td>963</td>
<td>2754</td>
<td>4200</td>
<td>Jocqué</td>
</tr>
<tr>
<td>Tetragonathidae</td>
<td>47</td>
<td>955</td>
<td>2731</td>
<td>1910</td>
<td>Álvarez-Padilla</td>
</tr>
<tr>
<td>Theraphosidae</td>
<td>121</td>
<td>932</td>
<td>2665</td>
<td>900</td>
<td>Raven</td>
</tr>
<tr>
<td>Oonopidae</td>
<td>88</td>
<td>857</td>
<td>2451</td>
<td>2500</td>
<td>Platnick</td>
</tr>
<tr>
<td>Clubionidae</td>
<td>17</td>
<td>579</td>
<td>1656</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictynidae</td>
<td>51</td>
<td>570</td>
<td>1630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philodromidae</td>
<td>29</td>
<td>535</td>
<td>1530</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysderidae</td>
<td>24</td>
<td>524</td>
<td>1498</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anyphaenidae</td>
<td>56</td>
<td>514</td>
<td>1470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ctenidae</td>
<td>39</td>
<td>468</td>
<td>1338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxyopidae</td>
<td>9</td>
<td>431</td>
<td>1232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miturgidae</td>
<td>28</td>
<td>359</td>
<td>1027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nemessidae</td>
<td>42</td>
<td>355</td>
<td>1015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pisauridae</td>
<td>49</td>
<td>331</td>
<td>946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idiopidae</td>
<td>22</td>
<td>313</td>
<td>895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barychelidae</td>
<td>44</td>
<td>304</td>
<td>869</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proctosidae</td>
<td>30</td>
<td>303</td>
<td>866</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amauridiidae</td>
<td>52</td>
<td>288</td>
<td>824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uloboridae</td>
<td>18</td>
<td>262</td>
<td>749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leptonetidae</td>
<td>22</td>
<td>261</td>
<td>746</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hahnidae</td>
<td>27</td>
<td>247</td>
<td>706</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selenopidae</td>
<td>10</td>
<td>237</td>
<td>678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scytodidae</td>
<td>5</td>
<td>228</td>
<td>652</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nesticidae</td>
<td>9</td>
<td>209</td>
<td>598</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamponidae</td>
<td>23</td>
<td>192</td>
<td>549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liocranidae</td>
<td>30</td>
<td>183</td>
<td>523</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desidae</td>
<td>38</td>
<td>181</td>
<td>518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orsolobidae</td>
<td>28</td>
<td>181</td>
<td>518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dipluridae</td>
<td>25</td>
<td>180</td>
<td>515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cybaeidae</td>
<td>10</td>
<td>177</td>
<td>506</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirsiliidae</td>
<td>15</td>
<td>175</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ochyroceratidae</td>
<td>14</td>
<td>161</td>
<td>460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphinectidae</td>
<td>32</td>
<td>159</td>
<td>455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mimetidae</td>
<td>13</td>
<td>156</td>
<td>446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trochanteriidae</td>
<td>19</td>
<td>152</td>
<td>435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Araneidae</td>
<td>38</td>
<td>150</td>
<td>429</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrablemmidae</td>
<td>30</td>
<td>142</td>
<td>406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stiphidiidae</td>
<td>22</td>
<td>135</td>
<td>386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyrtarachniidae</td>
<td>18</td>
<td>134</td>
<td>383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palpimanidae</td>
<td>15</td>
<td>131</td>
<td>375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ctenizidae</td>
<td>9</td>
<td>125</td>
<td>357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sicariidae</td>
<td>2</td>
<td>124</td>
<td>355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>Species</td>
<td>Genus</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mysmenidae</td>
<td>23</td>
<td>123</td>
<td>352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segestridae</td>
<td>3</td>
<td>118</td>
<td>337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trechaleidae</td>
<td>16</td>
<td>117</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filistatidae</td>
<td>17</td>
<td>113</td>
<td>323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oecobiidae</td>
<td>6</td>
<td>110</td>
<td>315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexathelidae</td>
<td>12</td>
<td>105</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eresidae</td>
<td>8</td>
<td>96</td>
<td>275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migidae</td>
<td>10</td>
<td>91</td>
<td>260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liphistidae</td>
<td>5</td>
<td>89</td>
<td>254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theridiosomatidae</td>
<td>16</td>
<td>89</td>
<td>254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoropsidae</td>
<td>14</td>
<td>86</td>
<td>246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caponiidae</td>
<td>15</td>
<td>84</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synotaxidae</td>
<td>14</td>
<td>82</td>
<td>234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoridae</td>
<td>14</td>
<td>79</td>
<td>226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micropholcommatidae</td>
<td>19</td>
<td>66</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symphytognathidae</td>
<td>7</td>
<td>66</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nephilidae</td>
<td>4</td>
<td>61</td>
<td>174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telemidae</td>
<td>7</td>
<td>60</td>
<td>172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyatholipidae</td>
<td>23</td>
<td>58</td>
<td>166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gallieniellidae</td>
<td>11</td>
<td>57</td>
<td>163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deinopidae</td>
<td>4</td>
<td>57</td>
<td>163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phyxelididae</td>
<td>12</td>
<td>54</td>
<td>154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archaeidae</td>
<td>3</td>
<td>54</td>
<td>154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tengellidae</td>
<td>8</td>
<td>51</td>
<td>146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Titanoeicidae</td>
<td>5</td>
<td>49</td>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atypidae</td>
<td>3</td>
<td>48</td>
<td>137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zorocratidae</td>
<td>5</td>
<td>42</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinopodidae</td>
<td>3</td>
<td>40</td>
<td>114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pimoidae</td>
<td>4</td>
<td>37</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclocenidae</td>
<td>5</td>
<td>36</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pararchaeidae</td>
<td>7</td>
<td>35</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antrodietae</td>
<td>2</td>
<td>33</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senoculidae</td>
<td>1</td>
<td>31</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plectreuridae</td>
<td>2</td>
<td>31</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psechridae</td>
<td>2</td>
<td>30</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nicodarmidae</td>
<td>9</td>
<td>29</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mecysmauchenidae</td>
<td>7</td>
<td>25</td>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ammoxenidae</td>
<td>4</td>
<td>18</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microstigmatidae</td>
<td>7</td>
<td>16</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gradungulidae</td>
<td>7</td>
<td>16</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drymusidae</td>
<td>1</td>
<td>15</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diguettidae</td>
<td>2</td>
<td>15</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synaphridae</td>
<td>3</td>
<td>13</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stenochilidae</td>
<td>2</td>
<td>13</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypochilidae</td>
<td>2</td>
<td>12</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malkaridae</td>
<td>4</td>
<td>11</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mecicothritidae</td>
<td>4</td>
<td>9</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austrochilidae</td>
<td>3</td>
<td>9</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penestomidae</td>
<td>1</td>
<td>9</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paratropididae</td>
<td>4</td>
<td>8</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cithaeronidae</td>
<td>2</td>
<td>7</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homalonymychidae</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periegopidae</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holarchaeidae</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chumnidae</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinopimoidae</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huttoniidae</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3859</td>
<td>42,751</td>
<td>122,243</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7 (left). A–B. The relatively constant, cumulative growth in genera and species over the last ten years (rates 1.05% and 1.3% per year, respectively, $R^2 > 0.98$). C. The number of new species described per year since 2001.

Figure 8 (below). Extrapolations of taxonomic progress based on various rates of taxonomic discovery. The dark line extrapolates linear increase based on the last 12 years, and the lighter lines extrapolate linear increase based on the last six years, with the line on the left representing a "best case scenario" based on the last three years.
Spider phylogenetics – practice and progress

Spiders are an old and diverse clade with known fossils dating from at least the Carboniferous (Penney & Selden, 2011), some 320 mya; previous reports of Devonian fossil spiders (Shear et al., 1989; Selden et al., 1991) concerned specimens recently classified in a new, extinct arachnid order Uraraneida (Selden et al., 2008: Dunlop & Penney, 2012). To put that in perspective, spiders are nearly twice the age of mammals. Resolving deep nodes in spider phylogenetics is thus challenging. Novel DNA sequencing techniques, however, hold promise. Also, new morphological synapomorphies are being steadily discovered (Griswold et al., 2005; Ramirez, in press), and certainly many areas of comparative morphology, broadly construed, have barely been touched. On the whole, the number, complexity, and consistency of morphological synapomorphies do not decrease with increasing age of divergence. Put another way, in morphological datasets nodes at any level are more or less equally likely to be strongly supported, or contentious. The opposite is often true for DNA data. The older the node, the less likely support will be strong and uncontroversial. Branch length is also important; old and short branches are particularly contentious.

Prior to this millennium spider phylogenetics relied almost exclusively on morphological data. The state of the art of morphological spider phylogenetics was reviewed by Coddington & Levi (1991), and again relatively recently...
by Coddington (2005), hence we offer only a short summary of Coddington’s reviews here, with notes on some notable progress since 2005.

Although taxonomic treatments without original phylogenetic analyses are widely published (Raven, 2008; Miller et al., 2009, 2010b; Kuntner & Agnarsson, 2010; Sanchez-Ruiz et al., 2010; Bayer, 2011; Jäger, 2011, 2012a–c; Platnick & Duperre, 2011a–d; Platnick et al., 2011, 2012ab; Zhang & Maddison, 2012ab), the field is slowly moving towards integration of phylogenetic data in taxonomic works. Ideally, all new species would be routinely placed in the global tree of life using quantitative data. There has also been a dramatic increase in original phylogenetic studies, which may or may not have a taxonomic component, in the past few years. Recently published phylogenies draw data from diverse sources. Some still purely rely on the traditional, morphology-based cladistic approach. However, the use of molecular phylogenetic approaches, which only a few years ago, was characterized as still “in its infancy” (Coddington, 2005), has grown so rapidly that it may now be called, charitably, a middle-aged infant. Thus, the majority of the recent large phylogenetic datasets are molecular, or a total evidence approach combining both types of data. Here, we give a brief overview of this literature.

Morphological phylogenetics

Dozens of morphological phylogenetic studies on spiders were published in the early 21st century with a clear push towards richer datasets and towards better illustrated, and thus more testable, homology inferences. Noticeable among these have been publications of several morphological ‘atlases’ based on thorough surveys of higher level taxa and emphasizing documentation through images. Whether these included their own detailed phylogenetic analysis or not, these are data rich papers that will be invaluable for future morphological phylogenetics (Hornig, 1994; Agnarsson, 2004; Griswold et al., 2005; Miller, 2007; Alvarez-Padilla & Hornig, 2011; Miller et al., 2012b). Recent cladistic papers that analyzed datasets well in excess of 100 morphological characters have focused on theridiids, in particular on the genus *Anelosimus* (Agnarsson, 2005, 2006b; Agnarsson & Kuntner, 2005; Agnarsson & Zhang, 2006), tetragnathids (Alvarez-Padilla, 2007; Alvarez-Padilla & Hornig, 2011), with some emphasis on the genera *Cyrtognatha* (Dimitrov & Hormiga, 2009) and *Atelidea* (Alvarez-Padilla & Benjamin, 2011), nephilids (Kuntner et al., 2008), and more specifically on the subclades *Herennia* (Kuntner, 2005), *Clitaetra* (Kuntner, 2006), *Nephilengys* (Kuntner, 2007), and *Nephila* (Kuntner & Coddington, 2009), on the Savignia group (Frick et al., 2010) and other linyphiids (Miller, 2005; Paquin et al., 2008; Tu & Hormiga, 2011), on anyphaenids (Lopardo, 2005; Izquierdo & Ramírez, 2008; Werenkraut &
Ramirez, 2009; Marquez & Ramirez, 2012), on paraphyletic (Rix, 2006), micropholcommatids (Rix & Harvey, 2010), pholcids (Huber, 2005, 2007, 2011a; Huber et al., 2005), phyxelidids (Griswold et al., 2012b), and on oonopids (Alvarez-Padilla et al., 2012). A number of studies have continued to test the monophyly within other spider clades through amassing smaller datasets, such as studies on theraphosids (Fukushima et al., 2005; West et al., 2008, 2012; Guadanucci, 2011; Yamamoto et al., 2012), scytodoids (Labarque & Ramirez, 2012), leptonetids (Ledford & Griswold, 2010), corinnids (Bonaldo & Brescovit, 2005; Jocque & Bosselaers, 2011), pisaurids (Santos, 2007), lycosids (Stratton, 2005; Framenau & Yoo, 2006; Yoo & Framenau, 2006; Langlands & Framenau, 2010), tetrablemmids (Tong & Li, 2008), gallieniellids (Haddad et al., 2009), zoropsids (Polotow & Brescovit, 2011), thomisids (Benjamin, 2010), philodromids (Muster, 2009), salticids (Ruiz & Maddison, 2012), linyphiids (Hormiga & Scharff, 2005), pimoids (Hormiga et al., 2005; Hormiga, 2008; Hormiga & Tu, 2008), anapids (Lopardo & Hormiga, 2008), araneids (Schmidt & Scharff, 2008; Framenau et al., 2010b) and deinopids (Coddington et al., 2012).

Throughout these morphological studies, data collection and standards remain mostly unsynchronized. By tradition only characters that are cladistically informative are mentioned, with the consequence that character lists overlap little from one study to the next. Even shared characters are often defined and treated differently. This makes it difficult to unite studies to analyze deeper patterns. Of course matrices are, necessarily, tuned to specific clades and relatively few characters are easily scored across all spiders. With as yet little evidence from evolutionary-developmental biology, dueling theories of homology persist. Scoring discrete data for character states can be subjective, and continuous data continues to be problematic (Hendrixson & Bond, 2009).

Methods to document characters (Ramirez et al., 2010; Ramírez, in press) and to recognize homologies (Agnarsson & Coddington, 2008) can be improved. Cladistic studies by different researchers on substantially the same clades often conflict. For example, the placement and circumscription of Nephiilidae remains controversial and subjective (see Kuntner, 2005, 2006, 2007; Kuntner et al., 2008 versus Dimitrov & Hormiga, 2009; Alvarez-Padilla & Benjamin, 2011).

Morphological systematics could benefit tremendously from more standardized imagery and protocols, leading to less subjectivity and improved communication among systematists (Ramirez et al., 2010; O’Leary & Kaufman, 2011). Also, Hennig’s (1960) ‘reciprocal illumination’ could guide effort to reevaluate morphological characters, for example in light of molecular or total evidence analyses. As an example, the homology of the ‘peg teeth’, a putative synapomorphy of the original Palpimanidea is rejected by molecular and total evidence analyses, which implies reevaluation of this character system. Such
an approach might help solve the debate concerning putative homologs in the palpal organs of Nephilidae and other Araneoidea.

Spider systematists also use behaviour, most notably webs and web building behaviours (reviewed in Kuntner et al., 2008 and Blackledge et al., 2011), building on the classical works from the 1980s (Eberhard, 1982b; Coddington, 1986ab). Behavioural characters do provide important synapomorphies for crucial clades (Coddington & Levi, 1991; Coddington, 2005; Kuntner et al., 2008), although intra- versus interspecific variation and phenotypic constancy should be further studied (Eberhard, 1990b; Gregorić et al., 2010).

Molecular phylogenetics

Molecular phylogenetic studies on spiders have proliferated in last decade. Certain clades such as Mygalomorphae, Salticidae, Orbiculariae and Lycosoidea have received more attention than some other parts of the tree, although, in retrospect, the available markers have been markedly inadequate to address fundamental questions in spider phylogeny – i.e. the relationships of families.

Genbank now contains considerable and diverse sequences for spiders. Estimating species cover in Genbank is difficult as many studies include sequences from specimens with uncertain taxonomic status, including undescribed species. In the following paragraph “species” thus refers to an estimate of species level taxa, described and undescribed, available on Genbank (Table 2) and as a percentage of described species. Over 50% of sequences come from seven families: Salticidae (~370 species, ~7%), Lycosidae (~370 species, ~11%), Linyphiidae (~250 species, ~6%), Theridiidae (~160 species, ~13%), Araneidae (~140 species, ~10%), Pholcidae (~125 species, ~8%), Thomisidae (~120 species, ~7%). However, these families contain the most described species, representing nearly 50% of spider diversity. Thus, molecular phylogenetic effort roughly mirrors diversity.

Nevertheless, there are several large spider groups that have received disproportionately little effort to date, such as the diverse ground dwelling spiders Clubionidae, Gnaphosidae and Oonopidae. As of November 2012, the first DNA sequence of the diverse Oonopidae (1016 species) has yet to be submitted to Genbank. The first large scale project to tackle spider diversity systemically, the spider tree of life project (http://research.amnh.org/atol/files/), aimed to remedy this problem. The goals of this project targeted the kind of data spider systematics urgently needs, a broad spectrum morphological and molecular survey of clades representing all spiders. However, this project has as yet not resulted in published phylogenetic articles or submission of substantial data to Genbank.

In total a little over 3000 spider species, rather less than 10% of known species, are represented by sequences in Genbank. The most common is a por-
Table 2. The usual suspects – availability of genes on Genbank, listing mostly the markers used routinely in spider phylogenetics. The number of species is estimated based on species counts, not total number of sequences on Genbank, in many cases there may be multiple sequences available per gene per species.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Species</th>
<th>Genome</th>
<th>Protein</th>
<th>Max utility level</th>
<th>Typical length bp</th>
<th>Typical primers (forward; reverse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome Oxylase COI</td>
<td>2233</td>
<td>mitochondrion</td>
<td>yes</td>
<td>population-genus</td>
<td>600–1200</td>
<td>LCO1490; C1-N-2776</td>
</tr>
<tr>
<td>28S rDNA</td>
<td>1356</td>
<td>nucleus</td>
<td></td>
<td>species–family</td>
<td>800–1800</td>
<td>28S0; 28SC</td>
</tr>
<tr>
<td>16S rDNA</td>
<td>1356</td>
<td>mitochondrion</td>
<td></td>
<td>population-genus</td>
<td>800</td>
<td>16SF; 16SR</td>
</tr>
<tr>
<td>Nitroge dehydrogenase ND1</td>
<td>817</td>
<td>mitochondrion</td>
<td>yes</td>
<td>species–family</td>
<td>350</td>
<td>LR-J-12864; LR-N-13398</td>
</tr>
<tr>
<td>18S rDNA</td>
<td>808</td>
<td>nucleus</td>
<td>family–order</td>
<td></td>
<td>800</td>
<td>18Sa; 18Sb</td>
</tr>
<tr>
<td>Histone 3</td>
<td>693</td>
<td>nucleus</td>
<td>yes</td>
<td>species–family</td>
<td>300</td>
<td>H3aF; H3aR</td>
</tr>
<tr>
<td>tRNA Leucine</td>
<td>518</td>
<td>mitochondrion</td>
<td></td>
<td>population–species</td>
<td>&lt;20</td>
<td>LR-J-12864; LR-N-13398</td>
</tr>
<tr>
<td>12S rDNA</td>
<td>305</td>
<td>mitochondrion</td>
<td></td>
<td>population–species</td>
<td>350</td>
<td>12S-ai; 12S-bi</td>
</tr>
<tr>
<td>Alpha elongation factor</td>
<td>119</td>
<td>nucleus</td>
<td>yes</td>
<td>species–family/order</td>
<td>~500</td>
<td>eff172; eff912</td>
</tr>
<tr>
<td>Internal Transcribes Spacer ITS2</td>
<td>100</td>
<td>nucleus</td>
<td></td>
<td>population–species</td>
<td>400</td>
<td>ITS4; ITS5.8</td>
</tr>
<tr>
<td>Cytochrome Oxylase COII</td>
<td>67</td>
<td>mitochondrion</td>
<td>yes</td>
<td>species–genus</td>
<td>400</td>
<td>COII-2635; COII-3131</td>
</tr>
<tr>
<td>Gamma elongation factor</td>
<td>64</td>
<td>nucleus</td>
<td>yes</td>
<td>species–family/order</td>
<td>600–1200</td>
<td>EF1gF78; EF1gR1261</td>
</tr>
<tr>
<td>Actin</td>
<td>41</td>
<td>nucleus</td>
<td>yes</td>
<td>species–family/order</td>
<td>900–1300</td>
<td>SC-F-229; SC-R-1057</td>
</tr>
<tr>
<td>Wingless</td>
<td>37</td>
<td>nucleus</td>
<td>yes</td>
<td>genus–order</td>
<td>352</td>
<td>SpwgF1; Spwgr1</td>
</tr>
</tbody>
</table>
tion of mitochondrial COI, the ‘DNA barcoding’ gene, from over 2200 species. Mitochondrial 16S (often including tRNA^cyt) and nuclear 28S are each known from over 1300 species (Table 2). Although single gene phylogenies are still published (Muster et al., 2007; Park et al., 2007; Bolzern et al., 2010; Framenau et al., 2010a), most of the phylogenies reviewed here draw from multiple markers. While some studies use only nuclear (Bond & Hedin, 2006; Hedin & Bond, 2006; Rix et al., 2008) or mitochondrial markers (Ayoub et al., 2005; Cheng et al., 2010), many use several nuclear (most frequently Histone 3, 18S rDNA, and 28S rDNA) and mitochondrial loci (typically 16S rDNA, and COI) (Agnarsson, 2006a).

These markers have been used so often (Amedo et al., 2007; Bidegaray-Batista et al., 2007; Bidegaray-Batista & Amedo, 2011; Ledford et al., 2011) that they have been referred to as the “usual suspects” (Dimitrov et al., 2010, oral presentation, see Table 2 for detail).

Other promising markers exist, especially for relatively recent nodes. The nuclear marker ITS2 is useful for shallow divergences (Amedo & Gillespie, 2006; Chang et al., 2007; Agnarsson, 2010). ND1 can resolve species or genus level phylogenies (Amedo & Gillespie, 2006; Maddison & Needham, 2006; Agnarsson et al., 2007; Chang et al., 2007; Maddison et al., 2007, 2008; Řezáč et al., 2008; Agnarsson, 2012; Ruiz & Maddison, 2012). Elongation factor-1 gamma has been used to resolve family level splits in mygalomorphs (Ayoub et al., 2007), wingless has been included in studies, for example of Orbiculariae (Blackledge et al., 2009), and actin has been proposed as a useful marker for spiders in general (Vink et al., 2008a). Elongation factor-1 alpha was found useful in species level phylogenetics of both Salticidae (Hedin & Maddison, 2001) and Thomisidae (Garb & Gillespie, 2009), but no general spider primers exist. A sprinkling of other markers are beginning to appear, a recent study of the trapdoor genus Myrmeckiaphila (Bailey et al., 2010) for example, used the mitochondrial 12S rRNA, tRNA-val, and 16S rRNA, and the nuclear glutamyl and prolyl tRNA synthetase genes. As with morphological research, the collection of DNA data across taxa is relatively unsynchronized, so that combining data from different studies to address more inclusive questions is impeded. Even though the “usual suspects” include only a few loci, different authors utilized different primers, often resulting in different length sequences and non-overlapping regions of the same locus. Generating matrices from several independent studies to span many taxa, as done here, is difficult (see below).

Total evidence phylogenetics

Combining genetic data from many independent nuclear and mitochondrial markers with morphology and/or behaviour, so called “total evidence" phy-
logenetics, is now widely used. Such analyses have been applied at varying phylogenetic scales, both lower levels (Astrin et al., 2006; Aagnarsson et al., 2007; Arneto et al., 2009a; Huber & Astrin, 2009; Bailey et al., 2010; Macias-Hernandez et al., 2010; Rix & Harvey, 2010; Bond et al., 2012a; Harvey et al., 2012; Richardson & Gunter, 2012), and higher levels (Arneto et al., 2009b; Blackledge et al., 2009b; Lopardo et al., 2011), even though the ability of available molecular data to reconstruct higher level spider phylogeny, from both theoretical and empirical perspectives, is problematic. The classical philosophical argument for total evidence is that a single analysis of all data maximizes explanatory power (Kluge, 1997). On the other hand, the danger is that one data source may overwhelm the phylogenetic signal of others. It is thus advisable to examine the results from each data source separately. Arguably the strongest support for a clade, regardless of standard measures such as bootstrap and posterior probability values, is independent support from different kinds of evidence. For example, the rich molecular dataset of Dimitrov et al. (2012) offered weak support for the monophyly of Orbiculariae and the orb web, but nonetheless satisfying, precisely since the result was independent of morphological and behavioral data that originally suggested this clade.

Molecular and morphological datasets often both agree and disagree, for example in the case of combined analyses of pholcids (Bruvo-Madaric et al., 2005). What is the source of this incongruence and can we judge which kind of data are closer to the true tree? In mygalomorphs Bond & Hedin (2006) concluded that morphology was not variable enough to resolve phylogenetic relationships and therefore relied on molecules. However, this is not really incongruence, but rather lack of resolution. In other cases authors have concluded that one source of data was more erroneous than the other, for example, calling “morphology to the rescue” of their molecular analyses (Lopardo et al., 2011). The combined results agreed better with traditional, morphology-based notions of familial monophyly and groupings at the familial level. Such evaluations presume knowledge of the ‘correct’ phylogenetic relationships. Alvarez-Padilla et al. (2009) found that morphology linked nephilids with tetragnathids (see above), but the combined dataset rejected that relationship. Which signal is ‘better’ in this case is difficult to say, but total evidence analysis does agree better with other recent studies (Kuntner et al., 2008; Blackledge et al., 2009b; Dimitrov et al., 2012), suggesting that the incongruence may be due to issues with the morphological datasets rather than a better molecular signal. In sum, total evidence analyses, no doubt, represent best practice, but it is clearly advisable to evaluate critically the results of independent analyses before data are combined, and after.
• clade holds

*to Araneoidea

*to RTA clade

*monophyly consistently rejected in DNA studies

● clade not recovered here

Unplaced families
- Chummidae✓
- Cybaeidae✓
- Cyclosternidae✓
- Hahniidae✓
- Halitidae✓
- Homalonychidae✓
- Synaphridae✓
State of the art – what do we know about spider phylogeny?

Coddington’s (2005) summary phylogeny based entirely on morphological data (Figure 10) serves well as a consensus view of phylogenetic knowledge of spiders at the beginning of this millennium. As reviewed above, numerous morphological and molecular phylogenies have been published since Coddington (2005), and what we have learnt from these can be summarized in a few words: most deep clades in spider phylogenetics are disputed, mainly by molecular results. Not only are new molecular studies incongruent with much of ‘traditional’ knowledge but they are often incongruent with one another. Cases in which molecular phylogenies strongly support deep clades proposed by morphology (or contributing to total evidence analyses) are thus relatively few. Hence we represent the, admittedly gloomy, state of the art knowledge with a recapture of Coddington’s cladogram, on which we highlight current areas of agreement, and major conflict (Figure 10).

Although it seems that few of the clades proposed by Coddington (2005) are consistently supported by recent molecular data, some of the major clades still hold (marked red in Figure 10). These represent our few ‘benchmark clades’, basically well supported hypotheses and the few things we truly can claim to ‘know’ about deep spider phylogeny. Notable among these is the basal dichotomy of Mesothelae, the small group of spiders primitively retaining external signs of abdominal segmentation, vs. Opistothelae the vast majority of spiders. Also consistently supported are the monophyly of Mygalomorphae (‘tarantulas’), Entelegynae (spiders with sclerotized female genitalia), Araneoidea (the classical ecribellate orb-weavers), and the RTA clade (spiders with a retrolateral tibial apophysis on the male pedipalp). The last of these, the RTA, has recently been modified based on molecular data to include the superfamily Titanoeocoidea (Miller et al., 2010a), a reasonable

Figure 10. Coddington’s (2005) consensus phylogeny based on morphological studies, with recent support and refutation from molecular and total evidence analyses marked. This is a summary highlighting some of the major conflicts and does not attempt to capture all disagreements and agreements in family placement. Red dots highlight clades that have either been supported or not strongly refuted by recent studies. Red and green stars indicate taxa that have recently been transferred to different clades. Blue stars indicate clades that have failed to be supported in prior molecular analyses. In orange we highlight clades that prior analyses have either supported, or not tested, but are not recovered in the analyses presented in Figures 11–12. In sum, all clades/taxa marked with stars represent serious challenges to the consensus morphological view. The monophyly of clades marked in orange is not recovered in our analysis (see text), but some, such as Orbiculariae, have been supported by molecular data previously. In general these represent clades with little or relatively weak support from molecular evidence, and thus these clades require further scrutiny from future studies. Of the seven families that had not been placed phylogenetically in 2005 (see list in Figure), four have now been placed and are indicated with a checkmark.
placement given that some of these spiders have an RTA, while others may have secondarily lost it. Other major clades that have not really been disputed include Araneomorphae (the so called ‘true’ spiders, all spiders other than Mesothelae and Mygalomorphae) and Haplogynae (sensu lato). However, one has to be cautious, as the lack of dispute for deep morphology-based clades may sometimes simply mean the lack of any phylogenetic tests. For example, while Griswold et al. (2005) tested the monophyly of Entelegynae, their test of Haplogynae was weak (see below). A novel test of the deep nodes in the spider topology is thus needed and is presented below.

Other traditional clades have been disputed since 2005. Araneoclada was rejected in a recent morphological study (Griswold et al., 2005) thus rejecting the dichotomy Haplogynae + Entelegynae. In their optimal tree Haplogynae is placed more basally than previously hypothesized, as the first offshoot of araneomorph spiders. Their study was designed to test the relationships within Entelegynae, and thus deep spider nodes could not have been recovered with much certainty.

The superfamily Palpimanoidea (Forster & Platnick, 1984) is another example. At the time it was proposed, it was a bold hypothesis that contradicted the traditional view of mimetids and their relatives as araneoids, based on newly discovered putative synapomorphies. It was also logical and well-argued cladistically, in an era when few phylogenetic hypotheses could make that claim. However, it fell as soon as it was critically re-examined, albeit 20 years later, in the slow-paced world of spider phylogenetics (Schütz, 2003; Griswold et al., 2005). It proved to be based on convergent traits interpreted as homologies. No recent studies, based on morphological and/or molecular data have supported it, but rather have shown the araneoid placement of the following lineages: Mimetidae (Blackledge et al., 2009a), Malkaridae (Rix et al., 2008), Micropholcommatidae (Schütz, 2003; Rix et al., 2008, 2010), Holarchaeidae (Rix et al., 2008) and Pararchaeidae (Rix et al., 2008).

According to Rix et al. (2008) Palpimanoidea should only include haplogyne spiders from the families Archaeidae, Palpimaniidae, Stenochilidae and Huttoniidae, and the phylogenetic status of Mecysmaucheniiidae remains to be settled. The most recent studies support this view (Dimitrov et al., 2012), and our own analysis also placed Mecysmaucheniiidae with this revised Palpimanoidea (Figure 11).

The superfamily Eresoidea is supported by morphology (Griswold et al., 2005), but not by molecules (Miller et al., 2010a; Dimitrov et al., 2012; Agnarsson et al., 2013). Morphology may suggest that the close relatives of orbicularians (orb-weavers) are among the Eresoidea, with some molecular analyses placing cersids sister to orbicularians (Figure 11), whereas ocobiids and hersiliids appear more closely related to the RTA clade. A different topology was recovered by Dimitrov et al. (2012), where the RTA clade plus Hersi-
liidae and Oecobiidae group with Orbicularia, while eresids are a more basal offshoot. Miller et al. (2010a) called for redefinition of several RTA families such as Agelemidae, Amaurobiidae, Cybaeidae, Dictynidae and Hahniidae, and proposed a new family, Penestomidae, as sister to Zodariidae. Agnarsson et al. (2013) concur suggesting that many of the RTA families, and deeper level clades, require new circumscriptions, as the current findings also suggest (Figure 11). While both studies undersample Dionycha, both are consistent with the monophyly of Dionycha but not that of the Lycosoida.

Nicodamids have since 2005 been hypothesized to nest inside the ecribellate araneoids (Blackledge et al., 2009a; Dimitrov et al., 2012), although the placement of the cribellate Megadictyna differs among analyses. Agnarsson et al. (2013) placed Megadictyna close to Eresidae, while the remaining nicodamids nested within Araneoidea. Thus, the group Araneoidea, or true orb weavers, has been significantly redefined during this first decade of the 21st century, with the inclusion of nicodamids, return of classical araneoids briefly exiled to Palpimanoidea (see above), and numerous internal rearrangements (Schütt, 2003; Blackledge et al., 2009a; Dimitrov et al., 2012).

Although the contents of Araneoidea may be clear, araneoid relationships are not. Nearly all clades above the family level differ among recent phylogenetic analyses from the hypothesis presented in Figure 10, and no consensus view has been established. For example, recent studies agree that Tetragnathidae does not contain nephilids (Kuntner et al., 2008), however, their exact placement differs among studies. Blackledge et al. (2008) placed them sister to “zygiellids”, Dimitrov et al. (2012) sister to a large “araneid” clade, Kuntner et al. (2008) sister to all other araneoids, and Agnarsson et al. (2013) sister to Araneidae+Anapidae. Instead, tetragnathids appear to be phylogenetically closer to mimetids and the ‘araneid’ genus Arkys (Dimitrov & Lazarov, 2002; Blackledge et al., 2009a; Agnarsson et al., 2013).

No studies including molecular data have supported the legacy clades ‘Derived Araneoids’, ‘Symphytognathoids’, ‘Araneoid Sheet Web Weavers’, or the ‘Spineless Femur Clade’. Similarly theridioids (Theridiidae plus Nesticidae) are consistently rejected in molecular analyses (Dimitrov & Lazarov, 2002; Blackledge et al., 2009a; Agnarsson et al., 2013). Linyphioids (Pimoidae plus Linyphiidae) are less controversial (Arnedo et al., 2009b; Agnarsson et al., 2013), although they are not consistently supported (Dimitrov et al., 2012). Current molecular phylogenies fail to support deeper nodes and are inconsistent, so that their rejection of morphological results is difficult to evaluate. Significantly, no one has attempted a similarly comprehensive morphological analysis of Araneoidea to superecede that of Griswold et al. (1998).

Within Mygalomorphae, recent studies are also revamping knowledge with marked changes in the phylogenetic hypothesis since 2005. Bond et al.
(2012b), proposed a new classification that recaptured some of the key clades in Figure 10 as previously defined, namely Atypoidea and Avicularioidea, however, most of the remaining major clades are either dramatically redefined, Crassitarsae, Theraphosodinea, Domiothelina, to preserve monophyly, or flatly rejected, such as Rastelloidina, while proposing novel clades such as Bipectina and Euctenizoidina.

Phylogenetic structure, and hence the validity of current genera, is completely unknown within the earliest spider branch, the Mesothelae, with the single extant family, Liphistiidae.

Since 2005 some of the ‘unplaced families’ (see Figure 10), have now been included in phylogenetic analyses. Chummidae, Hahniiidae, Homalonychidae, and Cybaeidae were all included in the studies by Miller et al. (2010) and Agnarsson et al. (2013), and are included here (Figure 11). All belong to the RTA clade, and Hahniiidae and Cybaeidae appear polyphyletic as currently circumscribed. The phylogenetic position of Cycloctenidae, Halidae, Synaphridae, and the recently discovered Trogloraptoridae (Griswold et al. 2012) remain unknown.

While the deeper level phylogenetic structure is thus poorly known in spiders, we may ask if the picture looks clearer at lower taxonomic levels, such as families and genera. This does not seem to be the case. There are certainly groups that have never seriously been contested since they were proposed, including morphologically highly distinct families such as Liphistiidae, Dysderidae, Scytodidae, Pholcidae, Salticidae, Deinopidae and others. Also, within groups that were the early focus of study of molecular phylogenetics, much of the ‘noise’ had already been sorted out; families within Araneoidea are probably by and large monophyletic as currently circumscribed. However, even within Araneoidea recent analyses again highlight historically controversial issues, such as the placement of the araneid Arkys with tetragnathids, the placement of mimetids among tetragnathids, the exclusion of nephilids from tetragnathids, the definition of Synaphridae, Anapidae, Symphytognathidae, and others.

Even more issues are apparent in groups that only recently are being analyzed using molecular evidence. For example within the RTA clade, Miller et al. (2010) found that the families Ageelenidae, Amaurobiidae, Cybaeidae, Dictynidae and Hahniiidae were not monophyletic as currently circumscribed. Further, in a recent first detailed molecular study of Mygalomorphae (Bond et al., 2012b) the monophyly of Mecicobothriidae, Hexathelidae, Cyrtarachnidae, Nemesiidae, Ctenizidae and Dipluridae was refuted, and the new family Euctenizidae was erected.

The same story may apply to even lower level studies at the genus level. Though we do not review that literature here, certainly many genera routinely turn out to be para- or polyphyletic. For example, Agnarsson (2004,
2006b) transferred species from the genus Anelosimus to no less than five other genera, and to Anelosimus from an additional three genera, following the first phylogenetic analyses of the genus. Similarly, reclassification routinely follows new revisionary and phylogenetic work at the genus level, as exemplified in recent papers, for example on Araneus (Framenau et al., 2010a), Pholcus (Huber, 2011b), and deinopids (Coddington et al., 2012), to name but a few. Historical taxonomy was often inconsistent, Jäger (2010), for example, discovered that from a series of specimens of a single species, collected from a single locality, Strand (1911) described two new monotypic genera of Pisauridae.

A novel molecular analysis of spider families

Rather than summarizing existing molecular phylogenies in a supertree, we chose here to assemble and analyze a family-level matrix from Genbank. This new phylogeny will be questionable due to missing data and the unavoidable use of loci that have performed poorly, judged either by replicability or independent corroboration, at reconstructing deeper, inter-familial phylogenetic nodes. It disputes some of the ‘indisputable’ clades mentioned above, such as Araneomorphae and Haplogynae. It implies, once again, that available molecular data remains inadequate to recover deep phylogenetic nodes in spiders, even though dense taxon coverage with molecular data is clearly more feasible than with morphology. Of course, comparative morphology may also mislead, but at least the empirical facts and their relative generality seem stable, even if interpretation and analysis are controversial. Molecular data, in contrast, may seem more problematic at the moment, but the potential to contribute more, and more sophisticated, data in the future is undeniable. Regardless, the list of deep spider clades robustly supported by both morphological and molecular data is very short indeed.

As noted above, most of the available loci are expected to be more informative at relatively shallow rather than deep genetic divergences, so that implied relationships will be relatively less reliable. The problem is exacerbated by poor overlap of genes and gene regions available for different groups of spiders in Genbank, resulting in molecular matrices with many missing data. Despite these drawbacks, the kind of analysis we present here, drawn from synthetic databases and relying on the work of dozens, if not hundreds of colleagues, represents the foreseeable future of comprehensive phylogenetic analysis. The current attempt highlights the shortcomings of available loci and unsynchronized sampling effort. The results do not strongly test morphological phylogeny. They do illustrate where we currently stand with respect to molecular data, and thus can guide us in the collection of new data to solve this deep phylogenetic question. It is our hope that repeating this practice some 5–10 years from now when larger datasets have become routine, will result in
Figure 11. Molecular phylogeny of spider families based on analysis of ten loci and 136 terminal taxa representing 98 spider families. Colours represent major groups based, in part, on morphology. Node numbers are posterior probability values, in many cases very low due apparently to missing data – analyses including fewer taxa with less missing data recovered largely the same results but with higher nodal support in general (Figure 12). The tree differs starkly from the consensus knowledge view (Figure 10), see text for discussion. In bright yellow are taxa that appear conspicuously misplaced based on recent molecular and total evidence analyses, in this case Uloboridae and Deinopis.
a more credible and strongly supported hypothesis, and thus demonstrate rapid progress in the phylogenomic era.

For a brief outline of phylogenetic methodology, see Appendix 1. Our analyses based on ten nuclear and mitochondrial loci were clearly constrained by missing data ranging from around 30% (*Aphonopelma*, Theraphosidae) to almost 94% (Cyatholipidae) in the full analysis. Accordingly, clade support was very low in the full analysis and highest in the 81 taxa analysis where the 55 taxa with the least data were removed. To repeat, the reconstruction of 400 million years of spider evolution at the family level with fragmental samples of loci more appropriate for much more recent divergences, such as relationships among species or genera, is obviously problematic. However the basic strategy of this kind of summary molecular analysis is sound, and will be repeated, no doubt, many times in the future. It behooves us therefore to assess available data, and what they tell, and do not tell us.

The results of the analysis of the entire dataset of 136 taxa by ten loci (Figure 11) differ dramatically from the consensus view expressed in Figure 10. Nodal support is very low for many of these contrasting clades, however, the same clades are in most cases recovered with higher support in analyses of the smaller datasets where missing data is less of a problem (Figure 12). We here highlight the few points of agreement between the consensus view and this phylogeny, as well as the most controversial disagreements. The phylogeny does not effectively test Mesotheleae vs. Opisthotheleae – to do that, one would need to sample taxa outside of spiders – but is at least consistent with these groups. Other clades in notable agreement with the consensus view are Entelegynae, Mygalomorphae, and the recently re-circumscribed Araneoidea, the RTA clade, and Palpimanioidea. Palpimanioidea here includes only palpimanids, araneoids and mecyssmauchenii, corroborating the transfer of the remaining members of the recent Palpimanioidea to Araneoidea. Some other smaller clades are in partial agreement with morphology such as higher lycosoids, albeit now containing Thomisidae, and Zodarioidea. Many, though not nearly all, of the tested families are also monophyletic.

This phylogeny differs from the consensus view, partly because the available DNA data are inadequate to resolve this deep level phylogeny, but also because the consensus view is doubtless incorrect in some aspects. It is intriguing to realize that we cannot say which of these two causes dominates. Clearly some of the relationships suggested in Figure 11 go against all prior evidence and can be presumed to be false, e.g. the placement of *Uloborus* as sister to the RTA clade. On the other hand, as highlighted in Figure 10, we

Figure 12. Results from molecular analysis of the 81 taxa set, where the 55 taxa with least available data from Figure 11 have been removed. In addition to *Deinopsis* and *Uloborus*, here Pholcidae is clearly misplaced, possibly as a result of long branch attraction; it moves within Haplogyneae when another pholcid is included in the analysis (Figure 11). See legend for Figure 11 and text for details.
also know that recent phylogenetic tests have contradicted the consensus view, broadly and in detail. We therefore discuss the differences between Figures 10 and 11 to highlight the most urgently needed tests in spider phylogeny.

Most dramatically, Figure 11 does not recover Araneomorphae nor Haplogynae. Haplogynae is a paraphyletic grade leading to Mygalomorphae. Although unexpected, to say the least, the question has never been asked at this scale with molecular data. Failure to recover Araneomorphae could readily be explained by root placement within Opisthotheleae. Root placement is a difficult issue in deep level phylogenetics, and here Mesothelae is represented only by two species, both of which lack available sequences for many of the genes. Thus, if the root is drawn between Mygalomorphae and the remaining Opisthotheleae, our results resemble much more the consensus view (Figure 13). Furthermore, the taxon thought to represent the basal split of Araneomorphae, Hypochilus, is placed with relatively high support (Figure 12) within the paraphyletic Haplogynae. A surprising placement as, even though Hypochilus has haplogyne genitalia and resembles in many other ways various members of the classical Haplogynae, it shares two pairs of book lungs with Mygalomorphae and Mesothelae, and lacks the hypothesized morphological synapomorphies of Haplogynae, such as fused chelicerae.

Another unexpected result is the recovered polyphyly of Austrochilidae, and the placement of the haplogyne Leptonetidae with part of Austrochiloidea. We note however, that Austrochilus is a long branch that was not stable in the analyses, and for example ‘attracted’ another long branch, Pholcidae, in the three smaller analyses (see Figure 12).

We failed to recover Orbiculariae, which is not surprising given that no prior molecular analysis has recovered it with strong support. The results do not support Eresoidea, in agreement with other recent molecular studies (Miller et al., 2010; Agnarsson et al., 2013). We do recover the RTA clade, however, within it no major ‘consensus clade’ is supported, with results differing dramatically between the current tree and morphology, as well as prior molecular studies, to the extent that results can be compared due to very different taxon sampling.

Perhaps most surprising is the apparent polyphyly of Dionycha, however, very few of the results within the RTA clade are well supported. We do, nonetheless, always recover Thomisidae within the higher Lycosoids, close to Oxyopidae with relatively high support (Figure 12). Given that within the RTA the results of the pruned matrix are in much greater agreement with morphology than those of the full matrix may indicate that missing data present a major problem in our analysis of the RTA clade. Yet, other recent molecular studies have also contrasted strongly with morphology, suggesting that the phylogenetic structure within the RTA clade will likely be dramatically changed as stronger molecular data accumulate (Miller et al., 2010; Agnarsson et al., 2013).
Figure 13. An overview of the results obtained here, when Opistothelae is rerooted between Mygalomorphae and the remaining Opistothelae. The particularly unstable placement of Eresidae and Deinopoidea is indicated with a polytomy. This rerooted version is more similar to the consensus view (Figure 10) and recent morphological analysis (Griswold et al., 2005), however, both Haplogynae and Austrochiloidea remain paraphyletic.

In sum, relatively few parts of the spider tree of life received robust support from morphological and molecular data, most notable among these being Mesothelae, Opistothelae, Mygalomorphae, Entelegynae, Araneoidea, and the RTA clade. Much of the remaining spider tree remains to be resolved. Currently available molecular data are insufficient to construct a robust spider tree of life, thus future work including stronger molecular data and revised morphology is urgently needed.

Prospects: choice of loci and synchronization of effort

Although the above results are questionable, we reiterate that this sort of analysis – all available markers for all available taxa – will be regularly repeated for the foreseeable future, even as less inclusive and comprehensive analytical strategies appear. For future progress, it is therefore important to make efforts to synchronize and standardize data collecting. This approach
will synergistically benefit all researchers interested and involved in spider phylogenetics. Such effort could, for example, start by providing a list and protocols of universal primers, and suggest priorities for primer choices for maximum overlap among studies (see e.g. http://www.islandbiogeography.org/uploads/6/6/8/0/6680387/primers.doc). The issue of data overlap will predictably grow exponentially as spider phylogenetics enter the genomics era and choice of data becomes broad. It would be unfortunate if the new injection of large datasets in spider phylogenies from NGS would come hand in hand with reduced overlap in data collected by different laboratories. However, this seems likely to happen given that it appears that laboratories in many cases are designing their own pipelines, and resulting phylogenetic datasets may have limited overlap among different labs. It is urgent to establish a dialogue and a resource centre where systematists can share experiences and maximize comparability of the data that will be generated in the coming decades. This kind of effort could be centralized around an international organization such as the International Society of Arachnology (ISA: http://www.arachnology.org/).

A more synergistic approach to data sampling could more strongly evaluate the utility of different loci or NGS pipelines. Given that it will be some time yet until phylogenomic studies become affordable on large datasets, such as biogeographic or phylogeographic studies involving hundreds, or thousands of specimens, it is likely that analyses of multiple loci obtained from standard genes (but new primers) will remain the best option. Evaluating the performance of loci is still important. Identifying the ‘magical’ set of genes depends on the taxonomic question. Some of the ‘usual suspects’ may robustly resolve lower level phylogenetics. As an example, ITS2 and COI, two readily amplified and sequenced markers representing the nuclear and mitochondrial genomes, appear to resolve population level phylogeographic (Kuntner & Agnarsson, 2011b), and species level biogeographic patterns (Kuntner & Agnarsson, 2011a) in Nephila and Nephilengys, respectively. At this shallow level nuclear and mitochondrial data strongly agree, perhaps implying that results may not change with additional data.

At intrageneric levels, individual gene trees sometimes conflict. In social Anelosimus Agnarsson et al. (2007) found that the placement of the highly social A. eximius, a key species, differed among gene trees, even differences between two protein coding mtDNA loci, with important implications for the evolution of sociality. Discordance among gene trees typically then continues to increase as studies tackle deeper level splits such as among genera within a family, or among related families (e.g. Arnedo et al., 2004, 2007, 2009b; Arnedo & Gillespie, 2006; Blackledge et al., 2009a; Miller et al., 2010a; Bodner & Maddison, 2012).

Combined analyses of multiple markers sometimes result in credible phylogenetic estimates as judged by congruence with morphological and
behavioural data (e.g. Arnedo et al., 2004; Blackledge et al., 2009a). But if the currently utilized loci continue to display the pattern of increasing discordance with phylogenetic depth, they may not be able to robustly solve the deepest level spider phylogenetics (see below and Figure 11). Thus, future progress will lie in adding to our current toolkit; the range of loci and other types of molecular data need to expand immensely in the next decade as we enter the era of next generation sequencing.

Utility of phylogenies

Considering only the literature since 2000, spider phylogenetic studies are increasingly used to test evolutionary hypotheses. Examples from the spider literature of the use of phylogenies to pose and test phylogeographic (Bidegaray-Batista et al., 2007; Starrett & Hedin, 2007; Su et al., 2007; Cooper et al., 2011; Kuntner & Agnarsson, 2011b), biogeographic (Crews & Hedin, 2006; Garb & Gillespie, 2006; Hendrixson & Bond, 2007; Wood et al., 2007; Dimitrov et al., 2008; Macias-Hernandez et al., 2008; Wang et al., 2008; Crews & Gillespie, 2010; Rix et al., 2010; Bidegaray-Batista & Arnedo, 2011; Kuntner & Agnarsson, 2011a; Su et al., 2011; Rix & Harvey, 2012), ecological (Arnedo & Gillespie, 2006; Opell, 2006; De Busschere et al., 2010; Spagna et al., 2010; Satler et al., 2011), behavioural (Kuntner & Agnarsson, 2009; Kuntner et al., 2009; Agnarsson, 2012; Kralj-Fišer & Kuntner, 2012), palaeontological (Penney et al., 2003; Penney, 2004) and evolutionary hypotheses and scenarios (Garb & Hayashi, 2005; Stratton, 2005; Agnarsson, 2006a; Agnarsson et al., 2006, 2010; Johannesen et al., 2007; Blackledge et al., 2009b; Hedin & Lowder, 2009; Kuntner & Coddington, 2009; Cheng et al., 2010; Kuntner et al., 2010; Higgins et al., 2011; Dimitrov et al., 2012; Ledford et al., 2012). Additionally, several recent studies have used molecular or combined data in order to discover and describe unknown diversity, or delimit species (Griffiths et al., 2005; Hendrixson & Bond, 2005; Johannesen et al., 2005; Bond & Stockman, 2008; Vink et al., 2008b; Duncan et al., 2010; Macias-Hernandez et al., 2010; Hamilton et al., 2011; Hedin & Carlson, 2011; Kuntner & Agnarsson, 2011a). Phylogenies are powerful tools, particularly if they draw from several independent sources and if they are time calibrated. In the near future, the power of phylogenetics will rapidly increase through transcriptomics and phylogenomics. Early indication of this is already seen in the spider literature (Garb et al., 2007; Prosdocimi et al., 2011; Mattila et al., 2012a).

Conclusions

Although we have made enormous progress since 2005 in taxonomic discovery and in morphological, molecular and combined phylogenetic research on
spiders, most major suprafamilial clades of spiders have yet to survive rigorous testing. A few pillars of knowledge emerge: clades supported by morphological and molecular analyses alike. We highlight the large task ahead by identifying the major instabilities in spider phylogenetic research, and by initiating the workhorse analysis of the future: all Araneae and all markers, filtered as seems wise by current standards of data quality.

We further hope to inspire laboratories to work towards synchronizing the global efforts and towards making the data comparable, alignable, and thus usable. We may all be moving towards phylogenomics, but we believe there is still room for rapid improvement of the utility of the usual suspects, the genes commonly used for amplification in population, species, and higher taxon phylogenetics. Hence, spider molecular phylogenetics will mature from the “middle-aged infant” to a teenager. We also urge researchers not to entirely bypass and neglect the more classical character systems, because morphology, ecology and behaviour will continue to supply us with critical evidence of homology versus convergence. Furthermore, obtaining a robust ‘tree’ is a goal in itself, however, cladograms are only truly useful to the extent that we know something about the morphology and the biology of its constituent taxa, and thus can use them to study evolutionary and ecological processes. Spider systematics remains largely contentious and in need of intensive and synchronized research. A robust understanding of spider phylogeny will, eventually, provide us with opportunities to test a myriad of ecological and evolutionary questions.

Acknowledgements

We thank the following colleagues for their estimation of known and unknown diversity of spiders: Holger Frick, Charles Haddad, Bernhard Huber, Peter Jäger, Rudy Jocqué, Wayne Maddison, Fernando Alvarez-Padilla, Norman Platnick, Robert Raven, Adalberto Santos and Joseph Spagna. We also thank Jose Sanchez and Angela Alicea for data mining. Nadine Dupérré kindly supplied the art appearing in Figure 9. Thanks to Federico Lopez-Osorio, Anne McHugh, Heine Kiesbüty, Zamira Yusseff Vanegas, Matjaž Gregorič, and Ren-Chung Cheng for comments on a draft of the manuscript. Funding for this research came from NSF (DEB-1314749) to IA, and ARRS (P1 – 0236 to MK and BI-US/09-12-016 to MK and JC).

References


of California (Mygalomorphae, Antrodiaetidae, Aliatypus coylei, sp. nov.), including consideration of mitochondrial phylogeographic structuring. Zootaxa, 2963: 55–68.


Jäger, P. 2010. *Papakula* and *Hesydrimorpha*: how two spider genera were described from the same species collected from the same locality (Araneae: Pisauridae). Zootaxa, 2551: 65–68.


Linné, C. 1758. Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species cum characteribus differentiis, synonymis, locis. Editio decima, reformata. [Publ. not given], Holmiae, tomos 1; 1–821.
Lopardo, L. 2005. Phylogenetic revision of the spider genus Negryan (Araneae, Anyphaenidae,


Russell Smith, A. & Stork, N.E. 1994. Abundance and diversity of spiders from the canopy of
Appendix 1. Methods for molecular analyses

Accession numbers for the ten genes most commonly used in spider systematics (COI, 28S, ND1, H3, 18S, 16S, 12S, Gamma, Actin, Wingless), for all spider species deposited in Genbank, close to 9000 entries, were assembled into an Excel spreadsheet. Taxon representatives of spider families were then chosen based on gene coverage. A family level taxon represents a single taxon, or a chimera, constructed to maximize gene coverage (Supplementary Table and further detail available from the authors). All families of spiders with any representation in Genbank were included. For each family, a second taxon was added if a single terminal, or single genus chimera could be constructed to contain at least four of the ten genes. Data were downloaded from Genbank through the evolutionary analysis software Mesquite (Maddison & Maddison, 2011). Each gene matrix was aligned using MAFFT (Kazutaka et al., 2005) through the EMBL-EBI server (http://www.ebi.ac.uk/Tools/msa/mafft/) with gap opening and gap extension penalties set at default (1.53 and 0.123, respectively), except in protein coding genes where gap opening penalty was set to maximum (3). Alignment was run with 100 tree rebuilding replications and 100 maximum reiterations. The matrices were then pruned at 5’ and 3’ ends to eliminate regions shared by less than 10% of the taxa. The ND1 matrix was further trimmed eliminating a region of gappy alignment containing stop codons at the 5’ end, and the Gamma matrix with such a region was eliminated at the 3’ end. The matrices were then concatenated in Mesquite, and exported for analysis. We analyzed the full matrix, containing 136 terminals, as well as smaller subsets of the data (120, 100, and 81 taxa) pruned to exclude the taxa with most missing data. Phylogenetic analyses were partitioned by genes and within protein coding genes by codon position, and a partitioned Bayesian analysis was run for 40,000,000 generations for the larger matrices (136 and 120 taxa) and 30,000,000 for the smaller matrices (100 and 81 taxa). Burnin was determined through analyses of .p files in Tracer (Drummond & Rambaut, 2007). Stationarity was in no analysis reached before the typical analysis length of 10,000,000 generations, but after about 20,000,000 generations in the 81 taxa analysis, 23,000,000 generations in the 100 taxa analysis, 25,000,000 generations in the 120 taxa analysis, and 12,000,000 in the full analysis.