Biological Journal of the Linnean Society, 2004, **81**, 79–89. With 3 figures

Mitochondrial differentiation, introgression and phylogeny of species in the *Tegenaria atrica* group (Araneae: Agelenidae)

P. J. P. CROUCHER†, G. S. OXFORD* and J. B. SEARLE

Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

Received 20 August 2002; accepted for publication 28 July 2003

The relationships between the three members of the *Tegenaria atrica* group (*T. atrica*, *T. saeva* and *T. gigantea*) were examined with DNA sequence data from mitochondrial CO1, 16S rRNA, $tRNA^{leu(CUN)}$ and ND1 genes. Members of this group of large house spiders have overlapping distributions in western Europe and hybridize with each other to a variable degree. The close relatedness of all three species was supported by all analyses. *T. saeva* and *T. gigantea* are more closely affiliated than either is to *T. atrica*. Haplotypes clearly assignable to *T. gigantea* were also present in many specimens of *T. saeva*, suggesting asymmetrical introgression of mtDNA from *T. gigantea* into *T. saeva*. Molecular clock calibrations (CO1) suggest that deeper divisions within the genus *Tegenaria* may be in excess of 10 million years old, and that the evolutionary history of the *T. atrica* group has been moulded by Quaternary glacial–interglacial cycles.

ADDITIONAL KEYWORDS: 16S rRNA – cytochrome oxidase 1 – hybridization – large house spiders – molecular clock – nicotine adenine dinucleotide dehydrogenase 1 – speciation – $tRNA^{lue(CUN)}$.

INTRODUCTION

One way to understand better the processes of speciation is to investigate the interactions and relationships between closely related species or forms within species. Of particular interest are those taxa that are distinct enough to be defined as separate species on the basis of morphology yet which still hybridize. These cases represent the final stages of speciation according to the Biological Species Concept (BSC) (Mayr, 1942). The BSC is nowadays generally couched in terms of gene flow, such that species are considered to be groups of populations that are reproductively isolated from other such groups by genetically based traits that prevent gene exchange (isolating mechanisms) (Coyne & Orr, 1998). Molecular analyses of the relationships between forms that exhibit some history of hybridization can be used to estimate how long the 'species' have been diverging and provide insights into present and past interactions – hybridization may leave a molecular signature. Species with these characteristics are the large house spiders *Tegenaria atrica*, *T. saeva* and *T. gigantea*, which comprise the *Tegenaria atrica* group (*Tegenaria*, Agelenidae) (Merrett, 1980; Maurer, 1992). Hybridization between members of this group had long been suspected because of the occasional observation of individuals with intermediate morphologies (e.g. Locket, 1975; Merrett, 1980).

In continental Europe, *T. atrica* is widely distributed (Maurer, 1992) but is only very occasionally found, as an introduction, in Britain (Harvey, Nellist & Telfer, 2002). *T. saeva* and *T. gigantea* are both apparently confined to the west coast of continental Europe and both are frequently found in Britain. They occupy largely allopatric distributions across much of England and Wales but in northern England their ranges overlap to a greater extent (Merrett, 1980; Oxford & Chesney, 1994; Croucher, 1998; Harvey *et al.*, 2002). In the north-east English county of York-

*Corresponding author. E-mail: gso1@york.ac.uk
†Present address: 1st Department of Medicine, Christian-Albrechts University, Kiel, D-24105 Kiel, Germany.

shire the species are broadly sympatric and experience widespread hybridization (Oxford & Smith, 1987; Oxford & Plowman, 1991; Croucher, 1998). The greater overlap in the north of England is probably the result of a recent northwards colonization of the two species, which were first recorded in Yorkshire in the 1970s (Smith, 1985), and in the north-west English county of Cumbria at about the same time (Parker, 1984). Croucher (1998) has demonstrated an apparently long-standing boundary between the two species in central southern England (Dorset). Here, the frequency of hybrids is lower than in Yorkshire and the boundary appears to be a type of hybrid zone known as a 'tension zone' (Barton & Hewitt, 1985), with hybrid individuals experiencing significantly reduced fitness, as indicated by a marked increase in mortality rates in laboratory-generated hybrids compared with progeny from intraspecific matings reared under identical conditions (Croucher, 1998).

In order to understand fully the interactions between members of the *T. atrica* group, the significance of the tension zone in southern England and the dynamics within the area of sympatry in Yorkshire, it is important to clarify the phylogenetic relationships between the species. This is the purpose of the present paper. Spiders have been the subject of certain types of molecular analyses, notably phylogenetic and phylogeographical (e.g. Gillespie, Croom & Palumbi, 1994; Hedin, 1997, 2001; Piel & Nutt, 1997; Zehethofer & Sturmbauer, 1998; Garb, 1999; Gillespie, 1999; Hausdorf, 1999; Tan, Gillespie & Oxford, 1999), and taxonomic (Vink & Mitchell, 2002), but only rarely in the context of hybridization or speciation (e.g. Masta & Maddison, 2002). Here, two DNA sequence fragments from mitochondrial (mt) gene regions are examined: a fragment containing part of the cytochrome oxidase 1 (CO1) gene and a fragment consisting of the 3′ end of the 16S rRNA gene, the entire $tRNA^{leu(CUN)}$ gene and the 5′ half of the nicotine adenine dinucleotide dehydrogenase 1 (ND1) gene. Three additional *Tegenaria* species were included in the initial analyses to stabilize the phylogeny. The resulting phylogenetic information is used to elucidate asymmetries in hybridization between *T. saeva* and *T. gigantea* and to estimate divergence times for members of the *T. atrica* group.

MATERIAL AND METHODS

Samples and identification

All specimens used in this study were identified by P.J.P.C. and independently verified by G.S.O. Eighteen out of 24 samples of *T. saeva* and *T. gigantea* were collected from a broad area of southern England (Table 1, Fig. 1). In this region, the distribution of the two species is clearly defined with relatively allopatric populations of *T. saeva* occurring in the far south-west and allopatric populations of *T. gigantea* occurring in the far south-east (Fig. 1; Harvey *et al.*, 2002). Species identifications were made by examining the copulatory structures of the spiders (the shape of the conductor and tegulum of the male pedipalp and the shape and orientation of the receptacles and apophyses of the female epigyne) (see Merrett, 1980). In southern England, the separation of the two species based on these characters is straightforward and reliable and putative intermediates are only found at the species boundary (Fig. 1). Croucher (1998) demonstrated the robustness of this identification through discriminate function analyses trained on allopatric specimens and based on 20 character measurements in males (18 from the pedipalps) and 15 character measurements in females (11 from the epigyne). For 150 males and 150 females from southern England this classification schema concurred fully with the visual identification. Additionally, two specimens of *T. saeva* and one putative intermediate were sampled from the area of sympatry in Yorkshire, in the vicinity of the city of York. Three further specimens originated from other countries, two specimens of *T. saeva* from Nancy in eastern France and one specimen of *T. gigantea* from Washington State, USA.

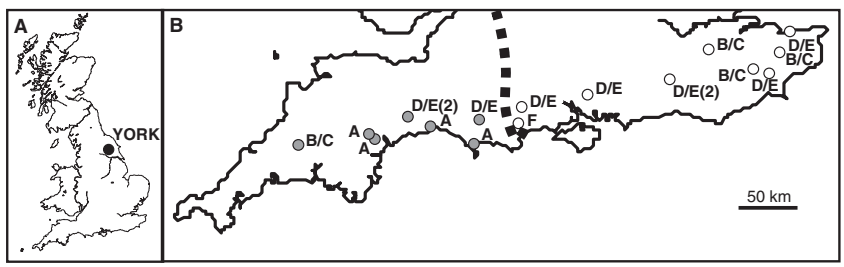

Figure 1. Map indicating (a) the location of York, Yorkshire (sympatry), and (b) the distribution of spider samples, their inferred haplotypes (A, B/C, D/E, F), and the location of the species boundary in southern England (dashed line). Open circles = *Tegenaria gigantea*. Shaded circles = *T. saeva*.

Table 1. List of *Tegenaria* specimens, with locality, haplotype and gene regions sequenced

Species	Collection locality	Haplotype	CO1	ND1
T. agrestis	Liverpool, N England	*T. agrestis*	✓	–
T. domestica	Axmouth, Devon, SW England	*T. domestica*	✓	–
T. domestica	Honiton, Devon, SW England	*T. domestica*	✓	–
T. domestica	York, N England	*T. domestica*	✓	–
T. parietina	Oare, Kent, SE England	*T. parietina*	✓	–
T. parietina	Pavia, Italy	*T. parietina*	✓	–
T. parietina	Pavia, Italy	*T. parietina*	✓	–
T. atrica	York, N England	*T. atrica*	✓	–
T. atrica	Castleknock, Co. Dublin, Eire	*T. atrica*	✓	✓
T. atrica	S France	*T. atrica*	✓	✓
T. saeva	Nancy, NE France	A	✓	–
T. saeva	Axmouth, Devon, SW England	A	✓	–
T. saeva	Weymouth, Dorset, S England	A	✓	–
T. saeva	Shobrooke, Devon, SW England	A	✓	✓
T. saeva	Nancy, NE France	A	–	✓
T. saeva	Newton St. Cyres, Devon, SW England	A	–	✓
T. saeva	Trewen, Cornwall, SW England	B	✓	✓
T. saeva	Cheselbourne, Dorset, S England	D/E	✓	–
T. saeva	Tadcaster, York, N England	D/E	✓	–
T. saeva	Acomb, York, N England	D/E	✓	–
T. saeva	Church Green, Devon, SW England	D/E	✓	–
T. saeva	Church Green, Devon, SW England	E	–	✓
T. gigantea	Broadstone, Dorset, S England	F	✓	–
T. gigantea	Dunton Green, Kent, SE England	B/C	✓	–
T. gigantea	Biddenden, Kent, SE England	C	–	✓
T. gigantea	Godmersham, Kent, SE England	C	✓	✓
T. gigantea	Snohomish Co., WA, USA	C	✓	✓
T. gigantea	Whitstable, Kent, SE England	D	✓	✓
T. gigantea	Southwater, W. Sussex, S England	D/E	✓	–
T. gigantea	Southwater, W. Sussex, S England	D/E	✓	–
T. gigantea	Swanmore, Hampshire, S England	D/E	✓	–
T. gigantea	Horton, Dorset, S England	D/E	✓	–
T. gigantea	Woodchurch, Kent, SE England	E	✓	✓
T. gigantea/*saeva*	Wheldrake, York, N England	D/E	✓	–

Haplotypes for *T. saeva* and *T. gigantea* were designated A–F. Haplotypes B and C, and D and E, were only resolvable with the ND1 (16S rRNA-tRNA$^{leu(CUN)}$–ND1) data. The specimen *T. gigantea/saeva* was a putative hybrid on the basis of morphology.

MOLECULAR PROCEDURES

DNA was extracted from specimens preserved in 70% or 95% ethanol, or frozen at –80°C. Mitochondrial DNA was recovered from a single leg (leg IV) using a modification of the NaCl-extraction technique of Medrano, Aasen & Sharrow (1990), as used in spiders by Gillespie *et al*. (1994). A 325-bp sequence of the cytochrome oxidase 1 gene (CO1), corresponding to positions 1777–2101 of the *Drosophila yakuba* mtDNA (Clary & Wolstenholme, 1985), was amplified using primers C1-J-1718 and C1-N-2191 (Nancy). A 543-bp sequence comprising the 3′ end of the 16S rRNA gene, the tRNA$^{lue(CUN)}$ gene and the 5′ half of the nicotine adenine dinucleotide dehydrogenase 1 gene (ND1), corresponding to positions 12295–12913 of the *D. yakuba* mtDNA (Clary & Wolstenholme, 1985), was amplified using primers N1-J-12261 and LR-N-12945 (N116S). Primers were as described by Simon *et al*. (1994) except N1-J-12261, which followed Hedin (1997). LR-N-12945 differed from that given by Simon *et al*. (1994) in that the initial G from the 5′ end was deleted (M. Hedin, pers. comm.). The specimens sequenced for each of the mtDNA regions are indicated in Table 1.

The PCR reactions were performed in an M.J. Research PTC-100 programmable thermal cycler. The optimized temperature profile for the CO1 fragment

consisted of 35 cycles of 30 s denaturation at 95°C, 1 min annealing at 40°C, a further 30 s annealing at 50°C and 2 min extension at 72°C. For the 16S rRNA–tRNA$^{lue(CUN)}$–ND1 fragment the temperature profile consisted of 35 cycles of 30 s denaturation at 95°C, 30 s annealing at 47°C and 90 s extension at 72°C. In both cases there was an initial denaturation step of 2 min at 95°C. PCR products were checked for size and lack of multiple bands by agarose gel electrophoresis and purified with the Wizard PCR Preps DNA Purification System (Promega) according to the manufacturer's protocols.

Sequences were determined directly, in both the 5′ and 3′ directions, using the same primers as above. Sequencing was either manual, using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical/Amersham International plc.) and ^{35}S dATP (Du Pont NEN) with autoradiographic visualization, or automated, using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS (Perkin Elmer) with separation on an ABI PRISM 377 DNA Sequencer (Perkin Elmer).

Sequences were aligned using PILEUP (Genetics Computer Group, 1994) and CLUSTAL W version 1.7 (Thompson, Higgins & Gibson, 1994). Sequences were numbered following the system for the *D. yakuba* mtDNA (Clary & Wolstenholme, 1985). Comparison with the *D. yakuba* mtDNA and with sequences deposited in the databases suggested that the sequence fragments were from functional CO1, ND1, tRNA$^{lue(CUN)}$ and 16S rRNA genes and no missense or stop codons were detected when the coding regions of CO1 and ND1 were translated using the invertebrate mitochondrial translation table. Alignment of the 16S rRNA–tRNA$^{lue(CUN)}$–ND1 sequence (hereafter referred to simply as ND1) with published sequences for other spiders (*Nesticus tennesseensis*, AF004653, and *Habronattus pugillis*, AF239933) showed good agreement and indicated that the *T. atrica* group share the same reduction in the TΨC arm of the tRNA$^{lue(CUN)}$ as has been described in these species (Masta, 2000). More detailed alignments and a prediction of the secondary structure of the tRNA$^{lue(CUN)}$ molecule are available from the authors or via the internet (pcroucher.net/tegenaria). The overall consistency of the sequence data, absence of multiple bands in the PCR reactions, and strong alignment at both the DNA and amino acid sequence level with published sequences indicate that our data reflect genuine mitochondrial sequences and not nuclear pseudogenes as have now been described in many species (Bensasson *et al*., 2001). All sequences are available from GenBank, with accession numbers: ND1, AY138816–AY138822; CO1, AY138827–AY138836.

Phylogenetic analysis

All phylogenetic analyses used PAUP* (Swofford, 1998). Maximum parsimony (MP) analyses were conducted using exhaustive searches to find the shortest trees. The empirically derived transition : transversion (TS : TV) ratio was used for phylogenetic reconstruction. This was evaluated by using the 'state changes and stasis' option of MacClade (Maddison & Maddison, 1992) to count the average number of TS and TV events on 1000 equiprobable random trees (Halanych, 1996; Halanych & Robinson, 1997; Halanych *et al*., 1999). In addition, equal (1 : 1), 2 : 1 and 10 : 1 weighting strategies were also tried. For maximum likelihood (ML) analyses, the most appropriate model of DNA evolution for the data was established using the nested hierarchical likelihood tests of MODELTEST (Posada & Crandal, 1998) on an initial neighbour-joining tree. For all data, CO1, ND1 and combined CO1 and ND1 data, the Tamura–Nei model (TrN: Tamura & Nei, 1993) was chosen. Gamma corrections of 0.2277, 0.0088 and 0.0160 were utilized for the CO1, ND1 and combined CO1 and ND1 data, respectively. Additionally, the commonly used Kimura two-parameter model of substitution (K2P: Kimura, 1980) was employed on the basis that more sophisticated models do not necessarily result in better phylogenetic inference (Nei & Kumar, 2000). Various transition : transversion ratios were employed as for the MP analyses.

For all trees, bootstrap support for the nodes was estimated using 1000 pseudo-replicates using a branch-and-bound algorithm.

The CO1 gene was tested for clock-like behaviour by calculating the log-likelihood score of the selected model of substitution both with and without the clock enforced using PAUP* (Swofford, 1998). The scores were compared in a likelihood ratio test with 'number of taxa – 2' degrees of freedom using the calculator function of MODELTEST (Posada & Crandall, 1998).

RESULTS

The distribution of mtDNA haplotypes among the individuals collected is given in Table 1 and Figure 1. The variable positions in each haplotype for the CO1 and 16S rRNA–tRNA$^{lue(CUN)}$–ND1 gene fragments are shown in Table 2. Each of the phylogenetic analyses that follow (CO1, ND1 and combined CO1 and ND1 data) resulted in single unambiguous ML and MP trees. The ML phylograms, with bootstrap values from ML and MP, are presented. In all cases, ML and MP trees had identical topologies and the alternative weighting strategies and models of substitution did not alter these topologies, or generally improve the bootstrap values and goodness of fit measures.

Table 2. Variable sites for all the CO1 and ND1 haplotypes detected in *Tegenariaa*

D. yakuba	11111111111111111111111111111111111111111111111111
numbering	77777777888888888888888888888888888888999999999999
	78899999000011222233344456677788888999900122233445
	92512458013689145706925873635812479014645736928470
	CO1
Tegenaria atrica	GATTCAATATATAGTAAGTAAAATTGTATTTTTTAATGTCTGGTAATGAT
A	....T...G....A...A..G...C.........G....T..A.G..A..
B	....T...G........A...G..C.........G......AA.G..AG.
C	....T...G........A...G..C.........G......AA.G..AG.
D	....T...G.......GA...G..C.........G......AA.G..A..
E	....T...G.......GA...G..C.........G......AA.G..A..
F	....T...G........A...G..C.........G......AA.G..AC.
T. agrestis	..G.TGTAG.G....G.A.GTG...A........G.........GT.AT.
T. domestica	.G.GT.TAGCT.G.GT.AGTTT.A....GA.AAGG.GAA...AAGG...G
T. parietina	A.A.TGTAGC.CG.GT.AA.G.G...GC.GCAAAGT.AG.GAA.GTAAG.
D. yakuba	11111111111111111122222222222222222222222222222222
numbering	99999999999999999900000000000000000000000000000001
	55566667778888999900011222333344455666667777889990
	13635892473679027846716258145704928124703692814 70
Tegenaria atrica	ATTTTTGTGTATTAAACTGAACGGATTAATTGAAGTTATAAATTAAGAT
A	..................A....A.....C.A.GA..............
B	........AC........A....A.......A.GA......G...G...
C	........AC........A....A.......A.GA......G...G...
D	..................A.............A.GA..........G...
E	..................A.............A.GA..........G...
F	........AC........A....A.......A.GA..........G...
T. agrestis	......A.T...GT......G.AA........G........T..T.A..
T. domestica	GG.A.GA.A.T.G.GGGATGGT.TGGAGG.CA.G...G...G..G..GG
T. parietina	..C.CAAAT.TGACG.GA..G.AT.......AG.AACTATGTAGGG.G.
D. yakuba	11111111111111111111111
numbering	22222222222222222222222
	33333333333444555667788 8
	00222345569334368044523 9
	03017751939452717281149 26
	ND1-tRNA$^{leu(CUN)}$-16SrRNA
Tegenaria atrica	CTCTTTATTTTTTCTACACTATACC
A	TCT.C...C.CC.TA.T.TCTCTGT
B	.CTC.CGC.C..C.AG.TTCTC..T
C	.CTC.C.C.CC.C.A..TTCTC..T
D	.CTC.CGC.CC.CTA..TT.TC..T
E	.CTC.CGC.CC.C.A..TT.TC..T

Position numbers for CO1 and ND1–tRNA$^{leu(CUN)}$–16S rRNA fragments follow the *D. yakuba* mitochondrial DNA nomenclature (Clary & Wolstenholme, 1985). Dots represent positions identical in sequence to *T. atrica*. CO1 haplotypes B and C, and D and E, are identical and are only resolved in conjunction with the ND1 data.

CO1

The aligned CO1 dataset from 30 individual spiders consisted of eight haplotypes. Of the 325 bp, 99 (30.5%) were variable and 38 (11.7%) were parsimony informative. Of the parsimony-informative sites, 73.7% occurred in the third position, 23.7% in the first position and 2.6% in the second position. When only the five haplotypes corresponding to the *T. atrica* group are considered, there were 25 (7.7%) variable

positions and six (1.8%) were parsimony informative (all in the third position).

The ML tree for the CO1 data is shown in Figure 2A. The equivalent MP tree had 137 steps, a consistency index of 0.883 and the TS : TV ratio was 1.2 : 1. Bootstrap support was 88% or more for all clades. The five CO1 haplotypes for the *T. atrica* group fell as a discrete group. We argue later (see Discussion) that the haplotypes shared by *T. gigantea* and *T. saeva* are actually of *T. gigantea* origin. On this assumption, pairwise distances among the species of the *T. atrica* group (TrN model of substitution) range from 0.0069 to 0.0916. *T. agrestis* fell as the sister taxon to the *T. atrica* group (the mean pairwise distance to the *T. atrica* group was 0.2040; range 0.1902–0.2179). Pairwise distances between the more basal taxa (*T. domestica* and *T. parietina*) and all other taxa were high (mean 0.8238; range 0.7019–0.9387). Because the *T. atrica* haplotype fell as a discrete sister group to those haplotypes associated with *T. saeva* and *T. gigantea*, it was consequently used as the outgroup for the ND1 and combined CO1 and ND1 analyses. Four haplotypes, A, B/C, D/E and F, were identified from the CO1 data in individuals of *T. saeva* or *T. gigantea*. The mean pairwise distance of these four haplotypes from that of *T. atrica* was 0.0819 (range 0.0685–0.0916). Haplotype

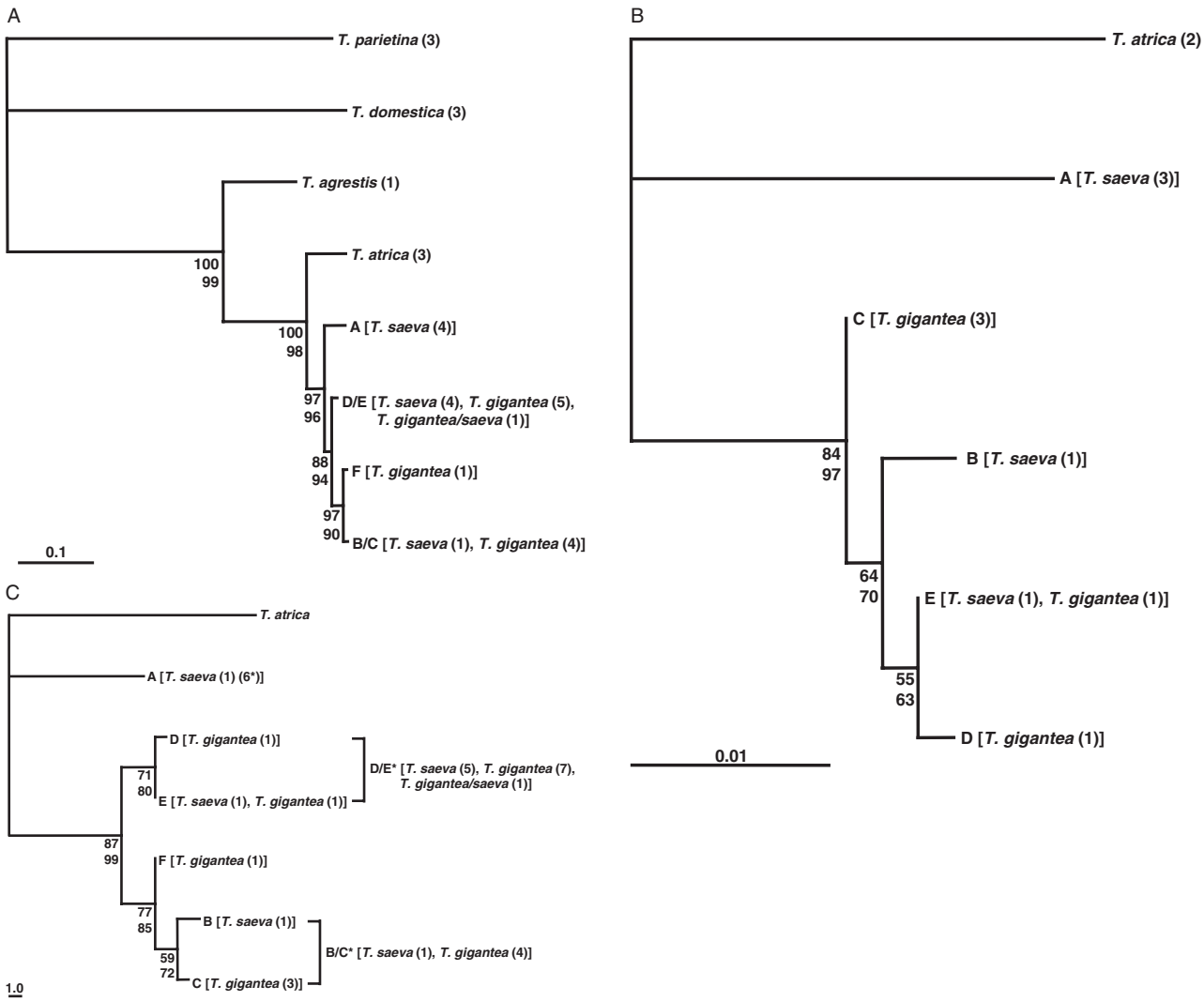

Figure 2. Maximum likelihood trees based on (A) CO1 sequences for the three members of the *Tegenaria atrica* group plus three other species of *Tegenaria*; (B) ND1 sequences for the three members of the *Tegenaria atrica* group; (C) combined CO1 and ND1 sequences for the three members of the *Tegenaria atrica* group. Numbers below the nodes indicate the percentage of 1000 bootstrap replicate samples recovering each clade, above for maximum likelihood, below for parsimony. For the *T. saeva* and *T. gigantea* haplotypes A–F the number of individuals of each species possessing that sequence is indicated after the haplotype. In (C), the total number of individuals of each species that can be inferred to possess haplotypes A, B/C or D/E is indicated (*).

A was uniquely observed in specimens of *T. saeva* and the mean pairwise distance of this haplotype from B/C, D/E and F was 0.0413 (range 0.0354–0.0458). Haplotypes B/C, D/E and F clustered together and the mean pairwise distance between them was only 0.0158 (range 0.0069–0.0219). Haplotype F was only observed once, in a specimen of *T. gigantea*.

ND1

The aligned ND1 dataset from 12 individuals of the *T. atrica* group consisted of six haplotypes. Of the 543 bp, 25 (4.6%) were variable, and ten (1.8%) were parsimony informative. Of the nine parsimony-informative sites occurring within the protein-coding region, six (66.7%) occurred in the third position, one (11.1%) in the first position and two (22.2%) in the second position.

The ML tree for the *T. atrica* group based on the ND1 data is shown in Figure 2B. The equivalent MP tree had 28 steps, a CI of 0.893 and the TS : TV ratio was 4.9 : 1. The ND1 data revealed a similar pattern to the CO1 data. Haplotype A (*T. saeva* only) appeared quite distinct from a cluster of four more closely related haplotypes: B, C, D and E. Bootstrap support for this arrangement was very strong. Pairwise distances (TrN model of substitution) among the haplotypes in the B, C, D and E clade were low, with a mean of 0.0110 (range 0.0023–0.0166). However, the mean pairwise distance between this clade and haplotype A was high, with a value of 0.3414 (range 0.2201–0.4890).

Combined CO1 and ND1 analyses

The combination of the CO1 and ND1 data, for the *T. atrica* group, resulted in seven composite haplotypes of 868 bp (with the exception of haplotype F for which no ND1 data were available and was therefore coded as missing for these bases). Of these 868 bp, 50 (5.8%) were variable and 19 (2.2%) were parsimony informative.

The ML tree for the combined CO1 and ND1 data is shown in Figure 2C. The equivalent MP tree had 56 steps, a CI of 0.911 and the TS : TV ratio was 9.2 : 1. The combined tree corroborated the arrangement of the *T. atrica* group that was revealed by the separate analyses, with high bootstrap support. The *T. saeva* haplotype A was again clearly distinct from the five haplotypes B, C, D, E and F. This latter cluster of haplotypes was further divided into two groups, illustrating the similarity between haplotypes D and E and haplotypes B, C and F. Indeed, haplotypes B and C, and D and E, were not separable on the basis of the CO1 data alone. The lineage that includes B, C, D and E is shared by both *T. saeva* and *T. gigantea*.

Figure 3 shows a parsimony network derived from the MP tree for the combined CO1 and ND1 data. The number of mutational steps between the haplotypes is indicated on the branches. This network clearly illustrates the difference between haplotype A, which was only found in specimens of *T. saeva*, and haplotypes B, C, D, E and F, which (with the exception of F which was only found in a single specimen of *T. gigantea*) were common to both *T. saeva* and *T. gigantea*. Haplotype A differed from these other haplotypes by an average of 31 substitutions.

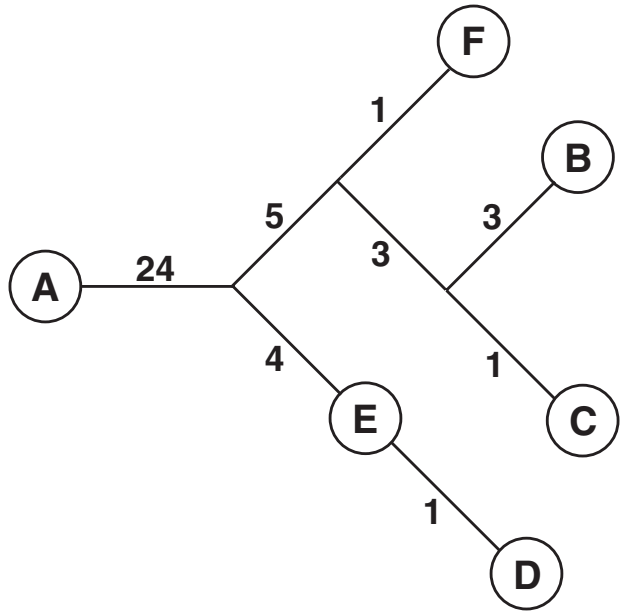

Figure 3. Unrooted parsimony network indicating the relationships between the six haplotypes (A–F) of the *T. gigantea* and *T. saeva* lineage based on the most parsimonious tree for the combined CO1 and ND1 data. The number of mutational steps between the haplotypes is marked on the branches and highlights the large difference between haplotype A (*T. saeva*) and all other haplotypes. Note that for haplotype F only CO1 data were available.

Molecular clock

Tests for clock-like evolution of the CO1 sequences indicated that the molecular-clock hypothesis could not be rejected and that the use of a molecular-clock calibration to estimate haplotype divergence times is statistically valid. The log likelihood ratio test yielded $\chi^2 = 1.36$ (6 d.f., $P = 0.968$).

DISCUSSION

Introgressive hybridization

Several closely related mitochondrial DNA haplotypes (B, C, D and E) have been identified that can be ascribed to *T. gigantea*, yet that are also present in specimens of *T. saeva*. They are markedly divergent in

sequence to haplotypes that are unique to *T. saeva* (haplotype A) and *T. atrica*. This strongly suggests that these haplotypes are of *T. gigantea* origin and result from the introgression of *T. gigantea* mitochondrial genomes into *T. saeva* populations. Figure 1 illustrates the distribution of the haplotypes among the *T. gigantea* and *T. saeva* specimens from southern England. The restriction of haplotype A to *T. saeva* populations is clear, as is the spill-over of *T. gigantea*-associated haplotypes into *T. saeva* individuals to the west of the hybrid zone.

An alternative explanation for this observation would be that mitochondrial genes sequenced here cannot be used to separate *T. gigantea* and *T. saeva*. The divergent *T. saeva* haplotype (A) could simply represent an ancestral haplotype that has been retained in *T. saeva* and has not been lost from the gene-tree by lineage-sorting following the divergence of these two species. Eliminating the possibility that our data may represent an ancestral polymorphism would be difficult, even with more extensive sampling. However, haplotype A was only found in *T. saeva*, never in *T. gigantea*, and was the only haplotype observed in the two specimens of *T. saeva* from Nancy, France, a region in which *T. gigantea* is very rare or absent. The asymmetric introgressive hybridization hypothesis is supported by observations comparing the hybrid zone in southern England and the area of sympatry in Yorkshire. In Yorkshire, the profiles of both allozyme and morphometric data indicate that hybridization is largely asymmetrical – with *T. saeva* experiencing the most introgression such that it more closely resembles *T. gigantea* here than it does in southern England (Croucher, 1998; P. J. P. Croucher, G. S. Oxford & J. B. Searle, unpubl. data). All individuals from Yorkshire analysed in this paper (two *T. saeva* and one putative hybrid) exhibited the '*T. gigantea*' D/E haplotype (Table 1). Although it is not possible to differentiate between past or recent/current introgression (because the rates of DNA evolution are not fast enough to elucidate such fine time-scales), the fact that six out of 12 individuals identified as *T. saeva* apparently possessed mitochondrial DNA of *T. gigantea* origin suggests that such introgression has occurred frequently via backcrossing to *T. saeva* of female hybrids derived from crosses between *T. gigantea* females and *T. saeva* males.

The currently observed introgressed mitochondrial genes may reflect one, or a combination of, historical hybridization event(s), current and recent gene flow from areas of contact or the transportation of colonists into allopatric regions of the other species. For synanthropic spiders such as *T. saeva* and *T. gigantea*, repeated, human-mediated transportation of small numbers of individuals between allopatric areas is highly likely. Here morphological characters and nuclear genetic variation of the colonist(s) may be 'hybridized away' but the maternally inherited, non-recombining mitochondrial genome would remain as an intact signal of the event – the transportation of one female and/or an egg-sac would be sufficient. The discovery of the *T. gigantea* B/C mitochondrial haplotype in a specimen of *T. saeva* from the extreme south-western English county of Cornwall, some 175 km west of the contact zone, may be evidence for this phenomenon. The B/C haplotype was otherwise only found in specimens of *T. gigantea* from Kent, the most south-eastern English county, and in the *T. gigantea* specimen from North America (Fig. 1, Table 1). *T. gigantea* is known to have been introduced into North America sometime before the 1920s (Chamberlin & Ivie, 1935). The possibility of the loss of species identity as a result of hybridization is illustrated in Yorkshire where the two species have apparently only been present since the 1970s (Smith, 1985). Here hybridization is extremely common and both morphological and genetic evidence suggests that the species are less differentiated from one another than they are further south (Oxford & Smith, 1987; Oxford & Plowman, 1991; Croucher, 1998; P. J. P. Croucher, G. S. Oxford & J. B. Searle, unpubl. data).

Evidence of unidirectional mitochondrial introgression from *T. gigantea* to *T. saeva* strongly supports inferences concerning the asymmetrical patterns of interaction between these species from laboratory studies on mating behaviour. These suggest that crosses between males of *T. gigantea* and females of *T. saeva* are less likely to result in progeny than the reciprocal cross. This may, to some extent, be a result of stronger mechanical barriers to copulation between males of *T. gigantea* and females of *T. saeva*, than vice versa. Observations of between-species copulations indicate that, compared with the reciprocal cross, the differences in the shape of copulatory organs of the two species appear to impede more greatly the ability of males of *T. gigantea* to engage successfully the genitalia of female *T. saeva*. Additionally, egg-clutches from such crosses always appeared yolky and unfertilized (Croucher, 1998; P. J. P. Croucher, G. S. Oxford & J. B. Searle, unpubl. data).

Asymmetrical patterns of mitochondrial introgression have been well documented in several other species. For example, the bank vole (*Clethrionomys glareolus*) has received mitochondrial DNA from the northern red-backed vole (*C. rutilus*) (Tegelström, 1987) and in house mice extensive introgression of *Mus musculus domesticus* mitochondrial DNA into *M. m. musculus* populations has occurred (Prager *et al*., 1993). Sota *et al*. (2001) have suggested that one reason for asymmetrical mitochondrial introgression in Japanese carabid beetles may be a physical barrier

to copulation in one direction but not the other, as suggested above for *Tegenaria*. In field crickets, crosses between males of *Gryllus firmus* and females of *G. pennsylvanicus* produce viable and fertile hybrids whereas the reciprocal cross does not, resulting in introgression of the *G. pennsylvanicus* mitochondrial genome (Harrison, Rand & Wheeler, 1987). It has been suggested that the incompatibility between *G. pennsylvanicus* males and *G. firmus* females may result from an ineffectiveness of sperm transfer in this cross or from variation in the accessory glands of the males. The *G. firmus* females act like virgins and lay few eggs. This suggests either a lack of fertilization or the lack of an oviposition stimulus from the male (Mandel, Ross & Harrison, 2001), a situation that may parallel in some respects that seen in *T. saeva* and *T. gigantea*.

Phylogeny and divergence

The molecular analyses presented here indicate the close relationships between *Tegenaria atrica*, *T. saeva* and *T. gigantea* and support their inclusion in a *T. atrica* 'group'. This confirms previous conclusions based on morphological similarities (Merrett, 1980; Maurer, 1992). The European distributions, with *T. atrica* widespread throughout most of Europe, and *T. saeva* and *T. gigantea* sharing largely concordant distributions in western Europe (Croucher, 1998; P. J. P. Croucher, G. S. Oxford & J. B. Searle, unpubl. data), already indicated that the latter two species may be more closely related to one another than either is to *T. atrica*. This is confirmed by the mitochondrial DNA analyses. The CO1 data did not fail the maximum likelihood ratio test for a molecular clock and the CO1 pairwise distances can therefore be used to obtain crude estimates of speciation times during the evolution of the *T. atrica* group. Brower (1994), in an assessment of mitochondrial CO1 divergence in a variety of invertebrate taxa, inferred a rate of 2.3% divergence per million years. This appears to be a fairly robust estimate and agrees with that suggested previously in *Dolichopoda* cave crickets by Venanzetti *et al*. (1993). This rate of divergence has been applied to arachnids by Wilcox *et al*. (1997). Although acknowledging that our genetic distances are based on a relatively short sequence, and assuming that Brower's rate also applies to *Tegenaria*, application of this rate to our data yields a divergence time for *T. saeva* and *T. gigantea* of approximately 1.5 million years ago (Mya) [uncorrected distance *p*: 1.3 Mya (range 1.2–1.5 Mya); K2P: 1.4 Mya (1.2–1.5 Mya), TrN+G: 1.8 Mya (1.5–2.0 Mya)]. The mean divergence time for *T. atrica* from *T. gigantea* and *T. saeva* is placed at approximately 3.0 Mya [*p*: 2.4 Mya (2.1–2.7 Mya); K2P: 2.5 Mya (2.2–2.8 Mya), TrN+G: 3.6 Mya (3.0–4.0 Mya)]. The error in divergence time estimates increases greatly with genetic distance; however, the maximal divergence among the basal taxa (*T. parietina* and *T. domestica*) suggests that the genus *Tegenaria* has probably existed since the Pliocene, in excess of 10 Mya.

In the absence of a suitable palaeontological record with which to calibrate a spider mitochondrial clock, these dates should be treated with caution. However, even allowing for a wide error, they strongly indicate that the radiation of the *T. atrica* group occurred in the (probably early) Pleistocene. This is the modern age of glacial–interglacial cycles that began more than 2.4 Mya (Webb & Bartlein, 1992). It is therefore possible that speciation in the *T. atrica* group and current species distributions have been partly driven by allopatric divergence in refugia during glacial advances combined with expansion from these refugia (and the possible formation of hybrid zones) during interglacial periods (Hewitt, 1996, 1999). Similar patterns of postglacial range expansion from southern European or eastern European/Asian refugia have been described in a variety of organisms (Hewitt, 1999), e.g. the brown bear *Ursus arctos* (Taberlet & Bouvet, 1994); the bank vole *Clethrionomys glareolus*, pigmy shrew *Sorex minutus* and common shrew *Sorex araneus* (Bilton *et al*., 1998); and the grasshopper *Chorthippus parallelus* (Hewitt, 1990, 1996). The current geographical confinement of *Tegenaria saeva* and *T. gigantea* to the western European fringe might suggest that the refugia in this case could have been located in the Iberian peninsula.

In conclusion, *Tegenaria saeva* and *T. gigantea* are sister taxa that have apparently been diverging since 1–2 Mya. Mitochondrial DNA haplotypes present in *T. saeva*, yet assignable to *T. gigantea*, almost certainly represent the molecular signature of hybridization events. The asymmetrical pattern of introgression probably reflects true asymmetries and barriers in the sexual interactions of these species at behavioural, physical and genetic levels. This study therefore provides valuable data on the nature and dynamics of the species boundary in these hybridizing taxa.

ACKNOWLEDGEMENTS

We thank R Leborgne (University of Nancy), J. O'Connor (National Museum of Ireland) and G. Binford (University of Arizona) for supplying specimens, and also R. G. Gillespie and G. K. Roderick (University of California, Berkeley) for advice. We also thank S. Mascheretti for her help. The comments of two anonymous referees were extemely helpful in shaping the final version of the paper. P.J.P.C. was supported by a BBSRC studentship.

REFERENCES

Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 133–164.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16:** 314–321.

Bilton D, Mirol P, Mascheretti S, Fredga K, Searle JB. 1998. The Mediterranean regions of Europe as an area of endemism for small mammals rather than a major source in the post-glacial colonisation of central and northern Europe. *Proceedings of the Royal Society of London B* **265:** 1219–1226.

Brower AVZ. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA* **91:** 6491–6495.

Chamberlin RV, Ivie W. 1935. Miscellaneous new American spiders. *Bulletin of the University of Utah* **26:** 1–79.

Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evoution* **22:** 252–271.

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London B* **353:** 287–305.

Croucher PJP. 1998. Evolutionary interactions of two colonizing species of large house spider (Araneae: *Tegenaria* spp.) – testing the reinforcement hypothesis. DPhil Thesis, University of York, York, UK.

Garb JE. 1999. An adaptive radiation of Hawaiian Thomisidae: biogeographic and genetic evidence. *Journal of Arachnology* **27:** 71–78.

Genetics Computer Group. 1994. *Program manual for the Wisconsin package*, Version 8. Madison, WI: Genetics Computer Group.

Gillespie RG. 1999. Comparison of rates of speciation in web-building and non-web-building groups within a Hawaiian spider radiation. *Journal of Arachnology* **27:** 79–85.

Gillespie RG, Croom HB, Palumbi SR. 1994. Multiple origins of a spider radiation in. Hawaii. *Proceedings of the National Academy of Sciences, USA.* **91:** 2290–2294.

Halanych KM. 1996. Testing hypotheses of chaetognath origins: long branches revealed by 18S rDNA. *Systematic Biology* **45:** 223–246.

Halanych KM, Demboski JR, Jansen van Vuuren B, Klein DR, Cook JA. 1999. Cytochrome *b* phylogeny of North American hares and jackrabbits (*Lepus*, Lagomorpha) and the effects of saturation in outgroup taxa. *Molecular Phylogenetics and Evolution* **11:** 213–221.

Halanych KM, Robinson TJ. 1997. Phylogenetic relationships of cottontails (*Sylvilagus*, Lagomorpha): congruence of cytogenetic and 12S rDNA data. *Molecular Phylogenetics and Evolution* **7:** 294–302.

Harrison RG, Rand DM, Wheeler WC. 1987. Mitochondrial DNA variation in crickets across a narrow hybrid zone. *Molecular Biology and Evolution* **4:** 144–158.

Harvey PR, Nellist DR, Telfer MG. 2002. *Provisional atlas of British spiders (Arachnida, Araneae)*, Vol. 2. Huntingdon: Biological Records Centre.

Hausdorf B. 1999. Molecular phylogeny of araneomorph spiders. *Journal of Evolutionary Biology* **12:** 980–985.

Hedin MC. 1997. Molecular phylogenetics at the population/species interface in cave spiders of the Southern Appalachians (Araneae, Nesticidae, *Nesticus*). *Molecular Biology and Evolution* **14:** 309–324.

Hedin MC. 2001. Molecular insights into species phylogeny, biogeography, and morphological stasis in the ancient spider genus *Hypochilus* (Araneae: Hypochilidae). *Molecular Phylogenetics and Evolution* **18:** 238–251.

Hewitt GM. 1990. Divergence and speciation as viewed from an insect hybrid zone. *Canadian Journal of Zoology* **68:** 1701–1715.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58:** 247–276.

Hewitt GM. 1999. Post-glacial re-colonisation of European biota. *Biological Journal of the Linnean Society* **68:** 87–112.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16:** 111–120.

Locket GH. 1975. The identity of Blackwall's *Tegenaria saeva* (Araneae, Agelenidae). *Bulletin of the British Arachnological Society* **3:** 85–90.

Maddison WP, Maddison DR. 1992. *MacClade: analysis of phylogeny and character evolution*, Version 3.0. Sunderland, MA: Sinauer.

Mandel MJ, Ross CL, Harrison RG. 2001. Do *Wolbachia* infections play a role in unidirectional incompatibilities in a field cricket hybrid zone? *Molecular Ecology* **10:** 703–709.

Masta SE. 2000. Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TPSIC arm. *Molecular Biology and Evolution* **17:** 1091–1100.

Masta SE, Maddison WP. 2002. Sexual selection driving diversification in jumping spiders. *Proceedings of the National Academy of Sciences, USA* **99:** 4442–4447.

Maurer R. 1992. *Checkliste der Europäischen Agelenidae nach der Roewerschen Systematik. 1954 – unter Berücksichtigung Angrenzender Östlicher Gebiete 1:* 1–28. Holderbank, Switzerland. Privately published.

Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

Medrano JF, Aasen E, Sharrow L. 1990. DNA extracted from nucleated red blood cells. *Biotechniques* **8:** 43.

Merrett P. 1980. Notes on the variation, identification and distribution of British species of the *Tegenaria atrica* group (Araneae, Agelenidae). *Bulletin of the British Arachnological Society* **5:** 1–8.

Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford: Oxford University Press.

Oxford GS, Chesney HCG. 1994. Large house spiders *Tegenaria* spp. in Northern Ireland: previously overlooked species or recent introductions? *Irish Naturalists' Journal* **24:** 354–357.

Oxford GS, Plowman A. 1991. Do large house spiders *Tegenaria gigantea* and *T. saeva* (Araneae, Agelenidae) hybridise in the wild? – a multivariate approach. *Bulletin of the British Arachnological Society* **8:** 293–296.

Oxford GS, Smith CJ. 1987. The distribution of *Tegenaria gigantea* Chamberlain and Ivie, 1935. and *T. saeva* Blackwall, 1844 (Araneae, Agelenidae) in Yorkshire. *Bulletin of the British Arachnological Society* **7:** 123–127.

Parker JR. 1984. Synanthropic spiders and other things – Part 2. *Newsletter of the British Arachnological Society* **39:** 1–2.

Piel WH, Nutt KJ. 1997. *Kaira* is a likely sister group to *Metepeira*, and *Zygiella* is an araneid (Araneae, Araneidae): evidence from mitochondrial DNA. *Journal of Arachnology* **25:** 262–268.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC. 1993. Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society* **50:** 85–122.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87:** 651–701.

Smith CJ. 1985. The distribution of the larger *Tegenaria* spider species (Agelenidae) in the York area. *Newsletter of the British Arachnological Society* **42:** 5–6.

Sota T, Ishikawa R, Ujiie M, Kusumoto F, Vogler AP. 2001. Extensive trans-species mitochondrial polymorphisms in the carabid beetles *Carabus* subgenus *Ohomopterus* caused by repeated introgressive hybridization. *Molecular Ecology* **10:** 2833–2847.

Swofford DL. 1998. *PAUP*: phylogenetic analysis using parsimony (*and other methods)*, Version 4.0. Sunderland, MA: Sinauer.

Taberlet P, Bouvet J. 1994. Mitochondrial DNA polymorphism, phylogeography and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proceedings of the Royal Society of London B* **255:** 195–200.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512–526.

Tan A-M, Gillespie RG, Oxford GS. 1999. Paraphyly of the *Enoplognatha ovata* group (Araneae: Theridiidae) based on DNA sequences. *Journal of Arachnology* **27:** 481–488.

Tegelström H. 1987. Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *Journal of Molecular Evolution* **24:** 218–227.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22:** 4673–4680.

Venanzetti F, Cesaroni D, Mariottini P, Sbordoni V. 1993. Molecular phylogenies in *Dolichopoda* cave crickets and mtDNA rate calibration. *Molecular Phylogenetics and Evolution* **2:** 275–280.

Vink CJ, Mitchell AD. 2002. 12S DNA sequence data confirms the separation of *Alopecosa barbipes* and *Alopecosa accentuata* (Araneae: Lycosidae). *Bulletin of the British Arachnological Society* **12:** 242–244.

Webb T, Bartlein PJ. 1992. Global change during the last 3 million years: climatic controls and biotic responses. *Annual Review of Ecology and Systematics* **23:** 141–173.

Wilcox TP, Hugg L, Zeh JA, Zeh DW. 1997. Mitochondrial DNA sequencing reveals extreme genetic differentiation in a cryptic species complex of neotropical pseudoscorpions. *Molecular Phylogenetics and Evolution* **7:** 208–216.

Zehethofer K, Sturmbauer C. 1998. Phylogenetic relationships of central European wolf spiders (Araneae: Lycosidae) inferred from 12S ribosomal DNA sequences. *Molecular Phylogenetics and Evolution* **10:** 391–398.