

**Evolution of satellite DNAs in a radiation of endemic Hawaiian spiders: Does concerted evolution of highly repetitive sequences reflect evolutionary history?**

Joan Pons <sup>1,2,\*</sup> and Rosemary G. Gillespie <sup>1</sup>.

*1 Division of Insect Biology, ESPM, University of California, Wellman Hall 201 Berkeley  
CA 94720-3112, USA*

*2 The Natural History Museum, Department of Entomology, Cromwell Road, London  
SW7 5B, UK*

\*Corresponding author:

Joan Pons, Department of Entomology, The Natural History Museum, London SW7  
5BD, U.K., Telephone: +44 207 942 5048 Fax: +44 207 942 5229, Email:  
joap@nhm.ac.uk

Key words: satellite DNA, spider, *Tetragnatha*, Hawaii, concerted evolution, molecular  
drive

Running head: Phylogenetic signal of *Tetragnatha* satellites

## **Abstract**

Satellite DNAs are known for an unusual and non-uniform evolution characterised by rapid evolutionary change between species, and concerted evolution leading to molecular homogeneity within species. In this paper we use satellite DNAs for phylogenetic analysis of a rapidly evolving lineage of spiders, and compare the phylogeny with a hypothesis previously generated based on mitochondrial DNA and allozymes. The spiders examined include almost all species within a monophyletic clade of endemic Hawaiian *Tetragnatha* species, the spiny-leg clade. The phylogeny based on satellite sequences is largely congruent to those produced by mtDNA and allozymes, except that the satellite DNA yields much longer branches, with higher levels of support for any given node. Closely related species that have differentiated ecologically within an island, are well resolved with satellite DNA but much less so with mtDNA. These results suggest that *Tetragnatha* stDNA repeats seem to be evolving gradually and cohesively during the diversification of these endemic Hawaiian spiders. The study also reveals gain-loss of satellite DNA copies during species diversification. We conclude that satellite DNA sequences may potentially be very useful for resolving relationships between rapidly evolving taxa within an adaptive radiation. In addition, satellite DNA as a nuclear marker suggests that hybridization or peripatry could play a role in species formation that can not be revealed by mitochondrial markers due to its maternal inheritance.

## **Introduction**

Multicopy gene families such as nuclear ribosomal DNA (rDNA) exists in hundreds to thousands of copies / repeats that are reiterated tandemly at one or more chromosomal locations. Repeats generally evolve in a concerted manner leading to complete homogenization of repeats at the intra-specific level and fixation of species-specific repeats with diagnostic mutations during species diversification (Elder and Turner 1995). The large difference in rates of turnover within chromosomes (due to unequal crossing over) versus between chromosomes is the main factor driving the gradual and cohesive turnover of copies produced by mutation (molecular drive, Dover 2002). Concerted evolution ensures that all rDNA repeats are nearly identical in sequence at the intra-specific level, within and between loci (i.e. orthologs and paralogs are identical), and hence it does not generally matter which copy is used for inter-specific comparisons (Álvarez and Wendel 2003). Consequently, despite the high number of repeats, rDNA sequences generally reflect species diversification and hence have the potential to be more widely used as phylogenetic markers

Satellite DNA (stDNA) is a non-coding DNA composed of thousands to millions of tandemly arranged units that also show high intra-specific sequence identity despite their wide heterochromatic distribution. Nevertheless, the relaxed evolutionary constraints acting on stDNA monomers usually lead to marked sequence divergences at the inter-specific level, rendering the signal of the satellite DNA phylogenetically uninformative (Elder and Turner 1995; Ugarkovic and Plohl 2002). In a minority of cases a stDNA

family is conserved over considerable evolutionary time, evolving gradually (i.e. homogenization and fixation of mutations) and reflecting species diversification in a manner similar to rDNA markers (Arnason et al. 1992; Bachmann and Sperlich 1993; de la Herran et al. 2001). Rarely, the intermediate stages of that gradual process have been witnessed (i.e. coexistence of new and ancestral copies of a stDNA family in different stages of spreading throughout the genome, Strachan et al. 1985; Pons et al. 2002).

Adaptive radiation on islands has been used as a model system to study stDNA evolution because species tend to be very closely related and thus are more likely to conserve sequence similarity between their repeats. Examples include Hawaiian *Drosophila* (Miklos and Gill 1981), Canarian darkling beetles in the genus *Pimelia* (Pons et al. 2002), and Hawaiian *Tetragnatha* spiders (Pons and Gillespie 2003). Species in the genus *Tetragnatha* (Araneae:Tetragnathidae) constitute the largest known radiation of spiders in the Hawaiian Islands. All species are endemic and belong to one of two major clades, the “spiny leg” clade and a diverse web-building clade; there is also a small group of web-building species that represent a separate colonization (Gillespie et al. 1994). A previous study on Hawaiian *Tetragnatha* using one representative from each of the three main clades suggested that all three species share satellite sequences of common origin (Pons and Gillespie 2003). The length (183-189 bp) and nucleotide composition (58 % A+T rich on average) of repeats of the three *Tetragnatha* species studied is very similar at the intra- and inter-specific levels. However, they exhibit marked sequence divergence at the inter-specific level (up to 45%) but showing substantial similarity in particular DNA stretches. The results suggest a common origin of *Tetragnatha* repeats and an

evolutionary constraint on the length and nucleotide composition at the inter-specific level. At the intra-specific level, the three species show different degrees of sequence identity between repeats, probably as a result of specific historical processes. *T. hawaiiensis* stDNA exhibits strong homogenization of the monomeric sequences, as happens in *T. quasimodo*, although the actual repeats are higher-order units (dimers linking two divergent subfamilies of monomers). On the other hand, *T. acuta* shows little homogenization of repeats due to the spreading of the three non-linked and divergent subfamilies of repeats.

Our present study focuses on a single clade of Hawaiian *Tetragnatha*, the spiny leg clade, where, because they are closely related, we expect high sequence conservation of repeats to allow us to infer phylogenetic relationships from divergence of satellites, provided the evolution of the repetitive DNAs in these spiders is dictated by "gradual" processes.

Alternatively, if the repetitive DNAs evolve through "saltatory" processes such as the "library hypothesis" (Mestrovic et al. 1998; Pons et al. 2004), divergence of repeats will bear no connection to phylogenetic relationships. The phylogenetic signal of our stDNA sequences is compared to an independent and resolved phylogeny based on analysis of mitochondrial (12S, 16S and COI sequences) and allozyme data. Results to date based on these data suggest that species on any one island are generally most closely related to each other (Gillespie 2004). The clade includes four discrete "ecomorphs" (see fig. 1a) which appear to have evolved independently on the different islands (Gillespie 1991; Gillespie 2002), resulting in a one-to-one convergence of the same set of "ecomorph" types on each island. Across the archipelago, cladogenesis in the spiny leg clade shows

the typical pattern of lineage progression from older to younger islands (Wagner and Funk 1995; Hormiga et al. 2003). The most ancestral species, *T. kawaiensis* and *T. pilosa*, are on the oldest island (Kaua'i), with subsequent sequential colonizations to Oahu (*T. polychromata*, *T. tantalus*, *T. perrerai* and *T. quasimodo*), Maui Nui (*T. waikamoi*, *T. brevignatha*, *T. restricta* and *T. kamakou*) and Hawaii (mostly populations from Maui) (Gillespie 2004).

## **Material and methods**

### *Sampling and genomic DNA isolation*

Individuals from 12 endemic species of *Tetragnatha* “spiny-leg” spiders were collected from different localities in the Hawaiian Islands: *T. quasimodo* (Oahu, Mount Kaala, 600 m and Hawaii, Puu Waawaa, 1350 m), *T. perrerai* (Oahu, Mount Kaala, 1350 m), *T. polychromata* (Oahu, Mount Kaala, 1350 m), *T. tantalus* (Oahu, Poamoho, 600 m), *T. kamakou* (Maui, Kaumakani, 1250 m), *T. waikamoi* (Maui, Puu Kukui, 1600 m), *T. kawaiensis* (Kauai, Nualolo, 1150 m), *T. pilosa* (Kauai, Pihea, 1200 m), *T. brevignatha* (Hawaii, Saddle Road, 1650 m), *T. restricta* (Hawaii, Hakalau, 2000 m), and *T. anuenue* (Hawaii, Saddle Road, 1750 m). *T. hawaiiensis* (Kauai, Kahili, 650 m), which is also endemic to the Hawaiian Islands was used as an outgroup. Specimens of six species (*T. mohihi*, *T. kukuiki*, *T. kukuhaa*, *T. kikokiko*, *T. obscura*, and *T. macracantha*) were not available for the current analysis. DNA was isolated from two or three frozen adults specimens stored at -80° C by standard phenol extraction and ethanol precipitation procedures.

### *Isolation, cloning and sequencing of satellite DNA*

One to two micrograms of genomic DNA were digested with restriction enzymes according to the instructions of the manufacturer (Roche, New England), and the fragments obtained were electrophoresed on 2 % agarose gels. Several restriction enzymes were tested for isolation of stDNA sequences: *Acc* I, *Afl* II, *Alu* I, *Apa* I, *Bfa* I, *Bst* BI, *Dra* I, *Hae* III, *Hinc* II, *Hinf* I, *Mbo* I, *Nla* IV, *Pvu* I, and *Sal* I. Conspicuous DNA bands of about 190 bp and 380 bp were found. These were cut from the gel, purified with the Qiaquick Gel Extraction Kit (Qiagen), ligated into the *Eco* RV site of the pMOSBlue vector (Amersham Pharmacia Biotech), and used to transform Sure™ competent cells (Stratagene). Recombinant clones were sequenced on both strands by the dideoxy sequencing method using the Big Dye™ Terminator Cycle Sequencing Kit and the automatic sequencing system ABI PRISM™ 377 DNA Sequencer (Applied Biosystems).

### *Screening for presence of species-specific stDNA sequences in other species of the spiny leg clade by slot-blot*

100 ng of genomic DNA were blotted onto nylon membranes, denatured with 0.5 N NaOH and cross-linked to the membrane (Osmotics) by exposure to ultra-violet light. Labeling of the probe, filter hybridizations, and detection of the hybridization signals were performed as described in the manual for the DIG High Prime DNA Labeling and Detection Starter Kit I (Roche). Post-hybridization washes were performed allowing 60

% similarity (2 x 5 minutes in 2x SSC and 0.1 % SDS at RT, and 2 x 15 minutes in 2 x SSC and 0.1 % SDS at 37 ° C).

### *Sequence analysis*

We first performed multiple alignment using Clustal W v. 1.7, and aligned sequences were subjected to analysis using MODELTEST v. 3.06 (Posada and Crandall 1998) which performs a hierarchical test of likelihood (hLRT) fits under 56 different models of DNA substitution. Distance trees were built by the Neighbor-Joining method using the pair-wise nucleotide divergences estimated using the model and parameters previously estimated with MODELTEST. Bootstrap values were calculated on 1000 replicates in PAUP\* 4.05 (Swofford 2002). Details of the nucleotide composition and  $\chi^2$  test of homogeneity of base frequencies across sequences in the repetitive units were also estimated using PAUP\* 4.05. DNA polymorphism, DNA divergence, and putative gene conversion events were estimated using DnaSP v. 3.51 (Rozas and Rozas 1997). Sequence divergence between stDNA subfamilies or species was calculated as the average nucleotide substitution per site ( $D_{xy}$  value from DnaSP 3.51).

Bayesian phylogenetic analyses of the Clustal alignment were performed with MrBayes 2.0 (Huelsenbeck and Ronquist 2001) using the simplest model and the HKY+ $\Gamma$  model selected by hLRT in MODELTEST v. 3.06. Searches were performed based on 3 million generations with four Markov chain Monte Carlo simulations (MCMC) starting from random trees which were sampled every 100<sup>th</sup> generations. The log-likelihood scores of

sample points were plotted against generation time, and stationarity of Markov chains was assumed when the log-likelihood values reached a stable equilibrium (Huelsenbeck and Ronquist 2001). The stationarity of the chains was confirmed by plotting the remaining log-likelihood values (*sump* command). All sample points prior to stationarity were discarded as burn-in values (the first 40,000 generations), and remaining points were used to generate a strict consensus tree, with each clade posterior probability value represented by the proportion of nodes recovered in the sample of trees.

Clustal alignments were also analysed using parsimony criterion in PAUP\* 4.05 (Swofford 2002). All nucleotide positions in the matrix were considered equivalent and weighted equally, using gaps as both missing data and as a 5<sup>th</sup> character state. Searches were performed using 1000 random additions with TBR branch swapping and the “Multrees” option. Bootstrap values were calculated on 100 replicates.

The separation of alignment and cladogram searching in a two-step analysis may not be the most efficient way of discovering the globally optimal solution and often leads to incompatibility of the optimality criteria applied at different stages of the analysis. Accordingly, we also conducted direct character optimization analyses of the stDNA sequences as implemented in the computer program POY 2.6 (Gladstein and Wheeler 1997) because this program allows insertion and deletion events to be incorporated into the tree searches. POY implements the concept of dynamic homology with the intermediate alignment step avoided by direct assessment of the number of evolutionary events i.e. DNA sequence transformations. The method generates more efficient

explanations of sequence variation (shorter, optimal trees) than do multiple alignments and can produce multiple optimal results if more than a single optimal cladogram/alignment exists (Wheeler 1996, 2002). For the current study, the parameter space in POY was explored by varying gap costs (1, 2, and 4). For all searches reported, we present the best (lowest cost) trees for the 50% majority rule consensus sequences from 1000 random addition replicates and collecting no more than 3 shortest trees in each replicate (commands: *-random 1000 -maxtrees 3*). Bremer Support was calculated by an approximate method implemented in POY based on a search procedure using TBR swapping on a constraint file obtained with the program Jack2Hen available with the POY software. An “implied alignment” can be produced in POY which is reconstructed from the list of synapomorphies at internal nodes and is based on the initial cladogram (Wheeler 2003). The aligned matrix obtained from this procedure is a representation of the homologies established in the direct optimization, and, in contrast to most standard alignment procedures, does not represent homoplastic insertions in the same column. The resulting matrix of aligned sequences is a synapomorphy scheme which approximates the character transformation in the dynamic homology search of POY (Wheeler, 2003). The clustal W alignment for the stDNA consensus sequences and the “implied alignments” for gap cost =1, 2 and 4 are included as supplementary information.

## **Results**

### *StDNAs characterization*

Ten out of the eleven studied species showed bands of about 190 bp after their genomic DNA was digested with restriction enzymes and analyzed by agarose gel electrophoresis (table 1). The majority of species displayed a prominent band (weaker in *T. tantalus* and *T. polychromata*). *T. anuenue* genomic DNA revealed no visible bands despite of having been screened it with 20 different endonucleases. Some species showed a dim ladder of oligomers of the 190bp-band such as those described elsewhere in *T. quasimodo* and *T. hawaiiensis* (Pons and Gillespie 2003). We cloned the monomeric bands of these ten species and then randomly sequenced from 12 to 19 positive clones for each species for both strands (table 1). The genomic DNA digestions of *T. kauaiensis* and *T. kamakou* also induced conspicuous dimeric bands that could be cloned to analyze the variability of contiguous repeat units. We were able to sequence two dimers from *T. kauaiensis* (pdKIS) and eight from *T. kamakou* (pdKAM). In the analysis of the repeat unit sequences we also included those isolated from *T. quasimodo* and *T. hawaiiensis* (Pons and Gillespie 2003). As in *T. quasimodo* and *T. hawaiiensis*, in both *T. kauaiensis* and *T. kamakou* dimers the repeat unit exhibiting the restriction site *Bfa* I was labeled pdKIS A and pdKAM A respectively, and the unit lacking this site, pdKIS B and pdKAM B respectively. Species from the spiny leg clade have consensus repetitive sequences of similar length (188-191 bp). The consensus repeat unit from the outgroup species, *T. hawaiiensis*, was shorter (183 bp). Some repeat units are shorter due to the partial deletion of the sequence: pWAI04, pWAI14, pKIS08, pKIS15, pKIS17, and pBRE16H. All *Tetragnatha* repeat units show similar nucleotide composition (57.99 % A + T on average,  $p < 0.999$ ).

### *Phylogenetic analyses of stDNA sequences using Clustal alignments*

*Tetragnatha* repetitive sequences still display similarity despite their high divergence (table 2, alignment ALIGN\_000247, acc. num. AJ413315-559 and AJ413577-82).

Phylogenetic analysis of all 251 repeat unit sequences based on their similarity (HKY+Γ model) using Neighbor-Joining with *T. hawaiiensis* as the outgroup, placed *T. pilosa* and *T. kauaiensis* as monophyletic and basal (fig. 1B). These sequences are the least divergent relative to the outgroup sequences, *T. hawaiiensis* (70-65 % similarity). The tree displays two other groups, both monophyletic relative to the basal sequences: 1) sequences from Oahu (*T. tantalus*, *T. polychromata*, and *T. perreirai*, and *T. quasimodo* from Oahu and Hawaii) and 2) sequences from Maui and Hawaii (*T. kamakou*, *T. waikamoi*, *T. brevignatha* and *T. restricta*). Surprisingly, *T. quasimodo* from Hawaii only shows one of the stDNA variants found in *T. quasimodo* from Oahu. The cross-hybridization of the stDNAs of each species against the genomic DNA of the other taxon recovered the same groups of species as the Neighbor-Joining analysis based on sequence divergence (table 3). Interestingly, *T. anuenue* genomic DNA, in which we could not isolate any repetitive sequences, showed no hybridization with any of the stDNA sequences isolated in the other *Tetragnatha* species studied, even at low stringency conditions. The Neighbor-Joining tree revealed that, except for *T. waikamoi*, the repeat units have a monophyletic origin at the intra-specific level. We also performed Bayesian and parsimony searches on Clustal alignments. Since satellites were monophyletic and very similar at the intra-specific level we used the 50% majority rule consensus sequence obtained from the repeats for each species. In those taxa showing two divergent

subfamilies of monomers, *T. quasimodo* Oahu, *T. brevignatha*, and *T. waikamoi*, we built the 50% majority rule consensus of each subfamily to assess their monophyly. Both Bayesian and parsimony analyses of the 15 consensus sequences retrieved the same tree topology (fig. 2a) as the Neighbor-Joining tree based on the 251 sequences (fig 1b).

#### *Phylogenetic analyses using direct optimization*

Since POY requires high computational power, searches were again performed with the 15 consensus sequences. Searches were initially performed under equal weighting of all characters (*-gap 1 -change 1*), but alignment parameters were varied to explore the effects on topology. Under equal weighting (fig. 2b), the topology retrieved by the POY search was very similar to that obtained in all the different analyses performed with the Clustal alignments (fig. 1b and 2a), except that the Kauai clade (*T. pilosa* + *T. kauaiensis*) was paraphyletic. Increasing the gap cost = 2 relative to nucleotide changes had no effect on tree topology and cost = 4 had little effect. The discrepancies were the basal position of *T. restricta* within the clade composed of sequences from *T. kamakou*, *T. brevignatha*, and *T. waikamoi*, and the monophyly of basal sequences from Kauai (*T. pilosa* + *T. kauaiensis*). Interestingly, total Bremer support values decreased (132, 97, 92, see fig. 2b for specific values for each node) as gap opening cost increased (1, 2, and 4 respectively). Plots of uncorrected distances against total numbers of transitions or transversions plots, using both the Clustal alignment (see fig. 3) and the "implied alignment" from POY (gap cost = 1; not shown), displayed no evidence of saturation in the stDNA sequences. The

slope and r-square values for the "implied alignment" plot were very similar to those shown in the Clustal alignment plot (fig. 3).

#### *Analysis of nucleotide variation*

The presence of two non-monophyletic subfamilies explains the higher nucleotide diversity of *T. waikamoi* repetitive sequences (16.2 %). These subfamilies (WAI 1 and WAI 2) differ in length (188 and 191 bp) and by 49 fixed nucleotide positions (38.5 % divergence). The nucleotide diversity within these subfamilies (3.9 % and 1.7 % respectively) is much lower than the divergence between subfamilies. The variability within subfamilies is similar to the nucleotide diversity estimated in the other species (table 1) and the divergence of *T. waikamoi* subfamilies is similar to the inter-specific divergence of stDNA sequences (table 2). Nucleotide diversity in other species was low (table 1). The intra-specific polymorphism of repetitive sequences appears to be due to single nucleotide substitutions that are distributed throughout the monomer. The nucleotide substitutions are mainly autapomorphies (unique) or synapomorphies shared by few monomers. These species, as in *T. hawaiiensis*, showed no divergent subfamilies. The exception was *T. brevignatha* that possessed two monophyletic subfamilies (BRE 1 and BRE 2) supported by high bootstrap and c.a. 10 % divergence as in *T. quasimodo* from Oahu. *T. brevignatha* subfamilies differ in length (188 and 191 bp respectively, positions 155, 156, and 168) and by 17 fixed nucleotide substitutions. *T. brevignatha* sequences showed low nucleotide diversity within variants (1.4 % and 0.8 % respectively) and higher divergence between them (11.6 %), like those from *T.*

*quasimodo* Oahu. The analysis of the two variants of *T. brevignatha* reveals the presence of one sequence with a tract that has diagnostic mutations of the other variant (pBRE 05 positions 61-73), possibly caused by gene conversion. Other species such as *T. kamakou*, *T. tantalus* and *T. quasimodo* Hawaii have variants differing in length (one or two indels) but show neither fixed nucleotide substitutions nor are supported by bootstrap values.

At the inter-specific level some closely related species show the coexistence of ancestral and recent copies (i.e. intermediate stages of a gradual and continuous turnover of repeats in a multigene family through molecular drive (Dover 2002; Pons et al. 2002; Strachan et al. 1985). Several positions indicating the coexistence of two major type of repeats in different stages of spreading (i.e. shared synapomorphies) were found in the closely related species *T. polychromata*, *T. tantalus* and *T. perrerae* (e.g. position 62 indel/A variants, 130 T/C, 151 indel/G, 169 T/G, and 177 T/A). Similar results were found in other close related species: *T. kauaiensis* and *T. pilosa* (position 12 T/G variants, 52 A/G, and 110 G/A), *T. kamakou* and *T. restricta* (1 G/A, 37 T/A, 139 T/A, and 140 T/A), and *T. waikamoi*, *T. restricta*, and *T. brevignatha* (61 C/G).

To assess the relative usefulness of stDNA versus mtDNA in phylogenetic estimation we plotted the percentage nucleotide substitution change in stDNA versus Cytochrome Oxidase I mtDNA for pairwise comparisons (Fig. 4). As can be seen, the rate of change of stDNA is similar to that of mtDNA at very low levels of divergence. Above 9 % mtDNA sequence divergence, stDNA divergence is approximately four times higher than that of mtDNA, and increases linearly with mtDNA divergence ( $R^2 = 0.220$ ,  $p = 0.001$ ).

## Discussion

*Tetragnatha* stDNA sequences show large divergences between certain species (up to 60 % in the most distant cases) which could raise some questions about the homology of the satellites. We assume that if *Tetragnatha* repeats undergo concerted evolution (molecular drive, Dover 2002), making copies identical within species and thereby removing the possibly confounding effect of using multiple paralogous copies, then they should retrieve the same species diversification pattern (i.e. tree topology) as the unambiguous mitochondrial genes and allozymes. Despite the large sequence divergences, the phylogeny of the spiny leg *Tetragnatha* species based on stDNA retrieves strong phylogenetic signal that is mostly concordant with that based on classical mtDNA sequences. Therefore, we assume that *Tetragnatha* repeats evolve in a gradual and cohesive manner, fixing species-specific mutations during species formation.

Interestingly, the methods of tree building, i.e. a two step protocol versus direct optimisation, and different parameters used in the trees searches had little impact on the tree topology, indicating the unambiguous phylogenetic signal from the stDNA sequences.

The difficulty in establishing sequence identities in stDNA sequences at the inter-specific level has prohibited the study of their evolutionary dynamics and homology. Despite some difficulties (e.g. failure to detect stDNA sequences in *T. anuenue*, and the presence of two divergent variants / paralogs coexisting in *T. waikamoi*), the present study and

other similar ones in insects (e.g Strachan et al. 1985; Pons et. al 2002) illustrates the conditions under which an evolutionary model can be postulated (gradual evolution through molecular drive). This gradual evolutionary change has been detected over a limited evolutionary period only (usually quite short, < 5My). However, it may occur over longer periods of time, but cannot be detected due to the absence of sequence similarity, which also precludes phylogenetic analyses to assess homology and evolutionary dynamics. Previous studies on Canarian *Pimelia* (Pons et al. 2002) and Hawaiian *Tetragnatha* (Pons & Gillespie 2003) have suggested that the satellite sequences have a common origin and selective constraints act on the length and nucleotide composition of repeats at the inter-specific level (Pons and Gillespie 2003). The results of the current study support the existence of such constraints. The more relaxed selective pressure on the primary sequence of *Tetragnatha* repeats compared to *Pimelia*, where the periodical distribution of adenines is linked to a partial conservation of the primary sequence, could explain the large divergences in base composition found in the spider repeats.

The phylogeny of the spiny leg *Tetragnatha* species based on stDNA sequences and on classical mtDNA sequences is consistent with the progression rule of colonization from the oldest island, Kauai, to Oahu and down to Maui Nui (Molokai, Lanai, and Maui), and Hawaii. Both mtDNA and stDNA place *T. pilosa* and *T. kauaiensis* (Kauai) as basal species. MtDNA suggests that colonization from Kauai to Oahu gave rise to the lineage (O) that includes *T. perreirai*, *T. polychromata*, and *T. tantalus*. Then from this group, two sister lineages arose, the first (Q) including *T. quasimodo*, which has colonized each

island as it appeared. The second lineage (M) also shows a pattern of progression, but with multiple ecomorphs forming independently on Maui Nui: *T. waikamoi*, *T. restricta*, and *T. kamakou*; and *T. brevignatha* and *T. macracantha* (Gillespie et al. 1997; Gillespie 2004). The stDNA phylogeny suggests a similar scenario, although this nuclear marker retrieves an alternative hypothesis where the main lineages (O + Q + M) may have originated on Kauai and then colonized the other islands sequentially. Interestingly, the stDNA sequences in the current study gave a similar phylogenetic signal to the mtDNA markers. Moreover, the branch lengths and support for the overall tree topology are much higher for the stDNA than for the mtDNA. The reason for this is clearly the tendency for stDNA to show high levels of intra-specific conservatism (due to concerted evolution), yet equally high levels of inter-specific disparity between closely related species. This phenomenon is illustrated in Figure 4, which shows the relationship between percentage sequence divergence in stDNA versus COI mtDNA. The bimodality in divergence rates of stDNA sequences, and the linear relationship with divergence in mtDNA sequences indicates that stDNA can be useful as a marker for identifying non-recombining gene pools and for elucidating phylogenetic relationships of very closely related species such as those that have undergone recent adaptive radiation.

The detection of intermediate stages of turnover of repeats (i.e., coexistence of older and more recent stDNA copies) and gene conversion events (a molecular mechanism driving concerted evolution) also suggests "gradual" as opposed to "saltatory" evolution (Dover 2002; Pons et al. 2002; Bachmann and Sperlich 1993; Strachan et al. 1985). However, in contrast to the rather constant fixation rate of nucleotide substitutions found in the closely

related Iberian *Pimelia* beetles (Pons et al. 2002), Hawaiian *Tetragnatha* spiders show a non-constant rate. For instance, very closely related species, such as *T. polychromata*, and *T. tantalus* have identical sequences with no species-specific fixed mutations. But other closely related species pairs such as *T. kamakou* and *T. restricta* exhibit very divergent sequence repeats (35 % divergence) although their units still share few synapomorphies (i.e. mutations present in the common ancestor). This effect can be also deduced from Figure 4, which shows that species mostly exhibit stDNA divergence of either <10% or > 30%, with very little in the 10-30% range. Interestingly, complete replacement of repeat units, sometimes with very divergent repeats (e.g. *T. restricta*), can be achieved over short evolutionary periods since the species of *Tetragnatha* endemic to Maui and Hawaii cannot be older than the age of the islands (Maui ranges from 0.8-1.3 my; Hawaii from 0-0.4 my, Carson and Clague 1995). Similar non-constant rates have been described in a recent study comparing distant *Pimelia* species (Pons et al. 2004). However, it is unclear which particular factors are involved in this accelerated and static evolution of stDNAs in each case. Accelerated evolution of satellite nucleotide sequences might occur, for instance, in response to various forms of stress and/or retrotransposon-mediated adaptive evolution (e.g., Kalendar et al. 2000). On the other hand, biased gene conversion can lead to extreme sequence conservation if ancestral sequence variants are favored over recently generated ones during concerted evolution (Hillis et al. 1991).

Three spiny leg species (*T. waikamoi*, *T. brevignatha*, and *T. quasimodo* Oahu), as well as the previously described *T. acuta* (Pons and Gillespie 2003), harbor divergent subfamilies of repeats in their respective genomes. The evidence of high nucleotide

identity within subfamilies suggests that strong homogenization acts at this level. The presence of two variants in *T. brevignatha* and in *T. quasimodo* imposes no difficulty for phylogenetic analysis because the variants occur at the terminal branches. In the case of *T. waikamoi*, variant divergence is higher than some inter-specific divergence. One explanation for this is that historical processes occurring during speciation, such as hybridization (which is so common among island radiations, e.g. Jordan et al. 2003; Álvarez and Wendel 2003), could have given rise to the different variants. Alternatively, the two variants may be maintained in a large population of *T. waikamoi*, with one giving rise to *T. brevignatha* and *T. macracantha*, the other to *T. restricta* and *T. kamakou*. Preliminary results from ITS-2 (unpublished data) also show the coexistence within a species of different variants of ITS-2, suggesting similar evolutionary patterns. On the other hand, biases in any of the multiple factors involved in the concerted evolution, such as drastic changes in population size (Ohta and Dover 1984), or a lack of non-reciprocal transfer between non-homologous chromosomes (Trick and Dover 1984), could cause the formation and fixation of variants within species. However, such biases are an unlikely explanation for the presence of the different variants in the current study, as it would be expected to lead to incongruent phylogenetic signal between satellite and mtDNA, which we do not find.

The results presented here also suggest a scenario of gain-loss of stDNA copies through the evolutionary history of the *Tetragnatha* species in the spiny leg clade, as has been described in other studies of closely related species (e.g., Panzera et al. 1995). The cross-hybridization experiments, carried out with the same amount of genomic DNA for each

species, resulted in a weaker signal from the genomic DNA of *T. tantalus*, and *T. polychromata* as compared to *T. perreirai*, even using its own stDNA sequences as probes, suggesting a loss of repeats. Moreover, the loss of DNA sequences and weaker hybridization in Hawaii relative to Oahu populations of *T. quasimodo*, could be related to the absence of variant 2 of Oahu in Hawaii *T. quasimodo*. The impossibility of isolating stDNA sequences in another “small spiny-leg” species, *T. anuenue*, together with the fact that no signal was obtained in cross-hybridization experiments, suggests a complete loss of the stDNA, or enough of a loss as to make the copies undetectable by standard methods. The loss-gain and/or redistribution of heterochromatin, mainly composed of stDNA repeats, has been linked, as cause or consequence, with chromosome rearrangement and evolution (i.e. genome reorganization), and therefore with speciation (Brutlag 1980).

In summary, the stDNA sequences of the *Tetragnatha* spiny-leg spiders studied, despite their high sequence divergence, give a phylogenetic that is related to but much higher than that of mtDNA within the range of approx. 5-10% mtDNA sequence divergence (Fig. 4). The phylogenetic relationships recovered from stDNA are therefore very similar to those recovered from traditional makers, but tend to accentuate differences between segregated gene pools. Both phylogenies are compatible with the geological history of the Hawaiian Islands, suggesting sequential colonization from older to newer islands. Moreover, the repetitive DNA may reflect historical processes that have occurred during speciation, such as peripatry or hybridization, which are not usually reflected by mitochondrial phylogenies but are commonly recovered in nuclear phylogenies. Given

the notoriously problematic nature of phylogenetic reconstruction among species showing explosive diversification, stDNA demands some attention for its potential usefulness in these cases.

### **Acknowledgements**

This work was funded by a grant from the Obra Cultural i Social Sa Nostra Foundation (Spain) to J.P., with additional funds from the Schlinger Foundation, and a National Science Foundation (DEB 9726573) grant to RGG. For help with field work, we are indebted to Haleakala National Park, the Nature Conservancy of Hawaii, Maui Pineapple Company, the State Department of Land and Natural Resources, and the Hawaii Natural Areas Reserve System. For help with collecting we owe particular thanks to M.A. Arnedo, R. Bartlett, C.F. Davenport, J. Giffin, A.C. Medeiros, D. Preston, and G.K. Roderick. We also thank M.A. Arnedo for help with the POY analysis.

### **References**

- Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* 29:417-434.
- Arnason U, Gretarsdottir S, Widegren B (1992) Mysticetae (baleen whale) relationships based upon the sequence of the common cetacean DNA satellite. *Mol Biol Evol* 9:1018-1028.

- Bachmann L, Sperlich D (1993) Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. *Mol Biol Evol* 10:647-659.
- Brutlag DL (1980) Molecular arrangement and evolution of heterochromatin DNA. *Ann Rev Genet* 14:121-144.
- Carson HL, Clague DA (1995) Geology and biogeography of the Hawaiian Islands. *Hawaiian Biogeography: evolution on a hot spot archipelago*. W. L. Wagner and V. A. Funk. Washington, Smithsonian Institution Press: 14-29.
- de la Herran R, Rejon CR, Rejon MR, Garrido-Ramos MA (2001) The molecular phylogeny of the Sparidae (Pisces, Perciformes) based on two satellite DNA families. *Heredity* 87:691-697.
- Dover GA (2002) Molecular drive. *Trends Genet* 18:587-589.
- Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart Rev Biol* 70:297-323.
- Gillespie RG (1991) Hawaiian spiders of the genus *Tetragnatha*: I. Spiny leg clade. *J Arachnol* 19:174-209.
- Gillespie RG (1999) Comparison of rates of speciation in web-building and non-web-building groups within a Hawaiian spider radiation. *J Arachnol* 27:79-85.
- Gillespie RG (2002) Hawaiian Spiders of the Genus *Tetragnatha*: IV: New, Small Species in the Spiny Leg Clade. *J Arachnol* 30:159-172.
- Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian Spiders. *Science* 303:356-359.
- Gillespie RG, Palumbi SR, Croom HB (1994) Multiple origins of a spider radiation in Hawaii. *Proc Natl Acad Sci USA* 91:2290-2294.

- Gillespie RG, Croom HB, Hasty GL (1997) Phylogenetic relationships and adaptive shifts among major clades of *Tetragnatha* spiders (Araneae: Tetragnathidae) in Hawai'i. *Pac Sci* 51:380-394.
- Gladstein DS, Wheeler WC (1997) POY: the Optimization of Alignment Characters. American Museum of Natural History, New York.
- Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251:308-310.
- Hormiga G, Arnedo MA, Gillespie RG (2003) Speciation on a conveyor belt: sequential colonization of the Hawaiian Islands by *Orsonwelles* spiders (Araneae: Linyphiidae). *Syst Biol* 52:70-88.
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.
- Jordan S, Simon C, Polhemus D (2003) Molecular systematics and adaptive radiation of Hawaii's endemic Damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Syst Biol*. 52:89-109.
- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci USA* 97:6603-6607.
- Mestrovic N, Plohl M, Mravinac B, Ugarkovic D (1998) Evolution of satellite DNAs from the genus *Palorus* – experimental evidence for the "library" hypothesis. *Mol Biol Evol* 15:1062-1068.

- Miklos GL, Gill AC (1981) The DNA sequences of cloned complex satellite DNAs from Hawaiian *Drosophila* and their bearing on satellite DNA sequence conservation. *Chromosoma* 82:409-427.
- Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. *Genetics*. 108: 501-521.
- Panzer F, Perez R, Panzer Y, Álvarez F, Scvortzoff E, Salvatella R (1995) Karyotype evolution in holocentric chromosomes of three related species of triatomines (Hemiptera-Reduviidae). *Chromosome Res* 3:143-150.
- Pons J, Gillespie RG (2003) Common origin of the satellite DNAs of Hawaiian spiders of the genus *Tetragnatha*: evolutionary constraints on the length and nucleotide composition of the repeats. *Gene* 313: 169-177.
- Pons J, Petitpierre E, Juan C (2002) Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Mol Biol Evol* 19:1329-1340.
- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarkovic D, Juan C (2004) Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common satellite DNA library. *Heredity* 92:418-427.
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817-818.
- Rozas J, Rozas R (1997) DnaSP version 2.0 a novel software package for extensive molecular population genetic analysis. *Comput Applic Biosci* 13:307-311.

- Strachan T, Webb D, Dover GA (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. EMBO J 4:1701-1708.
- Swofford DL (2002) PAUP\*. Phylogenetic Analysis Using Parsimony (\* and other methods). Sinauer Associates, Sunderland, Mass.
- Trick M, Dover GA (1984) Unexpectedly slow homogenization within a repetitive DNA family shared between two subspecies of tsetse fly. J Mol Evol 20:322-329.
- Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles-causes and effects. EMBO J 21:5955-5959.
- Wagner WL, Funk V (1995) Hawaiian Biogeography: Evolution on a Hot Spot Archipelago. Smithsonian Institution Press, Washington, DC.
- Wheeler WC (1996.) Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12:1-9.
- Wheeler WC (2002) Optimization Alignment: down, up, error, and improvements: In; techniques in molecular systematics and evolution. DeSalle R, Giribet G and Wheeler WC, eds, pp 55-69. Birkhäuser, Basel.
- Wheeler WC (2003) Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. Cladistics 19:261-268.

**Table 1.** List of specimens and restriction enzymes (RE) used to isolate the stDNA sequences, and characteristics of repeat units sequenced.

<b>species</b>	<b>RE</b>	<b>code</b>	<b># repeats sequenced</b>	<b>nucleotide diversity</b>	<b>consensus unit length (bp)</b>
<i>T. brevignatha</i>	<i>Alu</i> I	pBRE	12	0.059	variantBRE1 188
			7		variantBRE2 191
<i>T. waikamoi</i>	<i>Bfa</i> I	pWAI	14	0.162	variantWAI1 188
			4		variantWAI2 191
<i>T. kamakou</i>	<i>Bfa</i> I	pKAM	28	0.057	189
<i>T. restricta</i>	<i>Hinf</i> I	pRES	14	0.076	188
<i>T. quasimodo</i> Oahu	<i>Bfa</i> I	pQUA	11	0.067	variantQUA1 189
			16		variantQUA2 189
<i>T. quasimodo</i> Hawaii	<i>Hinf</i> I	pLIS	17	0.026	189
<i>T. tantalus</i>	<i>Pvu</i> I	pTAN	16	0.071	190
<i>T. polychromata</i>	<i>Nla</i> IV	pPOL	15	0.064	190
<i>T. perreirai</i>	<i>Acc</i> I	pPER	15	0.051	189
<i>T. pilosa</i>	<i>Bfa</i> I	pPIL	18	0.056	189
<i>T. kauaiensis</i>	<i>Bfa</i> I	pKIS	19	0.059	189
<i>T. hawaiiensis</i>	<i>Acc</i> I	pHAW	45	0.028	183

Table 2. Matrix of pair-wise divergence values of stDNA sequences of the twelve endemic Hawaiian Tetragnatha spiders studied. The code for the species-sequences is indicated in table 1.

	pBRE	pRES	pWAI	pKAM	pQUA	pLIS	pTAN	pPOL	pPER	pKIS	pPIL	pHAW
pBRE	–	0.531	0.313	0.354	0.551	0.553	0.435	0.431	0.434	0.463	0.474	0.522
pRES		–	0.477	0.431	0.518	0.520	0.521	0.517	0.519	0.586	0.609	0.600
pWAI			–	0.327	0.517	0.527	0.463	0.475	0.462	0.539	0.541	0.559
pKAM				–	0.509	0.524	0.425	0.424	0.430	0.542	0.511	0.535
pQUA					–	0.070	0.382	0.375	0.382	0.522	0.481	0.469
pLIS						–	0.383	0.369	0.381	0.540	0.500	0.487
pTAN							–	0.070	0.102	0.518	0.492	0.512
pPOL								–	0.086	0.508	0.492	0.496
pPER									–	0.517	0.495	0.500
pKIS										–	0.256	0.350
pPIL											–	0.306
pHAW												–

**Table 3.** Summary of the pair-wise cross-hybridization results of the stDNA sequences isolated in each species (rows) with 100 ng of genomic DNA of the 12 *Tetragantha* species of the spiny leg clade included in this study plus the outgroup *T. hawaiiensis* (columns). Hybridizations were performed allowing 60 % similarity. The particular cloned unit used as probe is indicated in each case. The codes for species-sequences (top row) are indicated in table 1. Dim hybridizations are indicated with the symbol + d and strong ones with + s. Groups showing similar sequences are highlighted with different colors and boxes. Since *T. waikamoi* has two very divergent subfamilies, one representative of each subfamily was used as probe.

	BRE09	WAI21	WAI07	KAM01	RES29	QUA05	LIS17	TAN50	POL48	PER23	KIS03	PIL20	HAW02
<i>T. brevignatha</i>	+s	+	+ d	+ d	-	-	-	-	-	-	-	-	-
<i>T. waikamoi</i>	+ d	+ s	+ d	+	-	-	-	-	-	-	-	-	-
<i>T. kamakou</i>	+ d	+	+	+ s	-	-	-	-	-	-	-	-	-
<i>T. restricta</i>	-	-	-	-	+ s	-	-	-	-	-	-	-	-
<i>T. quasimodo</i> Oahu	-	-	-	-	-	+ s	+ s	-	-	-	-	-	-
<i>T. quasimodo</i> Hawaii	-	-	-	-	-	+	+	-	-	-	-	-	-
<i>T. tantalus</i>	-	-	-	-	-	-	-	+ d	+ d	+ d	-	-	-
<i>T. polychromata</i>	-	-	-	-	-	-	-	+ d	+ d	+ d	-	-	-
<i>T. perrerae</i>	-	-	-	-	-	-	-	+	+	+ s	-	-	-
<i>T. kauaiensis</i>	-	-	-	-	-	-	-	-	-	-	+	+ d	-
<i>T. pilosa</i>	-	-	-	-	-	-	-	-	-	-	+ d	+	+ d
<i>T. hawaiiensis</i>	-	-	-	-	-	-	-	-	-	-	-	+ d	+ s
<i>T. anuenue</i>	-	-	-	-	-	-	-	-	-	-	-	-	-

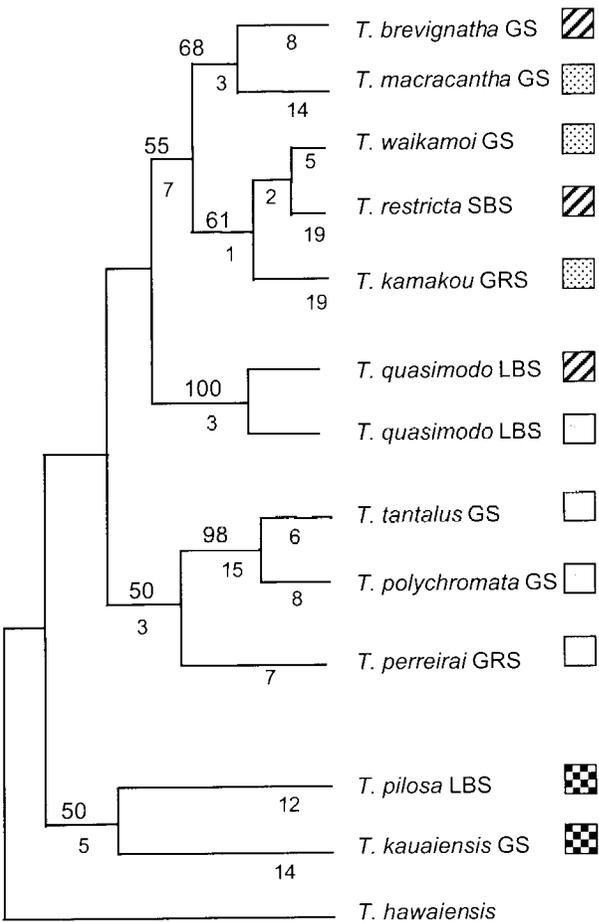
**Figure 1.** A. Cladogram representing the single shortest tree obtained from maximum parsimony analysis of mtDNA (12S and COI) for representatives of the spiny-leg clade of Hawaiian *Tetragnatha* (Gillespie et al. 1997). Numbers at each node indicate bootstrap support; numbers below branches indicate actual numbers of bases changes associated with that branch. B. Rooted Neighbor-Joining tree of all 251 monomer sequences based on their similarity and using *T. hawaiiensis* sequences as the outgroup. Bars represent genetic distance  $d=0.05$  of pair-wise comparisons calculated according to the HKY+ $\Gamma$  model. Numbers at each node indicate bootstrap support. Bootstrap values under 50% are not indicated. The island and ecomorph is indicated for each species: “Green Spiny” (GS), “Green and Red Spiny” (GRS), “Large Brown Spiny”(LBS), and “Small Brown Spiny” (SBS). In the center there is a map of the Hawaiian Islands with the estimated ages in million years.

**Figure 2.** A. Phylogram representing tree obtained from a Clustal W alignment of the consensus stDNA sequences of *Tetragnatha* species of the spiny leg clade (201 characters) using Bayesian inference with a HYK+ $\Gamma$  model. The 50% majority rule consensus sequence for each species was obtained from all clones. In those species showing two subfamilies of monomers, *T. quasimodo*, *T. brevignatha*, and *T. waikamoi*, we used the 50% majority rule consensus of each subfamily to assess the monophyly of those subfamilies. Parsimony analysis using gaps as missing data (28 constant characters and 149 informative ones for a tree length of 447 steps) or 5<sup>th</sup> state (19 constant characters and 156 informative ones for a tree length of 488 steps) retrieved identical topologies. Numbers above branches refer to Bremer Support values (parsimony gaps missing data/

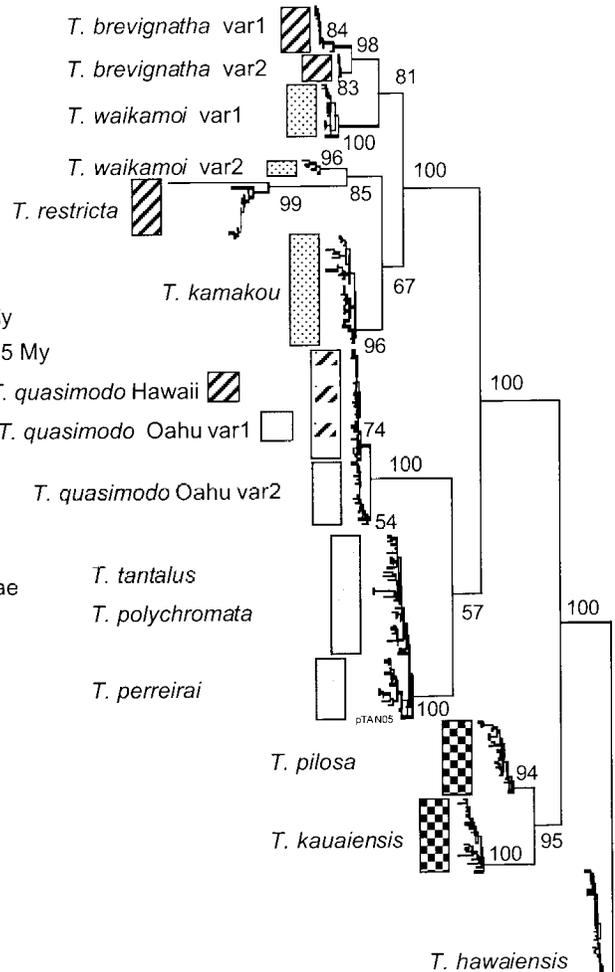
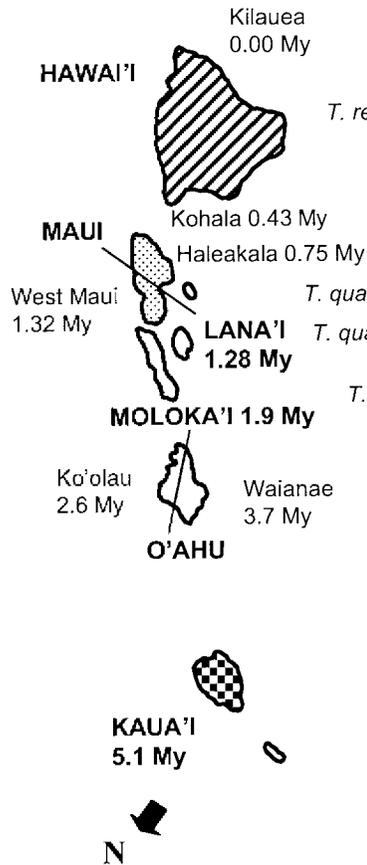
parsimony gaps 5<sup>th</sup> state), and numbers below branches are bootstrap support values (parsimony gaps missing data/ parsimony gaps 5<sup>th</sup> state/ Bayesian posterior probability). B. Cladogram (449 steps) obtained from same consensus sequences under gap cost = 1 using POY vers. 2.6. Gap cost = 2 retrieved the same topology and gaps cost = 4 differed in two nodes (\* *T. pilosa* and *T. kauaiensis* are monophyletic and \*\* *T. restricta* is basal to *T. kamakou*, *T. waikamoi* and *T. brevighatha*). Numbers at each node indicate Bremer Support values for the different gap costs (1/ 2 / 4).

**Figure 3.** Plots of total number of transitions (black diamonds) and transversions (grey squares) as a function of uncorrected divergence in *Tetraglatha* stDNA sequences.

**Figure 4.** Relationship between % sequence change in satellite DNA (from Table 2) and % sequence change in mitochondrial DNA. For the mtDNA, estimates are based on 560 base pairs of Cytochrome Oxidase (Gillespie 1999). The line indicates the expected level of base frequency change in stDNA if the rates were similar to those of mtDNA. There are only 3 points at very low mtDNA divergences at which the rates are comparable. Between the mtDNA divergence levels of 5-10%, stDNA divergences are approx. 4 times greater than those of mtDNA.



**A**



**B**

