Molecular Phylogenetics and Evolution 47 (2008) 950–959

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the eastern United States: Comparison of cuticular hydrocarbons, chromosome number, and DNA sequences

C. Everaerts [a], K. Maekawa [b], J.P. Farine [a], K. Shimada [b], P. Luykx [c], R. Brossut [a], C.A. Nalepa [d,*]

[a] CNRS, UMR 5548, Faculté des Sciences, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France
[b] Department of Biology, University of Toyama, 3190 Gofuku, Toyama 930-8555, Japan
[c] Department of Biology, University of Miami, Coral Gables, FL 33124, USA
[d] Department of Entomology, North Carolina State University, Campus Box 7613, Raleigh, NC 27695-7613, USA

ARTICLE INFO

Article history:
Received 14 June 2007
Revised 2 March 2008
Accepted 6 March 2008
Available online 16 March 2008

Keywords:
Cryptocercus punctulatus
Blattaria
Karyotype
Molecular analysis
Species complex
Cryptic species
Taxonomy

ABSTRACT

The goal of the current study was to determine if cuticular hydrocarbons could be used to empirically delimit taxa within the *Cryptocercus punctulatus* species complex in the eastern United States. Cockroaches were collected from rotting logs in 22 locations across four states. Hydrocarbon phenotypes and two mitochondrial (16S and COII) genes and one nuclear (ITS2) gene were independently analyzed to determine their relationship with chromosome number. Five distinct hydrocarbon phenotypes were found, but these were only partly congruent with chromosome number and thus with purported species descriptions. Molecular and cuticular hydrocarbon data each indicate that *Cryptocercus* with a male karyotype of $2n = 43$ belong to at least two distinct, distantly related lineages. One $2n = 43$ lineage is sister group to the $2n = 37$ and $2n = 39$ clade, and has a unique hydrocarbon profile. The other $2n = 43$ lineage is sister group to the $2n = 45$ samples, and its cuticular hydrocarbons group with four samples of the $2n = 45$ lineage. The cuticular hydrocarbons of two additional $2n = 45$ samples diverge from this assemblage. Results indicate cuticular hydrocarbons and chromosome number have some degree of evolutionary independence; neither is completely reliable in delineating historical lineages. Our data provide support for the parallel model of chromosome evolution in the species complex.

1. Introduction

The wingless, wood-feeding cockroach *Cryptocercus punctulatus* Scudder species complex in the Appalachian mountains of the eastern United States has been difficult to circumscribe taxonomically. Four karyotype groups (male $2n = 37, 39, 43, 45$) are known from within its geographical range. These have been described as separate species (Burnside et al., 2000), but the evidence to date is equivocal (summarized in Nalepa et al., 2002; Lo et al., 2006). Among other inconsistencies, the two criteria used to define species within the complex, chromosome number and molecular characteristics, are in conflict for at least one large population in the Appalachians in the northern half of North Carolina and part of southern Virginia. Molecular analyses of the karyotype groups have resulted in three competing phylogenetic trees: (43(39(45, 37))) (Kambhampati et al., 1996), (45(43(37, 39))) (Burnside et al., 2000; Lo et al., 2006), and (43(45(37, 39))) (Clark et al., 2001). The emerging consensus is that the $2n = 37$ and 39 groups are closely related and relatively apical. The $2n = 39$ group, however, may not be monophyletic (Lo et al., 2006) and appears to be divided into two geographically disjunct populations (Nalepa et al., 2002). The placement of the $2n = 43$ and 45 groups within the tree differs among existing analyses. Reliable morphological differences between karyotype groups are yet to be reported. All examined members of the genus are ecologically, socially, and behaviorally similar (Nalepa and Bandi, 1999; Nalepa et al., 2001; Park et al., 2002; Nalepa, 2003; Folkerts, 2006).

Cuticular hydrocarbons have been used as chemotaxonomic characters to help clarify relationships in a number of insect groups difficult to resolve with morphological or molecular techniques; their use has had some success in delimiting species boundaries in cockroaches. Three species of North American *Blattella,* for example, can be distinguished on the basis of their hydrocarbons, although a statistical technique is required to separate the closely related *B. germanica* and *B. asahinai* (Carlson and Brenner, 1988). Saïd et al. (2005) found that four species of *Periplaneta* have distinct chemical signatures. Tartivita and Jackson (1970), on the other hand, found a remarkably close correspondence in the hydrocarbons of two cockroach species belonging to different genera

* Corresponding author. Fax: +1 919 2338394.
E-mail address: christine_nalepa@ncsu.edu (C.A. Nalepa).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.03.011

(*Blatta orientalis* and *Rhyparobia* (=*Leucophaea*) *maderae*). Two studied species of Australian soil burrowing cockroaches may be more comparable to *Cryptocercus* than the above mentioned vagile, cosmopolitan taxa. *Macropanesthia rhinoceros* exhibits a great deal of population-level variability in cuticular hydrocarbons. No correlating biological differences were found among populations, however, and so the authors considered them as either a single, highly variable species or a complex of unrecognizable sibling species (Brown et al., 2000). Substantial variability was also found in *Geoscapheus dilatatus*, but two hydrocarbon phenotypes could be correlated with the presence or absence of a pair of tubercles on the male's pronotum, suggesting that the complex has some taxonomic subdivision (Brown et al., 1997).

Here we examine cuticular hydrocarbons in the four karyotype groups of *C. punctulatus* in the Appalachian Mountains of the eastern United States to determine if they may be used to infer species limits. We also analyze mitochondrial 16S rRNA and COII, and nuclear ITS2 genes to establish how closely gene sequences correspond to cuticular hydrocarbons and karyotypes.

2. Materials and methods

2.1. Samples

Adult *Cryptocercus* of both sexes were collected from rotting logs in 22 locations (Table 1) in the Appalachian Mountains. Collections ranged over a fairly large geographic area, primarily in North Carolina but also including five sites in Tennessee, three in Virginia, and one in Georgia. Samples were collected between March 2002 and May 2003, and included cockroaches from each of the four karyotype groups known in the eastern United States (male $2n = 37$, 39, 43, and 45). Karyotypes were determined by examining meiotic and mitotic chromosomes from the testes of two to six males from each site, using the method of Luykx (1983).

2.2. Preparation of hydrocarbon extracts

The insects were immobilized in a freezer, and then placed individually into clean glass vials with Teflon-lined caps. Three milliliters of HPLC grade *n*-pentane (Fisher Scientific) were added to cover the insect. The vial was periodically agitated at room temperature, and after 5 min the extracts were transferred via clean glass-pipette to new vials with Teflon-lined caps. The extracts were then kept at −20 °C until analysis. For most sites, 5 males and 5 females were extracted, but sample size ranged from 1 to 6 individuals of a sex per site. Extracts were analyzed in a blind study; researchers performing the analysis were unaware of the karyotypes and origin of the sampled insects.

Table 1
Sample localities

No.	Name, state	County	Chrom. No.	Latitude	Longitude
1	Log Hollow, NC[a,c]	Transylvania	37	35.329	82.821
2	Deep Low Gap, NC[b]	Swain	37	35.512	83.346
3	Noland Divide, NC[b]	Swain	37	35.563	83.471
4	Keg Drive Branch, NC[b]	Swain	37	35.576	83.450
5	Chilowee, TN	Blount	37	35.559	84.011
6	Woodfin Cascades, NC[a,c]	Jackson	39	35.455	83.106
7	Mauney Cove, NC	Haywood	39	35.508	83.024
8	Kanati Fork, NC	Swain	39	35.572	83.385
9	Cabin Flats, NC[b]	Swain	39	35.609	83.333
10	Black Rock Mtn, GA	Rabun	39	34.910	83.416
11	Puckett Cabin, VA[a]	Patrick	43	36.643	80.547
12	Mountain Lake, VA[c]	Giles	43	37.364	80.519
13	Sweet Briar, VA	Amherst	43	37.551	79.098
14	South Mountains, NC[c]	Burke	43	35.603	81.634
15	Clifton, NC[c]	Ashe	43	36.433	81.581
16	Shady Valley, TN[c]	Sullivan	43	36.539	81.951
17	Mt. Pisgah, NC[a]	Transylvania	45	35.416	82.747
18	Utah Mtn, NC	Haywood	45	35.548	82.988
19	Mt. Collins, NC[b]	Sevier	45	35.595	83.476
20	Sweat Heifer, TN[b]	Sevier	45	35.621	83.404
21	Laurel Top, TN[b]	Sevier	45	35.663	83.328
22	Camelback, TN[b]	Cocke	45	35.726	83.207

Location numbers correspond to those in the figures and in the Supplementary material.
[a] Blue Ridge Parkway.
[b] Great Smoky Mountains National Park.
[c] Karyotype previously reported in Nalepa et al. (2002).

2.3. Hydrocarbon chemical analysis

Gas chromatographic analyses of extracts were performed using a Varian 3900 gas-chromatograph (GC) fitted with a flame-ionization detector. A CP Sil 5CB (25 m × 0.25 mm i.d., 0.12 μm film thickness, Chrompack) fused silica capillary column was used. Just prior to injection, the volume of the solvent was reduced to about 500 μl under a light stream of nitrogen. Two microliters of each sample was injected into a split–splitless injection system, operating with a split flow of 60 ml/min and a septum purge of 3 ml/min. The split port was closed during injection, and then opened 30 s after injection. The column was held isothermally at 140 °C for 2 min, then programmed to increase at a rate of 5 °C/min to 280 °C. Helium was used as carrier gas (50 cm/s velocity at room temperature). The injector and detector temperatures were 260 and 280 °C, respectively.

2.4. Hydrocarbon statistical analysis

The data were automatically computed and recorded using PC software (Star 5.2, Varian). As no internal standard was added to the solvent, the peak areas were calculated and expressed as a percentage of the six prominent compounds (*P*1–*P*6), which represented at least 20% of the total extract in at least one sample (see online Supplementary material). Percentages were subjected to an ArcSin (squared root) transformation prior to analysis. The relative quantities of the six components were used for a principal component analysis (PCA type Pearson's correlation matrix (n-1)) with standardized values; all factors corresponding to the five retained compounds were used. Thereafter, the most meaningful variables were used in discriminant analyses (DA, quadratic model). Data were statistically analyzed using XLStat-pro 2007 (Addinsoft).

2.5. DNA extraction, amplification, purification, and sequencing

The samples used for molecular analyses were 22 individuals preserved in 80–100% ethanol, one from each of the 22 sampled locations. Total DNA was extracted from the leg tissue of each individual, using the DNeasy Tissue Kit (Qiagen, Tokyo, Japan). Each individual was considered representative of the population of its collection location. The fragments of mitochondrial 16S rRNA (approximately 410 bp) and COII (448 bp), and nuclear ITS2 (approximately 400 bp) were amplified using PCR. Primer sequences for the amplifications of 16S and COII are given in Kambhampati (1995) and Park et al. (2004), respectively. Primers for ITS2 were newly designed in 5.8S (forward: 5′-CGA TGA AGA ACG CAG CAA A-3′) and 28S (reverse: 5′-TCC TCC GCT TAT TGA TAT GC-3′) rRNA regions. The temperature profile for amplifying the 16S rRNA fragment was 94° for 3 min, followed by 35 cycles of 94° for 30 s, 50° for 1 min and 72° for 1 min, 30 s. For amplifications of COII and ITS2, the temperature profile was 94° for 3 min, followed by 35 cycles of 94° for 1 min, 50 for 1 min and 70° for 2 min. PCR products were purified by using the MagExtractor Kit (Toyobo, Osaka, Japan), then used as templates for sequencing using a DNA sequencer (ABI373; Applied Biosystems, CA, USA). We were unable to amplify and sequence ITS2 of samples #12

and #15 using the direct sequencing method; the amplified fragment was cloned into a pT7Blue-2 T-Vector (Novagen, CA, USA). The insert was amplified from a single bacterial colony using U19mer/SP6 primers, purified and sequenced as described above. Three randomly chosen clones were sequenced and were found to be identical. Nucleotide sequence data ($n = 66$) are available in the DDBJ/EMBL/GenBank databases under the Accession Nos. AB425841–AB425906.

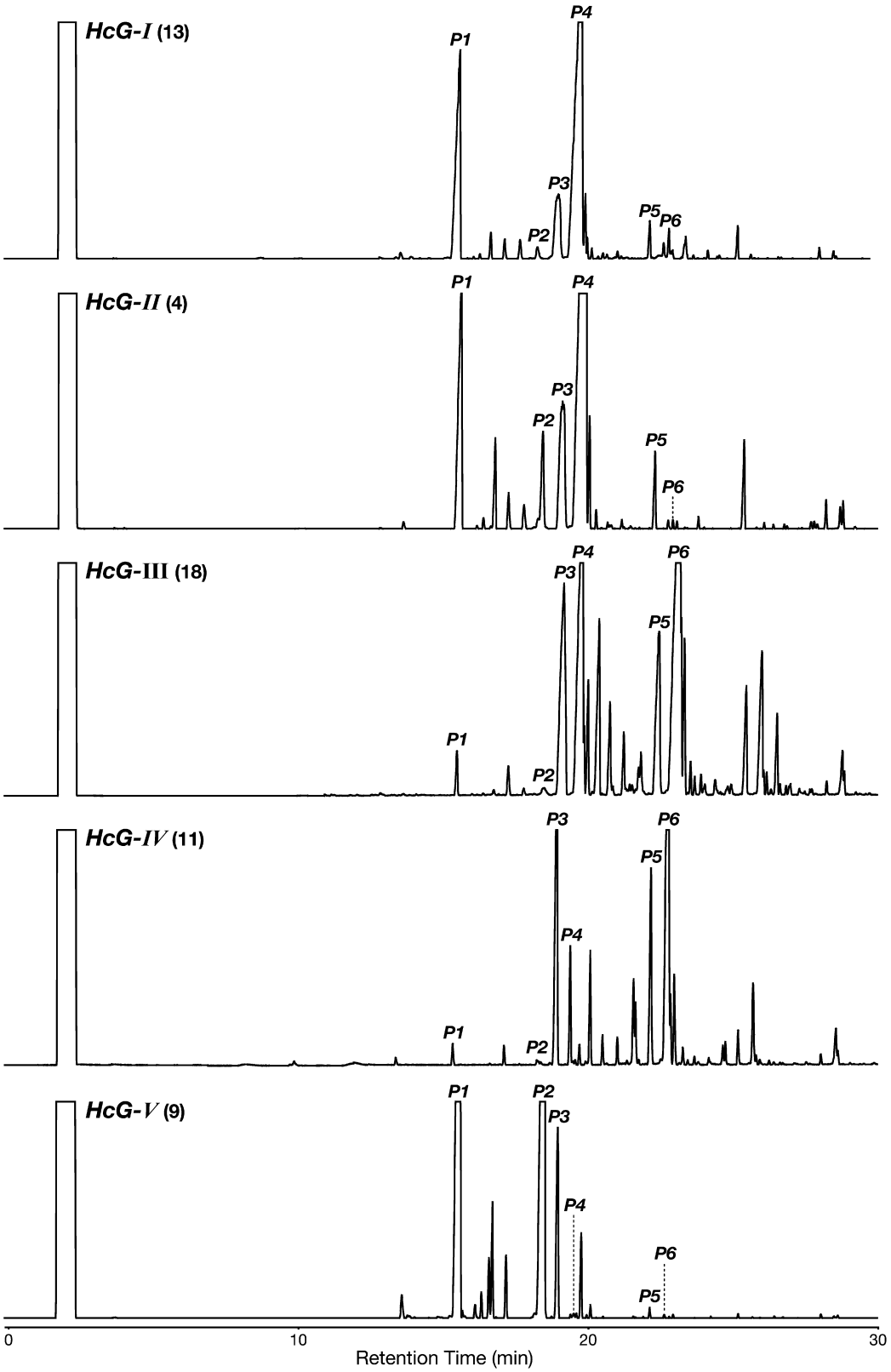

Fig. 1. Representative GC traces of cuticular hydrocarbons from adults of *C. punctulatus*. *HcG-I* to *HcG-V*, distinct groups of hydrocarbon profiles; the number between brackets corresponds to the location number (see Table 1); *P*1–*P*6, representative prominent hydrocarbons retained for analyses. Analysis conditions are described in Section 2.

2.6. Phylogenetic analyses

ClustalX 1.83 program package (Thompson et al., 1997) was used for the alignments of each gene. 16S regions were aligned based on the secondary structure shown in Park et al. (2004), and ITS2 regions were aligned visually (see online Supplementary material). Primer regions were excluded from the analyses. To consider the possibility of mitochondrial introgression, we analyzed mitochondrial (16S and COII) and nuclear (ITS2) datasets separately. First, we analyzed aligned 16S and COII individually. After confirming no contradiction in their topologies, they were integrated into the mitochondrial dataset. Asian *Cryptocercus* (*C. relictus* and *C. kyebangensis*), and *C. clevelandi* from the northwestern United States were also included in the mitochondrial dataset, and the termite *Mastotermes darwiniensis* was used as an outgroup (Accession Nos.: (16S) AB078588, AB078597, AB078558, U38385, AY380302; (COII) AB078548, AB005908, AB078517, AB078557, AB014071). For the nuclear dataset we attempted to use *C. clevelandi* (Accession Nos.: AF322488–9, AF371285–7) as an outgroup, but compared with cockroaches from the Appalachians, there were too many insertions and deletions for a precise alignment. Thus, the ITS2 tree is unrooted. Both mitochondrial and nuclear tree topologies were essentially identical to each other (see Section 3), thus the two datasets were combined and analyzed to infer the phylogenetic tree (*C. clevelandi* was used as an outgroup, but ITS2 data were treated as missing because of the reason mentioned above). Estimations of tree topology were obtained under maximum parsimony (MP) criteria, and by Bayesian Inference (BI), using the programs PAUP* 4.0b10 (Swofford, 2000) and MrBayes 3.0 (Huelsenbeck and Ronquist, 2001), respectively. For BI, the most appropriate model of sequence evolution was determined using Modeltest 3.06 (Posada and Crandall, 1998), and parameters for the selected model of substitution were estimated from the data. A total of 10,000 trees were obtained (ngen = 1,000,000, samplefreq = 100), and the first 1000 of these were considered as the 'burn in' and discarded. A 50% majority-rule consensus tree of the remaining trees was produced. Two independent runs under the same model of sequence evolution were performed. MP trees were estimated heuristically in PAUP* using default options with 100 random addition replicates. About 50% majority-rule bootstrap trees were also obtained using PAUP* (1000 replicates, 10 random addition replicates per bootstrap replicate). All characters were weighted equally, and gaps were treated as a fifth base. For the testing of alternative phylogenetic hypotheses, the Templeton test (MP) and Shimodaira–Hasegawa test (ML) were employed, with the parsimony or likelihood scores of competing topologies compared statistically at the 5% level of significance.

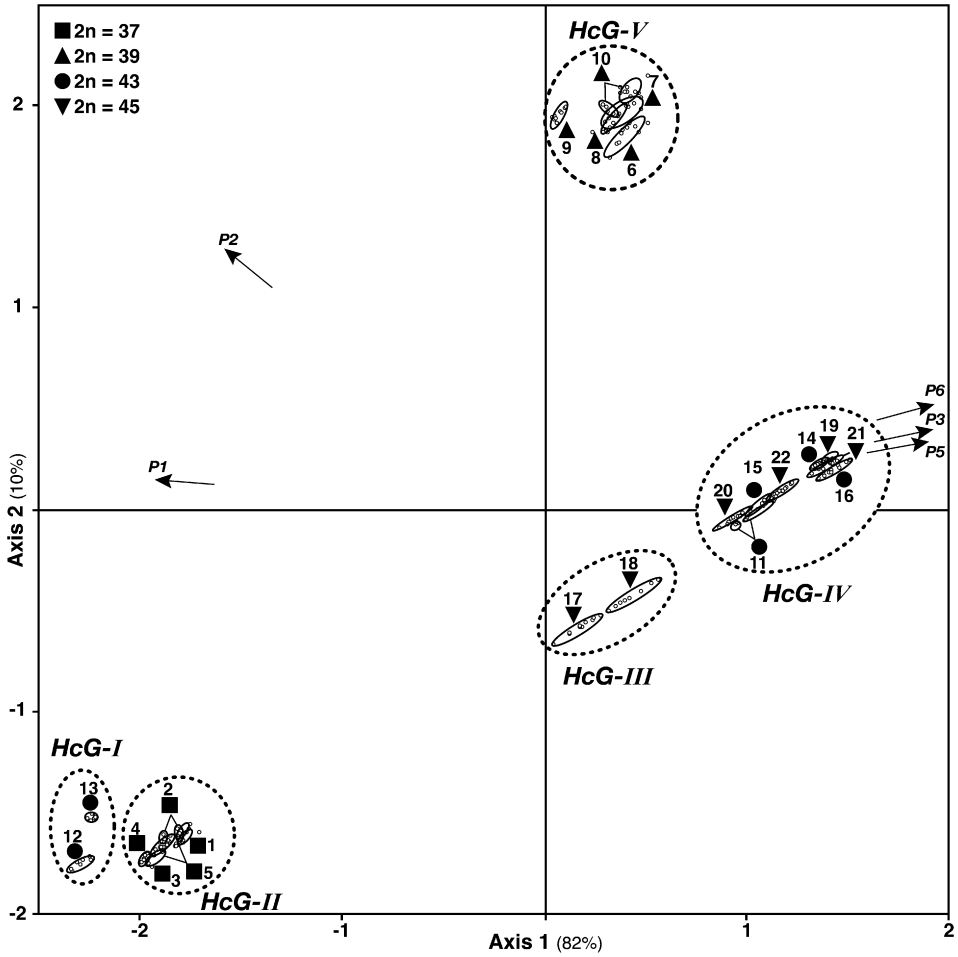

Fig. 2. First plane pf the Principle Component Analysis using the proportion of the six prominent compounds in extracts of cuticular hydrocarbons of the *Cryptocercus punctulatus* species complex from 22 locations. The two first axes explain 82% and 10% of the sample total variance (eigenvalues of 4.10 and 0.51). *P*1, *P*2, *P*3, *P*5, and *P*6 are the vectors corresponding to the prominent cuticular hydrocarbons selected during the preliminary PCA. The solid ellipses represent the $p = 0.90$ confidence limit for the samples from each locality (for the sample localities 2, 5, 10, and 11, one ellipse was drawn for each sex), and the dotted ellipses represents the *HcG* groups (for further information, see text).

3. Results

3.1. Cuticular hydrocarbon profiles

Typical GC traces revealed the presence of about 50 various unidentified compounds that constitute the cuticular hydrocarbon profiles of male and female adult *Cryptocercus*, and the PCA allowed the delineation of five distinct groups of profiles (Fig. 1, *HcG-I* to *HcG-V*). After measuring the elution times of alkane standards run under the same conditions, the chain lengths of these prominent compounds were estimated to range from eicosane ($C_{20}H_{42}$, MW: 282) to dotriacontane ($C_{32}H_{66}$, MW: 450). Comparison of the retention time of the compounds with those of the alkane standards indicate that *P*1, *P*3, and *P*5 have, respectively, the same retention times as the saturated hydrocarbons heneicosane ($C_{21}H_{44}$, MW: 296), tricosane ($C_{23}H_{48}$, MW: 324) and pentacosane ($C_{25}H_{52}$, MW: 352). Further chemical investigations are required to confirm the chemical nature of these peaks.

The PCA analysis indicated that the sampling adequacy of one out of the six major hydrocarbons, peak 4 (*P*4), reached only 0.154. Consequently, this peak was excluded from the next steps of the statistical analysis. The total matrix sampling adequacy for the five other compounds reached 0.864 and in none was variable sampling adequacy lower than 0.815. All the final estimates of communalities were higher than 0.883.

Fig. 2 presents the first plane of the PCA achieved with the five retained compounds (the variables are represented using a correlation biplot). The two axes, respectively, explained 82% and 10% of the sample total variance (eigenvalues of 4.10 and 0.51). The samples corresponding to each sampled location were very closely matched, without sexual dimorphism, except for locations 2, 5, 10, and 11, where adults of the two sexes differed slightly. The samples were distributed into 5 "hydrocarbon groups" (*HcG*). *HcG-I* and *HcG-III* each included only two locations (locations 12 and 13 and locations 17 and 18, respectively). Locations 1–5 and 6–10 constituted two other groups, *HcG-II* and *HcG-V*. The last group, *HcG-IV*, included locations 11 and 14–22. The horizontal segregation between all groups was due to antagonism between the *P*1 and *P*2 compounds (respective squared cosines equal to 0.89 and 0.60), versus the *P*3, *P*5, and *P*6 compounds (squared cosines higher than 0.86), while the vertical segregation was due primarily to the *P*2 compound (*P*2 squared cosines = 0.40; other compounds lower than 0.06).

3.2. Hydrocarbons, karyotypes, and geographic distribution

A comparison of the GC data with the insect karyotypes indicates that groups *HcG-I*, *HcG-II*, *HcG-III*, and *HcG-V* were "monokaryotypic" but the remaining group, *HcG-IV*, pooled samples with two different karyotypes. *HcG-I* consists of cockroaches from two locations in Virginia (12, Mountain Lake and 13, Sweet Briar) in which the chromosome number was $2n = 43$ (Fig. 3). These were the northernmost sites sampled. The *HcG-II* group, closely similar to *HcG-I*, consists of insects from all sampled sites from the southern half of North Carolina and eastern Tennessee where individuals had a karyotype of $2n = 37$. *HcG-III* was composed of samples from two $2n = 45$ sites (17, Mt. Pisgah and 18, Utah Mountain) that were located further east than the other four sampled sites of this karyotype. *HcG-IV* consists of the remaining samples with a karyotype of $2n = 45$, all located in the Great Smoky Mountains National Park, together with four sites from the geographically adjacent, more southern range of the $2n = 43$ karyotype. *HcG-V* is a coherent cluster of all five locations where individuals had a karyotype of $2n = 39$.

3.3. Validity of the defined groups

The validity of the cuticular and karyotype groups was tested in a DA using the data of the five retained peaks as quantitative variables, and the *HcG* or 2*n* as qualitative variables. For the two anal-

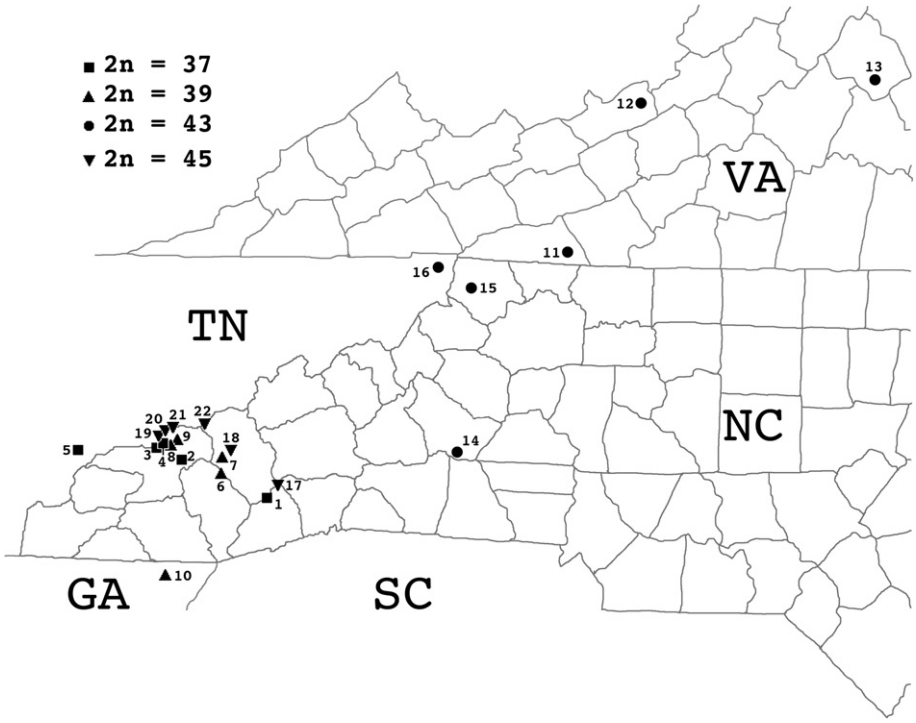

Fig. 3. Map of 22 locations sampled for *Cryptocercus punctulatus* in Virginia, North Carolina (VA and NC, respectively; county lines delineated), Tennessee (TN), and Georgia (GA) in the eastern United States (SC, South Carolina). For a map of geographic relationships of karyotype groups in 71 populations of *C. punctulatus*, see Nalepa et al. (2002: Fig. 2).

yses, the Box test (Chi^2 asymptotic approximation) confirmed that the covariance matrices were different between the groups (for *HcG*: Chi^2 = 1864 with 60 df, $p < 0.0001$; for 2*n*: Chi^2 = 2243 with 45 df, $p < 0.0001$). For both DAs, the Wilks' Lambda test indicated that the difference between the means vectors of the groups is significant (for *HcG*: $F = 1961$, with 20 and 631 df, $p < 0.0001$; for 2*n*: $F = 905$, with 15 and 128 df, $p < 0.0001$).

When using the *HcG* groups as qualitative variable, 59% and 40% of the variance was represented with the two first factors (eigenvalues: 403 and 275). While the horizontal segregation between all groups was due again to the antagonism between the *P*1 and *P*2 compounds, versus the *P*3, *P*5, and *P*6 compounds, the vertical segregation was due primarily to the antagonism between the *P*1 compound and the others. The DA achieved using the *HcG* as qualitative variables did not reclassify any individuals, and we can assume that the a priori hydrocarbon groups provided a 100% accurate classification.

When using the karyotypes as qualitative variables, 97% of the variance is explained with the first factor (eigenvalue: 370). The segregation along this first factor was due again to the antagonism between the *P*1 and *P*2 compounds versus the three other compounds. On the basis of their karyotype, the DA misclassified 10 insects collected from location 14; while they have a chromosome number of $2n = 43$, the DA classifies them as $2n = 45$.

See online Supplementary material for PCA communalities and the DA analysis.

3.4. Molecular analysis

Analyses of the aligned 16S and COII sequences from the 26 *Cryptocercus* taxa and *M. darwiniensis* (306 and 404 bp) yielded 124 and 171 variable sites, of which 80 and 120 are phylogenetically informative, respectively. Analysis of the aligned ITS2 sequences from the 22 *Cryptocercus* taxa (383 bp) yielded 99

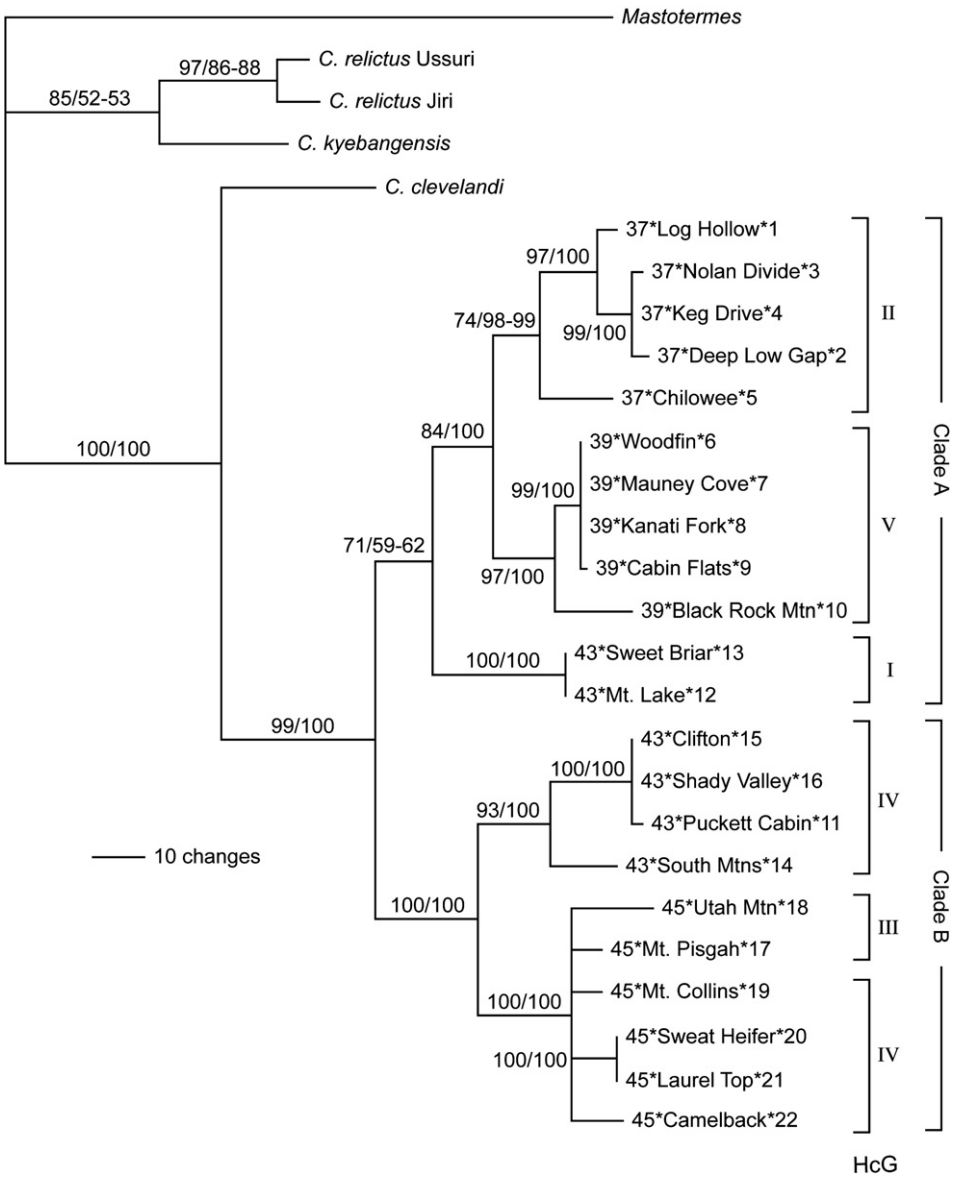

Fig. 4. Phylogenetic relationships of the *Cryptocercus punctulatus* species complex from 22 locations, derived from the combined data set of mitochondrial COII and 16S rRNA genes (710 bp) using parsimony analysis. Tree length, 574; CI, 0.66; RI, 0.81; RC, 0.54; HI, 0.34. The topologies found in MP and BI analyses (TrN + G model of substitution selected by Modeltest3.06) were consistently in agreement with each other. Numbers above or below nodes indicate the percentage of bootstrap support (1000 replicates) and BI posterior probabilities. Only one number is given if BI posterior probability was identical at that node in the two different runs. Each branch is labeled with the diploid male chromosome number, collection location, and sample number, respectively. HcG, hydrocarbon group, corresponding to Fig. 2.

variable sites, of which 85 are phylogenetically informative. Mean base composition was 16S (A = 0.272, C = 0.107, G = 0.191, T = 0.429), COII (A = 0.403, C = 0.181, G = 0.087, T = 0.328) and ITS2 (A = 0.201, C = 0.311, G = 0.304, T = 0.184). The topologies found in Bayesian and MP analysis were consistently in agreement with each other (Figs. 4 and 5). Moreover, inferred topologies from the mitochondrial and nuclear datasets were consistent with each other (Figs. 4 and 5). The results suggest that mitochondrial introgression had not occurred, and hybridization is unlikely in the sampled regions. Although the $2n = 37$, $2n = 39$, and $2n = 45$ karyotypes each formed a monophyletic group with strong support (MP bootstrap values (BV)/BI posterior probabilities (PP): 74/99%, 97/100%, 100/100%, respectively), two groups were represented in samples with a $2n = 43$ karyotype (Fig. 4). Samples 12 and 13 grouped with the $2n = 37$ and $2n = 39$ karyotypes in clade A, while samples 11 and 14–16 grouped with the $2n = 45$ group in clade B.

Analyses of the nuclear dataset also indicated that three of the four karyotypes were genetically more closely related to each other than to other karyotypes (Fig. 5). The individuals from location 12 (Mountain Lake) and 13 (Sweet Briar), however, were distantly related to the other $2n = 43$ individuals. The $2n = 37$ and $2n = 39$ groups in clade A, and the $2n = 43$ (excluding locations 12 and 13) and $2n = 45$ groups in clade B were closely related to each other, with high support values (BV/PP in mtDNA: 84/100% and 100/100%, BV/PP in nuclear DNA: 99/68% and 100/71%, respectively).

The phylogenetic tree inferred from the combined mtDNA and nuclear DNA datasets showed an almost identical topology (Fig. 6A). Tree length is 461, and the log likelihood score of the tree (selected model by Modeltest3.06 was HKY + G) was −3418.45. We tested alternative topologies, showing monophyly of each karyotype (Fig. 6B and C), based on the statistical comparisons of parsimony tree-lengths (Templeton test) and likelihood scores (Shimodaira–Hasegawa test). The resulting tree-lengths were significantly longer, and the likelihood scores were significantly lower than those of most likely topology (Fig. 6).

4. Discussion

Five distinct hydrocarbon phenotypes were represented in the 22 samples of *Cryptocercus* analyzed in this study, but these would be of limited utility in circumscribing purported species in the taxon. Among the locations we sampled, cuticular hydrocarbons from insects of the $2n = 39$ group formed a distinct, coherent, homogenous

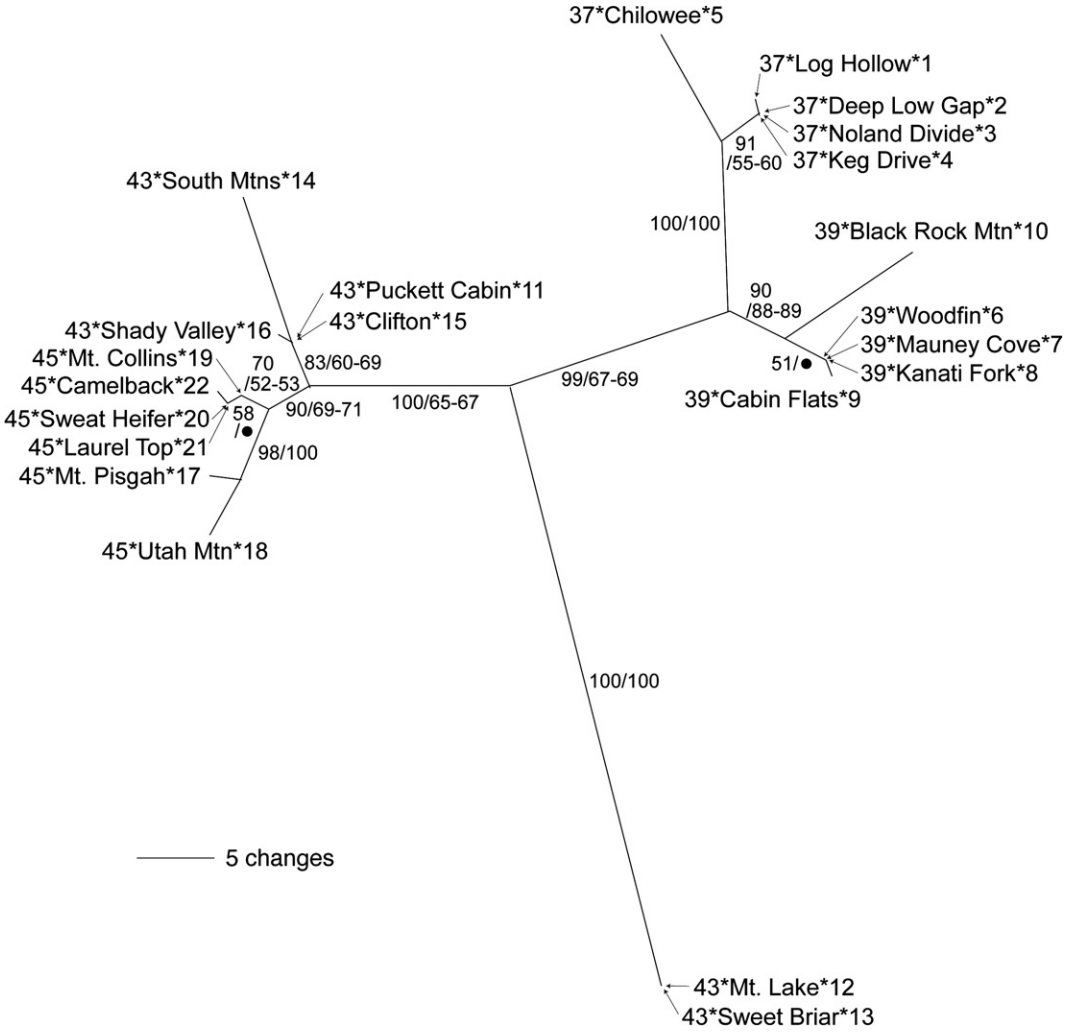

Fig. 5. Unrooted bootstrap parsimony tree of the *Cryptocercus punctulatus* species complex from 22 locations based on the nuclear of ITS2 sequences (383 bp including gaps). The topologies found in MP and BI analyses (HKY + G model of substitution selected by Modeltest3.06) were consistently in agreement with each other. Numbers above or below nodes indicate the percentage of bootstrap support (1000 replicates) and BI posterior probabilities. Only one number is given if BI posterior probability was identical at that node in the two different runs. Dot indicates that a node that was not supported in greater than 50% of BI posterior probabilities. Each branch is labeled with the diploid male chromosome number, collection location, and sample number, respectively.

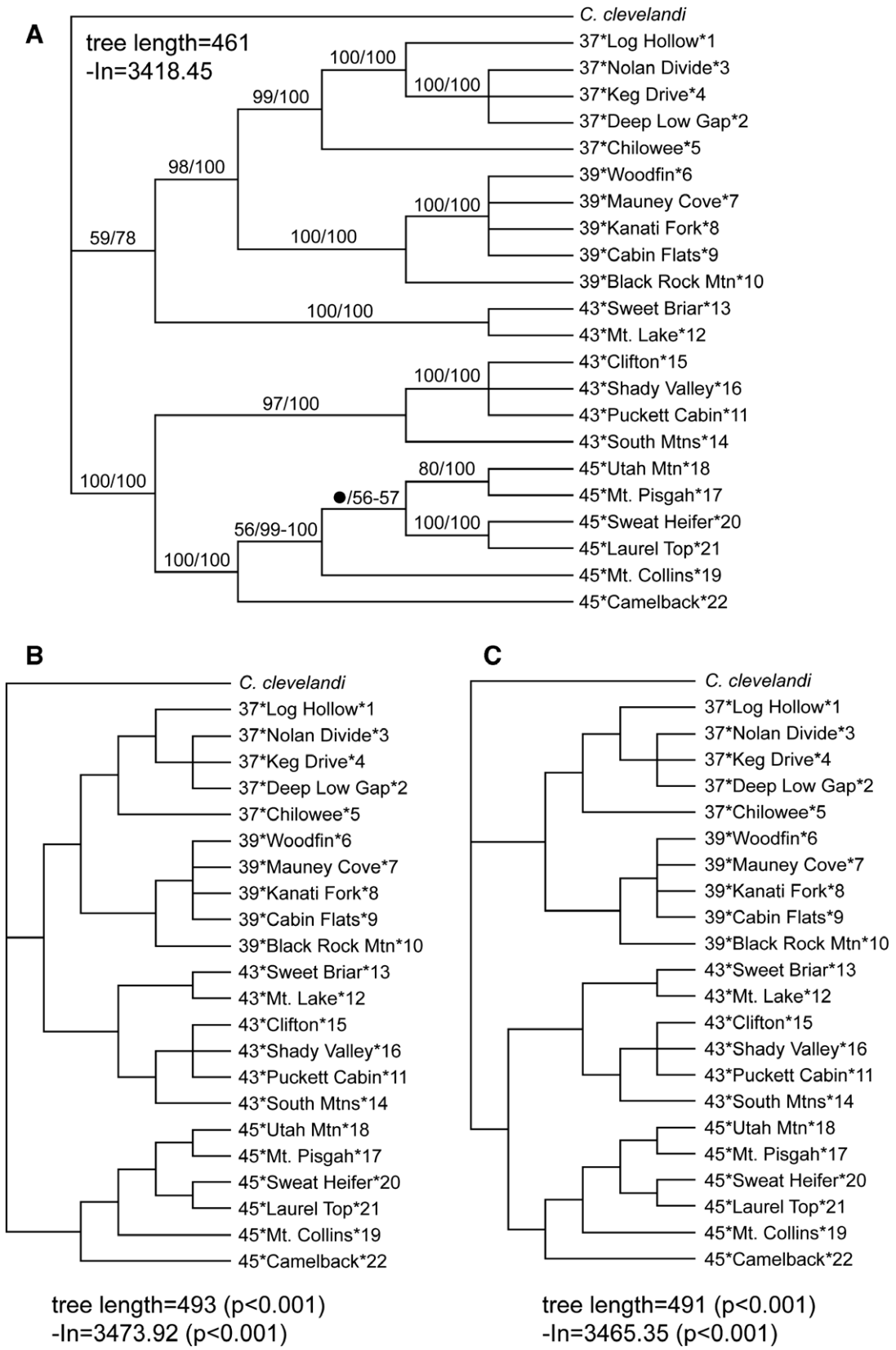

Fig. 6. (A) Strict consensus of three most parsimonious trees of the *Cryptocercus punctulatus* species complex from 22 locations derived from the combined datasets of mtDNA and nuclear DNA (1093 bp). Tree length, 461; CI, 0.74; RI, 0.89; RC, 0.66; HI, 0.26. The topologies found in MP and BI analyses (HKY + G model of substitution selected by Modeltest3.06) were consistently in agreement with each other. Numbers above or below nodes indicate the percentage of MP bootstrap support (1000 replicates) and BI posterior probabilities. Only one number is given if BI posterior probability was identical at that node in the two different runs. Each branch is labeled with the diploid male chromosome number, collection location name, and sample number, respectively. (B and C) Alternative topologies showing monophyly of each karyotype. Both tree-lengths and the likelihood scores were significantly different from those of most parsimonious one (A).

grouping (*HcG-V*). Cockroaches with a chromosome number of $2n = 37$ formed a distinct cluster (*HcG-II*), closely similar to *HcG-I* ($2n = 43$, two locations). *HcG-III* was made up of two of the $2n = 45$ samples, and so did not include all samples of this karyotype. Both $2n = 43$ (four locations) and $2n = 45$ (four locations) were included in *HcG-IV*.

The molecular analysis resulted in a phylogenetic tree divided into two well defined clades. There are three distinct sub-clades in clade A, with a close correspondence between DNA sequence data and male diploid chromosome number. Each of these subclades has a unique cuticular hydrocarbon phenotype. In clade B the results are more ambiguous. Although the molecular analyses separated the samples in concordance with chromosome number, the division is not mirrored in the hydrocarbon profiles. This is evident in the samples from locations 11, 14, 15, and 16 ($2n = 43$), whose hydrocarbons were indistinguishable from those in *HcG-IV* with a chromosome number of $2n = 45$. Possible explanations for these results include: first, the $2n = 43$ population in clade B was derived from a $2n = 45$ subpopulation with a distinct hydrocarbon profile; second, convergent evolution of hydrocarbons in the $2n = 45$ and $2n = 43$ populations; or third, some degree of interbreeding occurred between $2n = 45$ and $2n = 43$ sub-populations with different hydrocarbon profiles.

A key result of this study is that our samples of *Cryptocercus* with a chromosome number of $2n = 43$ do not form a monophyletic group. They fall into one of two different hydrocarbon profiles, and are confirmed by molecular analysis to belong to different clades (SH and Templeton tests; Fig. 6). These results are consistent with genetic variation previously described in *Cryptocercus* cockroaches with a chromosome number of $2n = 43$ (Nalepa et al., 2002: Fig. 5). The existence of at least two genetically distinct groups in an apparently geographically contiguous population with the same chromosome number helps explain the difficulty in past studies in determining whether the $2n = 43$ or the $2n = 45$ karyotype is basal in competing phylogenetic trees (Nalepa et al., 2002: Fig. 1).

Geographically, samples from clade B bisect samples from clade A. Locations 12 and 13 ($2n = 43$) are at the northern end of the range, while the remainder of the samples from clade A are at the southern extreme of the sampled range. The sample from location 14 ($2n = 43$), whose cuticular hydrocarbons cannot be differentiated from several samples of $2n = 45$ (*HcG-IV*), was collected in the South Mountains; these are geographic outliers east of the main chain of the Appalachians and separated from them by the Catawba River Valley (Nalepa, 2001). The remaining $2n = 43$ samples in *HcG-IV* are located in northwestern North Carolina (location 15, Clifton), northeastern Tennessee (location 16, Shady Valley), and in southern Virginia, near the North Carolina border (location 11, Puckett Cabin). At the southern extreme, it is of interest that two disjunct populations with a chromosome number of $2n = 39$ have not significantly diverged in either cuticular hydrocarbons or in the DNA sequences examined here. Sample 10, Black Rock Mountain, is from a population geographically disjunct from the remainder of the $2n = 39$ sites sampled in this study; a region occupied by $2n = 37$ cockroaches separates them (Nalepa et al., 2002; Nalepa and Luykx, unpublished data). In this case, geographic disconnection has not resulted in significant genetic divergence. On the other hand, the hydrocarbons of samples 17 and 18 (*HcG-III*) have diverged from the remainder of the $2n = 45$ sites sampled (*HcG-IV*). These two sites are the easternmost samples of the apparently contiguous $2n = 45$ group, and their unique hydrocarbons may indicate population level variation. The geographic ranges of the karyotype groups in the Southern Appalachians, however, particularly in the Great Smoky Mountains and adjacent Blue Ridge Parkway, are indicative of complex, interdigitating, parapatric populations (Nalepa et al., 2002; Nalepa and Luykx, unpublished data). Until the limits of geographical distribution of the different molecular, hydrocarbon, and karyotype groups are mapped and analyzed on a fine scale, it is premature to speculate on the taxonomic significance of these results. The area we sampled is limited in relation to the extensive range of the species complex in the eastern United States, and there are numerous small, geographically isolated populations near the range limits (Nalepa et al., 2002; Folkerts, 2006).

Nalepa et al. (2002) offered two different evolutionary scenarios that could account for the four karyotype groups known in *C. punctulatus*. The extreme version of the parallel scenario is that all the karyotype variants arose independently, and perhaps more than once, from an ancestral population via centric fusions. In the sequential scenario, an ancestral karyotype ($2n = 47$) gave rise to a second containing a chromosomal fusion ($2n = 45$), which in turn gave rise to a third with an additional chromosomal fusion ($2n = 43$), and so on. Analysis of the COII gene from cockroaches from 15 locations by Lo et al. (2006) upheld the serial reduction hypothesis, although support for the relationships among the chromosomal lineages was not high. They found that taxa with the same number of chromosomes formed monophyletic groups, with the exception of the two disjunct populations of the $2n = 39$ karyotype. Only two samples of their data set, however, overlapped with the locations used in this study. All methods of defining taxa are sensitive to sampling design (Sites and Marshall, 2004). This is particularly true for groups as geographically substructured as this one and should be a consideration in further studies of this cockroach.

In sum, our results indicate that a change in chromosome number is not necessarily accompanied by a change in the genes that code for cuticular hydrocarbons; samples of both $2n = 43$ and $2n = 45$ are in *HcG-IV*. On the other hand, differentiation of cuticular hydrocarbons can occur with minimal genetic and karyotype change (Marcillac et al., 2005); cockroaches within the $2n = 45$ clade may be either *HcG-III* or *HcG-IV*. Chromosome number and cuticular hydrocarbons, then, have some degree of evolutionary independence, and neither is completely reliable in delineating historical lineages. Crossing studies may be useful in exploring the relative influence of genic and chromosomal changes in determining reproductive compatibility. Cockroaches with a chromosome number of $2n = 43$ in clade A may not be meiotically compatible with those of the same chromosome number in clade B, because of additional chromosomal rearrangements such as inversions or translocations, because of expansion or loss of heterochromatin, or because populations with $2n = 43$ arose from an ancestral $2n = 45$ population more than once with different chromosomes involved in the fusions; this possibility was envisioned in our earlier study of geographic karyotype variability (Nalepa et al., 2002). Crosses of populations of different chromosome number with a shared hydrocarbon profile ($2n = 43$ and $2n = 45$ in *HcG-IV*), as well as crosses of the same chromosome number whose cuticular hydrocarbons have diverged ($2n = 45$ in *HcG-III* versus *HcG-IV*) would be of value in exploring the full scope of taxonomic complexity in *C. punctulatus*.

Acknowledgments

Insects were collected under permits BLRI-2003-SCI-0015 and GSRM-2005-SCI-0084 from the National Park Service, and yearly permits from the North Carolina Wildlife Resources Commission. We thank Linda Fink and Patrick Rand for assistance in the field. This study was supported in part by a Grant-in-Aid for Scientific Research (Nos. 16770014 and 199770012 to K.M.) from the Japan Society for the Promotion of Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.03.011.

References

Brown, W.V., Rose, H.A., Lacey, M.J., 1997. The cuticular hydrocarbons of the soil burrowing cockroach *Geoscapheus dilatatus* (Saussure) (Blattodea: Blaberidae: Geoscapheinae) indicate species dimorphism. Comp. Biochem. Physiol. 118B, 549–562.

Brown, W.V., Rose, H.A., Lacey, M.J., Wright, K., 2000. The cuticular hydrocarbons of the giant soil-burrowing cockroach *Macropanesthia rhinoceros* Saussure (Blattodea: Blaberidae: Geoscapheinae): analysis with respect to age, sex and location. Comp. Biochem. Physiol. 127B, 261–277.

Burnside, C.A., Smith, P.T., Kambhampati, S., 2000. Three new species of the wood roach *Cryptocercus* (Blattodea: Cryptocercidae) from the eastern United States. J. Kansas Entomol. Soc. 72, 361–378.

Carlson, D.A., Brenner, R.J., 1988. Hydrocarbon-based discrimination of three North American *Blattella* cockroach species (Orthoptera: Blatellidae) using gas chromatography. Ann. Entomol. Soc. Amr. 81, 711–723.

Clark, J.W., Hossain, S., Burnside, C.A., Kambhampati, S., 2001. Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. Proc. R. Soc. Lond. B 268, 393–398.

Folkerts, G.W., 2006. The range of the eastern woodeating cockroach, *Cryptocercus punctulatus* Scudder (Blattaria: Cryptocercidae) in the United States. Entomol. News 117, 101–108.

Huelsenbeck, J.P., Ronquist, F., 2001. Mrbayes: bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Kambhampati, S., 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc. Nat. Acad. Sci. USA 92, 2017–2020.

Kambhampati, S., Luykx, P., Nalepa, C.A., 1996. Evidence for sibling species in *Cryptocercus punctulatus*, the wood roach, from variation in mitochondrial DNA and karyotype. Heredity 76, 485–496.

Lo, N., Luykx, P., Santoni, R., Beninati, T., Bandi, C., Casiraghi, M., Lu, W.H., Zakharov, E.V., Nalepa, C.A., 2006. Molecular phylogeny of *Cryptocercus* wood-roaches based on mitochondrial COII and 16S sequences, and chromosome numbers in Palearctic representatives. Zool. Sci. 23, 393–398.

Luykx, P., 1983. XO-XX sex chromosomes and Robertsonian variation in the autosomes of the wood-roach *Cryptocercus punctulatus* (Dictyoptera: Blattaria: Cryptocercidae). Ann. Entomol. Soc. Amr. 76, 518–522.

Marcillac, F., Bousquet, F., Alabouvette, J., Savarit, F., Ferveur, J.F., 2005. A mutation with major effects on *Drosophila melanogaster* sex pheromones. Genetics 171, 1617–1628.

Nalepa, C.A., 2001. *Cryptocercus punctulatus* (Dictyoptera: Cryptocercidae) from monadnocks in the Piedmont of North Carolina. J. Entomol. Sci. 36, 329–334.

Nalepa, C.A., 2003. Evolution in the genus *Cryptocercus* (Dictyoptera: Cryptocercidae): No evidence of differential adaptation to hosts or elevation. Biol. J. Linn. Soc. 80, 223–233.

Nalepa, C.A., Bandi, C., 1999. Phylogenetic status, distribution, and biogeography of *Cryptocercus* (Dictyoptera: Cryptocercidae). Ann. Entomol. Soc. Amr. 92, 292–302.

Nalepa, C.A., Li, L., Lu, W., Lazell, J., 2001. Rediscovery of the wood-eating cockroach *Cryptocercus primarius* (Dictyoptera: Cryptocercidae) in China, with notes on ecology and distribution. Acta Zootax. Sin. 26, 184–190.

Nalepa, C.A., Luykx, P., Klass, K.-D., Deitz, L.L., 2002. Distribution of karyotypes of the *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the Southern Appalachians: relation to habitat and history. Ann. Entomol. Soc. Amer. 95, 276–287.

Park, Y.C., Grandcolas, P., Choe, J.C., 2002. Colony composition, social behavior and some ecological characteristics of the Korean wood-feeding cockroach (*Cryptocercus kyebangensis*). Zool. Sci. 19, 1133–1139.

Park, Y.C., Maekawa, K., Matsumoto, T., Santoni, R., Choe, J.C., 2004. Molecular phylogeny and biogeography of the Korean woodroaches *Cryptocercus* spp. Mol. Phylog. Evol. 30, 450–464.

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

Saïd, I., Costagliola, G., Leoncini, I., Rivault, C., 2005. Cuticular hydrocarbon profiles and aggregation in four *Periplaneta* species (Insecta: Dictyoptera). J. Ins. Physiol. 51, 995–1003.

Sites Jr., J.W., Marshall, J.C., 2004. Operational criteria for delimiting species. Ann. Rev. Ecol. Sys. 35, 199–227.

Swofford, D.L., 2000. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

Tartivita, K., Jackson, L.L., 1970. Cuticular lipids of insects: I. Hydrocarbons of *Leucophaea maderae* and *Blatta orientalis*. Lipids 5, 35–37.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL:X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.