Phylogenetic analysis of non-stereotyped behavioural sequences with a successive event-pairing method

Frédéric Legendre, Tony Robillard, Laure Desutter-Grandcolas, Michael Whiting, Philippe Grandcolas

Phylogenetic analysis of non-stereotyped behavioural sequences with a successive event-pairing method

FRÉDÉRIC LEGENDRE[1]*, TONY ROBILLARD[1], LAURE DESUTTER-GRANDCOLAS[1], MICHAEL F. WHITING[2] and PHILIPPE GRANDCOLAS[1]

[1]*Muséum national d'Histoire naturelle, Département Systématique et Évolution, UMR 5202 CNRS, CP 50 (Entomologie), 45 rue Buffon 75005 Paris, France*

[2]*Department of Integrative Biology, 693 Widtsoe Building, Brigham Young University, Provo, Utah 84602, USA*

A new method is proposed which uses transitions among acts in non-stereotyped behavioural sequences as phylogenetic characters. This method is derived from the event-pairing method designed for the phylogenetic study of developmental sequences and from ethological analyses of transition matrices. It is applied to study the phylogenetic relationships within a well-known group, the presocial Zetoborinae cockroaches. The analysis is carried out with three data sets: a behavioural data set with transitions among acts in behavioural dyadic sequences, together with morphological and molecular data sets. Non-stereotyped behaviour proved to be phylogenetically informative and to display low homoplasy. This new method opens an avenue for studying the evolution of behaviour in the framework of phylogenetic analysis, which was restricted until now to the study of stereotyped sequences and/or isolated features involved in courting or building activities.

ADDITIONAL KEYWORDS: behaviour – cockroaches – evolution – phylogeny – transition matrices – Zetoborinae.

INTRODUCTION

Studies on behaviour and phylogeny have a long common history. From the early times, comparative studies have shown that behaviour can be remarkably informative regarding the relationships of taxa, as reviewed by Hinde & Tinbergen (1958). These comparative studies focused on highly stereotyped and ritualized behaviours, such as courting or nest building (Wenzel, 1992). Being stereotyped, these behaviours were easily compared among species to assess their patterns of evolution (Lorenz, 1941; Hinde, 1955; Tinbergen, 1959). In this context, the homology of stereotyped behaviour has been repeatedly discussed by several seminal papers (Baerends, 1958; Atz, 1970; Hodos, 1976; Lauder, 1986; Wenzel, 1992; Greene, 1994) with reference to the classical criteria of homology proposed by Remane (1952) and especially the criteria of position and special quality. The rationale behind using behaviour for phylogenetic inference was that behavioural acts that are repeatedly observed in various individuals and populations without strong variation or evidence for learning can be presumptively considered as inherited, in the same way as other phenotypic traits, such as morphology or development (Wenzel, 1992). More recently, many studies confirmed the old perception (de Queiroz & Wimberger, 1993; Proctor, 1996; Brooks & McLennan, 2002) that behaviour is phylogenetically informative and that it can be studied within such an evolutionary perspective (e.g. Coddington, 1986; McLennan, Brooks & McPhail, 1988; Wenzel, 1993; Kennedy, Spencer & Gray, 1996; Johnson *et al.*, 2000; Cap, Aulagnier & Deleporte, 2002; Noll, 2002; Price & Lanyon, 2002; Desutter-Grandcolas & Robillard,

2003; Robillard, Höbel & Gerhardt, 2006a). However, treatments of behavioural characters greatly vary among the different studies. In the best case, particular behaviours are used as characters or attributes in phylogenetic analyses, coded 'present' vs. 'absent' (Deleporte, 1993; Grandcolas *et al.*, 2001). In the worst case, some very broad behavioural classes are mapped onto the phylogenetic trees, leading to potentially biased reconstructions (as reported by Proctor, 1996; Desutter-Grandcolas & Robillard, 2003; Grandcolas & D'Haese, 2004).

This way to treat behaviour in phylogenetic analyses ignores the fact that behavioural acts do not exist in isolation: they are expressed in a context, usually as an answer to a stimulation by a conspecific or the environment, and they can be combined or repeated. For example, a particular grooming act in a grooming sequence is not equivalent to a similar grooming act in a dyadic (i.e. involving two interacting individuals) agonistic sequence; this grooming act is said to be 'displaced' in the second case (McFarland, 1993). Behaviour is therefore better described and behavioural homology can be better established with reference to the criterion of position when the place of different behaviours is considered along the behavioural sequence (Robillard *et al.*, 2006b) and also with reference to the criterion of special quality, as reviewed by Wenzel (1992). When behavioural sequences are stereotyped, as in the case of highly ritualized behaviours, the comparisons among species are straightforward as there is only one sequence per species. These sequences may be aligned and analysed as was carried out by ethologists in non-phylogenetic and intraspecific comparisons (Abbott, 1995; Wilson, Harvey & Thompson, 1999; Abbott & Tsay, 2000; Schlich, 2001; Van der Aalst *et al.*, 2003; Hay, Wets & Vanhoof, 2004). Alignments of stereotyped behavioural sequences for phylogenetic analysis can be generated in a dynamic way (Robillard *et al.*, 2006b) via direct optimization (Wheeler, 1996) or used in a static way in a standard phylogenetic analysis (Wilson *et al.*, 1999). However, when sequences are not stereotyped within species there is presently no way to use them directly to study species phylogenetic relationships.

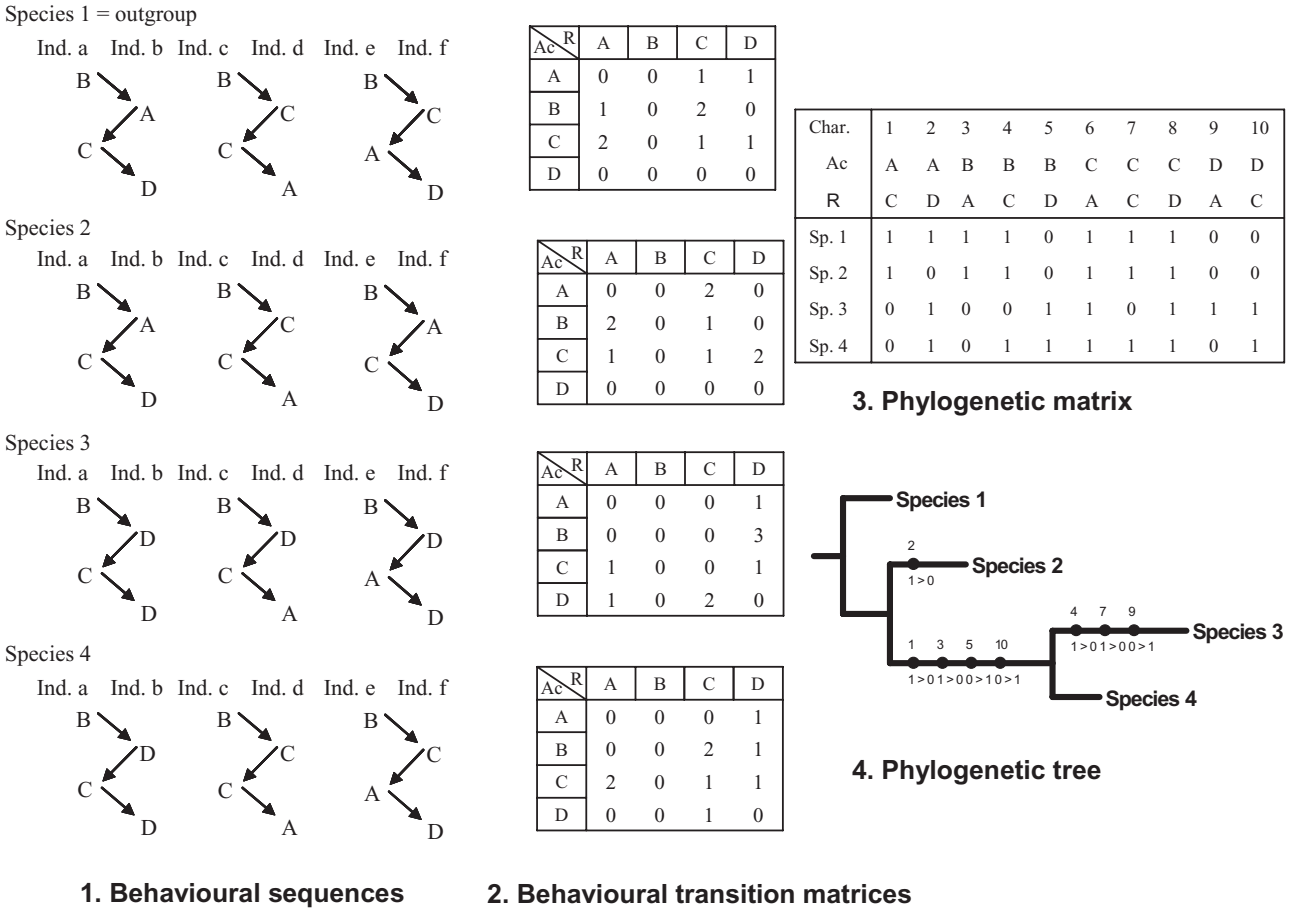
The aim of this paper is therefore to propose a new methodology allowing the study of non-sterotyped sequences in a phylogenetic framework and to show its potential by applying it to a typical case of sequences obtained from social interactions. This application is made for some social cockroaches, the behaviour of which has already been much studied (Gautier, 1974; Grandcolas, 1991; van Baaren & Deleporte, 2001; van Baaren *et al.*, 2002, 2003a).

ESTABLISHING A SUCCESSIVE EVENT-PAIRING METHOD TO STUDY BEHAVIOURAL SEQUENCES

The method presented here was inspired by the procedure of event-pairing, which was developed simultaneously by Mabee & Trendler (1996), Smith (1997) and Velhagen (1997) to study developmental sequences (Bininda-Emonds *et al.*, 2002; Jeffery *et al.*, 2002, 2005). Using that method, developmental sequences are recoded in all possible pairwise combinations of events, thereby encoding the relative position of each item in the sequence. Each developmental event is coded as possibly occurring before, simultaneously or after any other event (Jeffery *et al.*, 2005). Event-pairing coding has been recently challenged by Schulmeister & Wheeler (2004) who suggested that treating features of sequences as if they were independent can produce inconsistent reconstructions. This is mainly because developmental sequences are more constrained in terms of temporal linearity than are ethological sequences with respect to the biological process involved. For example, many developmental events cannot occur earlier or later within a sequence because the structures where they should take place are not yet developed or cannot develop twice (Schulmeister & Wheeler, 2004).

The limitations of event-pairing coding as applied to the study of development are not a problem for non-stereotyped behavioural sequences, where the same event can be expressed several times within the sequence. There is not a one-to-one relationship between a linear sequence and a species, but many different sequences for the same species. Our goal is to code in a phylogenetic context the occurrence and frequency of transitions between two acts among many differently ordered sequences. Only transitions between two successive events are considered here and not the relative position between all pairs of events. This methodology is named the successive event-pairing method to avoid confusion with the event-pairing method.

Establishing a matrix of characters coding the occurrence of two successive events in a behavioural sequence is already part of the current statistical analyses of behavioural sequences (Fig. 1). Ethologists build matrices of transition where each cell is filled with the frequency of a transition between two particular acts (Martin & Bateson, 1986; Gottman & Roy, 1990; Bakeman & Quera, 1995). These frequencies are then organized in flow charts generated by hand or according to correspondence analyses (van der Heijden, 1987) to investigate how different acts are organized in different kinds of sequences and to compare them between different species. To adapt this procedure to the successive event-pairing



Ac \ R	A	B	C	D
A	0	0	1	1
B	1	0	2	0
C	2	0	1	1
D	0	0	0	0

Ac \ R	A	B	C	D
A	0	0	2	0
B	2	0	1	0
C	1	0	1	2
D	0	0	0	0

Ac \ R	A	B	C	D
A	0	0	0	1
B	0	0	0	3
C	1	0	0	1
D	1	0	2	0

Ac \ R	A	B	C	D
A	0	0	0	1
B	0	0	2	1
C	2	0	1	1
D	0	0	1	0

Char.	1	2	3	4	5	6	7	8	9	10
Ac	A	A	B	B	B	C	C	C	D	D
R	C	D	A	C	D	A	C	D	A	C
Sp. 1	1	1	1	1	0	1	1	1	0	0
Sp. 2	1	0	1	1	0	1	1	1	0	0
Sp. 3	0	1	0	0	1	1	0	1	1	1
Sp. 4	0	1	0	1	1	1	1	1	0	1

Figure 1. A scheme for the establishment of a successive event-pairing method to analyse the evolution of behavioural sequences. In this theoretical example, four species with the same behavioural repertoire (acts A, B, C, D) are studied and three behavioural sequences were observed (1) which are used to build transition matrices (2). For the construction of such matrices, the individuals are alternatively the actor (Ac) and the receiver (R). The presence-absence of transitions within these matrices is used in a second step to characterize each species in a phylogenetic matrix (3) which in turn is analysed to build a phylogenetic tree (4). Char., character; Ind., individual; Sp., species.

method, we only need to consider that the cells of matrices of behavioural transitions can be used as phylogenetic characters (Fig. 1). This is justified on the basis of the classical criteria of homology applied to comparative ethology (Wenzel, 1992). Homology based on the cells of such matrices fits the criterion of position as it defines a particular succession of two acts, thus specifying the position of one act relative to another. In this context, events A occurring after an event B or after an event C are not considered homologous, strictly speaking: answering by A after B or after C is not the same behaviour and will not necessarily be shown by all species even if the fixed motor pattern involved in displaying A is the same in each case. This is easy to understand if one considers a real example where a species would tend to answer to conspecific aggression by escape, while another species would answer by reciprocal aggression. Ethologists have long known that this kind of difference can be species-specific or common to related species. The occurrence of transitions between two particular acts can be treated as presence–absence characters. The frequencies of transitions can also be used as it is very different to observe that a given transition is very rare or very common. Either a low or a high frequency can be considered characteristic of species and therefore used in phylogenetic analysis as characters. Frequencies can be discretized and coded in different character states using gap coding (Archie, 1985; Stevens, 1991). Recently, Goloboff, Mattoni & Quinteros (2006) argued that continuous characters need not to be discretized. However, their methodology treats continuous characters as additive characters, which requires an assumption of progressive evolution that we do not want to follow here. The study of quantitative characters has proved to be a

difficult problem in phylogenetics and the present work is not aimed at solving it. We will focus this work on the most commonly used approaches: discretization and gap coding.

Our method requires that all these behavioural patterns, both the acts and the trends of succession among acts, are largely heritable and that their plasticity and variability are low. This is the most general and necessary assumption that should be substantiated to legitimate phylogenetic studies of behaviour. It can be partly carried out in evaluating the congruence of the phylogenetic tree based on behavioural data with molecular and morphological data. Obviously, assessing the minimal sampling effort is needed to document correctly the behavioural acts, their transitions and frequencies and their independence. These assumptions are no different than for other phenotypic characters (morphology, cytology, etc.), as already argued by Wenzel (1992).

MATERIAL AND METHODS

An illustrative case study: gregarious behaviour in Zetoborinae cockroaches

Non-stereotyped behavioural sequences are most often observed in the context of social relationships. The observed behavioural sequences are not a series of acts successively emitted by the same individual, a situation which could occur with other behaviours such as territorial displays, grooming activities, etc., but a series of acts emitted by two individuals in alternation. These behavioural relationships are rarely stereotyped and there can be different answers to a particular act from a conspecific, depending on the context, and several different acts can initiate a sequence of interaction between two conspecifics.

Gregarious behaviour in cockroaches is a famous example of presocial behaviour (Schal, Gautier & Bell, 1984; Gautier, Deleporte & Rivault, 1988; Nalepa & Bell, 1997; Grandcolas, 1999; van Baaren *et al.*, 2002, 2003b). It has been recently analysed in a molecular and morphological comparative framework in the subfamilies Zetoborinae and Blaberinae (Grandcolas, 1991, 1993a, b, 1998; Pellens, Legendre & Grandcolas, 2007a; Pellens *et al.*, 2007b), which provided both a phylogenetic reference and a natural history context for the interpretation of social behaviour observed in the laboratory (Grandcolas, 1991; van Baaren & Deleporte, 2001; van Baaren *et al.*, 2002, 2003a). In a first attempt to understand the evolution of social behaviour, relevant categories such as 'gregarious', 'solitary' and 'subsocial', have been mapped onto a phylogenetic tree based on morphology (Grandcolas, 1993a, 1998). Ethological studies have shown that the gregarious behaviours should be analysed in detail and contrasted between species, not only considering broad behavioural categories (Grandcolas, 1991; van Baaren & Deleporte, 2001; van Baaren *et al.*, 2002, 2003a). A small group of closely related species has been already well studied and will be used as an example of how the successive event-pairing method can be applied to behavioural sequences to infer a phylogenetic tree congruent with other data and to propose hypotheses of behavioural evolution. We will not discuss the issues of behavioural plasticity and repertoire sampling which have been explored and controlled for in the specific papers cited earlier.

Dyadic interactions in four species – namely *Thanatophyllum akinetum* Grandcolas, 1991, *Schultesia lampyridiformis* Roth, 1973, *Lanxoblatta emarginata* Burmeister, 1838 and *Phortioeca nimbata* Burmeister, 1838 – have been observed among 11 to 17 groups of six nymphs placed in standard conditions for each species, according to the protocol described in Grandcolas (1991) and van Baaren *et al.* (2002). Each group has been placed in an open-field arena. The observations began 1 h later, lasted 15 min and have been recorded on a Samsung Digital Camcorder VP-D11. The observations have been carried out on nymphs in the middle of their development as this is the most characteristic and intense period of gregarious behaviour (Grandcolas, 1993b; van Baaren & Deleporte, 2001). One additional outgroup species *Eublaberus distanti* Kirby, 1903 from the closely related subfamily Blaberinae has also been observed. Transition matrices have been constructed using the behavioural sequences reported as described in Figure 1. Phylogenetic analyses have been carried out with these behavioural matrices and also by comparison with a morphological and molecular data set. The morphological data are taken from Grandcolas (1993a, 1998). Molecular data for all the species and behavioural data for *Eublaberus distanti* have been acquired for the present study.

The sampling effort for behaviour has been critically evaluated with respect to the previous behavioural studies that were conducted and published on the same insect species (e.g. van Baaren *et al.*, 2002). Accumulation curves for the occurrence of transitions according to the number of observations have been computed to show that the sampling effort is large enough to observe the transitions, either uncommon or frequent, occurring in each species.

Character independence has also been evaluated by checking whether frequencies of transitions involving a similar behavioural act are not misleadingly correlated. This can be easily tested with a χ^2 goodness-of-fit test (Chatfield & Lemon, 1970; Zar, 1999), which verifies whether the frequency of a transition between two acts can be determined by the total frequency

of each act involved. Basically, this test compares expected frequencies with observed frequencies. Following Zar (1999), some data have been pooled together in order to have an average expected frequency of at least six, which avoids bias in χ^2 computation.

Primers, PCR and sequencing

Leg muscle tissue was excised from roach specimens preserved in 100% ethanol. DNA was extracted using the Qiagen DNeasy protocol for animal tissue. Mitochondrial ribosomal DNA large subunit (16S, ~385 bp), nuclear ribosomal DNA small subunit (18S, ~1875 bp) and nuclear ribosomal DNA large subunit (28S) domains A (~360 bp) and C (~330 bp) were amplified. 18S was amplified and sequenced in four overlapping fragments corresponding to GA, AD1D2, BCE and EF domains. PCR reactions were lead on a DNA Engine DYAD™, Peltier Thermal Cycler with the following conditions: an initial heating step of 94 °C for 2 min followed by 40 cycles of 94 °C for 60 s, 55 °C for 60 s and 72 °C for 75 s. Then a final elongation at 72 °C was carried out over 7 min. The different primers used which have already been published are listed in Table 1. Electrophoresis gel was used to visualize PCR products and to check that there was no contamination as a result of a negative control. PCR products were purified via the Montage PCR_{96} Cleanup Kit (Millipore) and sequenced using ABI Big Dye 3.1 with the following sequence profile: 27 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 min. Sequencing reactions products were purified with Sephadex™ columns and fractionated on an ABI 3730 XL DNA sequencer. Each sequence was edited using Sequencher 4.0 (Genecodes, 1999) and blasted on GenBank (http://www.ncbi.nlm.nih.gov/blast/) to check for contamination. All the sequences (16S/18S/28SA/28SC, respectively) were deposited on GenBank under the following accession numbers: *Eublaberus distanti* (EU367504/EU367508/EU367511/EU367516), *Lanxoblatta emarginata* (EU367505/EU367509/EU367512/EU367517), *Phortioeca nimbata* (EU367506/EU367510/EU367513/EU367520), *Schultesia lampyridiformis* (EF363280/EF363251/EU367515/EU367520) and *Thanatophyllum akinetum* (EU367507/EU367503/EU367514/EU367519).

Phylogenetic analyses

Even if repertoires are basically identical among species and therefore cannot convey a great deal of phylogenetic information, their analysis has been carried out for a matter of comparison (analysis A). Using the successive event-pairing method, a second phylogenetic analysis has been performed on the presence–absence of transitions taken as characters (analysis B). In a third analysis (analysis C), characters based on the discretized frequencies of the transitions were added to the B data set. The marginal frequencies of each initiating act have been calculated. For instance, in species 1, one event A and two events C have been observed in answer to an event B (Fig. 1). Then, the frequencies for the transitions B/A, B/B, B/C and B/D are 0.33, 0.00, 0.67 and 0.00, respectively. All the frequencies for the same initiating act have been pooled throughout all the species to form a distribution. Each distribution of frequencies was discretized and coded using gap coding (Archie, 1985; Stevens, 1991).

The morphological data set has been taken from the matrix of 78 characters used by Grandcolas (1993a), considering only the five species of the present study.

Table 1. PCR primers names and sequences with the seven targeted portions of DNA

Genes	Primers	Sequences (5′-3′)	Sources
16S	16SAr	CGCCTGTTTATCAAAAACAT	Xiong & Kocher (1991)
	16SF	TTACGCTGTTATCCCTAA	Kambhampati (1995)
18S GA	1F	TACCTGGTTGATCCTGCCAGTAG	Giribet *et al.* (1996)
	b5.0	TAACCGCAACAACTTTAAT	Whiting *et al.* (1997)
18S AD1D2	2F	AGGGTTCGATTCCGGAGAGGGAGC	Hillis & Dixon (1991)
	b2.9	TATCTGATCGCCTTCGAACCTCT	Jarvis, Haas & Whiting (2004)
18S BCE	a1.0	GGTGAAATTCTTGGACCGTC	Whiting *et al.* (1997)
	7R	GCATCACAGACCTGTTATTGC	Whiting (2002)
18S EF	a3.5	TTGTGCATGGCCGYTCTTAGT	Whiting (2002)
	9R	GATCCTTCCGCAGGTTCACCTAC	Giribet *et al.* (1996)
28S A	Rd1.2a	CCCSSGTAATTTAAGCATATTA	Whiting (2002)
	Rd3b	CCYTGAACGGTTTCACGTACT	Jarvis *et al.* (2004)
28S C	28SA	GACCCGTCTTGAAGCACG	Whiting *et al.* (1997)
	28SB	TCGGAAGGAACCAGCTAC	Whiting *et al.* (1997)

As it represents few parsimony-informative characters (9), it has been analysed together with the molecular data (analysis D). The molecular data set represents portions of genes 16S, 18S and 28S. Molecular sequences, which present a very low variability in length, were aligned using Muscle 3.6 (Edgar, 2004). Finally, behavioural data (including frequencies) have been combined together with the morphological + molecular data set (analysis E).

All characters were equally weighted and coded as non-additive. All parsimony analyses have been performed using PAUP4.0b10 (Swofford, 1998) with an exhaustive search to ensure finding the most parsimonious tree. Consistency index (CI; Kluge & Farris, 1969), retention index (RI; Farris, 1989) and number of parsimony-informative characters were recorded. Bremer support values were computed with the help of TreeRot.v2b (Sorenson, 1999). Bootstrap values were computed for 1000 replicates using PAUP4.0b10. Character optimizations on the phylogenetic trees were performed with fast procedure (i.e. accelerated transformation ACCTRAN) using Winclada version 1.00.08 (Nixon, 2002).

Because the monophyly of the ingroup has been established previously (Grandcolas, 1993a, 1998), the ingroup was designated as monophyletic in tree visualization.

Analyses C and E include characters based on the frequencies of transitions and consequently include inapplicable characters. If a particular behavioural transition is not observed in one or several species, the character based on its frequencies is inapplicable for that or those species. Those characters were coded with a dash ('-') in the matrices but were interpreted as missing data during the tree search. This 'reductive coding' better reflects the information content of the data (Strong & Lipscomb, 1999).

The behavioural data were tested for significant structure using the permutation tail probability test (PTP; Faith & Cranston, 1991) and the g1 statistics (Hillis, 1991; Hillis & Huelsenbeck, 1992) in PAUP4.0b10 (Swofford, 1998). However, even if there is a structure in a phylogenetic tree based on behavioural data, this structure can be a reflection of common evolutionary ecological pressures rather than of phylogenetic relationships (Kennedy *et al.*, 1996). If the behavioural tree is congruent with a tree based on other data sets (morphological + molecular here), it is most likely to be as a result of a common phylogenetic signal in the different data sets. Behavioural and morphological + molecular trees were tested for congruence using the triplets tree comparison metric (Symmetric Difference of triplets, SDt hereafter) as implemented in the software Component v2.0 (Page, 1992); the lower the value, the more congruent the trees. This value can be compared with a null distribution calculated after generating all topologies with five leaves unrooted (options 'generate all' and 'tree-to-tree distances/triplets/SD').

RESULTS

Twenty-four different acts were identified from the behavioural sequences of the different species (Tables 2 and 3). *Eublaberus distanti* and *Schultesia lampyridiformis* have the largest behavioural repertoire (19 acts), whereas *Lanxoblatta emarginata* and *Phortioeca nimbata* have the smallest and the same one (15 acts). The outgroup *E. distanti* displays some autapomorphic acts and notably the sudden jump (SJ) and the act PS (when an individual puts its pronotum under the other individual and stands up suddenly), both of which are known in *Blaberus*, the sister genus of *Eublaberus* (Gautier, 1974).

Transition matrices (Appendix S1) have been constructed using the behavioural sequences. Accumulation curves showed that the total number of observations is large enough to observe all frequent or uncommon transitions, the number of which has reached a plateau (e.g. for *Eublaberus distanti*, Fig. 2). According to the matrices, *S. lampyridiformis* was the most active cockroach with more than 1300 behavioural transitions (Appendix S1) representing a mean of 65 interactions/h. Conversely, *P. nimbata* and *E. distanti* were the less active (*c.* 30 interactions/h). We found no correlation between the amount of activity and the number of kinds of transitions among the five species, which shows again that there is no bias related to a sampling effect depending on different species activities ($r = 0.70$, $P > 0.15$). According to the behavioural matrices (Appendix S1), some transitions were never observed

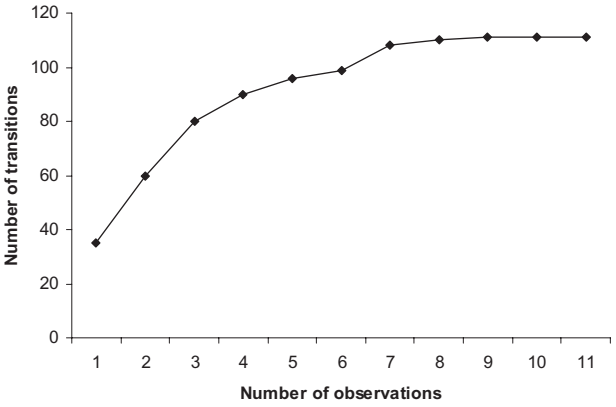


Figure 2. Accumulation curve of the number of different transitions according to the number of observation sessions for the species *Eublaberus distanti*.

Table 2. List of the different behaviours displayed by the cockroaches (also established according to Gautier, 1974; Grandcolas, 1991; van Baaren *et al.*, 2002, 2003a, b)

Behaviours promoting interactions	MT	Moving towards the other individual
	AC	Antennal contact with the body of the other individual
	MA	Mutual antennation
	CB	Climbing onto the body of the other individual with one to six legs
Behaviours favouring departure or a break in the interactions	PS	An individual puts its pronotum under the other and stands up suddenly
	KL	An individual kicks the other with one leg
	PP	An individual pushes the other with its pronotum
	SP	Stilt posture. An individual rises on its legs
	SA	Stilt posture combined with antennal movement
	SD	Rapidly after a stilt posture, an individual goes down
	BI	An individual bites the other
	SJ	An individual jumps suddenly towards the other
	GD	An individual goes down
	GA	An individual goes down and hides its antennae
	RO	Rotation: turning away from the other individual (without significant displacement of the centre of gravity of the body)
	WD	Withdrawal. An individual moves away from the other
	WP	An individual moves away from the other but stops in proximity
	ES	Escape. An individual moves quickly away from the other
	WA	Sudden withdrawal of the antenna(e)
	TP	Tilt posture. An individual gives way of legs on its stimulated side
	FP	Freezing posture. An individual does not move at all
Behaviours without particular significance	GA	Grooming behaviour of the antenna
	GL	Grooming behaviour of the leg
	NO	Nothing. The individual does nothing notable

Table 3. Behavioural repertoires of the five cockroaches used in this study (also established according to Gautier, 1974; Grandcolas, 1991; van Baaren *et al.*, 2002, 2003a, b)

	MT	AC	MA	CB	PS	PP	KL	SP	SA	SD	SJ	BI
Eublaberus distanti	X	X	X	X	X	X		X			X	
Lanxoblatta emarginata	X	X	X	X				X	X	X		
Phortioeca nimbata	X	X	X	X				X	X	X		
Thanatophyllum akinetum	X	X	X	X			X		X	X		X
Schultesia lampyridiformis	X	X	X	X		X	X	X	X			X

	GD	GA	TP	FP	RO	WD	ES	WP	GA	GL	WA	NO
Eublaberus distanti	X	X	X	X	X	X		X	X	X	X	X
Lanxoblatta emarginata	X	X			X	X		X	X	X		X
Phortioeca nimbata	X	X			X	X		X	X	X		X
Thanatophyllum akinetum	X	X			X	X	X	X	X	X		X
Schultesia lampyridiformis	X	X	X		X	X	X	X	X	X		X

(empty cells of the matrices), while others were rarely or commonly observed (cells filled with a small or a large integer, respectively). As a mean, the behavioural sequences include from 6 to 7 acts depending on the species, ranging from very short (two acts) to quite long (25 acts). Transition matrices show that many transitions are not observed in some species and that others are rare or frequent. Statistical independence among different transitions for a same species was assessed according to χ^2 goodness-of-fit tests which were highly significant (*Eublaberus*: $\chi^2 = 380.3$, ddl = 56, $\chi^2_{0.05} = 74.5$, $P << 0.05$; *Lanxoblatta*: $\chi^2 = 297.3$, ddl = 81, $\chi^2_{0.05} = 103.0$, $P << 0.05$; *Phortioeca*: $\chi^2 = 222.1$, ddl =

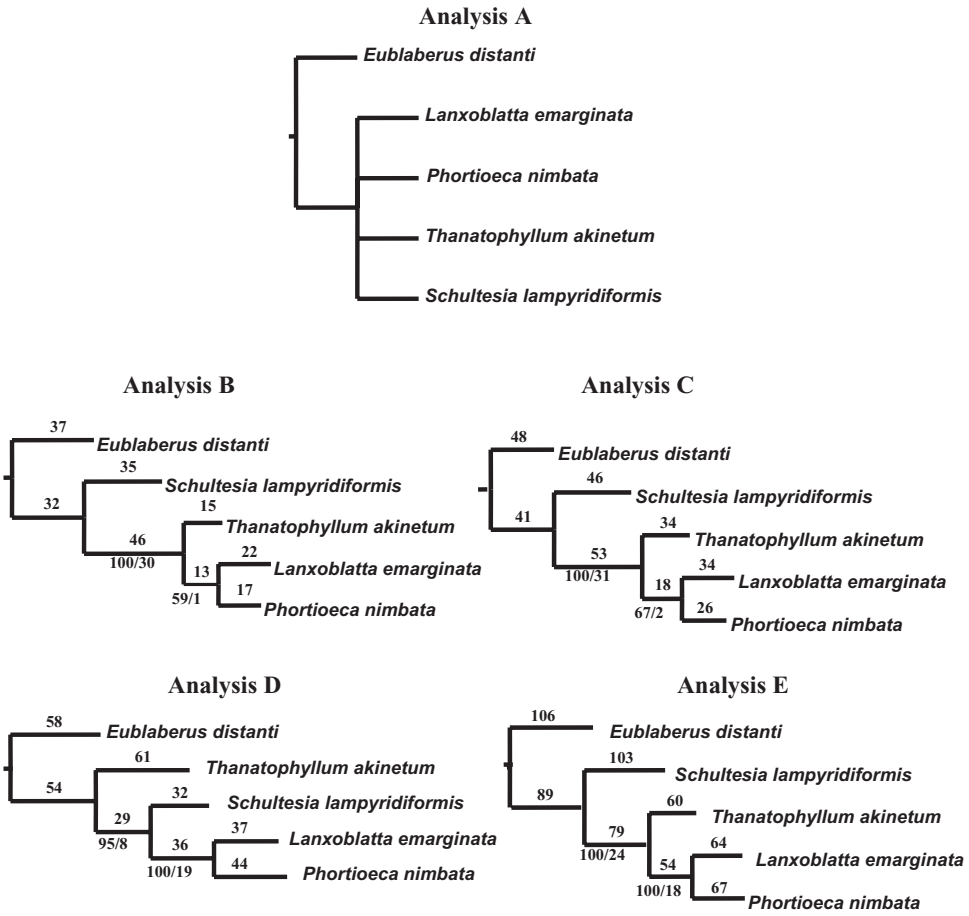


Figure 3. Analysis A, strict consensus of the four most parsimonious trees obtained (L = 15, CI = 0.80, RI = 0.50) with the data set based only on the behavioural repertoires. Analyses B–E, most parsimonious trees obtained with the four different data sets (B, presence–absence of behavioural transitions, L = 217, CI = 0.86, RI = 0.58; C, presence–absence and frequencies of behavioural transitions, L = 300, CI = 0.87, RI = 0.55; D, morphological and molecular data set, L = 351, CI = 0.88, RI = 0.50; E, molecular, morphological and behavioural data – including frequencies, L = 659, CI = 0.86, RI = 0.48). Numbers above and below branches are branch lengths (under fast optimization) and bootstrap/Bremer values, respectively.

72, $\chi^2_{0.05} = 92.8$, $P << 0.05$; *Schultesia*: $\chi^2 = 908.6$, ddl = 144, $\chi^2_{0.05} = 173.0$, $P << 0.05$; *Thanatophyllum*: $\chi^2 = 250.7$, ddl = 56, $\chi^2_{0.05} = 74.5$, $P << 0.05$). This trend can be easily observed in the transition matrices where the values within the same rows or lines are very different.

The phylogenetic analysis based on behavioural repertoires of the different species (Table 3) resulted in four most parsimonious trees (six informative characters, L = 15 steps, CI = 0.80, RI = 0.50), the strict consensus of which is totally unresolved (Fig. 3A). Even if the repertoires are rich, their phylogenetic analysis is not decisive enough for establishing a resolved tree.

Phylogenetic matrices were built up from the transitions matrices as explained in Figure 1. The data set (B) based on the presence-absence of transitions taken as characters comprised 576 characters, 74 of which were parsimony informative (i.e. 12.8 %). Its analysis resulted in a single most parsimonious tree (L = 217 steps; CI = 0.86, RI = 0.58; Fig. 3B). The addition of characters based on frequencies of the transitions (analysis C) gave also a single most parsimonious tree with the same topology and similar indices (L = 300 steps; CI = 0.87, RI = 0.55; Fig. 3C). The matrix comprised 1152 characters, with 89 parsimony-informative characters (7.7 %). In both trees, no zero-length branches were reconstructed which means that behavioural data bring information at different levels. The analysis of morphology + molecular data set yielded one most parsimonious tree with a slightly different topology: the positions of

Table 4. Phylogenetic analyses with their data sets, results and statistics

Analyses	data sets	Number of MPT	L (steps)	CI	RI	Number of informative characters	% of informative characters
A	Repertoire	4	15	0.80	0.50	6	25.0
B	Behaviour	1	217	0.86	0.58	74	12.8
C	Behaviour including frequencies	1	300	0.87	0.55	89	7.7
D	Molecular and morphology	1	351	0.88	0.50	88	2.9
E	Behaviour, molecular and morphology	1	659	0.86	0.48	177	4.3

CI, consistency index; L, length of the MPT; MPT, most parsimonious trees; RI, retention index.

Table 5. Partitioned Bremer values

Partitions	Node (L, P)	Node [T, (L, P)]	Σ Bremer	Number of informative characters	%	%
Behaviour	1	31	32	74	1.03	0.91
Behaviour (frequencies only)	1	1	2	15	0.32	
Morphology	6	1	7	9	1.85	0.22
16S	1	–5	–4	46	–0.21	
18S	4	–2	2	15	0.32	
28S	5	–2	3	18	0.40	
TOT	18	24	42	177		

'Σ Bremer' is the total Bremer support for each partition and '%' is the percentage of Bremer values that each partition supports normalized by the number of parsimony informative characters. Then, for each partition i, $\%_i = [(\Sigma \text{ Bremer }_i/\Sigma \text{ Bremer }_{TOT}*100)/N_i]$; with N_i the number of informative characters for the partition i. For the last column, partitions are: behaviour including frequencies and morphology + molecular

S. lampyridiformis and *T. akinetum* are inverted with regard to the analyses B and C (L = 351 steps, CI = 0.88, RI = 0.50; Fig. 3D). The matrix comprised 3034 characters among which 88 were parsimony informative (2.9 %). Finally, the combined analysis resulted in one most parsimonious tree similar to those based on behavioural characters (L = 659, CI = 0.86, RI = 0.48; Fig. 3E). Among the 4186 characters, 177 were parsimony informative (4.3 %). This latter topology showed maximal bootstrap values and very high Bremer values (18 and 24). Statistics of all those analyses have been summarized in Table 4. They demonstrated that behavioural data were not more homoplastic than molecules and morphology and that behaviour brought proportionally more informative characters (up to 12.8 % of informative characters vs. 2.9 %). Table 5 included the partitioned Bremer support values and the number of informative characters for each partition. It revealed that behaviour (including frequencies) brought more than 80% of the signal [(32 + 2)*100/42 = 80.95]. Even when the support of each partition was normalized by the number of informative characters, behavioural data appeared the most informative together with morphological data.

The g1 statistics and the PTP test indicated that behavioural data were highly structured (data set B: g1 = –1.406, $P < 0.01$; PTP test for 1000 replicates, $P = 0.001$) and the comparison between trees of the analyses B and D resulted in a SDt value of 0.2. Compared with the null distribution, this result suggested that those two trees (behavioural and morphological + molecular) were rather congruent. Only 12 pairwise comparisons between topologies out of 105 have a SDt value lower than 0.2.

The comparison of the trees and their statistics produced by the analyses B and C revealed to what extent presence–absence of behavioural transitions and their frequencies were informative. First, the B data set included 74 informative characters, whereas the C data set included 89 informative characters. Therefore, frequencies brought few supplementary informative characters (89 – 74 = 15) when compared with presence–absence of behavioural transitions.

Second, Bremer values increased (increment of one for both clades) which means that the two data sets are not contradictory. However, frequencies gave less support to the clade [*T. akinetum*, (*L. emarginata* – *P. nimbata*)] than did the presence–absence of behavioural transitions. The comparison of branch lengths confirmed this point. Indeed, the increase of branch length between the two topologies was, in proportion, the smallest for this clade. On the contrary, frequencies brought relatively more information than presence-absence of behavioural transitions for the clade (*L. emarginata* – *P. nimbata*) with a Bremer support value of one for 15 informative characters vs. a Bremer support value of one for 74 informative characters, respectively (see Table 5).

Using the topology retrieved in the combined analysis (Fig. 3), we have looked at the behavioural transitions supporting different nodes of the tree.

First, every branch was supported by a reasonable amount of changes. By comparison with repertoires which were basically identical for all species, this means that transitions offered a large amount of information with common states and differences among every species and group of species.

Second, acts supporting Zetoborinae with respect to the Blaberinae outgroup corresponded mainly to absences of agonistic or avoiding acts: slip ones pronotum under the other and stand up suddenly (PS), push the other with ones pronotum (PP), sudden jump (SJ) and sudden withdrawal of the antennae (WA), move away but stop in proximity (WP), rotation (RO) or freezing posture (FP), respectively. In the other way, Zetoborinae displayed other 'negative' acts in answer to 'positive' sollicitation (transitions which are not expressed by *Eublaberus*). Those 'negative' acts were: stilt posture with antennation (SA), goes down (GD), goes down and hides its antennae (GA) and withdrawal (WD).

Third, species with contrasted social behaviour (e.g. the solitary *T. akinetum* vs. other gregarious species) did not show a special amount of difference but particular changes. The behavioural repertoire of *T. akinetum* revealed only one autapomorphic behaviour: the absence of stilt posture (SP). Therefore, despite its solitary behaviour, *Thanatophyllum* did not display an especially idiosyncrasic repertoire with regard to gregarious species. However, according to the successive event-pairing analysis, we were able to determine that *T. akinetum* was the only species displaying the following behavioural transitions, all of which limiting inter-individual interactions: move towards/bite (MT/BI), antennal contact/leg kick (AC/KL), stilt posture with antennation/move away but stop in proximity (SA/WP), bite/escape (BI/ES), rotation/leg kick (RO/KL), move away but stop in proximity/goes down and hides its antennae (WP/GA), nothing/goes down and hides its antennae (NO/GA).

We also looked at its gregarious sister group, composed of two closely related species which were supposed to have very similar social gregarious behaviour, *L. emarginata* and *P. nimbata*. These species were not only very close morphologically, they also lived in similar habitats (under loose bark of trees, Grandcolas, 1993b) and exhibit the same behavioural repertoire. Thirteen transitions supported this clade and notably five transitions involving acts of antennation (antennal contact and mutual antennation, AC and MA, respectively). Moreover, on the five characters based on frequencies and supporting this clade, four involved antennation as an answer to an act: move towards (MT) twice, stilt posture with antennation (SA) and nothing (NO).

DISCUSSION

Behavioural sequences in phylogenetic analyses

Since Whitman (1898), Heinroth (1909), Lorenz (1941) and others pioneered the comparative study of behaviour, many phylogenetic analyses of behaviour have been carried out. The concept of behavioural homology has been discussed and finally found to be similar to the homology of other phenotypic characters (see Baerends, 1958; Atz, 1970; Hodos, 1976; Lauder, 1986; Wenzel, 1992). As morphological and molecular characters, behavioural ones proved to be phylogenetically informative when they were accurately defined and adequately sampled. As emphasized by Wenzel (1992), all these analyses took benefit and became more rigorous with the rise of phylogenetic methods, such as cladistics.

However, in spite of these very significant advances, the great majority of phylogenetic analyses of behaviour still considered behavioural acts in isolation (see Wenzel, 1992 for a review). Behavioural acts are regarded most often as equivalent if they are similar, even if they occur in different sequences or different places within the sequences. This is a much lower standard than for other kinds of characters (molecules, morphology) where the position of the feature is always carefully established (i.e. DNA sequence alignment, position criterion for morphological structures).

We recently emphasized that the most up-to-date phylogenetic algorithms designed to compare DNA sequences – such as direct optimization (Wheeler, 1996, Wheeler *et al.*, 2006) – can be successfully applied to the stereotyped behavioural sequences (Robillard *et al.*, 2006b). In the present paper, we make one more step and propose to compare non-

stereotyped and, therefore strictly speaking, not species-specific sequences. Thus, instead of focusing directly on the relationships among different behavioural sequences to infer relationships among species, we aimed at inferring relationships among characteristic parts of these sequences by applying a method of successive event-pairing which codes successions of acts.

The method is conceptually very straightforward by coding the occurrence of transitions between the acts in the sequences. It applies a better and more accurate concept of behavioural homology, each act being considered according not only to its special quality but also to its position within sequences.

Three main assumptions are implied and must be discussed with regard to behavioural characters: heritability, sampling and independence. All kind of characters used in phylogenetic analyses, either morphological, molecular or behavioural, should be checked from this point of view. In practice, this is usually made only for non-traditional characters such as behavioural or physiological ones, the variability of which is intuitively more questioned by scientists (Wenzel, 1992). Heritability is the first and most basic concern as non-heritable traits would be nonsensical if used in a phylogenetic context of descent with modification (see Grandcolas & D'Haese, 2003 for a review). With the exception of breeding and genetic studies, the only a priori way to assess the heritability of characters is to control for epigenetic effects by observing every species in the same conditions and by varying and repeating the conditions of observation. A posteriori, phylogenetic congruence between behaviour and other markers, including those molecular ones reputed to be neutral, is also a mean to assess heritability. In the present case, both criteria have been employed: repertoires and kinds of interactions have been found stable in repeated studies (Gautier, 1974; Grandcolas, 1991; van Baaren *et al.*, 2002, 2003a) and the present study) and the different data sets are reasonably congruent.

As for the sampling effort, behaviour is not more difficult to sample than morphology or molecules and it only requires to have living specimens placed in controlled and relevant conditions and to follow classical protocols (e.g. Martin & Bateson, 1986; Wenzel, 1992). In our case, accumulation curves were computed showing that the sampling of behavioural transitions has reached a plateau in every species, which allowed a sound interspecific comparison. Given that samples are large enough, the successive event-pairing method also has the advantage to bring more characters – potentially up to a square power more – than the acts considered in isolation.

Character independence has been mentioned for a long time as a potential problem in phylogenetics, but also as one without solution. The most obvious cases of dependence between characters must be checked for and discarded, but some dependence will necessarily exist between characters observed in a same organism which cannot be extirpated (Wiley, 1981; Simmons & Freudenstein, 2002). For instance, molecular phylogenetic studies usually do not consider that base pairs in the stem regions of ribosomal RNA are not truly independent. This problem does not occur with our case for the successive event-pairing method as shown by the χ^2 goodness-of-fit tests.

Finally, this method successfully proved to be efficient and informative according to the present example of sequences of dyadic interactions in cockroach groups taken from a well-studied case in the literature (van Baaren & Deleporte, 2001; van Baaren *et al.*, 2002, 2003a). In contrast with the undecisive analysis of the repertoires which were very similar, the analyses of the event-pairing data provided a high number of independent characters, which resulted in a phylogenetic tree fully resolved and highly consistent, not much different and not less consistent than the one retrieved with morphology and molecules. For example, 60 characters (or 89 with the data set including frequencies) potentially support the monophyly of the Zetoborinae, a subfamily which is well supported based on other data (Roth, 1970; Grandcolas, 1993a, 1998; Pellens *et al.*, 2007a, b). Our study also refuted the old and recurrent belief that behaviour is more homoplastic than other phenotypic traits, as already argued by McLennan *et al.* (1988), Wenzel (1992), De Queiroz & Wimberger (1993) and Proctor (1996). Consistency and retention indices and numbers of informative characters were similar between analyses C (behaviour) and D (morphology and molecules). In brief, our analysis suggested that even non-stereotyped sequences, which are usually considered more variable and less species-specific, also contain phylogenetic information. It is worth noting that phylogenetic analysis of habitats in the subfamily Zetoborinae revealed much less consistency than the present behavioural analysis (Grandcolas, 1998; Pellens *et al.*, 2007a), in that there was more homoplasy in habitat changes than in social behaviour for this clade.

Gregarious behaviour in Zetoborinae cockroaches

Behavioural transitions analysed via our method included many informative characters, few of which were homoplastic. These characters also supported relationships among species in a way which did not appear biased, as also shown by the comparison with morphological and molecular analyses. For example,

species did not cluster in broad behavioural categories or potentially non-natural classes, such as gregarious vs. solitary species. The solitary species *Thanatophyllum akinetum* was not the sister group of gregarious species, but it was nested within ingroup taxa, suggesting that some synapomorphic behavioural transitions were shared by two gregarious species and the solitary one, but not with the other gregarious species, including the outgroup. Additionally, the analysis identified some autapomorphic characters which made sense in the context of solitary life habits. With the same perspective, the acts supporting Zetoborinae – both the solitary and gregarious – with respect to the gregarious Blaberinae outgroup corresponded mainly to absences of agonistic or avoiding acts in response to varied acts such as antennations or moving away (PS, PP, SJ and WA, WP, RO or FP respectively). Zetoborinae displayed other negative acts in answer to positive sollicitation (transitions which are not expressed by *Eublaberus*). Those negative acts were SA (agonistic), GD, GA, RO, WD (avoiding). Thus, *Eublaberus* and the Zetoborinae studied here displayed different agonistic and avoiding acts in several particular situations. This suggests that Zetoborinae were more disposed than *Eublaberus* to display promoting acts (mostly antennation) in front of negative acts and that they gained the behaviour of displaying 'avoiding' acts (GD, GA, RO, WD) to stop interactions rather than agonistic behaviours (or that *Eublaberus* lost it and acquired aggressive behaviour). Zetoborinae did not appear to be prone to more or less social behaviour than *Eublaberus*; they displayed a different kind of social behaviour. Finally, *Lanxoblatta* and *Phortioeca*, two genera known to be closely related (Roth, 1970; Grandcolas, 1993a, 1998), clustered together on the basis of similar behavioural transitions involving antennation.

Using non-stereotyped behaviours in phylogenetic analyses opens new avenues for studies in behaviour evolution. In contrast to restricted stereotyped behaviours such as courtships or nest-building, non-stereotyped behaviours represent most of the behavioural activity of many species, such as feeding, foraging, playing, interacting, etc. (McFarland, 1993). Since long ago, these non-stereotyped behaviours were generally considered as less species-specific in their characteristics and composed of acts often widespread in related species, hence not chiefly adequate for comparative studies (e.g. Hinde & Tinbergen, 1958). Conversely, they were commonly analysed in psychological or sociological studies where the analysis of non-stereotyped sequences is a frequent and important matter (Abbott, 1995; Abbott & Tsay, 2000; Schlich, 2001; Elzinga, 2003; Van der Aalst *et al.*, 2003; Hay *et al.*, 2004).

The successive event-pairing methodology and the present case study show that non-stereotyped behaviours are informative from a phylogenetic and evolutionary point of view. This brings many characters potentially useful to study phylogenetic relationships, the adequacy of which can be established with a simple but careful preliminary statistical treatment. Once the phylogenetic tree is reconstructed, the occurrence of each kind of transition among two acts can be mapped on the tree to understand where and how it evolved, allowing for the more detailed elucidation of behavioural evolution.

ACKNOWLEDGEMENTS

This study has been developed in the course of a PhD granted by Ministère de la Recherche to Frédéric Legendre. Molecular work was partly made in the laboratory of Michael Whiting in Brigham Young University during a stay in Provo (Utah) supported by the programme pluriformation 'Etat et structure phylogénétique de la biodiversité actuelle et fossile' directed by Philippe Janvier and by NSF DEB-0120718 (MFW). Part of the behavioural observations has been carried out in Station biologique de Paimpont (UMR 6552 CNRS, Université de Rennes I) with facilities provided by Pierre Deleporte. The methodology has been presented in a talk at the Hennig XXV meeting in Oaxaca, Mexico owing to a grant of Ecole Doctorale 392 'Diversité du Vivant'. All the people having made possible this work in these different ways are warmly thanked. We are also indebted to Eric Guilbert, Cyrille D'Haese and two anonymous referees for their comments on previous versions of this manuscript.

REFERENCES

Abbott A. 1995. Sequence analysis: new methods for old ideas. *Annual Review Of Sociology* **21:** 93–113.

Abbott A, Tsay A. 2000. Sequence analysis and optimal matching methods in sociology. Review and prospect. *Sociological Methods & Research* **29:** 3–33.

Archie JW. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. *Systematic Zoology* **34:** 326–345.

Atz JW. 1970. The application of the idea of homology to behavior. In: Aronson LR, Tobach E, Lehrman DS, Rosenblatt JS, eds. *Development and evolution of behavior*. San Francisco, CA: W. H. Freeman, 53–74.

van Baaren J, Bonhomme A-S, Deleporte P, Pierre J-S. 2003a. Behaviours promoting grouping or dispersal of mother and neonates in ovoviviparous cockroaches. *Insectes sociaux* **50:** 45–53.

van Baaren J, Eckstein E, Deleporte P, Biquand V, Le Rouzo T, Pierre J-S. 2003b. Social interactions during

development in the ovoviviparous cockroach *Schultesia lampyridiformis*. *Behaviour* **140:** 721–737.

van Baaren J, Deleporte P. 2001. Comparison of gregariousness in larvae and adults of four species of zetoborine cockroaches. *Entomologia Experimentalis et Applicata* **99:** 113–119.

van Baaren J, Deleporte P, Grandcolas P, Biquand V, Pierre J-S. 2002. Measurement for solitariness and gregarism: analysing spacing, attraction and interactions in four species of Zetoborinae (Blattaria). *Ethology* **108:** 697–712.

Baerends GP. 1958. Comparative methods and the concept of homology in the study of behavior. *Archives Neerlandaises de Zoologie Supplement* **13:** 401–417.

Bakeman R, Quera V. 1995. *Analyzing interaction: sequential analysis with SDIS And GSEQ*. New York: Cambridge University Press.

Bininda-Emonds ORP, Jeffery JE, Coates MI, Richardson MK. 2002. From Haeckel to event-pairing: the evolution of developmental sequences. *Theory in Biosciences* **121:** 297–320.

Brooks DR, McLennan DA. 2002. *The nature of diversity. an evolutionary voyage of discovery*. Chicago, IL: The University of Chicago Press.

Burmeister H. 1838. *Handbuch der Entomologie. Zweiter Band, zweite Abteilung, Kaukerfe. Gymnognatha (Erste Hälfte: vulgo Orthoptera)*. Th. Ch. F. Enslin, Berlin.

Cap H, Aulagnier S, Deleporte P. 2002. The phylogeny and behaviour of Cervidae (Ruminantia Pecora). *Ethology, Ecology and Evolution* **14:** 199–216.

Chatfield C, Lemon RE. 1970. Analysing sequences of behavioural events. *Journal of Theoretical Biology* **29:** 427–445.

Coddington JA. 1986. Orb webs in non-orb weaving ogre-faced spiders (Araneae: Dinopidae): a question of genealogy. *Cladistics* **2:** 53–67.

Deleporte P. 1993. Characters, attributes and tests of evolutionary scenarios. *Cladistics* **9:** 427–432.

Desutter-Grandcolas L, Robillard T. 2003. Phylogeny and the evolution of calling songs in *Gryllus* (Insecta, Orthoptera, Gryllidae). *Zoologica Scripta* **32:** 173–183.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Elzinga CH. 2003. Sequence similarity. a non-aligning technique. *Sociological Methods and Research* **32:** 3–29.

Faith DP, Cranston PS. 1991. Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. *Cladistics* **7:** 1–28.

Farris JS. 1989. The retention index and the rescaled consistency index. *Cladistics* **5:** 417–419.

Gautier JY. 1974. Processus de différenciation de l'organisation sociale chez quelques espèces de blattes du genre Blaberus: aspects écologiques et éthologiques. PhD Thesis, Université de Rennes.

Gautier JY, Deleporte P, Rivault C. 1988. Relationships between ecology and social behavior in cockroaches. In: Schlobodchikoff CN, ed. *The ecology of social behavior*. San Diego, CA: Academic Press, 335–351.

Genecodes. 1999. *Sequencher*. Ann Arbor, MI: Genecodes.

Giribet G, Carranza S, Baguñà J, Riutort M, Ribera C. 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Molecular Biology and Evolution* **13:** 76–84.

Goloboff PA, Mattoni CI, Quinteros SA. 2006. Continuous characters analysed as such. *Cladistics* **22:** 589–601.

Gottman JM, Roy AK. 1990. *Sequential analysis: a guide for behavioral researchers*. New York: Cambridge University Press.

Grandcolas P. 1991. Les Blattes de la forêt tropical de Guyane Française: structure du peuplement et étude écoéthologique des Zetoborinae. PhD Thesis, Université de Rennes.

Grandcolas P. 1993a. Monophylie et structure phylogénétique des [Blaberinae + Zetoborinae + Gyninae + Diplopterinae] (Dictyoptera : Blaberidae). *Annales de la Société entomologique de France (N.S.)* **29:** 195–222.

Grandcolas P. 1993b. Habitats of solitary and gregarious species in the neotropical Zetoborinae (Insecta, Blattaria). *Studies in Neotropical Fauna and Environment* **28:** 179–190.

Grandcolas P. 1998. The evolutionary interplay of social behavior, resource use and anti-predator behavior in Zetoborinae + Blaberinae + Gyninae + Diplopterinae cockroaches: a phylogenetic analysis. *Cladistics* **14:** 117–127.

Grandcolas P. 1999. The origin of diversity in cockroaches: a phylogenetic perspective of sociality, reproduction, communication and ecology. In: Melic A, De Haro JJ, Mendez M, Ribera I, eds. *Evolucion y Filogenia de Arthropoda*. Zaragoza: Boletin de la Sociedad Entomologica Aragonesa, 397–420

Grandcolas P, D'Haese C. 2003. Testing adaptation with phylogeny: how to account for phylogenetic pattern and selective value together? *Zoologica scripta* **32:** 483–490.

Grandcolas P, D'Haese C. 2004. The origin of a 'true' worker caste in termites: mapping the real world on the phylogenetic tree. *Journal of Evolutionary Biology* **17:** 461–463.

Grandcolas P, Deleporte P, Desutter-Grandcolas L, Daugeron C. 2001. Phylogenetics and ecology: as many characters as possible should be included in the cladistic analysis. *Cladistics* **17:** 104–110.

Greene HW. 1994. Homology and behavioural repertoires. In: Hall BK, ed. *Homology: the hierarchical basis of comparative biology*. San Diego, CA: Academic Press, 369–391.

Hay B, Wets G, Vanhoof K. 2004. Mining navigation patterns using a sequence alignment method. *Knowledge and Information Systems* **6:** 150–163.

van der Heijden PGM. 1987. *Correspondence analysis of longitudinal categorical data*. Leiden: DSWO Press.

Heinroth O. 1909. Beobachtungen bei der zucht des zigenmelkers (*Caprimulgus europaeus* L.). *Journal of Ornithology* **57:** 56–83.

Hillis DM. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. In: Miyamoto MM, Cracraft J, eds. *Phylogenetic analysis of DNA sequences*. New York: Oxford University Press, 278–294.

Hillis DM, Dixon MT. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66:** 411–453.

Hillis DM, Huelsenbeck JP. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity* **83:** 189–195.

Hinde RA. 1955. A comparative study of the courtship of certain finches. *Ibis* **97:** 706–745.

Hinde RA, Tinbergen N. 1958. The comparative study of species-specific behavior. In: Roe A, Simpson GG, eds. *Behavior and evolution*. New Haven, CT: Yale University Press, 251–268.

Hodos W. 1976. The concept of homology and the evolution of behavior. In: Masterton RB, Hodos W, Jerison H, eds. *Evolution, brain, and behavior: persistent problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 153–167.

Jarvis KJ, Haas F, Whiting MF. 2004. Phylogeny of earwigs (Insecta: Dermaptera) based on molecular and morphological evidence: reconsidering the classification of Dermaptera. *Systematic Entomology* **29:** 359–370.

Jeffery JE, Bininda-Emonds ORP, Coates MI, Richardson MK. 2005. A new technique for identifying sequence heterochrony. *Systematic Biology* **54:** 230–240.

Jeffery JE, Richardson MK, Coates MI, Bininda-Emonds ORP. 2002. Analysing developmental sequences within a phylogenetic framework. *Systematic Biology* **34:** 326–345.

Johnson KP, McKinney F, Wilson R, Sorenson MD. 2000. The evolution of postcopulatory displays in dabbling ducks (Anatini): a phylogenetic perspective. *Animal Behaviour* **59:** 953–963.

Kambhampati S. 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. *Proceedings of the National Academy of Sciences of the USA* **92:** 2017–2020.

Kennedy M, Spencer HG, Gray RD. 1996. Hop, step and gape: do the social displays of the Pelecaniformes reflect phylogeny? *Animal Behaviour* **51:** 273–291.

Kluge AG, Farris JS. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **30:** 1–32.

Lauder GV. 1986. Homology, analogy, and the evolution of behavior. In: Nitecki MH, Kitchell JA, eds. *Evolution of animal behavior: paleontological and field approaches*. New York: Oxford University Press, 9–40.

Lorenz K. 1941. Vergleichende Bewegungsstudien an Anatinen. *Journal of Ornithology* **89:** 194–294.

Mabee PM, Trendler TA. 1996. Development of the cranium and paired fins in Betta splendens: intraspecific variation and interspecific comparisons. *Journal of Morphology* **227:** 249–287.

Martin P, Bateson P. 1986. *Measuring behaviour. An introductory guide*. Cambridge: Cambridge University Press.

McFarland D. 1993. *Animal behaviour*. Harlow: Longman

McLennan DA, Brooks DR, McPhail JD. 1988. The benefits of communication between comparative ethology and phylogenetic systematics: a case study using gasterosteid fishes. *Canadian Journal of Zoology* **66:** 2177–2190.

Nalepa CA, Bell WJ. 1997. Postovulation parental investment and parental care in cockroaches. In: Choe JC, Crespi BJ, eds. *The evolution of social behavior in Insects and Arachnids*. Cambridge: Cambridge University Press, 26–51.

Nixon KC. 2002. *Winclada version 1.00.08*. Trumansburg, NY: Author.

Noll FB. 2002. Behavioral phylogeny of corbiculate apidae (hymenoptera; apinae), with special reference to social behavior. *Cladistics* **18:** 137–153.

Page RDM. 1992. *COMPONENT* User manual (release 2.0). London: Natural History Museum.

Pellens R, Legendre F, Grandcolas P. 2007a. Phylogenetic analysis of social behaviour evolution in [Zetoborinae + Blaberinae + Gyninae + Diplopterinae] cockroaches: an update with the study of endemic radiations from the Atlantic forest. *Studies on Neotropical Fauna and Environment* **42:** 25–31.

Pellens R, D'Haese C, Bellés X, Piulachs MD, Legendre F, Wheeler W, Grandcolas P. 2007b. The evolutionary transition from subsocial to eusocial behavior: phylogenetic and ecological evidence for modification of the 'shift-independent-care' hypothesis with a new prototermite model. *Molecular Phylogenetics and Evolution* **43:** 616–626.

Price JJ, Lanyon SM. 2002. Reconstructing the evolution of complex bird song in the Oropendolas. *Evolution* **56:** 1514–1529.

Proctor HC. 1996. Behavioral characters and homoplasy: perception versus practice. In: Sanderson MJ, Hufford L, eds. *Homoplasy: the recurrence of similarity in evolution*. San Diego, CA: Academic Press, 131–149.

de Queiroz A, Wimberger PH. 1993. The usefulness of behavior for phylogeny estimation – levels of homoplasy in behavioral and morphological characters. *Evolution* **47:** 46–60.

Remane A. 1952. *Die Grundlagen des naturlichen Systems, der vergleichenden Anatomie und der Phylogenetik; Theoretische Morphologie und Systematik I*. Leipsig: Geest und Portig.

Robillard T, Höbel G, Gerhardt HC. 2006a. Evolution of advertisement signals in North American hylid frogs: vocalizations as end-products of calling behavior. *Cladistics* **22:** 533–545.

Robillard T, Legendre F, Desutter-Grandcolas L, Grandcolas P. 2006b. Phylogenetic analysis and alignment of behavioral sequences by direct optimization. *Cladistics* **22:** 602–633.

Roth LM. 1970. The male genitalia of Blattaria. III. Blaberidae: Zetoborinae. *Psyche* **77:** 217–236.

Schal C, Gautier JY, Bell WJ. 1984. Behavioural ecology of cockroaches. *Biological Review* **59:** 209–254.

Schlich R. 2001. Analysing intrapersonal variability of travel behaviour using the sequence alignment method. European Transport Research Conference, September, 2001, Cambridge.

Schulmeister S, Wheeler W. 2004. Comparative and phylogenetic analysis of developmental sequences. *Evolution and Development* **6:** 50–57.

Simmons MP, Freudenstein JV. 2002. Artifacts of coding amino acids and other composite characters for phylogenetic analysis. *Cladistics* **18:** 354–365.

Smith KK. 1997. Comparative patterns of craniofacial development in eutherian and metatherian mammals. *Evolution* **51:** 1663–1678.

Sorenson MD. 1999. *TreeRot*, version 2. Boston, MA: Boston University.

Stevens PF. 1991. Character states, morphological variation, and phylogenetic analysis: a review. *Systematic Botany* **16:** 553–583.

Strong EE, Lipscomb D. 1999. Character coding and inapplicable data. *Cladistics* **15:** 363–371.

Swofford DL. 1998. *PAUP* – phylogenetic analysis using parsimony (*and others methods)*. Sunderland, MA: Sinauer Associates.

Tinbergen N. 1959. Comparative studies of the behaviour of gulls (Laridae); a progress report. *Behaviour* **15:** 1–70.

Van der Aalst WMP, van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters AJMM. 2003. Workflow mining: a survey of issues and approaches. *Data and Knowledge Engineering* **47:** 237–267.

Velhagen WA. 1997. Analyzing developmental sequences using sequence units. *Systematic Biology* **46:** 204–210.

Wenzel JW. 1992. Behavioral homology and phylogeny. *Annual Review of Ecology and Systematics* **23:** 361–381.

Wenzel JW. 1993. Application of the biogenetic law to behavioral ontogeny: a test using nest architecture in paper wasps. *Journal of Evolutionary Biology* **6:** 229–247.

Wheeler W. 1996. Optimization alignment: the end of multiple sequence alignments in phylogenetics? *Cladistics* **12:** 1–9.

Wheeler W, Aagesen L, Arango CP, Faivovich J, Grant T, D'Haese CA, Janies D, Smith WL, Varon A, Giribet G. 2006. *Dynamic homology and phylogenetic systematics: a unified approach using POY*. New York: American Museum of Natural History.

Whiting MF. 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta* **31:** 93–104.

Whiting MF, Carpenter JC, Wheeler QD, Wheeler WC. 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology* **46:** 1–68.

Whitman CO. 1898. *Animal Behaviour*. Boston, MA: Ginn and Co.

Wiley EO. 1981. *Phylogenetics. the theory and practice of phylogenetic systematics*. New York: Wiley-Liss.

Wilson C, Harvey A, Thompson J. 1999. ClustalG: Software for analysis of activities and sequential events. Workshop on longitudinal research in social science: a Canadian focus. Windermere Manor, London, Ontario, Canada, October 25–27, 1999, 1–12.

Xiong B, Kocher TD. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). *Genome* **34:** 306–311.

Zar JH. 1999. *Biostatistical analysis*, 4th edn. Upple Saddle River, NJ: Prentice-Hall, Inc.

SUPPLEMENTARY MATERIAL

The following material is available for this article online:

Appendix S1. Transitions matrices of the five cockroaches.