Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Animal

Molecular and Acoustic Signal Evolution in Mediterranean Species of *Cicada* L. (Insecta: Cicadoidea)

Gabriela Alexandra Pinto Juma

DOUTORAMENTO EM BIOLOGIA
(BIOLOGIA EVOLUTIVA)
2008

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Molecular and Acoustic Signal Evolution in Mediterranean Species of *Cicada* L. (Insecta: Cicadoidea)

GABRIELA ALEXANDRA PINTO JUMA

Tese orientada por:

Professor Doutor José Alberto Quartau e Professor Doutor Michael W. Bruford

DOUTORAMENTO EM BIOLOGIA
(BIOLOGIA EVOLUTIVA)
2008

This work was carried out at Departamento de Biologia Animal da Faculdade de Ciências da Universidade de Lisboa, with support from Centro de Biologia Ambiental, and at School of Biosciences of Cardiff University.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

This study was supported by the Fundação Para a Ciência e Tecnologia: PhD Fellowship PRAXIS BD/18229/98)

This dissertation should be cited as:

Pinto-Juma, GA. 2008. Molecular and acoustic signal evolution in Mediterranean species of *Cicada* L. (Insecta: Cicadoidea). PhD Thesis, University of Lisbon, Portugal

To Imtiaz

and Madalena Noor

"Come to the edge, he said. They said: We are afraid.

Come to the edge, he said. They came. He pushed them and they flew"

Guillaume Apollinaire

Table of Contents

Nota Prévia i

Aknowledgements iii

Resumo vii

Abstract xiii

1. General Introduction 1

The species concept problem 3

Mitochondrial DNA 6

Population genetics 8

SMRS in cicadas 10

Cicada L. 1758 11

Key Hypotheses 13

Objectives and thesis structure 14

References 15

2. Mitochondrial DNA variation and the evolutionary history of the Mediterranean species of *Cicada* 27

Abstract 29

Introduction 30

Material and Methods 32

Collection of specimens 32

DNA extraction, PCR and sequencing 32

Sequence alignment 34

Alternative methods for haplotype detection (RFLP and SSCP 34

Phylogenetic analysis and population genetics 35

Results 36

Nucleotide substitutions 37

Haplotype and genetic diversity 41

Divergence among haplotypes 44

Phylogeny 45

Population hierarchical analysis and demography 47

Discussion 49

12S rRNA secondary structure 49

Independent evolutionary units 49

Population structure and demography 51

Expansion events 53
Final considerations 54
References 56

3. Calling song structure and geographic variation in *Cicada orni* **67**
Abstract 69
Introduction 69
Material and Methods 70
Results 73
Discussion and conclusions 77
References 81

4. Geographic variation in *Cicada barbara* **83**

4.1. Population structure of *Cicada barbara* from the Iberian Peninsula and Morocco based on mitochondrial DNA 85
Abstract 85
Introduction 85
Material and Methods 86
Molecular analysis 86
Results 88
Genetic variation 88
Population structure 89
Demographic analysis 90
Discussion 91
Population structure and genetic variation 91
Demographic expansions 92
Iberia Peninsula expansions 92
Moroccan expansions 93
Origin of Iberian cicadas and subspecies classification 93
References 93

4.2. Patterns of acoustic variation in *Cicada barbara* from the Iberian Peninsula and Morocco 97
Abstract 97
Introduction 97
Material and Methods 98
Acoustic analysis 98
Results 99

Calling song 100
Calling song and the differentiation between regions 101
Amplitude modulated signal 104
Discussion 105
Acoustic variation 105
Regional variation and subspecies 106
References 107

5. Unprecedented number of copies and insertion events of cyt b numts in *Cicada* L. (Hemiptera, Cicadoidea 111
Abstract 113
Background 114
Results 115
Detection of Numts 115
Assessment of Numts 115
Authentic mtDNA 120
Similar Numts vs. divergent mtDNA 120
Numts evolution and integration events 123
Discussion 125
Numt frequency 125
An ongoing process 126
Numt contamination in phylogenetic studies 126
Methods 128
Standard DNA extraction, PCR and sequencing 128
Cloning 128
MtDNA extraction method 129
Sequence analysis 129
Phylogenetic analysis and Numt integration events 130
References 131

6. General Discussion 147
MtDNA and *Numts* 149
Acoustic and molecular signal evolution in *Cicada* species 150
Population differentiation in *Cicada orni* and *C. barbara* 151
Dispersal barriers and postglacial re-colonization 153
Sympatric populations of *C. orni* and *C. barbara* 155
Further work 156

References 157

Appendix **163**

Calling songs in sympatric and allopatric populations of *Cicada barbara* and *C. orni* from the Iberian Península 165

List of Figures

Chapter 2

Figure 1. Collection sites of the *Cicada* specimens analysed in the Mediterranean region: inserts are given for Portugal and Greece, where most sites were sampled 31

Figure 2. Sequence alignment of the 12S rRNA fragment analysed in this study 37

Figure 3. Frequencies of transitions and transversions over K81 sequence divergences obtained among *Cicada* specimens for (a) the total length of 12S rRNA sequences; (b) loop regions based on the secondary structure of the 12S rRNA sequences; and (c) stem regions based on the secondary structure of the 12S rRNA 39

Figure 4. Minimum spanning network for *Cicada* in the Mediterranean area, based on 12S rRNA haplotypes 42

Figure 5. Neighbour-Joining (NJ) phylogenetic tree (maximum likelihood (ML) consensus tree had the same topology so it is not shown) obtained for *Cicada* in the Mediterranean area, based on 12S rRNA haplotypes. For each clade, an oscillogram (time *vs*. amplitude) of the calling song of one male of *C. orni*, *C. mordoganensis*, *C. cretensis*, *C. lodosi and C*, *barbara* is shown 44

Figure 6. Post-glacial colonization routes of *Cicada orni* (black arrow) and *C. barbara* (crosshatched arrow) in Europe and distribution of *Cicada* haplotypes in the Aegean area 53

Appendix 1. Distribution of 12S rRNA gene (~390bp) haplotypes among *Cicada* species from the Mediterranean area 61

Appendix 2. Secondary structure of domain III of 12S rRNA gene based on the template described by Hickson *et al.* (1996) obtained for (a) the consensus sequences of the *Cicada orni*, *C. mordoganensis* and *C. cretensis* haplogroups, *C. lodosi* substitutions are shown at the boxes (+ stands for insertion, D stands for deletion), and (b) the consensus sequence of the *C. barbara* haplogroup 63

Appendix 3. Kimura 3 parameter (proportion of invariables I=0.8116) distance matrix, obtained by comparison of all pairs of haplotypes analysed in this study, based on an approximately 394 bp fragment of the 12SrRNA gene 64

Chapter 3

Figure 1. Map showing sampling localities for *Cicada orni* acoustic signals 68

Figure 2. Calling song of *Cicada orni*. (A) Oscillogram (amplitude vs. time); (B) sonagram or spectrogram (frequency vs. time); and (C) mean amplitude spectrum (frequency vs. amplitude) 70

Figure 3. Histograms of distributions of frequency variables of the calling song of *Cicada orni* for which there were statistical differences between the recordings performed with Sony and with UHER equipment in the French populations at Molitg, St. Hippolyte, and Narbonne 72

Figure 4. Bidimensional diagrams of relationships between specimens of *Cicada orni* (176 OTUs) of the principal component analysis based on a correlation matrix between 13 acoustic characters 73

Figure 5. Boxplots of the 13 acoustic variables of the calling song of *Cicada orni* investigated for the three regions (71 specimens from the Iberian Peninsula, 29 specimens from southern France, and 74 specimens from continental Greece) 74

Figure 6. Boxplots of the 13 acoustic (including five time and eight frequency) variables of the calling song of *Cicada orni* investigated for each population 77

Chapter 4

4.1

Figure 1. Distribution and relative frequency of cyt b haplotypes among populations of *Cicada barbara* in the Iberia Peninsula and Morocco 86

Figure 2. Medium joining and minimum spanning networks for *Cicada barbara* in the Iberia Peninsula and Morocco based on cyt *b* haplotypes [black and white] 87

[colour] 94

Figure 3. Mismatch distributions within *Cicada barbara* 89

4.2

Figure 1. Recording sites in the Iberian Peninsula and North Africa of the *Cicada barbara* specimens studied 96

Figure 2. Oscillograms (time *vs*. amplitude) and sonagrams (time *vs*. frequency) of the calling songs of one male of *Cicada barbara lusitanica* from Crato, Portugal (a) and one male of *C. barbara barbara* from Fès, Morocco (b). A fragment of 0.2 seconds is shown, with an individual syllable pointed out, which is further magnified in (a) to show the pulses 98

Figure 3. Variation of the syllable rate with the temperature in specimens of *Cicada barbara* 99

Figure 4. Boxplots of the acoustic variables analysed for *Cicada barbara* at two regions (Iberian Peninsula + Ceuta and Morocco), and three regions (Ceuta, Iberian Peninsula and Morocco) 100

Figure 5. Component scores obtained from the Principal Component Analysis (PCA) based on a correlation matrix between all nine acoustic variables analysed for *Cicada barbara*. a) Population level and b) Region level 101

Figure 6. Oscillogram and sonagram of the amplitude modulated signal of (a) one male of *Cicada barbara lusitanica* from Moura and (b) one male of *C. barbara barbara* from Fès 103

Appendix I. Boxplots of the acoustic variables analysed for *Cicada barbara* at the sampled sites 107

Appendix II. Values of Average ± Standard deviation and coefficient of variation for each acoustic variable per population 108

Chapter 5

Figure 1. Minimum joining network between the clones from a PCR fragment of cyt *b* gene of five specimens of *Cicada barbara* and two of *C. orni* 119

Figure 2. Neighbour-joining tree of clones from a PCR fragment of cyt *b* for five *Cicada barbara* and two of *C. orni* and authentic mtDNA for *C. barbara*, *C. orni*, *C. mordoganensis* and *C. cretensis* 124

Figure 3. Frequency of haplotypes per number of clones 126

Appendix 1. Description of clones from a PCR fragment of cyt *b* gene of five specimens *Cicada barbara* and two specimens *C. orni* 136

Appendix 2. Amino-acid substitutions within mtDNA-like clones from a section of cytochrome *b* gene (477 bp, 159 amino-acids translated) of *Cicada barbara* and *C. orni* 141

Appendix 3. Summary statistics of clones from a cyt *b* gene PCR fragment per specimen cloned 145

List of tables

Chapter 2

Table 1. Summarized distribution of 12S rRNA gene (~390bp) haplotypes among *Cicada* species from the Mediterranean area 36

Table 2. Nucleotide substitutions in an approximately 390 bp fragment of the 12S rRNA gene within and among *Cicada* haplogroups 38

Table 3. 12S rRNA gene and nucleotide diversity of the cicada samples at each locality within the Mediterranean area 40

Table 4. Mean Kimura 3 parameter (proportion of invariables I=0.8116) distance matrix, obtained by comparison of all pairs of haplogroups and sequences of cicadas analysed in this study, based on an approximately 394 bp fragment of the 12S rRNA gene. 43

Table 5. Hierarchical analysis of molecular variance (AMOVA) design and results for different tests of differentiation among groups. a) Within *C. orni* and *C. barbara* haplogroups b) Other tests within *C. orni* and *C. barbara* haplogroups 45

Table 6. Estimated parameters from the sudden expansion model 46

Chapter 3

Table 1. Sampled populations of *Cicada orni* with abbreviation names (Abbrev.), number of specimens recorded for sound analyses (*n*), dates of recording, and ambient temperatures 69

Table 2. Description of the acoustic variables analyzed for *Cicada orni* calls 70

Table 3. Descriptive statistics of the acoustic variables of *Cicada orni* specimens 71

Table 4. Factor loadings of the principal component analysis (PCA) based on a correlation matrix between 13 acoustic variables for *Cicada orni* (using176 OTUs) 73

Table 5. Coefficients of variation for each acoustic variable analyzed within specimens (CVind) within populations (CVpop) (average ± standard deviation) and within regions (CVreg) for the three areas investigated 76

Table 6. Percentage of specimens presenting CVind > CVpop and percentage of populations presenting CVpop > CVreg 78

Chapter 4

4.1

Table 1. Sampled populations of *Cicada barbara* with the distribution of cyt *b* gene haplotypes (364bp) and genetic diversity among localities 85

Table 2. HKY pairwise distances between all pairs of haplotypes based in a 364 bp fragment of the Cyt *b* gene within *Cicada barbara* 88

Table 3. Slatkin's linearised F_{st} values (Slatkin, 1995) between all pairs of samples based in a 364 bp fragment of the Cyt *b* gene within *Cicada barbara*, below diagonal 88

Table 4. AMOVA results for the two groups of *Cicada barbara* maximizing the Φ_{CT} values analysed for cyt *b* gene (Morocco vs Iberian Peninsula and Ceuta) 88

Table 5. Diversity, neutrality and expansion parameters for *Cicada barbara* based on a section of cyt *b* gene 90

4.2

Table 1. Sampled populations of *Cicada barbara* 97

Table 2. Descriptive statistics of the acoustic variables of *Cicada barbara* 98

Table 3. Mann-Whitney U tests comparing two regions for each acoustic variable in *Cicada barbara* 99

Table 4. Spearman correlation (r_S) between latitude and acoustic variables analysed for *Cicada barbara* 101

Table 5. Mean, standard deviation (SD) and coefficient of variation (CV) for each acoustic variable of *Cicada barbara* on each region 101

Table 6. Multivariate analyses based on acoustic variables for *Cicada barbara*. a) Principal Component Analysis (PCA) based on nine acoustic variables: eigen analysis and component loadings. b) Discriminant Function Analysis (DFA) results based on seven acoustic variables 102

Table 7. Descriptive statistics for the acoustic variables measured in the phrases of the amplitude modulated signal of *Cicada barbara barbara* and *C. barbara lusitanica* 104

Chapter 5

Table 1. *Cicada barbara* and *C. orni* specimens used in this study 116

Table 2. Amino-acid substitutions within mtDNA-like clones (no deletions, insertions or stop codons) from a PCR fragment of cyt *b* gene (477 bp translated 159 aa) for *Cicada barbara* and *C. orni* specimens 117

Table 3. Summary of Numts and mtDNA clones of a cyt *b* PCR fragment from five specimens *Cicada barbara* and two specimens *C. orni* 121

Table 4. Mean, minimum and maximum *p* distances between all groups of haplotypes obtained from cyt *b* PCR clones from five *Cicada barbara* specimens and two *C. orni*, and mtDNA sequences from *C. barbara*, *C. orni*, *C. mordoganensis*, *C. cretensis* and *Tettigetta josei* 122

Table 5. Pairwise *P* distances between mtDNA-like clone sequences from a PCR fragment of cyt *b* gene 123

Nota Prévia

Na elaboração da presente dissertação, e nos termos do disposto no nº 1 do Artigo 40, Capítulo V, do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no *Diário da República – II Série Nº 153*, de 5 de Julho de 2003, esclarece-se que foram usados integralmente artigos científicos já publicados (4), aceites para publicação (1) e submetidos (1), em revistas indexadas de circulação internacional, os quais integram os capítulos da presente tese. Tendo os referidos trabalhos sido realizados em colaboração, a candidata esclarece que participou integralmente no planeamento, na elaboração, na análise, discussão e escrita dos resultados de todos os trabalhos apresentados.

Foreword

Accordingly to the terms referred to in nº 1 paragraph 40, Chapter V, of the Regulations for Post-Graduate studies of the University of Lisbon, published in *Diário da República – II Série Nº 153*, of 5 of July of 2003, the present thesis integrates manuscripts already published (4), accepted for publication (1) and submitted (1) to international scientific journals. The candidate declares that she contributed integrally in the planning, execution, analyses and writing of all manuscripts presented here.

Acknowledgements

I would like to express my deepest appreciation to the invaluable help and collaboration of many friends and colleagues, without them this work would not have been possible or as satisfying. To all of them, and to the ones that I might have not mentioned below due to an unfortunate mental-gap, my sincere THANKS!

To my supervisor Professor José Alberto Quartau for having introduced me to the world of cicadas and having suggested the theme of this thesis. I also appreciate his numerous inspiring stories, his enthusiasm and advice along the project.

To my co-supervisor Professor Michael Bruford for his guidance, support and confidence which were crucial to the right development and finalization of this work. I am also very grateful for all the opportunities he gave me to continue working in this field.

To "Fundação para a Ciência e Tecnologia" (FCT, Lisboa, Portugal) for granting me a PhD scholarship, PRAXIS XXI/BD/21479/99, which financed the work here presented.

To Sofia Seabra, for being such a wonderful collaborator in the field, lab and also in acoustic analysis. For her companionship and support during all this work both in Cardiff and Lisbon. Also, I will never forget her care and sympathy whenever I needed it most!

To Genage André for his great skills in field work, which were largely responsible for the success of many field trips. For his good humour and enormous generosity and his great help in many bureaucratic issues. And especially, for being "always present" no matter what, his friendship and support has been truly inspiring for me!

To Paula Simões for her collaboration in one acoustic paper, for her valuable help in field work and especially for collecting and providing cicadas from Greece. Many thanks to Mónica Ribeiro, Teresa Fernandes and Teresa Rebelo who helped collecting cicadas in previous years and to Dr. Tomy Trilar and Prof. Michel Boulard for providing some dried specimens. Also, to the many non-professional collaborators that helped me in the field (particularly Imtiaz Juma) and also to the ones that helped my colleagues Sofia and Paula.

I also owe special thanks to the collaboration of Stephen Funken for receiving me in the Institute of Zoology of London and supervise the mitochondrial DNA extraction of cicadas; to Nicola Antony for supervising the development of the technique SSCP and to Douda Bensasson, for her accessibility, providing important bibliography and discussions about *Numts*.

To all staff from School of Biosciences in Cardiff University, but especially Joan Hubbard (autoclave room) and Samantha Lawrence (finance office), their great competence and thoughtfulness helped simplifying this work. Also, to the staff from the Department of Animal Biology (Faculty of Sciences, the University of Lisbon), and especially Leonor Florentino, Eugénia Alfredo and Lurdes Saramago for being so helpful but also for their kindness and friendship.

To Dr. Jo Cable for her incentive, advice, humanity and highly valuable comments in one of the papers and other chapters from this thesis.

To Dr. John Morgan, for leading me to other interesting fields and for his encouragement and understanding on the last writing stage allowing me to deviate temporarily from the "attractive worms".

To Prof. Rúben Capela for initiating me in Entomology, for his good humour and encouragement which led me to work with insects.

To Gareth Lewis and John Morgan, my lab mates during the earlier stages in Cardiff University. For their great sense of humour and companionship.

To the friends I met in Cardiff University: Trinidad Perez, Fairus Jalil, Rhys Jones, Benoit Goosens, Simon Shayler, Patricia Faria, Renata Medeiros, Mireille Johnson-Bawe, Elisabete Pires, Farina Othman, Milena Salgado and Yoshan Moodley - all present in different moments but so important for my mental health. The moments of laugh, affection, doubt and so much more created a delightful link between us!

To my dear friends from the Entomology lab of FCL: Carla Rego, Anabela Cardoso, Sónia Santos, Carlos Aguiar, Mário Boieiro, Diana Marques and Teresa Rebelo, it feels like I left so long ago but I will always remember their companionship and support. Again some moments we shared were unforgettable! Especial thanks to Carla for always trying to keep in touch and give me the news from all the others.

To my friends Helen and Hassan Dawah for their warm welcoming and support since my first visit to Cardiff. Also, Fernanda Amaral and Florbela Afonso contributed to make this city warmer and kept me in touch with my cultural roots. To my special friend Firoz Nanji, for being ever so generous and caring and showing me there is also other things in the world. Their support, friendship and care have contributed to make Cardiff my second home.

To my soul-mates: António Couto, Ana Bessa, Paula Trindade, Fátima Pereira - it does not matter how often we speak or see each other, just knowing they exist fills me up with joy and hope.

To my family and especially my parents, for their unlimited support and care! To my mother I also thank her words of wisdom whenever I need them most and to my father his restless help looking after my things in Lisbon. To Ricardo and Sara, they will always occupy a very especial place in my life and heart. And to all of them a special thanks for taking care of my "treasure" during one of the months of writing up!

To Imtiaz… To him I owe the most beautiful discoveries of my life and without his motivation, support and understanding I would not even have started this PhD. His immense knowledge, creativity, humanity and constant will to learn more and more will always be an inspiration in my life, and I am so happy he chose me to be part of his life!

To Madalena, my companion and literally my Noor (light) during this entire "journey". In this "little person" I found support, understanding and an immeasurable love which gave me strength to carry on. Thanks to her for remind me the wonders of life whenever she smiles!

Imtiaz and Madalena: both mean the world to me, and to both I dedicate this thesis.

RESUMO

Evolução de sinais moleculares e acústicos em espécies de *Cicada* L. (Insecta: Cicadoidea) na área do Mediterrâneo

As cigarras (Insecta: Hemiptera) pertencem à familia Cicadidae, caracterizada pela capacidade dos machos em produzir sinais acústicos de alta intensidade através de órgãos localizados dorsolateralmente no primeiro e segundo segmentos abdominais – os tímbalos. Estes sinais acústicos são usados na comunicação intra-específica e, particularmente os sinais de chamamento, utilizados para a atracção de fêmeas conspecíficas (bem como para a agregação de outros machos e a formação de coros), são tipicamente característicos de cada espécie. Constituem os chamados sistemas de reconhecimento sexual específico (specific mate recognition systems – SMRSs). O estudo e caracterização destes sistemas de reconhecimento têm sido particularmente importantes na separação de espécies morfologicamente semelhantes, como é o caso das cigarras do género *Cicada* Linnaeus.

Na região Mediterrânica, o género *Cicada* é composto por um complexo de seis espécies actualmente reconhecidas (ou sete, incluindo *C. cerisyi*, uma espécie de identidade taxonómica duvidosa e apenas referida para o Egipto e a Líbia) com base em algumas diferenças na genitália dos machos e nos perfis dos seus sinais acústicos de chamamento. *Cicada orni* é a espécie mais comum e com a maior área de distribuição ao longo do Mediterrâneo, estendendo-se desde a Península Ibérica até a Europa de leste e Médio Oriente. *C. barbara* encontra-se amplamente distribuída pelo noroeste de África, em algumas ilhas Mediterrânicas e na Península Ibérica onde pode ocorrer em simpatria com *C. orni*. Na Turquia encontram-se as três espécies *C. lodosi*, *C. permagna* e *C. mordoganensis*, distribuindo-se também esta última por várias ilhas gregas localizadas perto do continente turco. *C. cretensis* foi recentemente descrita como nova para a Ciência e pensa-se que será endémica para Creta. Estas espécies têm ecologias semelhantes e ocorrem frequentemente associadas a oliveiras, pinheiros, carvalhos e até a eucaliptos.

Estudos anteriores baseados em alozimas revelaram uma divergência pequena entre o par de espécies *C. orni versus C. barbara*, ou nula entre o par *C. orni versus C. mordoganensis,* nunca tendo sido encontrados sinais de hibridação entre *C. orni* e *C. barbara*, nem mesmo em áreas de simpatria. Dado que os sinais de chamamento produzidos pelos machos destas espécies são consideravelmente distintos, este facto poderia indicar que os mecanismos de isolamento pré-copulatório evoluíram com pouca divergência genética, ou

seja, que a divergência acústica poderia ter ocorrido mais rapidamente do que a divergência genética neutral. Contudo, o grau de divergência genética e as relações filogenéticas entre as espécies do género *Cicada* estavam ainda praticamente desconhecidos.

O estudo apresentado nesta tese baseia-se na análise de sequências de ADN mitocondrial de espécies de cigarras do género *Cicada* da área Mediterrânica a fim de se testar a utilidade dos sistemas de reconhecimento sexual específico na delimitação e separação destas espécies, bem como de se determinarem as relações filogenéticas e a história evolutiva deste grupo. A estrutura populacional das duas espécies mais comuns e com ocorrência em simpatria na Península Ibérica, *C. orni* e *C. barbara*, foi também analisada aos níveis molecular e acústico para tentar explicar o padrão actual de distribuição destas espécies e testar a eventual congruência entre ADN mitocondrial e o comportamento acústico.

Exemplares de cigarras foram colectados principalmente na Península Ibérica, o ponto de distribuição mais ocidental e onde *C. orni* e *C. barbara* ocorrem em simpatria. Para investigar a hipótese da recente introdução de *C. barbara* na Península Ibérica, vinda através do noroeste de África, exemplares desta espécie foram também colectados em Marrocos e Ceuta. Várias amostras de cigarras na área do mar Egeu, um dos ponto mais orientais da distribuição de *C. orni* e local de ocorrência de outras espécies dentro deste género, foram também incluídas neste estudo. Material adicional doado, como exemplares de *Cicada* de França e de outras áreas do Mediterrâneo, foi especialmente importante para demonstrar o efeito de potenciais barreiras geográficas.

Os sinais de chamamento das cigarras *C. orni* e *C. barbara* foram gravados no campo e digitalizados através do software Avisoft. Para cada indivíduo foram produzidos oscilogramas, sonogramas e um espectro de amplitude média com base em um minuto de gravação. As características temporais e de frequência dos sinais de chamamento foram então analisadas estatisticamente.

A detecção de múltiplas cópias de ADN mitocondrial do gene Citocromo *b* (Cit. *b*) presente no genoma nuclear (*Numts*) das cigarras *C. orni* e *C. barbara* foi crucial para a subsequente análise de dados. A ocorrência de *Numts* é actualmente conhecida na maioria dos eucariotas, mas se não for detectada, a sua co-amplificação acidental via PCR pode comprometer os dados analisados e induzir conclusões erróneas. A clonagem de produtos de PCR de um fragmento do gene Cit. *b* revelou a existência de 87 haplótipos em 5 indivíduos *C. barbara* e 26 em dois indivíduos *C. orni.* A identificação dos haplótipos de origem

mitocondrial só foi possível com a sequenciação de ADN mitocondrial purificado. Em *C. barbara*, o uso de primers universais na reacção de PCR do fragmento do gene mitochondrial Cit. *b* analisado inicialmente neste estudo, amplifica quase exclusivamente cópias de origem nuclear. A frequência de *Numts* nesta espécie poderá ser uma das maiores reportada até ao presente e o processo de integração de ADN mitocondrial no núcleo destas cigarras aparenta ser frequente e possivelmente contínuo.

A autenticidade dos dados obtidos nas análises de ADN mitocondrial neste estudo foi no entanto devidamente confirmada através da sequenciação do ADN extraído directamente das mitocondrias de *C. orni* e *C. barbara*. Foram também concebidos "primers" específicos para amplificação de ADN exclusivamente mitocondrial do gene Cit. *b* em *C. barbara*. O uso de métodos padronizados para evitar ou detectar precocemente a co-amplificação de *Numts* deveria ser uma prioridade no início de qualquer estudo envolvendo análises de ADN mitocondrial.

Para a análise molecular das cigarras, o ADN total foi extraído de cada indivíduo e o domínio III do gene mitocondrial 12SrRNA foi amplificado via PCR em 462 exemplares. Através da sequenciação directa de produtos de PCR ou da identificação dos fragmentos de ADN mitocondrial usando as técnicas RFLPs e SSCPs, foram identificados diversos haplótipos. O alinhamento dos haplótipos foi efectuado com base na estrutura secundária do gene 12SrRNA e procedeu-se à sua análise usando "software" apropriado para genética populacional e estudos filogenéticos.

Com base no ADN mitocondrial, cinco haplogrupos, correspondentes à sua denominação específica, foram identificados: *C. orni*, *C. barbara*, *C. mordoganensis*, *C. cretensis* e *C. lodosi*. Cada um daqueles corresponde a um grupo evolutivo distinto. *C. barbara* é a espécie mais divergente dentro das espécies analisadas, apresentando distâncias genéticas superiores a 10% relativamente a todas as outras espécies analisadas, enquanto *C. orni* e *C. mordoganensis* são as mais semelhantes (divergência de 2,3%). Contrariamente aos dados de alozimas, estes resultados não sugerem que a especiação tenha ocorrido com pouca diferenciação genética, especialmente tendo em consideração que o gene estudado é consideravelmente conservativo. Comparando os oscilogramas dos sinais de chamamento das espécies analisadas com a sua posição na árvore filogenética, as espécies mais basais *C. barbara* e *C. lodosi* apresentam sinais de chamamento com uma série de pulsos contínuos, sem pausas, enquanto as espécies mais recentes apresentam em crescendo pausas mais longas,

sugerindo que a introdução de pausas nos sinais de chamamento é um evento mais recente, hipótese que ,contudo, não foi testada neste estudo. Estes resultados confirmam que estas espécies discriminadas essencialmente com base nos perfis acústicos correspondem de facto a linhagens evolutivas distintas, reforçando assim a importância dos sinais acústicos em estudos evolutivos e de sistemática envolvendo cigarras. Processos biogeográficos insulares possivelmente favoreceram a divergência populacional e especiação na área do mar Egeu. Dado que *C. orni* é a espécie que divergiu mais recentemente no género *Cicada* e como as espécies mais estreitamente aparentadas se encontram na área do mar Egeu, esta é provavelmente a região de origem desta espécie.

A estrutura populacional e alguns parâmetros demográficos em *C. orni* e, particularmente, em *C. barbara* não foram completamente resolvidos com este gene devido à pouca variabilidade encontrada. Contudo, a maior divergência intra-específica observada em *C. orni*, pode ser explicada pela divergência da maioria dos haplótipos da Grécia. As análises populacionais ao nível molecular em *C. orni* evidenciaram a separação em seis grupos: um incluindo quase todas as localidades amostradas na Europa (desde a Península Ibérica, França até à Macedónia), outro grupo separando a Grécia continental, e os restantes separando individualmente as ilhas Gregas Lesbos, Kithira, Naxos e Skyros. Com a excepção das populações *C. orni* gregas, todos os restantes haplogrupos apresentaram um padrão de expansão demográfica provavelmente relacionadas com as mudanças climáticas durante o Pleistocénio (período Quaternário). A diferenciação entre a Grécia e o resto da Europa sugere que as populações gregas permaneceram isoladas e não terão contribuído grandemente para a recolonização pós-glacial de outras regiões Europeias. Por outro lado, a falta de estrutura geográfica nas restantes populações não permite excluir a possibilidade de uma recolonização da Europa proveniente da Península Ibérica apesar da barreira geográfica dos Pirinéus. É mesmo possível que estas montanhas tenham constituído um obstáculo à passagem de cigarras, tendo a sua dispersão ocorrido ao longo das zonas costeiras. Os sinais de chamamento de machos *C. orni* da Europa ocidental e da Grécia continental constituem um grupo relativamente homogéneo. Contudo, os perfis acústicos do sudeste Europeu (Grécia) tendem a agrupar separadamente dos da Europa ocidental (Península Ibérica e França). A variável acústica que mais contribuiu para esta separação foi a duração dos intervalos entre equemas (grupos de pulsos), se bem que quase todas as variáveis apresentaram diferenças significativas entre a Grécia e o resto da Europa. Esta diferenciação acústica corrobora os resultados obtidos ao nível molecular. Ao contrário da duração dos intervalos entre equemas,

a duração dos equemas foi consideravelmente constante ao longo da distribuição geográfica de *C. orni*, indicando que provavelmente é uma variável importante no reconhecimento intra-específico destas espécies.

A estrutura populacional e padrões demográficos em *C. barbara* foram determinados com base na análise de um fragmento mitocondrial do gene Cit. *b*. Os resultados desta análise sugerem que as populações de Marrocos e da Península Ibérica encontram-se geneticamente isoladas, mas dado que a população de Ceuta no norte de África se assemelha mais às populações da Península Ibérica, o estreito de Gibraltar não parece ter constituído uma barreira intransponível à dispersão das cigarras, ao contrário das montanhas Rif que separam as populações de Marrocos. A maior variabilidade genética encontrada em Marrocos mostra que estas populações têm historicamente uma maior estabilidade demográfica do que as da Península Ibérica. Contudo, estes resultados não confirmam a hipótese de uma introdução recente desta espécie na Península Ibérica vinda de Marrocos. De facto, as regiões mais a noroeste de África, como Ceuta, poderão ter constituído um refúgio para as cigarras Ibéricas durante as condições climáticas mais severas da era Quaternária. Ao nível acústico, as variáveis de frequência em *C. barbara* encontram-se correlacionadas com a latitude, diminuindo de sul para norte. Contudo, os coeficientes de variação na maioria destas variáveis de frequência são reduzidos sugerindo que estas variáveis têm propriedades inerentes estáticas. A maioria das variáveis acústicas dos sinais de chamamento apresenta diferenças significativas entre Marrocos e Península Ibérica (mais Ceuta), corroborando as análises moleculares. A divisão de *C. barbara* em duas subespécies - *C. barbara lusitanica* em Portugal e Espanha (incluindo Ceuta) e *C. barbara barbara* em Marrocos - encontra fundamento nestes resultados, tanto ao nível molecular como acústico.

Em áreas de simpatria na Península Ibérica, onde *C. orni* e *C. barbara* ocorrem simultaneamente, não foram encontrados quaisquer híbridos com características de sons intermédios, nem foi detectado "desvio de caracteres" nos sinais de chamamento. *C. barbara* não apresentou diferenças nas propriedades acústicas das populações, quer em simpatria quer em alopatria, mas em *C. orni* a variável acústica "frequência mínima" mostrou diferenças significativas entre populações simpátricas e alopátricas. É provável que estas espécies já se encontrassem consideravelmente diferenciadas quando entraram em contacto na Península Ibérica.

De acordo com todas as análises efectuadas neste estudo, a evolução dos sinais moleculares e acústicos aparenta ser um processo paralelo. Contudo, as espécies mais

intimamente relacionadas apresentam uma sobreposição em vários caracteres acústicos. Dado que estas espécies nunca se encontram em simpatria, as populações geneticamente isoladas poderão reter os sinais de chamamento ancestrais através de selecção estabilizadora.

O presente estudo constitue uma contribuição para o melhor conhecimento da filogenia das espécies de *Cicada* L. na área do Mediterrâneo e para uma melhor compreensão dos seus padrões de especiação e de evolução. Esperamos que as incertezas e as novas hipóteses aqui sugeridas constituam uma base sólida para futuras investigações.

Palavras-chave: *Cicada*, ADN mitocondrial, filogenia, variabilidade acústica, cópias nucleares.

ABSTRACT

Cicadas are widely distributed in the Mediterranean area and are mainly identified by male acoustic signals, which act as specific mate recognition systems (SMRSs), attracting only conspecific females. Within the genus *Cicada* Linnaeus several species have diverged in their calling songs without showing external morphological differences. Yet, the degree of genetic differentiation among these species was still mostly undetermined and their phylogenetic relationships were practically unknown. The present study assesses mtDNA sequence variation in a complex of six closely related species of *Cicada* from the Mediterranean area in order to test the applicability of the SMRS species concept and to determine the phylogenetic relationships and the evolutionary history of this group. The population structure of the two most common species, *C. orni* and *C. barbara* was analysed in some detail at the molecular and acoustic levels in order to help explain the present patterns of distribution and to test the congruence between mtDNA and acoustic behaviour divergence. The detection of nuclear copies of mtDNA (*Numts*) of the cytochrome *b* gene was crucial for subsequent data analyses. The number of *Numts* found here is to our knowledge unprecedented in the literature for such a small DNA fragment analysed. The subsequent sequence analysis proved to be of exclusively mitochondrial origin. These analyses confirmed a general congruence between mtDNA and acoustic behaviour divergence at the species and the population levels in the genus *Cicada*. Each nominal species here analysed constitutes a distinct evolutionary group. Acoustic and mtDNA sequence data also supported the separation of the *C. orni* Greek populations from the rest of Europe and the separation of Moroccan and Iberian populations of *C. barbara*. Pleistocene climatic changes and several dispersal barriers, such as mountain ranges and sea barriers, were identified as important factors influencing the life history of these cicadas.

Keywords: *Cicada*, phylogeny, mtDNA, acoustic variability, *Numts*.

Chapter 1.

General Introduction

Our greatest glory is not in never falling, but in getting up every time we do.
Confucius

Chapter 1.

GENERAL INTRODUCTION

Cicadas (Insecta: Hemiptera) belong to the family Cicadidae in which males have the ability of producing striking airborne acoustic signals used in intraspecific communication (e.g. Alexander 1967; Claridge 1985; Bennet-Clark 1998). Within these acoustic signals, calling songs are used mainly for mate recognition, pair formation and courtship (as well as for male aggregation and chorusing) and have been shown to attract only conspecific females (Villet 1992; Cooley & Marshall 2001), therefore avoiding interspecific mating.

In such species, which use non-visual mating signals, speciation is not always accompanied by morphological change (Bickford *et al*. 2007), resulting in cryptic species. According to Bickford *et al*. (2007) "*Cryptic species are either differentiated by non-visual mating signals and/or appear to be under selection that promotes morphological stasis (lack of change in characteristics of gross external anatomy)*". In fact, the occurrence of morphologically cryptic species within cicadas is not unusual (*e.g.* genus *Cicada*, Quartau *et al.* 2000, 2001; genus *Tettigetta*, Quartau & Boulard 1995; and *Platypleura stridula* species complex, Price *et al*. 2007), and strong selection for predator avoidance associated with the long-range mate attraction by means of loud species-specific acoustic signals have been highlighted as possible factors leading to the formation of sibling species (Quartau *et al*. 2000; Price *et al*. 2007). So, constraints such as the physiology of sound production/reception organs and the need for concealment to avoid predators may favour morphological similarities within cicadas.

Since cryptic species, by definition, are difficult to differentiate on morphological grounds, other characters are required, such as molecular (Herbert *et al*. 2003; Bickford *et al*. 2007) and/or behavioural data (Heyer & Reid 2003; Angulo & Reichle 2008). Therefore, acoustic mating signals have been used to discriminate between closely related cryptic species in a variety of insects (e.g. lacewings, Henry 1994; Wells & Henry 1998; orthopterans, Huang *et al*. 2000; hemipterans, Quartau & Boulard 1995; Quartau & Simões 2005), but also in vertebrates including frogs, bats and birds (e.g. Barlow & Jones 2004; Bickford *et al*. 2007; Angulo & Reichle 2008; Jones 2008).

DNA sequence data have also revealed numerous cryptic species which are often subsequently confirmed with morphological, behavioural and/or ecological data. Thus, several authors suggest that molecular data should be incorporated in the description of new species

as a matter of routine (Tautz *et al*. 2003; Bickford *et al*. 2007). Identifying new cryptic, insect species, based on molecular data, has been essential for meaningful epidemiological studies and effective vector control programmes (e.g. Malaria-vector *Anopheles* species complex, Besansky 1999), as well as for pest management (e.g. beetles considered dietary generalists are in fact complexes of dietary specialists, Blair *et al*. 2005) and studies of coevolution and species interaction (Bickford *et al*. 2007).

According to integrative taxonomy, evidence for separating species should appear from concordant changes in more than one type of characters of an organism, e.g. molecules, morphology or mating signals (Bickford *et al*. 2007). This has been implemented in several organisms, and, for instance, some species initially identified by male calling songs have been subsequently corroborated with genetic data, such as North American crickets from the genus *Gryllus* (see Huang *et al*. 2000) and Hawaiian crickets, genus *Laupala* (see Shaw 2000; Shaw *et al*. 2007).

The speciation events associated with the divergence of mating signals in cryptic species, such as occurs in cicadas, makes these species good candidates for speciation studies (Quartau *et al*. 1997; Wells & Henry 1998; Quartau & Simões 2006). During the growth of evolutionary theory, especially over the last century, different types of speciation model have been developed by biologists namely allopatric, parapatric, sympatric (including polyploidy), and hybrid speciation (Mayr & Ashlock 1991). In this context, knowledge of the mechanisms involved in the acquisition of variation leading to present-day biodiversity is the key to understanding the evolutionary history of any population, species or clade.

In comparative analyses of the processes of adaptation or molecular evolution, and in studies of historical biogeography the resolution of the relationships between species is required. However, more problems exist in understanding the species concept than in any other basic theoretical idea in biology (Bock 2004). As Angulo & Reichle (2008) stated "*Difficulties in reaching a consensus are not surprising in view of the inherently diverse, variable, and dynamic nature of biological systems, which makes it difficult to identify clear-cut boundaries*".

The species concept problem

Many concepts of species exist (e.g., Mayr 1940, 1942, 1963, 1985, 1997; Mayr & Ashslock 1991; Wilson 1992; Claridge *et al*. 1997), however, the biological concept has been the most widely accepted. In the classical biological species concept developed particularly by Mayr (*op. cit*.), reproductive isolation is the critical criterion for establishing the species status

of populations of sexually reproducing organisms. Mayr (*e.g.* 1963) defined species as groups of actually, or potentially, interbreeding populations which are reproductively isolated from other groups. Dobzhansky (1937, 1970) stated that such isolation is maintained by species isolating mechanisms. Thus, according to Dobzhansky, Mayr and many succeeding authors, the phenomenon of speciation is seen as the evolution of reproductive isolation (*e.g.* Claridge & De Vriijer 1993; Claridge & Boddy 1994; Claridge 1995). However, biological species cannot be applied to organisms that do not undergo sexual reproduction or to parthenogenetic organisms. Even for sexually reproducing organisms, it is only possible to determine if two populations are reproductively isolated if they naturally co-exist in the field, that is if the populations are truly sympatric. Due to these problems many taxonomists have consistently rejected the biological concept often in favour of an operational approach such as the morphological/phenetic species (*e.g.* Sokal & Crovello 1970; Quartau 1983a, 1983b, 1988).

Paterson (1978, 1980, 1981, 1985) resolved some of the difficulties imposed by the biological species concept when he argued that species are primarily characterized by unique specific mate recognition systems (SMRSs). These serve to ensure the mating of conspecifics and only result secondarily in reproductive isolation (Claridge & De Vriijer 1993). Thus, Paterson (*op. cit.*) emphasises the importance of recognition rather than isolation. SMRS and the classical biological species concepts have much in common and together may still be regarded as the biological species concept (*e.g.* Mayr 1985; Claridge & Boddy 1994; Claridge 1995). In fact, biological species do have characteristic mate recognition systems which result in the reproductive isolation of species (Claridge & De Vriijer 1993), thus, SMRSs should allow the discrimination of species. However, when two divergent species are secondarily in contact, these concepts have different implications concerning the isolating mechanisms. While in the Dobzhansky-Mayr concept natural selection will reinforce the pre-zygotic isolation, Paterson's concept, based in recognition mechanisms (pre-zygotic), refutes any selective reinforcement at this level.

Cracraft (1983), inspired by Simpson (1951, 1961) and Wiley (1978), defined an alternative species concept that follows a phylogenetic and evolutionary approach, as the "*smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent*". According to this evolutionary concept, there is no gene flow between different species, therefore, genetic isolation can confirm their separation. But according to Bock (2004), evolutionists confused genetic and reproductive isolation. Although species reproductively isolated are genetically isolated, the reverse is not necessarily true (e.g.

horse and donkey). Lee (2004) also emphasized the loose correlation between genetic divergence and reproductive isolation and emphasized that it is not possible to associate a certain level of divergence with reproductive isolation. Templeton (1994) defended that in organisms with sexual reproduction, evolutionary lineages described as a species must be phylogenetically distinct, show no recent gene flow and demonstrate ecological or demographic limitations to reproduction.

There is no single consensus species concept that can be applied to all organisms (Bock 2004; Angulo & Reichle 2008) and, more than just being a part of systematics theory, the species concept is an integral part of evolutionary biology (Szalay & Bock 1991). Mayden (1997) suggested the use of several rational species concepts organized hierarchically as primary and secondary concepts based on their theoretical significance, generality and operationality. Primary concepts would help structure ideas and perceptions on species, whereas secondary concepts, subordinate to primary concepts, would work as operational tools in identifying species. In this respect, the SMRS can be a useful secondary operational species concept to consider in species where acoustic communication systems are extremely important SMRSs (e.g. anurans - Angulo & Reichle, 2008).

Using a combination of concepts may therefore help to apply a theoretical framework as well as a working guide to determine the status of populations of morphologically cryptic groups (Mayden 1997; Angulo &Reichle, 2008).

Mitochondrial DNA

The evolution and relationships among organisms has been traditionally investigated by comparing their morphological, ecological and behavioural characters but the development of molecular techniques now allows the use of elemental data to infer phylogenetic relationships based on the concepts of natural selection and neutral evolution. Analyses of DNA sequences within and between species are a powerful approach for determining the evolutionary mechanisms in specific gene regions but also for determining relevant aspects of the evolutionary history of the species (Hewitt 1998; Ramos-Onsis & Rozas 2002). Mitochondrial DNA has been widely and successfully used in the study of phylogenetic and phylogeographic relationships in numerous taxa including insects (*e.g.* Roderick 1996; Caterino & Sperling 1999; Caterino *et al*. 2000, 2001; Lin & Danforth 2004; Rubinoff & Holland 2005). It has also been recommended to use mtDNA as a sequence tag or barcode in taxonomic studies (Herbert *et al*. 2003; Tautz *et al*. 2003).

Mitochondrial DNA has a simple sequence organization, maternal inheritance and in most cases no recombination (Avise *et al*. 1987; Moritz *et al*. 1987; Harrison 1989; Moritz 1994; Simon *et al*. 1994; Rubinoff & Holland 2005; but see Piganeau *et al*. 2004). Haploid maternal inheritance reduces the effective population size and thus increases the sensitivity of this molecule to genetic drift (Birky *et al.* 1989; Moritz 1994). MtDNA is thus expected to be a more sensitive indicator of population subdivision than nuclear genes (Crozier 1990). In animals, mtDNA generally evolves much faster than nuclear DNA (Avise 2000; Hey & Machado 2003). Rapid rates of sequence divergence allow the discrimination of recently diverged lineages (Harrison, 1989). Such characteristics, besides its relatively simple isolation and manipulation, make mtDNA analysis a powerful tool in molecular systematics and evolutionary biology (Moritz 1994; Hewitt 1998), especially when looking at closely related taxa and intraspecific relationships.

The rates and patterns of mtDNA evolution vary among taxa and among different regions of the mtDNA (Arbogast *et al*. 2002; Ballard & Whitlock 2004). In insects, mutation rates in mtDNA are estimated to be two to nine times higher than in nuclear DNA (DeSalle *et al*. 1987; Moriyama & Powell 1997; Monteiro & Pierce 2001; Ballard & Whitlock 2004).

The development of PCR to obtain sufficient amounts of specific DNA, avoiding the cloning process, has facilitated population comparisons and the molecular systematics of insects, including cicadas (Martin & Simon 1990; Simon *et al.* 1994). Mitochondrial genes can be easily amplified with the use of universal conserved primers (*e.g*. Kocher *et al*. 1989; Simon *et al*. 1994, 2006). Mitochondrial genes, such as the cytochrome *b* and the 12S ribosomal RNA subunit, are increasingly used in phylogenetic studies (*e.g*. Crozier *et al*. 1989; Kocher *et al.* 1989; Irwin *et al.* 1991; Crozier & Crozier 1992; Simon *et al*. 1996; Tautz *et al*. 2003; Monteiro *et al*. 2004), and for evolutionary analysis of their structure and function (*e.g*. Howell 1989; Kocher *et al.* 1989; Irwin *et al.* 1991; Crozier & Crozier 1992). Cytochrome *b* has a rapid evolutionary rate suitable for comparative studies of species within the same genus or the same family (Simmons & Weller, 2001), and has been widely used for elucidating the phylogenetic relationships of insects at both species and intraspecific levels (e.g. Aikhionbara & Mayo 2000; Huang *et al*. 2000; Koulianos & Schmid-Hempel 2000).

The use of mitochondrial DNA has considerable advantages, but the role of mtDNA sequences in taxonomy and phylogenetic inference has been questioned (Ballard & Whitlock 2004; Hurst & Jiggins 2005; Rubinoff & Holland 2005). High evolutionary rates for single nucleotide positions can lead to homoplasy due to saturation (Lin & Danforth 2004; Ballard &

Whitlock 2004). This leads to a reduced phylogenetic signal in deeper nodes compared with nuclear DNA data. Furthermore, since mtDNA is maternally inherited, if male and female histories differ in a species due to sex-biased dispersal or fecundity, this can lead to discordance between nuclear DNA and mtDNA genealogies (Monsen & Blouin 2003; Lin & Danforth 2004; Hurst & Jiggins 2005).

Another criticism refers to the "gene tree versus species tree" problem: when studying any single locus one is merely studying the evolutionary history of this gene, rather than the organism (Nichols 2001) which may or may not agree with the species phylogeny. Though incongruence among gene trees is expected under certain circumstances such as horizontal transfer (including hybridization), lineage sorting (deep coalescence) and gene duplication coupled with extinction, due to their smaller effective population sizes mitochondrial genes may track short internodes for recent divergences more reliably than nuclear autosomal genes (Caterino *et al*. 2000).

Increasing evidence concerning mtDNA recombination (reviewed in Rokas *et al*. 2003) and the occurrence of mtDNA inherited symbionts such as *Wolbachia* (see Hurst & Jiggins 2005) which can confound analysis of population, biogeographic and clade history also highlight problems in the use of mtDNA. Moreover, the inadvertent amplification of paralogous nuclear copies of mitochondrial genes (pseudogenes) confounding phylogenetic analyses has been increasingly reported (Bensasson *et al*. 2001; Richly & Leister 2004). However, the incidence and distribution of these phenomena among different species is still mostly unknown.

Despite these problems, studies involving very closely related taxa are generally dependent on the higher overall rates of nucleotide substitution found in mitochondrial genes. Therefore, mitochondrial loci are usually suitable tools for resolving close relationships (Avise 2000; Lin & Danforth 2004; Rubinoff & Holland 2005). Additionally, potential problems including effects on selection, co-amplification of pseudogenes with mtDNA, recombination, etc., can often be avoided or solved by thoughtful molecular analysis (Bermingham & Moritz 1998) and by testing congruence with other markers.

Population genetics

The basis for all evolutionary change is genetic variation and DNA sequences can provide remarkably detailed views of patterns of variation within species and phylogenetic relationships among species (Harrison 1991; Hewitt 1998). When speciation occurs, each

descendent population has a sample of the genes present in the parental population (Crozier 1990). Speciation events do not leave unique signatures and it would be impossible to prove that a certain series of historical events has occurred. However, from patterns of molecular genetic variation, together with detailed knowledge of the biology and ecology of organisms, it is often possible to eliminate certain models or at least assign high probabilities to some scenarios and low probabilities to others (Harrison 1991).

The current distribution of a species is not necessarily a reliable indicator of its historical range (Losos & Glor 2003). In fact, the range of a species is dynamic and repeated extinctions followed by recolonization and secondary contacts are now recognized as important determinants of the structure and divergence of populations (Widmer & Lexer 2001). Major climatic oscillations have different consequences across the world and species react in response to local geographic and climatic conditions depending on their ecology (Hewitt 2000). Fossil remains of plants and animals, as well as molecular data from various organisms have revealed shifts in community composition and structure of species (Hewitt 2000, 2004). The range expansion and retreat of species has been well demonstrated and the Quaternary glaciations have been decisive in shaping current distributions of many species (*e.g.* Taberlet *et al*. 1998; Hewitt 1996, 1999, 2000, 2004; Avise 2000; Besnard *et al*. 2002). Many studies have recognized the Iberian Peninsula, Italy and the Balkans as three major southern refugia in Europe during the severe effects of the glaciations and from where populations expanded during interglacial periods (Hewitt 1996, 1999, 2000, 2004; Lunt *et al*. 1998; Kutnik *et al*. 2004). The patterns of genetic differentiation among extant populations are often due to historical events, such as range expansion, range fragmentation, population bottlenecks, survival in different refugial zones combined with genetic drift and founder effects during re-colonization (Newton *et al*. 1999; Besnard *et al*. 2002). In the Iberian Peninsula, the expansion and contraction of ice sheets have been implicated as generators of intraspecific diversity and substructure within Iberian taxa (e.g. Paulo *et al*. 2001; Hewitt 2004).

Demographic events have been shown to leave signatures on DNA sequence data: a population decline often causes a loss of sequence diversity whereas population growth can cause the retention of sequences that otherwise might have been lost (Harpending 1994; Rogers & Harpending 1992; Ramos-Onsis & Rozas 2002). In fact, DNA sequence diversity may provide an instrument for examining prehistoric demography (Harpending 1994).

SMRS in cicadas

According to Claridge & Boddy (1994), whatever the nature of the markers, the only certain way of establishing species boundaries is the investigation of SMRSs and the resulting reproductive isolation. It is common to find that the most striking differences between closely related sympatric species concern characters used in mate recognition (*e.g*. Maynard-Smith 1989). SMRS are well exemplified by classical exchanges of signals in courtship sequences described by ethologists (*e.g.* Tinbergen 1951), including in cicadas. Male cicadas produce loud and distinctive acoustic signals during pair formation and courtship, by means of a tymbal mechanism (Bennet-Clark 1998a), and females are only attracted to the calls of conspecific males (*e.g*. Claridge 1985, 1995; Quartau & Rebelo 1994; Quartau 1995). The tymbals are cuticle membranes located dorsolaterally in the first and second segments of the abdomen that are distorted by the action of powerful muscles (*e.g.* Pringle 1954; Popov 1975; Bennet-Clark 1998a).

There are different kinds of acoustic signals that cicada males can produce (*e.g.* Boulard 1995; Fonseca 1991), but the calling song is the most common one. This is typically regularly patterned, species-specific and is involved in mate attraction (as shown in *Magicicada* spp., Cooley & Marshall 2001) as well as in male aggregation and chorusing as referred to, for example, by Villet (1992), Simões *et al*. (2000) and Fonseca & Revez (2002).

Singing insects are under selective pressure to optimize the range, and to maintain the specificity of their calls because of the relationship between calling success and reproductive success (Bennet-Clark 1998b; Oberdörster & Grant 2007). Therefore, according to the Paterson's species recognition concept, the calling song constitutes a distinct specific-mate recognition system (SMRS) that would be maintained relatively constant by stabilizing selection across the distribution range of the species (Paterson 1985). However, according to the reinforcement model of speciation derived from Mayr-Dobzhansky species concept (Dobzhansky 1937, 1970; Mayr 1940, 1942, 1963), if two incompletely reproductively isolated species come into contact after a period of allopatry, selection against heterospecific mating reinforces or accentuates their differences. According to this, sympatric populations are expected to show more marked reproductive character displacement in mating mechanisms compared to allopatric populations as found in *Magicicada* species, i.e., selection reduces the possibility of cross-species sexual interactions by accentuating differences in sexual signals (Cooley *et al*. 2001, 2003, 2006). Alternatively, sympatric populations can affect each other through hybrid sexual interactions forming a "hybrid-zone" or, formerly

separated populations could be so differentiated that, after coming into contact, they co-exist without hybridizing at all, or with strong selection against hybrid viability. Generally, the greater the differences between populations upon contact, the greater the likelihood they will retain their distinctiveness.

Intraspecific variation in the calling song can occur at a variety of levels. It has been suggested that acoustic features of the calling song of swordtail crickets in Hawaii (Otte 1994, *in* Shaw & Herlihy 2000) and of *Drosophila willistoni* complex (Gleason & Ritchie 1998) are among the earliest characters to change in the speciation process. The opposite was found in the frog species complex *Leptodactylus fuscus* which shows a strong genetic differentiation but no salient differences in advertisement calls (Heyer & Reid 2003).

In addition to reproductive character displacement, evolutionary changes in calling songs may be generated by sexual selection (e.g. for some trait indicative of male fitness), a change of habitat (e.g. forested habitat becoming an open habitat), or also, by predator pressure (Gerhardt 1994).

Cicada L. 1758

Cicadas spend most of their life cycle underground feeding on plant xylem (Quartau 1995; Price *et al*. 2007). Only adults live above the surface for a short period of time for reproductive purposes. Female cicadas lay their eggs inside the branches of trees or shrubs and when the immature forms hatch they fall into the ground and dig tunnels where they live until they mature to adults. Cicadas are hemimetabolous insects, meaning they lack complete metamorphosis, so increasingly larger nymphs approach adult form without going through a pupal stage. The exact amount of time they live underground is still not known for many species. In fact, in *Cicada* immature stages may last between two to three years (Quartau 1995; Boulard & Mondon 1996; Giralda *et al*. 1998).

Within the genus *Cicada* Linnaeus, several species have diverged in their calling songs, without showing clear-cut external morphological differences, such as *Cicada barbara* Stål, *C. orni* Linnaeus, *C. mordoganensis* Boulard and *C. cretensis* Quartau & Simões (Boulard 1982; Quartau *et al.* 1997, 2000, 2001; Simões *et al.* 2000; Seabra *et al.* 2000; Quartau & Simões 2005).

In the Mediterranean area, the genus *Cicada* comprises a complex of six species (seven with *C. cerisyi*, a species of doubtful taxonomic identity and referred to Egypt and Lybia) which are currently recognised based on acoustic divergence and some differences in

the male genitalia. *C. orni* is one of the commonest and widespread cicadas occurring from the Iberian Peninsula along south central Europe to eastern Mediterranean Europe, western Asia and the Middle East (Popov 1975; Quartau & Fonseca 1988; Quartau & Simões, 2006). *C. barbara* is widely distributed in northwestern Africa, some Mediterranean islands and also occurs in Portugal and Spain where it may occur in sympatry with *C. orni* (Boulard 1982; Quartau 1988). These two sibling species are very difficult to distinguish based on morphologic characters alone (Quartau 1988) but have distinct calling songs, which are easily recognised in the field. Whilst *C. barbara* males emit a continuous series of pulses without pauses (Boulard 1982, 1995; Fonseca 1991; Quartau & Rebelo 1994), *C. orni* males produce a repetitive series of pulses (called echemes) alternating with silent pauses (Popov 1975; Boulard 1982; Fonseca 1991; Quartau *et al*. 1999).

Boulard (1982) described the populations in the Iberian Peninsula as a new subspecies, *C. barbara lusitanica*, differing slightly from the nominal subspecies from North Africa based on a few characteristics of the male genitalia and the female ovipositor. Three species are known to occur in Turkey, namely *C. lodosi* Boulard 1979, *C. permagna* Haupt 1917 (Boulard 1979) and *C. mordoganensis* which is also present on a few Greek islands close to the Turkish mainland (Boulard 1979; Simões *et al.* 2000). Finally, *C. cretensis* has been recently described by Quartau & Simões (2005) and is thought to be endemic to Crete. *Cicada orni*, *C. mordoganensis* and *C. cretensis*, all present in the Aegean area (but not in sympatry), are morphologically very similar and can be differentiated only by small differences in the temporal pattern of their calling songs (Simões *et al*. 2000; Quartau & Simões 2005, 2006). On the contrary, *C. lodosi* which can also be found in sympatry with *C. mordoganensis*, can be easily distinguished morphologically due to its larger size and on its continuous calling song. These species have a similar ecology, occurring in high shrubland and woodland and associated with olive, pine, oak and introduced eucalyptus.

At the molecular level, allozyme studies revealed low genetic distances between *C. orni* and *C. barbara* in Portugal (Quartau *et al.* 2000), whilst *C. mordoganensis* could not be separated from *C. orni* (see Seabra *et al.* 2000). *C. barbara* was found a genetically homogeneous species in Portugal, especially when compared to *C. orni* (see Quartau *et al*. 2000, 2001) and it was suggested that *C. barbara* is a relatively recent immigrant to the Iberian Peninsula from North Africa. More recently, microsatellite markers confirmed that the isolating barriers between *C. orni* and *C. barbara* are efficient (Seabra 2007).

With the exception of the above studies, no other molecular studies have been carried out on the genus *Cicada*. The degree of genetic differentiation among these species was still mostly undetermined and their phylogenetic relationships were practically unknown. Due to their peculiar specific mating recognition systems associated with their great abundance and wide geographical distribution, cicadas can potentially be useful models to test a variety of ecological and evolutionary hypotheses (*e.g.* Cooley *et al*. 2001; Quartau & Simões 2006; Price *et al*. 2007; Suer *et al*. 2007). The hypotheses described below will be tested using a range of molecular, morphometric and acoustic signal analysis.

Key Hypotheses

1. Within the genus *Cicada,* species which have been described mainly based on acoustic characters constitute independent evolutionary units when analysed using other approaches.
2. Since *C. orni* and *C. barbara* have a Mediterranean distribution, occurring along recognised refugial areas during the Pleistocene glaciations within Europe and north Africa, they have avoided the most severe effects of the climatic changes during the Pleistocene and as such comprise diverse and stable populations. Alternatively, since these species are highly thermophilic and dependent on environmental factors (such as tree distribution), they were in fact susceptible to the Pleistocene climate changes and suffered accentuated bottlenecks during glaciations.
3. MtDNA corroborates *C. barbara* as a genetically homogeneous species in the Iberian Peninsula, suggesting it as a relatively recent immigrant in this area.
4. The calling songs of *C. orni* and *C. barbara* are maintained relatively constant by stabilizing selection across the distribution range of the species (accordingly to Paterson's (1985) recognition concept of species).
5. When populations of *C. orni* and *C. barbara* are in sympatry their calling songs diverge more than when populations are allopatric due to reproductive character displacement.
6. MtDNA and acoustic similarity patterns within populations of *C. orni* and *C. barbara* have evolved in a similar way or alternatively there is no congruence between genetic and acoustic divergence.
7. MtDNA and acoustic signals markers can elucidate the hypothesis described above.

Objectives and thesis structure

This thesis aims to:

1. Assess the phylogenetic history of the genus *Cicada* in the Mediterranean using mtDNA markers, focusing on the population genetics of the two occasionally sympatric species: *C. orni*, which is widely spread in Europe, and *C. barbara*, which is thought to be recently introduced in the Iberian Peninsula from North Africa. The phylogenetic component aims at resolving the interspecific relationships in the genus *Cicada*, and thus, can test the applicability of the SMRS biological species concept. The population genetics component will reveal patterns of population historical demography to help explain the present pattern of distribution and genetic structure (testing hypotheses 1, 2 and 3).
2. Investigate acoustic recordings from specimens of *C. orni* and *C. barbara* along a wide distribution range of these species (testing hypothesis 4).
3. Compare the calling songs from specimens of *C. orni* and *C. barbara* from allopatric and sympatric populations (testing hypothesis 5).
4. Test the congruence between mtDNA and acoustic behaviour divergence by comparing the patterns of geographical variation at the molecular and acoustic levels from populations of *C. orni* and *C. barbara* (testing hypothesis 6).
5. Analyse the utility of mitochondrial DNA and acoustic signals as markers for species discrimination and population structure in cicadas based on the success of the results, and also, in the case of mtDNA, considering the co-amplification with mtDNA of mitochondrial nuclear copies (*Numt*s) (testing hypothesis 7).

Thesis structure

These hypotheses and objectives were explored and developed by individual manuscripts which have already been published, accepted or submitted for publication. For this reason, with the exception of the General Introduction and General Discussion chapters, each chapter contains one or two manuscripts, self-contained with its own references. The hypotheses within this thesis are interconnected so they were not addressed separately by each chapter. Instead the separation of chapters follows the subsequent criteria:

Chapter 1 – Presents a general introduction.

Chapter 2 - Addresses the phylogenetic history of the genus *Cicada* and provides evidences for population structure and geographical variation at the molecular level for *C. orni*.

Chapter 3 - Focuses on the population structure and geographical variation based on acoustic data in *C. orni* (the population structure at the molecular level is determined in the previous chapter).

Chapter 4 - Addresses the population structure and geographical variation based on molecular and acoustic data in *C. barbara* (includes two manuscripts).

Chapter 5 - Addresses the problem of mitochondrial nuclear copies found in this study.

Chapter 6 - Presents a general discussion integrating the results of the different chapters.

Appendix - Compares the calling songs of sympatric and allopatric populations of *C. barbara* and *C. orni* on the Iberian Peninsula (includes one manuscript in which the PhD candidate is the second co-author).

References

Aikhionbara FO, Mayo ZB. 2000. Mitochondrial DNA sequences of greenbug (Homoptera: Aphididae) biotypes. *Biomolecular Engineering* **16**: 199-205.

Alexander RD. 1967. Acoustical communication in arthropods. *Annual Review of Entomology* **12**: 495-526.

Angulo A, Reichle S. 2008. Acoustic signals, species diagnosis, and species concepts: the case of a new cryptic species of *Leptodactylus* (Amphibia, Anura, Leptodactylidae) from the Chapare region, Bolivia. *Zoological Journal of the Linnean Society* **152**: 59-77.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* **33**: 707-740.

Avise JC. 2000. *Phylogeography. The history and formation of species*. Cambridge: Harvard.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetic and systematics. *Annual Review of Ecology and Systematics* **18**: 489-522.

Ballard JW, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* **13**: 729-744.

Barlow KE, Jones G. 2004. Cryptic species of echolocating bats. *In* JA Thomas, CF Moss & M Vater (Eds) *Echolocation in Bats and Dolphins*. Chicago, The University of Chicago Press, pp. 345-349.

Bennet-Clark HC. 1998a. How cicadas make their noise. *Scientific American* May: 36–39.

Bennet-Clark HC. 1998b. Size and scale effects as constraints in insect sound communication. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **353**: 407-419.

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

Bermingham E, Moritz C. 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**: 367-369.

Besnard G, Khadari B, Baradat P, Bervillé A. 2002. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* **104**: 1353 - 1361.

Besansky NJ. 1999. Complexities in the analysis of cryptic taxa within the genus *Anopheles*. *Parassitologia* **41**: 97-100.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**: 148-155.

Birky CW, Fuerst P, Maruyama T. 1989. Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**: 613-627.

Blair CP, Abrahamson WG, Jackman JA, Tyrrell L. 2005. Cryptic speciation and host-race formation in a purportedly generalist tumbling flower beetle. *Evolution* **59**(2): 304-316.

Bock WJ. 2004. Species: the concept, category and taxon. *Journal of Zoological Systematics and Evolutionary Research* **42**: 178-190.

Boulard M. 1979. Cigales du genre *Cicada* Linné, originaires de Turquie (Homoptera, Cicadidae). *Türkiye Bitki Koruma Dergisi* **3**(2): 67-74.

Boulard M. 1982. Les cigales du Portugal, contribution à leur étude (Homoptera, Cicadidae). *Annales de la Sociètè Entomology de France* (N.S.) **18**(2): 181-198.

Boulard M. 1995. Postures de cymbalisation, cymbalisations et cartes d'identité acoustique de cigales. 1. Généralités et espéces méditerranéennes (Homoptera, Cicadoidea). *EPHE, Biologie et Evolution des Insectes* **7**: 1-72.

Boulard M, Mondon B. 1996. *Vies et mémoires de cigales*. Equinoxe, Barbentane.

Caterino MS, Sperling FAH. 1999. *Papilio* Phylogeny Based on Mitochondiral Cytochrome Oxidase I and II Genes. *Molecular Phylogenetics and Evolution* **11**: 122-137.

Caterino MS, Cho S, Sperling FAH. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology* **45**: 1-54.

Caterino MS, Reed RR, Kuo MM, Sperling FAH. 2001. A Partitioned Likelihood Analysis of Swallowtail Butterfly Phylogeny (Lepidoptera: Papilionidae). *Systematic Biology* **50**: 106-127.

Claridge MF. 1985. Acoustic signals in Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology* **30**: 297–317.

Claridge MF. 1995. Species concepts and speciation in insect herbivores: planthopper case studies. *Bolletim of Zoology* **62**: 53–58.

Claridge MF, Boddy L. 1994. Species recognition systems in insects and fungi. *In* DL Hawksworth (Ed.) *The Identification and Charaterization of Pest Organisms*. CAB International, Wallingford, pp. 261–274.

Claridge MF, De Vriijer P. 1993. Reproductive behaviour: the role of acoustic signals in species recognition and speciation. *In* RF Denno & TJ Perfect (Eds) *Planthoppers their Ecology and Management*. Chapman & Hall, New York, pp. 216–233.

Claridge MF, Dawah, HA, Wilson, MR (eds.).1997. *Species. The units of biodiversity*. The Systematics Association Special Volume Series 54, Chapman & Hall, London.

Cooley JR, Marshall DC. 2001. Sexual signalling in periodical cicadas, *Magicicada* spp. (Hemiptera: Cicadidae). *Behaviour* **138**: 827-855.

Cooley JR, Simon C, Marshall DC. 2003. Temporal separation and speciation in periodical cicadas. *Bioscience* **53**(2): 151-157.

Cooley JR, Marshall DC, Hill KBR, Simon C. 2006. Reconstructing asymmetrical reproductive character displacement in a periodical cicada contact zone. *Journal Compilation European Society for Evolutionary Biology* **19**: 855-868.

Cooley JR, Simon C, Marshall DC, Slon K, Ehrhardt C. 2001. Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *Magicicada* spp.): genetic, morphological, and behavioural evidence. *Molecular Ecology* **10**: 661-671.

Cracraft J. 1983. Species concepts and speciation analysis. *In* R Johnston (Ed.) *Current Ornithology*. Plenum, New York, pp. 159-187.

Crozier RH. 1990. From population genetics to phylogeny: uses and limits of mitochondrial DNA. *Australian Systematic Botany* **3**: 111–124.

Crozier RH, Crozier YC. 1992. The Cytochrome *b* and ATPase Genes of Honeybee Mitochondrial DNA. *Molecular Biology and Evolution* **9**(3): 474–482.

Crozier RH, Crozier YC, Mackinlay AG. 1989. The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Molecular Biology and Evolution* **6**(4): 399-411.

DeSalle R, Freeman T, Prager EM, Wilson AC. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *Journal of Molecular Evolution* **26**: 157-164.

Dobzhansky T. 1937. *Genetics and the origin of species*. Columbia University Press, New York.

Dobzhansky T. 1970. *Genetics of the evolutionary process*. Columbia University Press, New York.

Fonseca PJ. 1991. Characteristics of the acoustic signals in nine species of cicadas (Homoptera, Cicadidae). *Bioacoustics* **3**: 173-192.

Fonseca PJ, Revez MA. 2002. Song discrimination by male cicadas *Cicada barbara lusitanica* (Homoptera, Cicadidae). *Journal of Experimental Biology* **205**: 1285-1292.

Gerhardt HC. 1994. The evolution of vocalization in frogs and toads. *Annual Review of Ecology and Systematics* **25**: 293-324.

Giralda AA, Rodriguez JMP, Pintor FG. 1998. La cigarra del olivo o Magrebí- *Cicada barbara* (Stal.). Junta de Extremadura, *Boletin Fitosanitario de Avisos e Informaciones* **1/98**: 1-4.

Gleason, J.M. & Ritchie, M.G. (1998) Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* complex: do sexual signals diverge the most quickly? *Evolution* **52**: 1493-1500.

Harpending HC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**(4): 591-600.

Harrison RH. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution* **4**(1): 6–11.

Harrison RH. 1991. Molecular changes at speciation. *Annual Review of Ecology and Systematics* **22**: 281–308.

Henry CS. 1994. Singing and cryptic speciation in insects. *Trends in Ecology & Evolution* **9**: 388-392.

Herbert PDN, Cywinska A, Ball SL, DeWaard JR. 2003. Biological identification through DNA barcodes. *Proceedings of the Royal Society B* **270**: 313-321.

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

Hewitt GM. 1998. Diversity in insect species using DNA sequences. *In* A Karp, PG Isaac & DS Ingram (Eds.) *Molecular tools for screening biodiversity*. Published by Chapman & Hall, London, pp. 419–425.

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

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B* **359**: 183-195.

Hey J, Machado CA. 2003. The study of structured populations - new hope for a difficult and divided science. *Nature Reviews Genetics* **4**: 535-543.

Heyer WR, Reid YR. 2003. Does advertisement call variation coincide with genetic variation in the genetically diverse frog taxon currently known as *Leptodactylus fuscus* (Amphibia: Leptodactylidae)? *Anais da Academia Brasileira de Ciências* **75**: 39-54.

Howell N. 1989. Evolutionary conservation of protein regions in the protonomotive cytochrome *b* and their possible roles in redox catalysis. *Journal of Molecular Evolution* **29**: 157–169.

Huang Y, Ortí G, Sutherlin M, Duhachek A, Zera A. 2000. Phylogenetic relationships of north american field crickets inferred from mitochondrial DNA data. *Molecular Phylogenetics and Evolution* **17**(1): 48-57.

Hurst GDD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B* **272**: 1525-1534.

Irwin DM, Kocher TD, Wilson AC. 1991. Evolution of cytochrome *b* gene of mammals. *Journal of Molecular Evolution* **32**: 128–144.

Jones G. 2008. Sensory Ecology: Echolocation calls are used for communication. *Current Biology* **18**(1): 34-35.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villeblanca FX, Wilson AC. 1989. Dynamics of mitocondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences USA* **86**: 6196–6200.

Koulianos S, Schmid-Hempel P. 2000. Phylogenetic relationships among bumble bees (*Bombus*, Latreille) inferred from mitochondrial cytochrome *b* and cytochrome oxidase I sequences. *Molecular Phylogenetics and Evolution* **14**(3): 335-341.

Kutnik M, Uva P, Brinkworth L, Bagnères A-G. 2004. Phylogeography of two European *Reticulitermes* (Isoptera) species: the Iberian refugium. *Molecular Ecology* **13**: 3099-3113.

Lee MSY. 2004. The molecularisation of taxonomy. *Invertebrate Systematics* **18**: 1-6.

Lin CP, Danforth BN. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution* **30**: 686-702.

Losos JB, Glor RE. 2003. Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution* **18**(5): 220-227.

Lunt DH, Ibrahim KM, Hewitt GM. 1998. MtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity* **80**: 633-641.

Martin A, Simon C. 1990. Differing levels of among-population divergence in the mitochondrial DNA of periodical cicadas related to historical biogeography. *Evolution* **44**(4): 1066-1080.

Mayden RL. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. *In* MF Claridge, HA Dawah & MR Wlison (Eds) *Species: the Units of Biodiversity*. New York, Chapman & Hall, pp. 381-424.

Maynard-Smith MJ. 1989. *Evolutionary Genetics*. Oxford University Press, Oxford.

Mayr E. 1940. Speciation phenomena in birds. *American Naturalist* **74**: 249-278.

Mayr E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press, New York.

Mayr E. 1963. *Animal species and evolution*. Harvard University Press, Cambridge, Massachussetts.

Mayr E. 1985. The species as a category, taxon and populations. *In* S Atran *et al*. (Eds.) *Histoire du Concept d'Espéce dans les Sciences de la Vie*. Fondation Singer - Polignac, Paris.

Mayr E. 1997. *This is Biology. The Science of the Living World*. Harvard University Press, Cambridge.

Mayr E, Ashlock PD.1991. Principles of Systematic Zoology, 2nd edit., McGraw-Hill, Inc, New York

Monsen KJ, Blouin M. 2003. Genetic structure in a montane ranid frog restricted gene flow and nuclear-mitochondrial discordance. *Molecular Ecology* **12**: 3275-3286.

Monteiro A, Pierce NE. 2001. Phylogeny of Bicyclus (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1 a gene sequences. *Molecular Phylogenetics and Evolution* **18**: 264-281.

Monteiro FA, Donnelly MJ, Beard CB, Costa J. 2004. Nested clade and phylogeographic analyses of the chagas disease vector *Triatoma brasiliensis* in northeast Brazil. *Molecular Phylogenetics and Evolution* **32**: 46-56.

Moritz C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology* **3**: 401-411.

Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**: 269-292.

Moriyama EN, Powell JR. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *Journal of Molecular Evolution* **45**: 378-391

Newton AC, Allnutt ACM, Gillies AJ, Ennos RA. 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology & Evolution* **14**(4): 140-145.

Nichols R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16**: 358-364.

Oberdörster U, Grant PR. 2007. Acoustic adaptations of periodical cicadas (Hemiptera: Magicicada). *Biological Journal of the Linnean Society* **99**: 15-24.

Paterson HEH. 1978. More evidence against speciation by reinforcement. *South African Journal of Science* **74**: 369-371.

Paterson HEH. 1980. A comment on mate recognition systems. *Evolution* **34**: 330- 331.

Paterson HEH. 1981. The continuing search for the unknown and unknowable: a critique of contemporary ideas on speciation. *South African Journal of Science* **77**: 113-119.

Paterson HEH. 1985. The recognition concept of species. *In* ES Vrba (Ed.) *Species and Speciation*. Transvaal Museum, Pretoria, pp. 21–29.

Paulo OS, Dias C, Bruford MW, Jordan WC, Nichols RA. 2001. The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **268**: 1625-1630.

Piganeau G, Gardner M, Eyre-Walker A. 2004. A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution* **21**: 2319-2325.

Popov AA. 1975. The structure of the tymbals and the characteristics of the sound signals in singing cicadas (Homoptera, Cicadidae) in the southern regions of the USSR. *Entomological Review* **54**: 7-35.

Price BW, Barker NP, Villet MH. 2007. Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology*, **16**: 2574-2588.

Pringle JWS. 1954. A physiological analysis of cicada song. *Journal of Experimental Biology* **31**: 525-556.

Quartau JA. 1983a. A comparison of different numerical methods applied to the taxonomy of *Batracomorphus* Lewis (Insecta, Homoptera, Cicadellidae, Iassinae): Ordination techniques. *Journal of the Entomological Society of Southern Africa,* **46** (1): 9-35.

Quartau JA 1983b. A numerical classification of the Afrotropical species of the genus *Batracomorphus* Lewis (Homoptera, Cicadellidae). *Arquivos do Museu Bocage* (sér. A), **II (4)**: 41-59.

Quartau JA. 1988. A numerical taxonomic analysis of interspecific morphological differences in two closely related species of *Cicada* (Homoptera, Cicadoidea) in Portugal. *Great Basin Naturalist Memoirs* **12**: 171-181.

Quartau JA. 1995. Cigarras esses insectos quase desconhecidos. *Correio da Natureza* **19**: 33-38.

Quartau JA Boulard M. 1995. *Tettigetta mariae* n. sp., nouvelle Cigale lusitanienne (Homoptera, Cicadoidea, Tibicinidae). *École Pratique des Hautes Études, Biologie et Évolution des Insectes*, **7/8**: 105-110.

Quartau JA, Fonseca PJ. 1988. An annotated check-list of the species of cicadas known to occur in Portugal (Homoptera: Cicadoidea). *In Proceedings of the 6th Auchenorryncha Meeting*, Turin, Italy pp: 367-375.

Quartau JA, Rebelo MT. 1994. Sinais acústicos em Cicadidae e Cicadellidae (Homoptera, Auchenorryncha) que ocorrem em Portugal. *In* V Almada & R Correia (Eds.) *Biologia e Comportamento*, *Actas do I Congresso Nacional de Etologia*, Lisboa 28, 29 e 30 de Novembro de 1991, Publicado por Instituto Superior de Psicologia Aplicada, Lisboa, pp: 137-142.

Quartau JA, Simões PC. 2005. *Cicada cretensis* sp. n. (Hemiptera, Cicadidae) from southern Greece. *Biologia Bratislava* **60**(5): 489-494.

Quartau JA, Simões PC. 2006. Acoustic evolutionary divergence in cicadas: the species of *Cicada* L. in Southern Europe. *In*: S Drosopoulos & MF Claridge (Eds). *Insect sounds*

and communication. *Physiology, Behaviour, Ecology and Evolution*. Taylor & Francis Group, Boca Raton, Chapter 17, pp. 227–237.

Quartau JA, Ribeiro M, Simões PC, Coelho MM. 2001. Genetic divergence among populations of two closely related species of *Cicada* Linnaeus (Hemiptera: Cicadoidea) in Portugal. *Insect Systematics & Evolution* **32**: 99-106.

Quartau JA, Ribeiro M, Simões PC, Crespo A. 2000. Taxonomic separation by isozyme electrophoresis of two closely related species of *Cicada* L. (Hemiptera: Cicadoidea) in Portugal. *Journal of Natural History* **34**: 1677-1684.

Quartau JA, Rebelo MT, Simões PC, Fernandes TM, Claridge MF, Drosopoulos S, Morgan JC. 1999. Acoustic signals of populations of *Cicada orni* L. in Portugal and Greece (Homoptera: Auchenorrhyncha: Cicadomorpha: Cicadidae). *Reichenbachia*, 33: 71–80.

Quartau JA, Coelho MM, Viegas-Crespo AM, Rebelo AM, Ribeiro M, Pinto GA, Simões P, André G, Claridge M, Hemingway J, Morgan JC, Drosopoulos S. 1997. A behavioural, molecular and morphometric approach to population divergence in cicadas (Homoptera, Auchenorrhyncha): some introductory results. *9th International Auchenorrhyncha Congress, Sydney, 17-21 February 1997*, pp.: 47-48 (abstract).

Ramos-Onsins S, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**: 2092-2100.

Richly E, Leister D. 2004. NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21**: 1081-1084.

Roderick GK. 1996. Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual Review of Entomology* **41**: 325-352.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**(3): 552-569.

Rokas A, Ladoukakis E, Zouros E. 2003. Animal mitochondrial recombination revisited. *Trends in Ecology & Evolution* **18**(8): 441-417.

Rubinoff D, Holland BS. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54**(6): 952-961.

Seabra SG. 2007. Patterns of divergence in populations of two Mediterranean species of genus *Cicada* L. (Hemiptera, Cicadidae) based on microsatellite genetic markers and acoustic data. *PhD Thesis*, Faculdade de Ciências da Universidade de Lisboa.

Seabra S, Simões PC, Drosopoulos S, Quartau JA. 2000. Genetic variability and differentiation in two allopatric species of the genus *Cicada* L. (Hemiptera, Cicadoidea) in Greece. *Mitteilungen aus dem Museum für Naturkunde in Berlin - Deutsche Entomologische Zeitschrift* **47**(2): 143-145.

Shaw KL. 2000. Interspecific genetics of mate recognition: inheritance of female acoustic preference in Hawaiian crickets. *Evolution* **54**: 1303-1312.

Shaw KL, Herlihy DP. 2000. Acoustic preference functions and song variability in the Hawaiian cricket *Laupala cerasina*. *Proceedings of the Royal Society B: Biological Sciences* **267**: 577-584.

Shaw KL, Parsons YM, Lesnick SC. 2007. QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*. *Molecular Ecology* **16**: 2879-2892.

Simmons RB, Weller SJ. 2001. Utility and evolution of cytochrome *b* in insects. *Molecular Phylogenetics and Evolution* **20**: 196-210.

Simões PC, Boulard M, Rebelo MT, Drosopoulos S, Claridge MF, Morgan JC, Quartau JA. 2000. Differences in the male calling songs of two sibling species of *Cicada* (Hemiptera: Cicadoidea) in Greece. *European Journal of Entomology* **97**(3): 437–440.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach A. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics* **37**: 545–79.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annales of the Entomological Society of America* **87**(6): 651–701.

Simon C, Nigro L, Sullivan J, Holsinger K, Martin A, Grapputo A, Franke A, Mcintosh C. 1996. Large differences in sustitutional pattern and evolutionary rate of 12S ribosomal RNA genes. *Molecular Biology and Evol*ution **13**(7): 923–932.

Simpson GC. 1951. The species concept. *Evolution* **5**: 285-298.

Simpson GC. 1961. *Principles of animal taxonomy*. Columbia University Press, New York.

Sokal RR, Crovello T. 1970. The biological species concept: a critical evaluation. *American Naturalist* **104**: 127–153.

Suer J, Vanderpool D, Simon C, Ouvrard D, Bourgoin T. 2007. Molecular phylogeny of the genus Tibicina (Hemiptera, Cicadidae): rapid radiation and acoustic behaviour. *Biological Journal of the Linnean Society* **91**: 611-626.

Szalay FS, Bock WJ. 1991. Evolutionary theory and systematics: relationships between process and patterns. *Zeitschrift fuer Zoologische Systematik und Evolutionsforschung* **29**: 1-39.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453-464.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**: 70-74.

Templeton AR. 1994. The role of molecular genetics in speciation studies. *In* B Schierwater, B Streit, GP Wagner & R DeSalle (Eds) *Molecular Ecology and Evolution: Approaches and Applications*. Birkhäuser Verlag Basel, Switzerland, pp. 455-477.

Tinbergen N. 1951. *The study of instinct*. Oxford University Press, Oxford.

Villet M. 1992. Responses of free-living cicadas (Homoptera: Cicadidae) to broadcasts of cicada songs. *Journal of the Entomological Society of Southern Africa*, **55**: 93-97.

Wells MM, Henry CS. 1998. Songs, reproductive isolation, and speciation in cryptic species of insects: a case study using green lacewings. *In* DJ Howard & SH Berlocher (Eds) *Endless Forms: Species and Speciation*. Oxford, Oxford University Press pp: 217-233.

Widmer A, Lexer C. 2001. Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution* **16**(6): 267-269.

Wiley EO. 1978. The evolutionary species concept reconsidered. *Systematic Zoology* **27**: 17-26.

Wilson EO. 1992. *The diversity of life*. The Penguin Press, London.

Chapter 2.

Mitochondrial DNA variation and the evolutionary history of the Mediterranean species of *Cicada* L. (Hemiptera, Cicadoidea)

Accepted for publication on 03/10/2007: Pinto-Juma GA, Quartau JA, Bruford MW. 2007. Mitochondrial DNA variation and the evolutionary history of the Mediterranean species of *Cicada* L. (Hemiptera, Cicadoidea). *Zoological Journal of the Linnean Society* in press

Chapter 2.

Mitochondrial DNA variation and the evolutionary history of the Mediterranean species of *Cicada* L. (Hemiptera, Cicadoidea)

G. A. PINTO-JUMA, [1,2‡] J. A. QUARTAU [1] & M. W. BRUFORD [2]

(1) Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Portugal

(2) School of Biosciences, Cardiff University, Cardiff CF10 3US, Wales, U.K.

(‡) To whom correspondence should be sent: Gabriela A. Pinto Juma

Present address: Cardiff University, School of Biosciences, Park Place, Cardiff CF10 3US, U.K.

Tel: +44 (0)2920875776. Fax: +44 (0)2920874305. Electronic mail: jumag@cf.ac.uk

Abstract

Cicadas are widely distributed in the Mediterranean area and are mainly distinguished by male acoustic signals, which act as specific mate recognition systems. Within the genus *Cicada* Linnaeus several species have diverged in their calling songs without showing external morphological differences, so acoustic recordings and genetic studies are particularly useful for systematic, biogeography and evolutionary studies. This study assesses sequence variation in closely related species of *Cicada* from the Mediterranean area, using domain III of the 12SrRNA mitochondrial gene in order to determine the phylogenetic relationships and the evolutionary history of this group, as well as the population structure of the two most common species, *C. orni* and *C. barbara*. Five distinct haplogroups were identified as *C. orni*, *C. barbara*, *C. mordoganensis*, *C. cretensis* and *C. lodosi*, each corresponding to a distinct evolutionary group. *C. barbara* was the most divergent species within this group, while *C. orni* and *C. mordoganensis* were the most similar. The population structure and demographic parameters of the species were not completely resolved. However, there is evidence for the separation of the *C. orni* Greek populations from the rest of Europe and also for demographic expansions probably related to Pleistocene climate changes.

Key words: Cicadoidea, 12sRNA gene, phylogeny, Mediterranean.

Introduction

Cicadas are hemipterous insects characterized by the ability of males to produce loud and distinctive acoustic signals during pair formation and courtship, using a tymbal mechanism (*e.g.,* Bennet-Clark, 1998). The females are only attracted to the calls of conspecific males, therefore these acoustic signals are thought to act as specific mate recognition systems (SMRSs) that ensure the mating of conspecifics, resulting in reproductive isolation and genetic divergence (*e.g.*, Paterson, 1985; Claridge, 1985, 1995; Quartau & Rebelo, 1994; Quartau, 1995). Within the genus *Cicada* Linnaeus, there are several instances where species have diverged in their calling songs, without showing clear-cut external morphological differences, *e.g. C. barbara* Stål, 1866, *C. orni* Linnaeus, 1758, *C. mordoganensis* Boulard, 1979 and *C. cretensis* Quartau & Simões, 2005 (Boulard, 1982; Quartau *et al.,* 2000, 2001; Simões *et al.,* 2000; Seabra *et al.,* 2000; Quartau & Simões, 2005a).

In the Mediterranean area, six cicada species are currently recognised based on acoustic divergence and some eventual genitalia differences. *Cicada orni* is one of the commonest and most widely distributed cicadas in Europe. In contrast, *C. barbara*, which is widely distributed in northwestern Africa only occurs in a few scattered warm habitats in Portugal and Spain, where it may occur in sympatry with *C. orni*. *C. mordoganensis* is known to occur in Turkey and also on a few Greek islands close to the Turkish mainland (Boulard, 1979; Simões *et al.,* 2000). Two other species, *C. lodosi* Boulard, 1979 and *C. permagna* Haupt, 1917, are known to occur in Turkey (Boulard, 1979) while *C. cerisyi* Guérin-Méneville, 1844, a species of doubtful status, has been described from Egypt. *C. cretensis* is thought to be endemic to Crete and has recently been described by Quartau & Simões (2005a) on the basis of the distinct male acoustic signals. These species have a similar ecology, being mainly known to occur in closed high shrubland and woodland and being associated with olive, pine, oak and introduced eucalyptus.

At the molecular level, previous allozyme studies revealed that three of 19 loci were diagnostic for the separation of *C. orni* and *C. barbara* in Portugal (Quartau *et al.*, 2000), however genetic distance estimates between this species pair were very low and comparable to the subspecies level in well-studied insects, such as *Drosophila* (see Quartau *et al.*, 2001). On the other hand, similar comparisons between *C. orni* and *C. mordoganensis* did not reveal any diagnostic loci (Seabra *et al.,* 2000). Since the male calls of the first species pair are quite

distinct, this could indicate that pre-mating isolation might have occurred with little genetic divergence, *i.e.*, acoustic divergence could have proceeded more rapidly than neutral genetic divergence. In *C. orni* and *C. barbara*, no indication of hybridization has been found so far, even in areas of sympatry (Quartau *et al.*, 2001), although a much higher allozyme genetic diversity was found in *C. orni* than in *C. barbara*. The comparatively low genetic variability of *C. barbara* suggests that it have been a relatively recent immigrant to the Iberian Peninsula from North Africa (Quartau *et al.*, 2001). With the exception of these allozyme studies no other molecular studies have been carried out. The degree of genetic differentiation among these species is still mostly undetermined and their phylogenetic relationships are unknown.

Evolutionary biologists aim to recognize taxa as evolutionary entities. According to Hey (2001), taxa and evolutionary groups may correspond when all organisms placed in a category collectively constitute an evolutionary group. Crandall *et al.* (2000) noted that both genetic and ecological information should be used to define evolutionary significant units and that acknowledging historical population structure, as defined by molecular genetic techniques, is an imperative in such studies. If reliable phylogenies are produced, they may clarify the evolutionary events that generated present day diversity in genes and species and help us to understand the mechanisms of evolution as well as the history of organisms (Kumar *et al.,* 2001).

Cicadas have a particular mating behaviour associated with low morphological differentiation and wide distribution providing excellent models for speciation and evolutionary biology studies. The present study assesses sequence variation in cicadas of genus *Cicada* from the Mediterranean area, using domain III of the 12S rRNA mitochondrial gene, to test if currently recognised species constitute divergent evolutionary units, corroborating the role of reproductive isolation by the SMRSs. We also aimed to determine the phylogenetic relationships and the evolutionary history of this group in the Mediterranean area, focusing in the population structure of two occasionally sympatric species, *C. orni*, the most widely spread species in Europe within this genus, and *C. barbara*, a species widely distributed in North Africa but thought to be recently introduced in the Iberian Peninsula from North Africa. In addition, we intended to develop a rapid and low cost method for screening different haplotypes in large samples, within and between these two species, using Restriction Fragment Length Polymorphism (RFLP) and Single-Stranded Conformation Polymorphism (SSCP). Finally, we assess the usefulness of this mitochondrial gene, for determining relationships or variability at the population and species levels.

Material and Methods

Collection of specimens

Adult male cicadas of genus *Cicada* were collected during the summers of 1995-2002. In most cases males were located and identified in the field through their calling songs, then collected either by hand or using nets and preserved in liquid nitrogen, in 96% ethanol or kept dry. Specimens were collected at different localities in the Mediterranean area focusing on Portugal (the most westerly point of the distribution of *C. orni* and *C. barbara)* (Fig. 1). *C. orni* and *C. barbara* were also collected from Spain (including Algeciras, the most south-westerly point of the distribution of *C. orni*), with specimens of the latter species also sampled from North Africa (Ceuta and Morocco). Specimens from France, Croatia, Slovenia and Macedonia (*C. orni*), Greece (*C. orni* and *C. mordoganensis*), Crete (*C. cretensis*) and Turkey (*C. lodosi*) were obtained by donation. Males of *Lyristes plebejus* (Cicadidae) and *Tettigetta josei* and *Tympanistalna gastrica* (Tibicinidae) collected in Portugal were used as outgroups in the phylogenetic analyses.

DNA extraction, PCR and sequencing

Total DNA was isolated either from a small section of the thoracic muscle (frozen specimens) or from one leg (ethanol preserved and dried specimens). Each tissue sample was homogenised and DNA extracted using one of three methods: (i) phenol-chloroform (Pinto *et al*. 1998), (ii) salt precipitation (Livak, 1984), or (iii) the QIAamp DNA mini kit (QIAgen) following the manufacturer's protocol. The resulting pellet was suspended in H_2O and visualised on 0.8% ethidium bromide stained agarose gels by UV-illumination.

Total DNA extract was used as template for the amplification of an approximately 390 bp fragment of the mitochondrial ribosomal subunit 12sRNA gene by PCR using the primers SR-J-14233 and SR-N-14588 (numbers refer to the *Drosophila yakuba* sequence location, Simon *et al.,* 1994). PCR reactions were performed in 25µl containing 1x *Taq* buffer, 1.25mM $MgCl_2$, 0.2mM of dNTPs, 50pmol of each primer, 1unit of *Taq* polymerase (Invitrogen) and approximately 50ng of genomic DNA. The amplification conditions comprised 3min at 95ºC, followed by 30 cycles of 95ºC for 30s, 52ºC for 30s and 72ºC for 45s, with a final extension cycle of 72ºC for 7min. Negative controls were run simultaneously. PCR product sizes were estimated from 1.4% agarose gels and purified using the Geneclean turbo for PCR kit (QIAgen, BIO101) following the manufacturer's protocol.

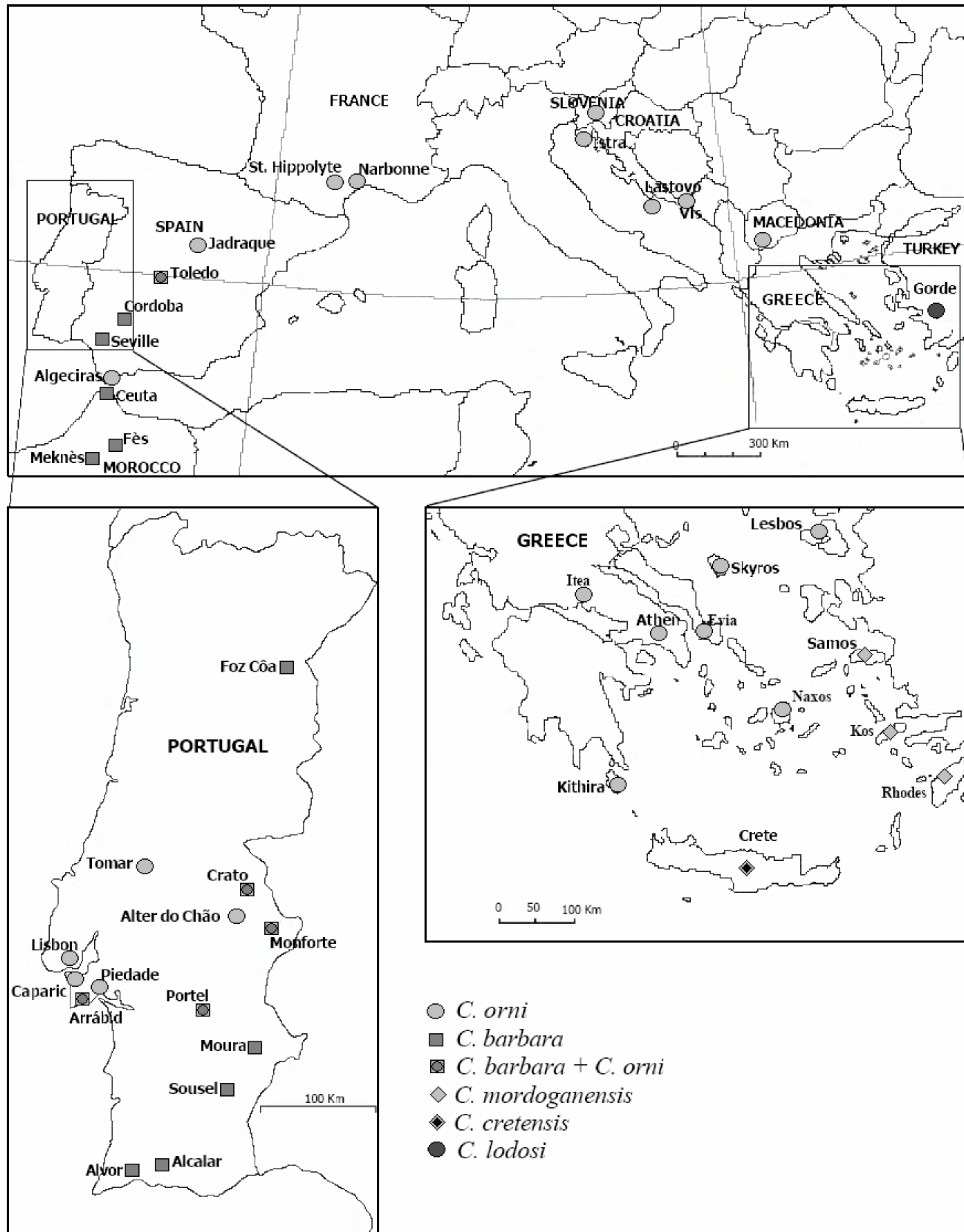


Fig. 1. Collection sites of the *Cicada* specimens analysed in the Mediterranean region: inserts are given for Portugal and Greece, where most sites were sampled.

Direct sequencing of purified double-stranded PCR products was firstly performed using an ABI PRISM™ dye terminator cycle sequencing-ready reaction kit with Ampli*Taq*® DNA polymerase (Perkin-Elmer), according to the manufacturer's instructions. Templates

were cycle sequenced with each PCR primer (1.6pmol) separately using standard cycling profiles on a Perkin Elmer GeneAmp PCR System 2400, then purified and electrophoresed in a Perkin Elmer ABI Prism™ - 377 DNA sequencer. Sequencing reactions were also performed using the ABI Big Dye version 2 with Better Buffer (WebScientific Ltd.) and sequences were separated in a Perkin Elmer 3100 DNA sequencer. All samples were sequenced at least once using both DNA strands to avoid ambiguities.

Sequence alignment

Chromatogram analysis was performed using Chromas version 1.4 (Conor McCarthy). Editing and alignment of the DNA sequence data were performed using DNAstar (Lasergene, USA), Editseq and MegAlign, respectively. Sequence alignments were made through the CLUSTAL routine in MegAlign with adjustments made by hand, and individual haplotypes were identified. The secondary structure prediction for the sequences obtained for the 12S rRNA was drawn manually based on alignments with conserved sequence motifs and the secondary structure model for the third domain of 12S rRNA for cicadas (see Hickson *et al*., 1996). The secondary structure of part of domain II was also determined manually (see Page, 2000).

Alternative methods for haplotype detection (RFLP and SSCP)

Alternative methods were tested for rapid identification of haplotypes in large samples. The sequences obtained for five specimens of each locality from *C. barbara* and *C. orni* were used to produce a restriction map to find restriction sites capable of separating both species. The most unequivocal restriction site identified was cut by the enzyme *Nla*III (New England Biolabs) (recognition site: 5'…CATG▼…3' / 3'…▲GTAC…5'). RFLPs were performed in 20µl reactions including 18µl of PCR product, 1x BSA (New England Biolabs), 1U *Nla*III and 1x NEBuffer 4 (New England Biolabs) by digesting at 37ºC for 1h, followed by an inactivation step at 65ºC for 5-10min. RFLP fragments were analysed directly in 2% agarose gels. This method was used to distinguish between species, but was incapable of distinguishing among haplotypes within species, so SSCP (Single-Stranded Conformation Polymorphism) was implemented to detect single mutations. For SSCP, vertical polyacrylamide gels were run at 4ºC: 40ml gels were made using 0.5x MDE (Gel solution 2x concentrate from Biowhittaker Mol. Ap.), 0.6x TBE, 20mM HEPES, 240µl 10% APS, 24µl TEMED, 5% glycerol for *C. orni* specimens or 10% glycerol for *C. barbara* specimens and dH_2O to 40ml. HEPES (20mM) was always added to 0.6x TBE buffer. Four microlitres of

each PCR sample was mixed with 1µl dye mix (0.5% Bromophenol blue and Xylene cyanol solution) and 5µl of SSCP mix (1 ml formamide, 1µl 10M NaOH, 40µl 0.5M EDTA), this mixture was denatured at 95ºC for 2min, immediately placed on ice and then loaded into the gels. Runs were carried out for 14h at 300V for *C. orni* and for 20h at 400V for *C. barbara*. The gels were dyed using a standard silver staining protocol described by Jordan, Foley & Bruford (1998) and then visualized under white light and photographed. Reference samples were run simultaneously with new samples to assist in identification of haplotypes. All samples showing a new pattern were sequenced; random samples were also sequenced to confirm correct identification of haplotypes.

Phylogenetic analysis and population genetics

Haplotypes detected by sequencing and SSCP were analysed using several population and phylogenetic analyses. Genetic diversity and nucleotide diversity indices within all sequences, stems and loops per locality and per species, were analysed using ARLEQUIN 2.000 (Schneider, Roessli & Excoffier, 2000). Modeltest vs. 3.6 (Posada & Crandall, 1998) was used to determine the most appropriate model of nucleotide substitution. Using the Akaike Information Criterion test (AIC) (see Posada & Buckley, 2004) the most likely model of DNA evolution for the data was K81uf+I (Kimura, 1981) with unequal base frequencies and a proportion of invariable sites of I = 0.8116. DAMBE vs.4.2.13 (Xia & Xie, 2001) was used to plot transition and transversion variation over the K81 sequence divergence obtained within *Cicada* specimens. Minimum spanning network was drawn manually representing the most parsimonious connections/relationships between the OTUs (unique individual haplotypes). Connection lengths between haplotypes were based on the number of pairwise differences using the adaptation of the algorithm described in Rohlf (1973) implemented by ARLEQUIN. The model of nucleotide substitution chosen by Modeltest was applied to the phylogenetic and molecular evolutionary analyses conducted by MEGA version 2.1 (Kumar *et al.,* 2001) and PAUP. A Neighbor-Joining (NJ) tree was computed using MEGA based on the K81 (I = 0.8116) distance matrix estimated by PAUP beta version (Swofford, 1999), and a maximum likelihood tree based on this model was obtained by PAUP, using an heuristic search (10 random addition-sequence replicates and tree-bisection-reconnection (TBR) swapping algorithm). For both phylogenetic methods, trees were rooted using *Lyristes plebejus*, *Tettigetta josei* and *Tympanistalna gastrica*, and to test the nodal support of the inferred trees bootstrap resampling using 1000 and 250 replicates for the NJ and ML analyses, respectively.

Analysis of molecular variance (AMOVA) was used to examine mtDNA variation within and among populations of *C. orni* and *C. barbara* separately. This, produced estimates of variance components and *F*-statistic analogs, designated as Φ-statistics, reflecting the correlation of haplotypic diversity at different levels of hierarchical subdivision (Excoffier, Smouse & Quattro, 1992). The significance of the variance components associated with the different possible levels of genetic structure (within populations, within groups of populations and among groups) was tested using non-parametric permutation procedures (1000 permutations). Several partitioning tests were made for each species where samples from each locality were considered different populations and these were grouped into different regions according to their geographical proximity and/or the presence of unusual haplotypes. Regions with the highest significant Φ_{CT} value in an AMOVA should reflect the most probable geographical subdivisions (Excoffier *et al*., 1992). ARLEQUIN and DnaSP4 (Rozas *et al*., 2003) were both used to perform the Neutrality test Fu's (1997) F_s statistics, in order to detect population demographic expansions. This test has been shown to be more consistent in assessing population growth than mismatch statistics (Fu, 1997; Ramos-Onsins & Rozas, 2002; Mes, 2003). F_s examines low frequency alleles in an expanding population compared to the that expected in a stationary one. The expansion coefficient S/d was also calculated (S is the number of variable sites and d is the mean pairwise differences among sequences). Large S/d values may indicate a recent population expansion (showing a high frequency of unique sites in sequences) while small values indicate relatively constant long-term population sizes.

We attempted to estimate the date of demographic expansion (*t*) based on the expansion time in mutational units (τ) generated by ARLEQUIN and using the estimators described by Rogers & Harpending (1992), and Harpending (1994): $\tau = 2ut$ where *t* is the number of generations elapsed between initial population and current population and $u = 2\mu k$, where μ is the mutation rate per million years and *k* is the length of the sequence. Several rRNA gene evolutionary rates described in the literature for different taxa, including insects (e.g. Simon *et al.* 1996; Caccone *et al.* 1997; Alves-Gomes, 1999; Baratti, Khebiza & Messana, 2004; Satoh *et al.* 2004), varying from 0.07% to 2.3% per million years, were used.

Results

A total of 462 male cicadas were analysed: 242 specimens were directly sequenced (usually five per locality, except where five were not available and for Greek and Crete samples where all individuals were sequenced), the remaining 220 specimens (identified as *C.*

barbara and *C. orni*) were screened by RFLP and SSCP. The enzyme *Nla*III recognized one restriction site for *C. orni* specimens and none for *C. barbara* allowing 100% discrimination of these two species. An additional, 49 specimens were sequenced to confirm SSCP haplotype recognition or to identify new haplotypes. Using RFLP and SSCP, it was possible to increase the size of most of the *C. orni* and *C. barbara* samples analysed to 15 specimens per locality. The SSCP method accurately identified previously obtained haplotypes by sequencing (100% match) and detected five new haplotypes with only one false new pattern detected among the 220 samples analysed. The distribution of the haplotypes among sampled localities is shown in Table 1 (and Appendix 1). The different haplotypes were designated into haplogroups which corresponded to the species identified. The sequences analysed were confirmed to be exclusively mitochondrial by amplifying and sequencing the same fragment directly from purified mtDNA from 10 *C. orni* specimens and 30 *C. barbara* specimens (Pinto-Juma, unpublished data). In addition, the secondary structure features and well-conserved sequence motifs proved to be in agreement with previous descriptions for other insects (e.g. Hickson *et al.,* 1996; Simon *et al.,* 1996; Page, 2000).

The secondary structure described for domain III of the sequences analysed (Appendix 2) did not differ substantially from the model for cicada found by Hickson *et al.* (1996). Consensus sequences were compiled for each haplogroup, with degeneracy/ambiguity codes representing variable nucleotide positions within the structure (Fig. 2). Some insertion/deletions were observed after helix 38, increasing the complexity of alignment for the following helices. Helix 42 revealed most of the stem differences between the haplogroups analysed.

Nucleotide substitutions

Nucleotide sequences were A+T rich (approximately 75%) as commonly observed for insects (Hickson *et al.* 1996). A Chi-square (χ^2) test of homogeneity of base frequencies across taxa did not show significant differences ($\chi^2 = 6.15$, P = 0.999).

Over the 394 nucleotides analysed for all *Cicada* haplogroups, 345 characters were constant and 49 were variable (when including outgroups: 277 were constant and 117 were variable). Within each haplogroup, only *C. orni* showed transversions (Tv) (Table 2), although transitions (Ti) outnumbered these by 2.9:1. In the *C. barbara* haplogroup indels were also detected.

Table 1. Summarized distribution of 12S rRNA gene (~390bp) haplotypes among *Cicada* species from the Mediterranean area (Pt-Portugal, Sp-Spain, Mr-Morocco, Fr-France, Ct- Croatia, Md- Macedonia, Sl-Slovenia, Gr-Greece, Cr-Crete, Tk-Turkey; HCo stands for *C. orni*, HCm for *C. mordoganensis*, HCc for *C. cretensis* and HCb for *C. barbara* haplogroups, HCl stands for *C. lodosi*)

Haplogroup	Haplotype	Frequency	Localities
C. orni	HCo1	**140**	Iberian Peninsula, France, Croatia, Macedonia and Slovenia
	HCo2	**41**	Greece, Caparica Pt and Lisbon Pt
	HCo3	**1**	Monforte Pt
	HCo4	**1**	Crato Pt
	HCo5	**1**	Portel Pt
	HCo6	**1**	St. Hippolyte Fr
	HCo7	**1**	St. Hippolyte Fr
	HCo8	**11**	Lesbos Gr and Skyros Gr
	HCo9	**3**	Skyros Gr
	HCo10	**11**	Naxos Gr
	HCo11	**4**	Kithira Gr
	HCo12	**1**	Kithira Gr
C. mordoganensis	HCm1	**25**	Kos Gr, Rhodes Gr and Samos Gr
	HCm2	**1**	Samos Gr
	HCm3	**1**	Kos Gr
	HCm4	**1**	Kos Gr
C. cretensis	HCc1	**16**	Heraklion Cr
	HCc2	**1**	Heraklion Cr
C. barbara	HCb1	**176**	Iberian Peninsula and Morocco
	HCb2	**1**	Toledo Sp
	HCb3	**1**	Fès Mr
	HCb4	**3**	Mèknes Mr
	HCb5	**13**	Sevilla Sp
	HCb6	**1**	Moura Pt
	HCb7	**1**	Alcalar Pt
	HCb8	**1**	Foz Côa Pt
C. lodosi	HCl	**1**	Gordes Tk
Total		**459**	

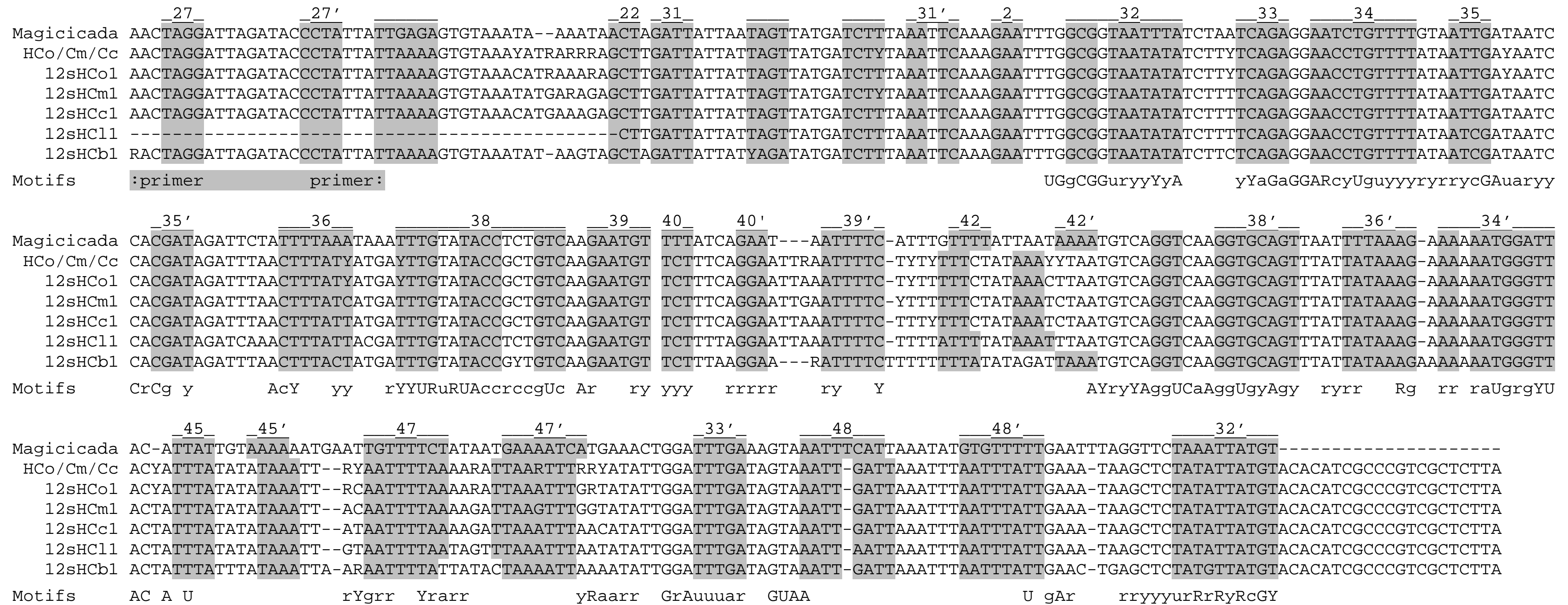


Fig. 2. Sequence alignment of the 12S rRNA fragment analysed in this study. The location of the stem regions is shaded: stem numbers (above the sequences) and the sequence motifs (below the sequences) are based on Hickson *et al.* (1996) and Page (2000). The sequence of *Magicicada* was obtained from Hickson *et al*. (1996). HCo/Cm/Cc is the consensus sequence for the *C. orni*, *C. mordoganensis* and *C. cretensis* haplogroups.

The Ti/Tv ratio decreased considerably when these haplogroups were compared to the *C. lodosi* sequence (Ti/Tv= 1.8 to 2.8) and with *C. barbara* (Ti/Tv= 0.6 to 1.1) (Table 2). The observed transversions between the haplogroups, including the outgroups, were mainly due to A↔T substitutions (82%), as observed in other insect species (*e.g.* Wolstenholme & Clary, 1985; Beckenbach, Wei & Liu, 1993).

Table 2. Nucleotide substitutions in an approximately 390 bp fragment of the 12S rRNA gene within and among *Cicada* haplogroups (loops/stems values are shown below the total fragment value).

		Transitions (Ti)		Transversions (Tv)				Proport.	Ti/Tv
	Taxa	CT	AG	GT	AT	CG	AC	differen.	Ratio
Within Haplogroups	*C. orni* haplogroup	1.5 1/1.3	1.4 1.4/0	0.3 0.3/0	0	0	0	0.009 0.011/0.01	2.9 2.2/-
	C. mordoganensis haplogroup	1 0.7/1	0.5 0.7/0	0	0	0	0	0.005 0.003/0.01	-
	C. cretensis haplogroup	1 1/0	0	0	0	0	0	0	-
	C. barbara haplogroup	0.5 0.3/0.7	1.3 1.1/0.7	0	0	0	0	0.007 0.01/0.01	-
Among Haplogroups	*C. orni / C. mordoganensis*	4.6 4/1.2	3.1 2.2/1	0.2 0.2/0	0	0	0	0.02 0.03/0.013	7.3 5.7/-
	C. orni / C. cretensis	6.4 5.3/1.3	2.8 2.8/0	0.2 0.2/0	0	0	0	0.025 0.038/0.01	8 7.5/-
	C. orni / C. lodosi	9 6.8/2	3 2.6/1	1 1.1/0	4 4/0	0	0	0.05 0.086/0.017	2.4 1.9/-
	C. mordoganensis / C.cretensis	5 3.8/1.5	4.3 3.3/1	0	0	0	0	0.024 0.033/0.015	-
	C. mordoganensis / C. lodosi	8 6.3/2.5	6 4/2	1 1/0	4 4/0	0	0	0.06 0.09/0.025	2.8 2.1/-
	C. cretensis / C. lodosi	7 6.5/1	2 1/1	1 1/0	4 4/0	0	0	0.04 0.075/0.010	1.8 1.5/-
	C. orni / C. barbara	9.1 3.8/4.3	7.3 6.2/1.3	1.8 1.8/0	9.1 5.1/3	0.9 0.9/0	3.1 3.1/0	0.082 0.107/0.046	1.1 0.9/1.9
	C. barbara / C. mordoganensis	7.9 2.5/4.8	8.6 5.9/2.3	2 2/0	9 5/3	0.9 0.9/0	3.1 3.1/0	0.081 1.1/0.053	1.1 0.8/2.4
	C. barbara / C. cretensis	7.9 3.6/3.3	4. 7 3.6/1.3	2.9 2.9/0	8.1 4.1/3	0	4 4/	0.071 0.094/0.04	0.8 0.7/1.6
	C. barbara / C. lodosi	5 3.1/2.3	5 3.1/2	3 2.9/0	12 9.1/3	0	1 1/0	0.08 0.113/0.043	0.6 0.5/1.4

Figure 3 plots transition and transversion variation against level of divergence for the total length of the 12S rRNA fragment and for the loops and stems separately. The most accentuated saturation of transitions with divergence level was observed for the loops, indicating the occurrence of multiple hits and a higher evolutionary rate. Also the loops revealed the highest number of transversions, outnumbering the transitions for increasing sequence divergence. In contrast, stems revealed a distinctive pattern of low evolutionary rate,

transitions always outnumbered transversions and there was no evidence of substitution saturation.

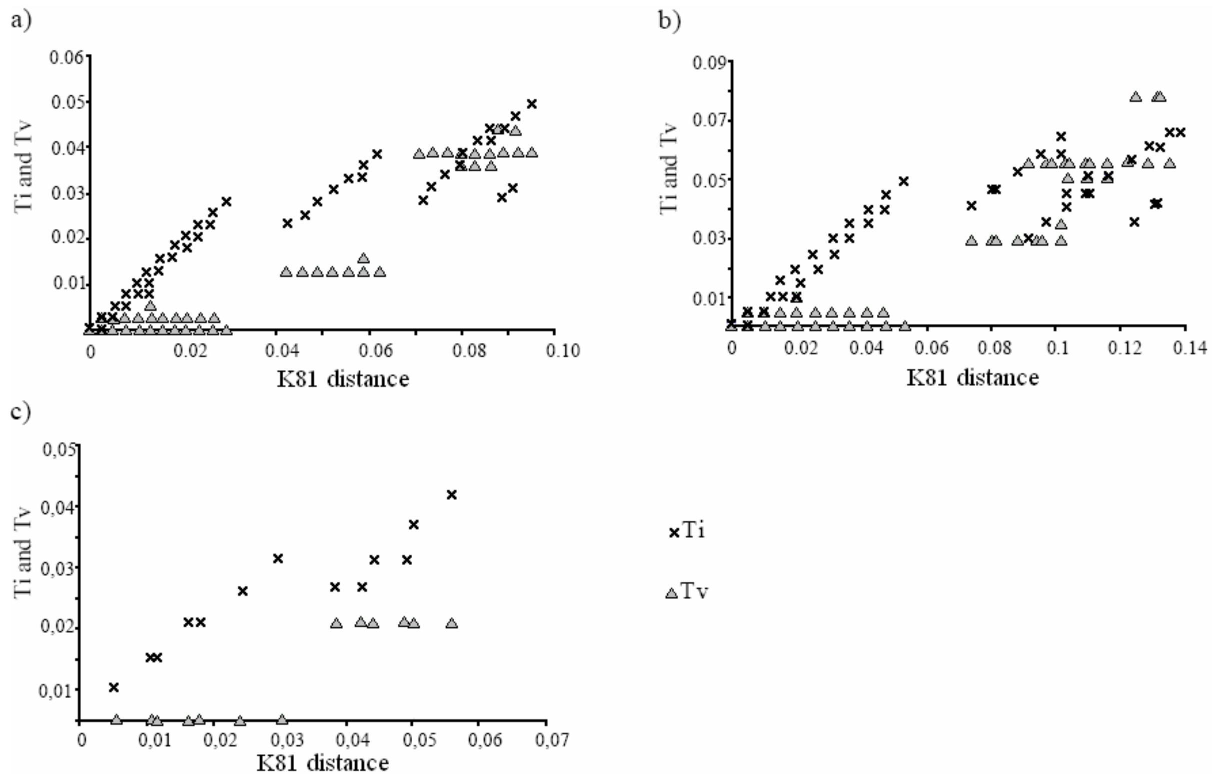


Fig. 3. Frequencies of transitions and transversions over K81 sequence divergences obtained among *Cicada* specimens for (a) the total length of 12S rRNA sequences; (b) loop regions based on the secondary structure of the 12S rRNA sequences; and (c) stem regions based on the secondary structure of the 12S rRNA.

Haplotype and genetic diversity

Sequencing and SSCP screening of the 12sRNA fragment revealed 27 haplotypes (Table 1), 12 for *C. orni* (216 specimens), 4 for *C. mordoganensis* (28 specimens), 2 for *C. cretensis* (17 specimens) and 8 for *C. barbara* (197 specimens). Only *C. orni* and *C. barbara* haplogroups were found in sympatry in the Iberia Peninsula.

All haplogroups possessed one haplotype more abundant, present in almost all sampling locations in their distribution. The only cases which presented sampled localities lacking a dominant haplotype occurred in *C. orni*: Caparica and Lisbon from Portugal, and no Greek samples showed HCo1, the most common and abundant haplotype from this haplogroup. With the exception of these localities, some samples had exclusive site specific haplotypes (Crato, Monforte and Portel samples from Portugal each had a single exclusive haplotype, and St. Hippolyte in France had two) in addition to haplotype HCo1. Greek samples revealed a different pattern: continental localities shared the HCo2 haplotype, but each island possessed unique haplotypes. In the *C. barbara* haplogroup, all localities shared

the predominant haplotype HCb1, however, Seville in Spain revealed a second, more abundant, haplotype (HCb5). Apart from HCb1, no other haplotypes were shared between different localities.

Table 3. 12S rRNA gene and nucleotide diversity of the cicada samples at each locality within the Mediterranean area (*D stands for deletion).

Localities	Haplo Group	No. of Haplot.	Polymorp. sites	Gene diversity	Ti	Tv	Mean no. of pairwise differences	Nucleotide diversity (%)
Alter do Chão Pt	orni	1	0	0	-	-	0	0
Arrábida Pt		1	0	0	-	-	0	0
Caparica Pt		1	0	0	-	-	0	0
Crato Pt		2	1	0.13 +/- 0.11	1	-	0.13 +/- 0.21	0.0003 +/- 0.0006
Lisbon Pt		1	0	0	-	-	0	0
Monforte Pt		2	1	0.20 +/- 0.15	1	-	0.20 +/- 0.27	0.0005 +/- 0.0008
Piedade Pt		1	0	0	-	-	0	0
Portel Pt		2	1	0.13 +/- 0.11	1	-	0.13 +/- 0.21	0.0003 +/- 0.0006
Tomar Pt		1	0	0	-	-	0	0
Algeciras Sp		1	0	0	-	-	0	0
Jadraque Sp		1	0	0	-	-	0	0
Toledo Sp		1	0	0	-	-	0	0
Narbonne Fr		1	0	0	-	-	0	0
St. Hippolyte Fr		3	2	0.26 +/- 0.14	2	-	0.27 +/- 0.31	0.0007 +/- 0.0009
Athens Gr		1	0	0	-	-	0	0
Evia Gr		1	0	0	-	-	0	0
Itea Gr		1	0	0	-	-	0	0
Croatia		1	0	0	-	-	0	0
Macedonia		1	0	0	-	-	0	0
Slovenia		1	0	0	-	-	0	0
Lesbos Gr		1	0	0	-	-	0	0
Kithira Gr		2	1	0.40 +/- 0.24	1	-	0.40 +/- 0.44	0.0010 +/- 0.0013
Naxos Gr		1	0	0	-	-	0	0
Skyros Gr		2	4	0.57 +/- 0.12	3	1	2.29 +/- 1.42	0.0059 +/- 0.0042
Kos Gr	mordoganensis	3	2	0.42 +/- 0.19	2	-	0.44 +/- 0.43	0.0011 +/- 0.0013
Rhodes Gr		1	0	0	-	-	0	0
Samos Gr		2	1	0.20 +/- 0.15	1	-	0.20 +/- 0.27	0.0005 +/- 0.0008
Crete	cretensis	2	1	0.18 +/- 0.10	1	-	0.18 +/- 0.20	0.0003 +/- 0.0006
Alcalar Pt	barbara	2	1	0.13 +/- 0.11	1	-	0.13 +/- 0.21	0.0004 +/- 0.0006
Alvor Pt		1	0	0	-	-	0	0
Arrábida Pt		1	0	0	-	-	0	0
Crato Pt		1	0	0	-	-	0	0
Foz Côa Pt		2	1	0.13 +/- 0.11	1	-	0.13 +/- 0.21	0.0003 +/- 0.0006
Monforte Pt		1	0	0.00	-	-	0.00	0.00
Moura Pt		2	1	0.13 +/- 0.11	1D*		0.13 +/- 0.21	0.0003 +/- 0.0006
Portel Pt		1	0	0	-	-	0	0
Sousel Pt		1	0	0	-	-	0	0
Ceuta Sp		1	0	0	-	-	0	0
Cordoba Sp		1	0	0	-	-	0	0
Sevilla Sp		2	2	0.25 +/- 0.13	2	-	0.50 +/- 0.45	0.0013 +/- 0.0013
Toledo Sp		2	1	1.00 +/- 0.50	1D*		1.00 +/- 1.00	0.0026 +/- 0.0036
Fès Mr		2	2	0.13 +/- 0.11	2	-	0.27 +/- 0.31	0.0007 +/- 0.0009
Mèknes Mr		2	2	0.34 +/- 0.13	2	-	0.69 +/- 0.55	0.0018 +/- 0.0016

As shown in Table 3, haplotype diversity within *C. orni* and *C. mordoganensis* ranged from one to three haplotypes per locality, however, most of the localities revealed only one haplotype for *C. orni* (75% of the localities). *Cicada cretensis* haplogroup had two haplotypes both occurring in Crete. In *C. barbara*, one to two haplotypes were found per locality, but the proportion of localities with one or two haplotypes was similar. *C. mordoganensis* revealed the highest average of genetic (h) and nucleotide (π) diversity per locality, 0.21 and 0.0005 respectively, followed by *C. barbara*, h= 0.14 and π=0.0005, and the mean values observed in *C. orni* haplogroup were the lowest: h=0.07 and π=0.0004. Nevertheless, the locality showing the highest diversity indices was Skyros, one of the Greek islands where only *C. orni* specimens were found.

Gene diversity within each haplogroup (ignoring localities) was considerably higher in *C. orni* (0.54) than within the other haplogroups (for *C. mordoganensis* gene diversity was 0.21, for *C. barbara* 0.20 and for *C. cretensis* 0.12). However, the gene diversity of *C. orni* declined considerably when excluding all the Greek samples: 0.24 (0.39 when excluding only the island samples). Conversely, Greek samples revealed the highest gene diversity (0.72).

Divergence among haplotypes

Figure 4 shows a minimum spanning network between all haplotypes. Five distinct haplogroups were detected and identified as *C. orni* (HCo), *C. barbara* (HCb), *C. mordoganensis* (HCm), *C. cretensis* (HCc) and *C. lodosi* (HCl). The haplogroup *C. barbara* was the most distinct, differing by at least 31 steps from the next closest haplotype (HCc1, from Crete). The single *C. lodosi* museum specimen also proved to be distinct from all other haplogroups, differing at 14 nucleotides when compared to HCc1. The remaining haplogroups were more similar, differing by less than seven nucleotides. The *C. orni* and *C. mordoganensis* haplogroups were the most similar and the *C. cretensis* haplogroup was very similar to a marginal haplotype of *C. orni*.

The average K81 distances observed within the *C. orni* haplogroup was 0.9% (ranged between 0.3% and 1.7%) (Table 4), while within the *C. cretensis*, *C. mordoganensis* and *C. barbara* haplogroups the mean distances was between 0.3 and 0.5%. The higher mean distance observed between the haplotypes of the *C. orni* haplogroup may be explained by the divergence of most Greek haplotypes. Between *C. orni* and *C. barbara* the average distance observed was 12.8%, which implies a clear divergence between these two haplotype groups. Similar distances were observed between *C. barbara* and the remaining *Cicada* haplogroups.

Fig. 4. Minimum spanning network for *Cicada* in the Mediterranean area, based on 12S rRNA haplotypes (HCo- *C. orni* haplogroup; HCm- *C. mordoganensis* haplogroup; HCc- *C. cretensis* haplogroup; HCl- *C. lodosi* and HCb- *C. barbara* haplogroup). The size of the circles representing each haplotype is proportional to the number of specimens showing that particular haplotype, the colours are representative of the locality.

Conversely, *C. orni* and *C. mordoganensis* proved to be very similar (2.3%); comparing these last haplogroups with *C. cretensis* both revealed 2.8 and 2.7% divergence, respectively. The single specimen of *C. lodosi* diverged 6.8% from *C. orni*, 7.9% from *C. mordoganensis* and only 5.2% from *C. cretensis*. The mean K81 divergence found for all the haplogroups within stems was 4.1% whilst within loops was 15.6%. The divergence in loops was usually significantly higher than in stems, for all levels of comparisons (within and between haplogroups).

Table 4. Mean Kimura 3 parameter (proportion of invariables I=0.8116) distance matrix, obtained by comparison of all pairs of haplogroups and sequences of cicadas analysed in this study, based on an approximately 394 bp fragment of the 12S rRNA gene. Outgroup comparisons are based on Kimura 2 parameter distances since K81 distances were undetermined. Within haplogroup mean divergence is shown on the diagonal (dark shaded cells), when applicable, and the intra and interspecific divergence range is shown below the mean.

	1	2	3	4	5	6	7
1 *C· orni*	0.009 (0.003-0.017)						
2 *C. mordoganensis*	0.023 (0.017-0.026)	0.004 (0.003-0.005)					
3 *C. cretensis*	0.028 (0.020-0.034)	0.027 (0.023-0.030)	0.003				
4 *C. lodosi*	0.068 (0.059-0.075)	0.079 (0.076-0.082)	0.052 (0.050-0.54)	-			
5 *C· barbara*	0.128 (0.099-0.142)	0.129 (0.118-0.142)	0.100 (0.092-0.109)	0.137 (0.131-0.141	0.005 (0.003-0.008)		
6 *L. plebejus*	0.219	0.217	0.203	0.229	0.216		
7 *T. josei*	0.165	0.171	0.155	0.156	0.177	0.181	
8 *T. gastrica*	0.195	0.201	0.184	0.182	0.194	0.220	0.107

L. plebejus outgroup had higher mean distances when compared to *Cicada* spp. than did *T. josei* or *T. gastrica* (approximately 22%, 17% and 19%, respectively, K80 distances, since for K81 these values were undetermined). These results are unexpected since the former species belongs to the same family as the genus *Cicada* (Cicadidae), while *T. josei* and *T. gastrica* belongs to a different family (Tibicinidae).

Phylogeny

Both trees obtained with the NJ and ML methods were very similar and showed reciprocal monophyly between the *C. barbara* haplogroup and all remaining haplogroups

(Fig. 5). Generally, bootstrap values obtained for the NJ tree were higher than those for ML, although the distinctiveness of the *C. barbara* haplogroup was strongly supported with 100% bootstrap values in both trees. *C. orni* and *C. mordoganensis* were the closest and most recently diverged haplogroups. The bootstrap value supporting the *C. mordoganensis* haplogroup was high, indicating strong support (96% NJ; 91% ML), however, the *C. orni* group did not possess high support (probably due to the higher variability among haplotypes within this haplogroup). The *C. cretensis* haplogroup was separated basally to these two haplogroups (96% NJ; 75% ML) and *C. lodosi* was even more basal.

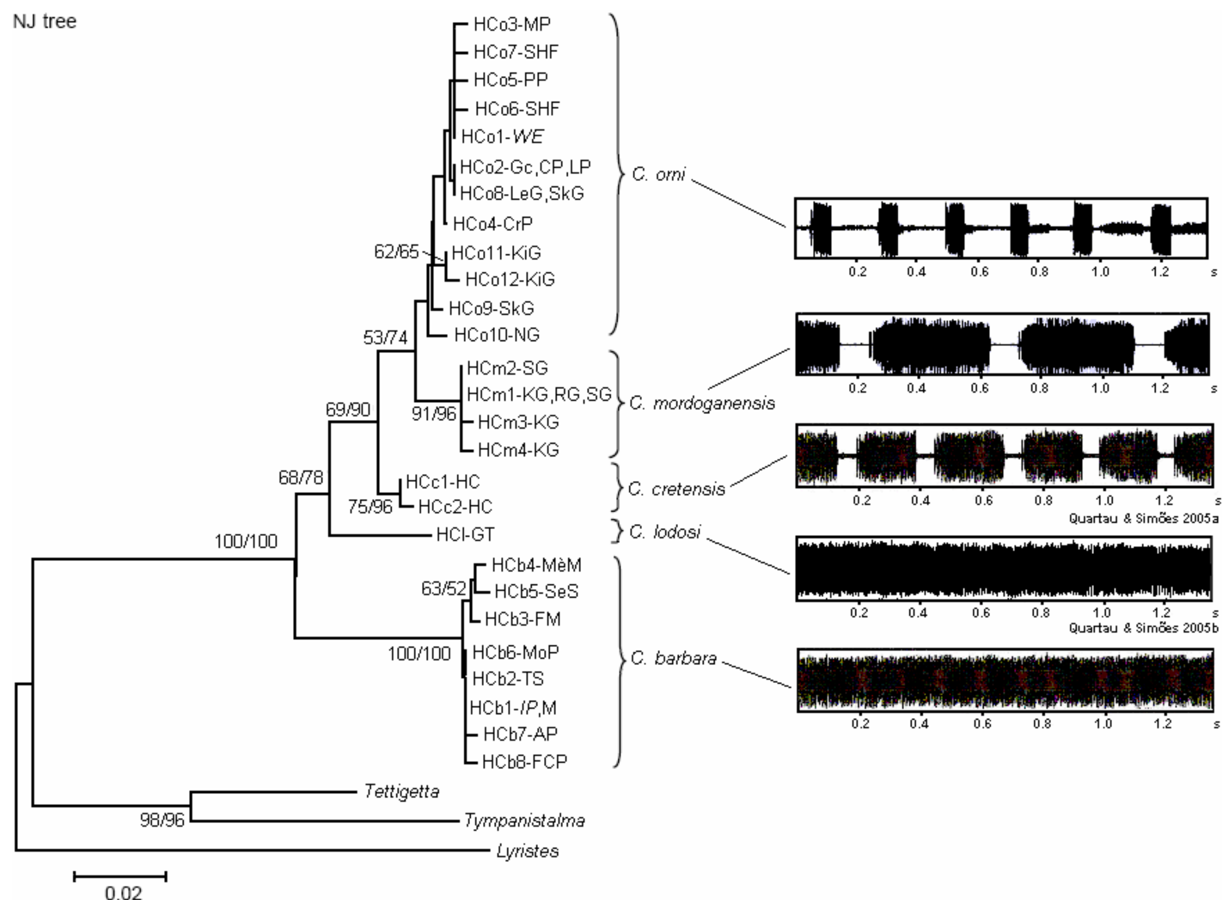


Fig. 5. Neighbour-Joining (NJ) phylogenetic tree (maximum likelihood (ML) consensus tree had the same topology so it is not shown) obtained for *Cicada* in the Mediterranean area, based on 12S rRNA haplotypes. 200 Bootstrap replicates for the ML tree and 1000 for the NJ tree were performed to assess the statistical significance of internal nodes (only bootstrap values > 50% are shown, ML bootstrap values followed by the NJ ones). For each clade, an oscillogram (time *vs*. amplitude) of the calling song of one male of *C. orni*, *C. mordoganensis*, *C. cretensis*, *C. lodosi and C*, *barbara* is shown. A locality code follows the haplotype: *WE*-Western Europe (including most Iberian Peninsula and France specimens but also Croatia, Macedonia and Slovenia), *IP*-Iberian Peninsula, *Gc*-Greece continental; in all other codes the last letter stands for country: P-Portugal, S-Spain, M-Morocco, F-France, G-Greece, C-Crete, T-Turkey, and the first letter(s) stands for locality: M-Monforte, SH-St. Hippolyte, P-Portel, , C-Caparica, L-Lisbon, Le-Lesbos, Sk Skyros, Cr-Crato, Ki-Kithira, N-Naxos, S-Samos, R-Rhodes, K-Kos, H-Heraklion, G-Gordes, Mè-Mèknes, Se-Sevilla, F-Fès, Mo-Moura, T-Toledo, A-Alcalar & FC-Foz Côa.

Each haplogroup clustered separately suggesting that they are independent evolutionary lineages. However, the within-haplogroup relationships were poorly defined, lacking consistent statistical support. This lack of resolution is an artefact of low variability and a lack of differentiation, as evidenced by short branch lengths in the NJ tree. Nonetheless, a 65-62% bootstrap value supported the separation of both haplotypes from Kithira (HCo11 and HCo12) and some *C. barbara* haplotypes from Mèknes, Fès and Seville grouped together with 52-63% bootstrap support.

Table 5. Hierarchical analysis of molecular variance (AMOVA) design and results for different tests of differentiation among groups.

a) Within *C. orni* and *C. barbara* haplogroups

	6 groups *C. orni*: continental Greece, Lesbos, Kithira, Naxos and Skyros vs. all others					2 groups of *C. barbara*: Seville vs. all others				
Source of variation	d.f.	Variance Component	%	Φ	*P*	d.f.	Variance Components	%	Φ	*P*
Among groups of localities	5	1.01Va	86.87	Φ_{CT}0.87	++	1	0.73 Va	90.40	Φ_{CT} 0.90	NS
Among localities within groups	18	0.09Vb	7.62	Φ_{SC}0.58	++	13	0.01 Vb	0.68	Φ_{SC} 0.07	+
Within localities	192	0.06Vc	5.51	Φ_{ST}0.94	++	187	0.07 Vc	8.93	Φ_{ST} 0.91	++

b) Other tests within *C. orni* and *C. barbara* haplogroups

	AMOVA	Φ_{CT}	*P*
C. barbara haplogroup	4 groups: North Portugal, South Portugal, Spain and North Africa	0.44	NS
	2 groups: Iberian Peninsula vs. North Africa	-0.16	NS
C. orni haplogroup	4 groups: Iberia Peninsula, France, eastern Europe and continental Greece	0.55	S
	2 groups: Continental Greece vs. all others (excluding Greek islands)	0.76	++
	2 groups: Greece (including islands) vs. all others	0.56	++
	5 groups: (Greece) continental Greece, Lesbos, Kithira, Naxos and Skyros	0.90	++

d.f., degrees of freedom; %, percentage of variation; Φ, fixation indices; *P*, significance of percentage variation and fixation indices estimated from permutation tests (1000 permutations); ΦCT correlation of random haplotypes within a group of populations relative to that drawn from the entire species; ΦSC correlation of random haplotypes within populations relative to that within a regional grouping of populations within a region; ΦST correlation of random haplotypes within a population relative to that from the whole species (Excoffier *et al.*, 1992); NS, not significant ($P>0.05$); ++ $P<0.001$, + $P<0.005$, S $P<0.05$.

Population hierarchical analysis and demography

A hierarchical analysis of molecular variance was performed for the *C. orni* and *C. barbara* haplogroups, in order to test for population structure (Table 5). For the *C. barbara* haplogroup, the highest Φ_{CT} value was found when Seville was tested vs. all the other

localities (Φ_{CT} = 0.9), but all other Φ_{CT} values obtained for this haplogroup were non-significant at the significance level of 0.5. For the *C. orni* haplogroup several tests were carried out. The most statistically significant (P<0.01) structure was revealed for six groups, continental Greece, Lesbos, Kithira, Naxos and Skyros vs. all remaining groups: 86.9% of variation was partitioned among groups, 7.6% among individual populations within groups and 5.5% within localities. When comparing the Greek samples, continental Greece vs. each island (Naxos, Kithira, Lesbos and Skyros), a Φ_{CT} value of 0.9 was found revealing high differentiation among groups, also with a P<0.01. In this case, no variation among populations within groups was shown because most of the groups corresponded to a single population.

The neutrality tests supported the hypothesis of demographic expansions revealing negative Fu's Fs for all groups analysed (most Fs values were statistically significant), with the exception of the Greek samples alone (Table 6). The expansion coefficient (S/d) supported these findings, showing the lowest value for Greek samples alone. The estimations of the date of demographic expansions (*t*) in the current study revealed *t* values greater than 100,000 years before present (Table 6).

Table 6. Estimated parameters from the sudden expansion model: S/d is an expansion coefficient (S is the number of polymorphic sites and d is the mean number of pairwise nucleotide differences); τ parameter represent the mean value accordingly to the percentile method (confidence level: α=0.01, based on 1000 replicates); t is the time since the expansion assuming a mutational rate of 1.15% per million years, or 2.3% per million years*, and one generation per two years/ or one generation each three years.

	S/d	**Fs**	**τ**	**t ya**
a) *C. barbara* haplogroup	22.04	-3.32	2.79	313,448 / 470,171 156,724 / 235,086*
b) *C. barbara* haplogroup excluding Seville	53.43	-5.82^{+}	2.75	
c) *C. cretensis* haplogroup	8.47	-0.75	2.42	
d) *C. orni* haplogroup	12.21	-3.94	0.81	
e) *C. orni* haplogroup excluding Greece	23.90	-6.36^{+}	2.48	276,478 / 414,716 138,238 / 207,358*
Greek samples only	4.07	1.62	5.07	
f) *C. mordoganensis* haplogroup	14.02	-3.27^{+}	2.57	

Discussion

The results confirm that each cicada taxon investigated, originally discriminated on the basis of acoustic profiles, corresponds to a distinct evolutionary lineage. This reinforces the importance of the acoustic signals in systematic and evolutionary studies involving cicadas since such signals are believed to act as specific mate recognition systems (SMRSs) (*e.g.*, Paterson, 1985; Claridge, 1985, 1995; Claridge & De Vriijer, 1993; Quartau & Rebelo, 1994; Quartau, 1995). SMRSs are involved in the reproductive isolation of species, and consequently in their genetic differentiation. The results here also allow a better understanding of the genetic relationships within and among these species.

12S rRNA secondary structure

As described by Sullivan, Holsinger & Simon (1995), the degree of sequence conservation can vary widely within helices (stems) and unpaired regions (loops) and the degree of variability between taxa can change among regions of the domain III. Therefore, the general assumption that stems are always more conserved than loops may not be completely reliable. In fact, some loops in the sequences analysed here were highly conserved in all haplogroups, moreover, some stems showed variability. Hickson *et al.* (1996) suggested that some unpaired regions can be highly conserved due to their involvement in protein, tRNA or rRNA interactions. Although not all substitutions in stems result in compensatory mutations (changes preserving base-pairing), these are still the most frequent substitutions, generally contributing to a lower level of variation as shown in the present analysis.

Independent evolutionary units

All analyses performed in the current study strongly suggest that each haplogroup forms an independent evolutionary unit supporting the acoustic and some morphologic differentiation known between these species (e.g. Boulard, 1982; Quartau, 1995; Simões *et al.,* 2000; Quartau & Simões, 2005a;) and the allozyme data (Quartau *et al.,* 2000, 2001). The Crete samples also formed an independent lineage in agreement with recent acoustic signals (Quartau & Simões, 2005a). Contrary to the allozyme data (Quartau *et al.,* 2000, 2001; Seabra *et al.,* 2000), these results do not suggest that speciation has occurred with small genetic divergences, especially taking into account the fact that the section of the gene analysed here is normally highly conserved (Simon *et al.,* 1994, 1996; Noor & Larkin, 2000). The genetic divergence between the closest pair of species analysed in this study, *C. orni* and *C. mordoganensis,* was 2.3% with a high ratio of transitions versus transversions (Ti/Tv)

suggesting a relatively recent divergence. However, the K81 distances between *C. orni* haplotypes from the Greek islands and the Western Europe varied from 0.5% to 1.7% and this upper limit is equivalent to the minimum distance found between *C. mordoganensis* and *C. orni* haplotypes (range of 1.7- 2.6% divergence). Though the range of intra- and interspecific sequence divergence hardly overlaps, the delimitation of very close species or even the identification of new species in the Aegean area may be problematic. The implementation of new DNA-based methods for delimitation of species (e.g. Sites & Marshall, 2003; Pons *et al.*, 2006) may simplify the recognition of new species. Conversely, *C. lodosi* and, particularly, *C. barbara* revealed lower Ti/Tv ratios and higher divergence levels compared to the previous species, indicative of a more ancient divergence. The level of divergence found between *C. barbara* and the other *Cicada* species was high (10-13.7%) when compared to the mtDNA divergence observed between other insect sibling species. For instance, Simon *et al.* (1996) found 7.8% of variable nucleotide sites in domain III of the 12S rRNA in two periodical cicadas, *Magicidada septendecim* and *M. cassini*, while in this study the variable nucleotide sites was approximately 11% between the cicadas analysed. Furthermore, Cameron & Williams (2003) found that the genetic divergence between bumble bee species of the subgenera *Fervidobombus* and *Thoracobombus* ranged from 0.75% to 8.6% within this same gene region. Within species of *Drosophila*, the third domain of 12S rRNA was found to have a much slower evolutionary rate than in other insects, with very little differentiation, inhibiting the identification of several species (Simon *et al.*, 1996; Noor & Larkin 2000). Nevertheless, several studies refer to marked differences in the evolutionary rates among different species (e.g. Simon *et al.*, 1996; Chiba, 1999; Michaux *et al.*, 2002), such as among cicadas from New Zealand (Arensburger *et al.,* 2004). In the present study, given the poor resolution of the more recent nodes of the inferred trees, this gene is probably more useful for species or higher-level phylogenies. The lack of resolution and occurrence of widespread haplotypes makes more difficult its application at the population level “phylogeny”.

In the Greek islands, where the most closely related species were found, molecular approaches may provide a more reliable method for identification than acoustic data. Acoustic signals are generally quite reliable, but can reveal some variation within species, which in a few cases can make identification problematic. In a recent study involving cicadas from the species complex *Platypleura stridula* (Price, Barker & Villet, 2007) it was possible to distinguish six clades based on mtDNA analysis whilst at the acoustic level only four clades could be identified. Nevertheless, the knowledge of the pattern of variation of acoustic signals

in the Greek islands might help clarifying the population substructure within *C. orni* in Greece. It is noteworthy that the most basal *Cicada* species, *C. barbara* and *C. lodosi*, as denoted by 12S rRNA, are characterized by continuous calling songs, in contrast to the remaining species where the signal is made up of a series of repetitive echemes (groups of pulses). In fact, the most recent diverged lineages possess increasingly longer pauses (Fig. 5). This might suggest that the continuous signal might be plesiomorphic and that the discontinuous state present in most *Cicada* species is a derived evolutionary innovation within the genus (Quartau & Simões, 2005b).

Population structure and demography

The typical star-shape patterns observed in the minimum spanning network of *C. barbara* and *C. orni* (excluding the Greek samples) are indicative of expanding populations, which was also supported by the neutrality tests. The τ parameter, which measures the time since the expansion (Rogers and Harpending, 1992; Harpending, 1994), did not vary greatly between the different species, suggesting the occurrence of a contemporaneous event leading to the population expansions of the species analysed. Only for *C. orni*, the τ parameter was considerably lower than the value obtained for the other species. However, excluding the Greek populations, this value became similar to the other species, suggesting that the Greek populations have experienced a different pattern of expansion.

In the case of *C. orni*, the network also revealed two star-shaped groups, one including all Greek populations together with a few Portuguese specimens from Caparica and Lisbon, and the other including all the remaining European populations. The partitioning of mtDNA variation in this species supports the separation of Greek populations from the rest of Europe a result in keeping with an acoustic analysis (Pinto-Juma *et al.,* 2005). Analyses of 13 characters of the calling songs of *C. orni* along continental Europe revealed that songs from south-eastern Europe (Greece) tended to group apart from western Europe (Iberian Peninsula and France), and the inter-echeme interval duration (the duration of pauses within the calling song) was the variable that contributed most to this separation (Pinto-Juma *et al.,* 2005). In fact, the most significant structure for all European *C. orni* samples is at the level of six groups: continental Greece, Lesbos, Kithira, Naxos, Skyros and a last group including all remaining European cicadas. Greek samples alone, i.e., continental Greece *versus* each of the islands Naxos, Kithira, Lesbos and Skyros showed a highly structured data set (Φ_{CT}=0.9, P<0.01) and the hypothesis of a sudden expansion was not supported by the neutrality tests.

This is perhaps not surprising, considering that the colonization of distant islands by *C. orni* must have been a slow process, with founder effects and other island biogeographical processes leading to more divergent populations (Cox & Moore, 2005). Only the islands of Lesbos and Skyros shared a haplotype (HCo8), while the other islands showed only exclusive haplotypes (Fig. 7). These results also suggest that the colonization of the Greek islands by *C. orni* occurred prior to the expansion events observed in continental Europe. In contrast, continental Greek samples did not show any variability, the three sampled localities revealing one single common haplotype. Nonetheless, the partition between continental Greece and the rest of Europe was also strongly supported ($\Phi_{CT}=0.76$, $P<0.01$). Similar patterns of genetic differentiation were found between western and eastern Mediterranean trees with which *Cicada* specimens are predominantly found associated, such as the maritime pine (Burban & Petit, 2003) and olive trees (Besnard & Bervillé, 2000) suggesting that the life history of these trees and the cicadas might be related.

Cicada orni specimens from Caparica and Lisbon, two nearby localities in Portugal, did not reveal the common haplotype observed in Western Europe. Instead they revealed the same haplotype found in continental Greece (HCo2). It would seem unlikely that the presence of this haplotype in western Europe might be due to migration from Greece since this haplotype is absent from intermediate localities, however, occasional long-distance migration or convergent evolution may occur. On the other hand, HCo2 might represent an ancestral haplotype that was spread at a lower frequency than HCo1 during the pre- expansion time, and subsequently largely eliminated by bottlenecks. Also, a human/ accidental translocation might be possible, especially taking into account that cicadas are usually associated with olive trees, which are known to be cultivated since ancient times and its dispersal was greatly influenced by humans (Besnard *et al*., 2002). Further studies on these populations and other intermediates using faster evolving genes might help to clarify this issue.

There was no evidence of any geographical structuring of *C. barbara* which is characterised by a higher average allelic and nucleotide diversity per locality than *C. orni*. Nonetheless, the fact that many localities have unique haplotypes, apart from the common haplotype, suggests that colonization was not recent, or that populations have undergone considerable bottlenecks and lineage sorting since the colonization. Acoustic and cytochrome *b* gene analysis within *C. barbara* specimens from the same localities analysed here suggest a divergence between the Iberian Peninsula and Morocco populations (Pinto-Juma, Seabra &

Quartau, in press; Pinto-Juma, Quartau & Bruford, in press). As pointed out before, 12S rRNA might not be the most suitable for population level analysis.

Small sample sizes of *C. mordoganensis* and *C. cretensis* did not allow a detailed analysis of their population structure. Yet, considering that their populations exist on islands, it is expected that differentiation might have occurred through isolation. Crete is relatively distant from the Greek and Turkish mainlands or other islands, which naturally might have favoured the speciation of *C. cretensis*. On the other hand, the samples of *C. mordoganensis* analysed here are located on islands very close to Turkey, where this species seems to be largely distributed (Boulard, 1979). As opposed to the *C. orni* populations sampled in islands, the most common haplotype of *C. mordoganensis* found in this study was present in all populations, suggesting a rapid and more recent colonization, and/or a higher degree of gene flow between localities, since such islands are very close to the Turkish mainland.

Expansion events

The population structure of a species is the result of both present and historic events. During the Pleistocene climate changes induced adjustments of most species distributions, often leading to severe reductions of population size, or even extinction (Hewitt, 1999, 2000). The decline of population distribution and size, followed by expansions and secondary contact are responsible for much of extant population structure and divergence (e.g. Hewitt, 1999, 2000, 2001, 2004; Widmer & Lexer, 2001; Buckley, Simon & Chambers, 2001; Neiman & Lively 2004). This process seems the most likely explanation for the fast expansions observed in all cicada species analysed in this study. Even though we can not assess accurately the dates of the expansion events, the estimated dates based on some conservative and moderate mutational rates for the 12S rRNA fall within the Pleistocene period ($t > 100,000$ ya).

Many studies have recognized the Iberian Peninsula, Italy and the Balkans as three major southern refugia in Europe during the severe effects of the glaciations and from where populations expanded during interglacial periods (Hewitt, 1996, 1999, 2000, 2004; Lunt, Ibrahim & Hewitt, 1998; Kutnik *et al.*, 2004). No specimens from Italy were available in this study, yet samples from the Iberian Peninsula and the Balkans were well represented. One might expect to find higher levels of genetic diversity in these regions than in the areas colonized during the inter-glacial period (Hewitt, 1996; 1999; Neiman & Lively 2004), because these populations would be older and would not have experienced such severe bottlenecks (Neiman & Lively 2004). However, such genetic gradients are not evident in the

present study. Mainland Greek populations of *C. orni* did not show higher levels of variability than in other localities, and within the rest of Europe no clear geographical structure was found in respect of genetic diversity. Widmer & Lexer (2001) pointed out that assuming higher genetic diversity within refugia might be too simplistic. Furthermore, transitory reductions in effective population size (bottlenecks) lead to a reduction of genetic diversity and this is the pattern typically expected in most phylogeographic analyses. This seems to better explain the genetic pattern observed in this study. All demographic parameters indicated an expansion from small populations. The differentiation between Greece and the rest of the Europe suggests that Greek populations remained isolated and did not contribute greatly to the re-colonization of the other European habitats. On the other hand, the lack of geographic structure among the other localities does not exclude the possibility of re-colonization from the Iberian Peninsula, in spite of the Pyrenean barrier (Fig. 6). It would be very interesting to analyse Italian samples in order to check if the same pattern is maintained.

Final considerations

These analyses confirmed the congruence between mtDNA and acoustic behaviour divergence at the species and the population levels in the genus *Cicada*. In the Aegean area island biogeographical vicariant processes have probably led to divergence between populations and the formation of new species. Since *C. orni* is the most recent diverged species within the genus *Cicada* and the closest related species are in the Aegean area, this is likely to be the region of origin of *C. orni*.

The origin of *C. barbara* in Portugal and Spain remains obscure. The hypothesis of this species being an immigrant to the Iberian Peninsula from North Africa (Quartau *et al.*, 2001) remains possible, but the presence of *C. barbara* in the Iberian Peninsula does not seem a recent event, as discussed above. Though no evidence of geographic structure was found within 12S rRNA, some genetic variability was detected within the Iberian Peninsula as well as within Morocco. However, as the main haplotype found was common to both areas, it is difficult to draw conclusions.

The fragment of the mitochondrial 12S rRNA gene analysed in the current study proved to be reliable for species discrimination, as well as to determine the phylogeny of the genus *Cicada* in the Mediterranean area. However, the population structure and demographic parameters were not completely resolved and some assumptions should be considered with caution due to the insufficient information concerning the evolutionary rate of this gene

within these species, as well as to the heterogeneous sampling. Use of larger and more complete sampling coupled with analyses of faster evolving genes will certainly help to clarify the unresolved issues.

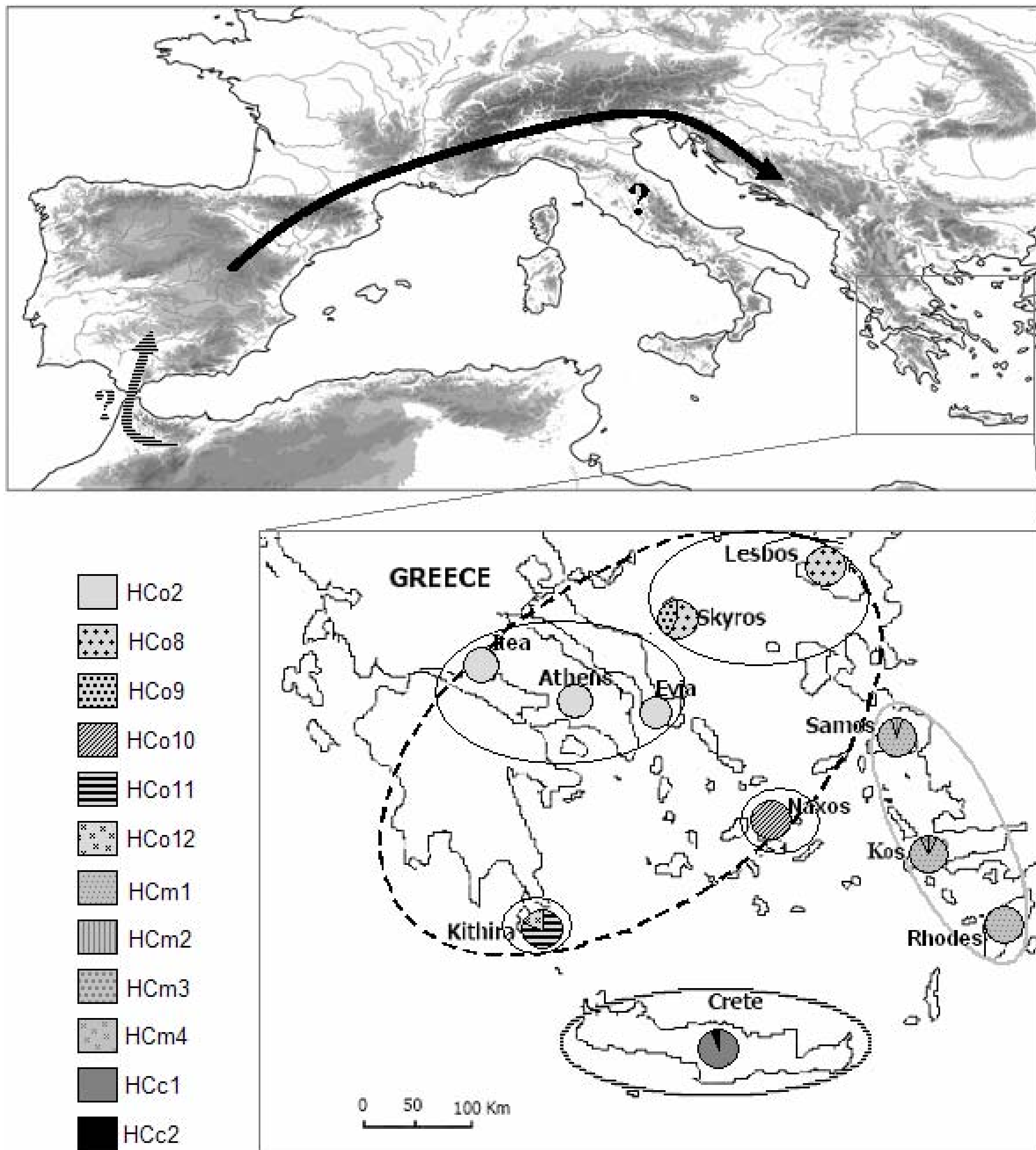


Fig. 6. Post-glacial colonization routes of *Cicada orni* (black arrow) and *C. barbara* (crosshatched arrow) in Europe (top map) and distribution of *Cicada* haplotypes in the Aegean area (map below). Pie graphics represent the proportion of each haplotype in each locality. Hatched circle surrounds *C. orni* haplotypes, grey circle groups *C. mordoganensis* and crosshatched circle groups *C. cretensis*. Smaller black circles highlights populational substructure within *C. orni*.

Acknowledgments

We express our thanks to Sofia Seabra, Genage André, Paula Simões, Teresa Rebelo, Mónica Ribeiro and Teresa Fernandes (Faculty of Science, University of Lisbon), for their help with field work. We are grateful to Dr. Tomy Trilar (Invertebrate Department, Slovenian Museum of Natural History) and Prof. Michel Boulard (Muséum National d`Histoire Naturelle, Paris) for providing some valuable dried specimens. We also thank Dr. Nicola Anthony for her help in optimizing SSCP conditions and Dr. Jo Cable (School of Biosciences, Cardiff University) as well as one anonymous reviewer for reading and commenting on this manuscript. We appreciate the collaboration of Prof. Janet Hemingway and John Morgan during the initial stage of this work. This research was supported by a PhD grant (PRAXIS BD/18229/98) from the Fundação para a Ciência e Tecnologia (Lisbon, Portugal).

References

Alves-Gomes JA. 1999. Systematic biology of Gymnotiform and Mormyriform electric fishes: phylogenetic relationships, molecular clocks and rates of evolution in the mitochondrial rRNA genes. *Journal of Experimental Biology* **202**: 1167-1183.

Arensburger P, Buckley TR, Simon C, Moulds M, Holsinger KE. 2004. Biogeography and phylogeny of the New Zealand cicada genera (Hemiptera: Cicadidae) based on nuclear and mitochondrial DNA data. *Journal of Biogeography*, **31**: 557-569.

Baratti M, Khebiza MY, Messana G. 2004. Microevolutionary processes in the Stygobitic genus *Typhlocirolana* (Isopoda, Flabellifere Cirolamidae) as inferred by partial 12S and 16S rDNA sequences. *Journal of Zoological Systematics and Evolutionary Research* **42**: 27-32.

Beckenbach AT, Wei YW, Liu H. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Molecular Biology and Evolution* **10** (3): 619-634.

Bennet-Clark HC. 1998. How cicadas make their noise. *Scientific American* **5**: 36-39.

Besnard G, Bervillé A. 2000. Multiple origins for Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. C.R. Acad. Sci. Paris*, Sciences de la Vie/ Life Sciences* **323**: 173-181.

Besnard G, Khadari B, Baradat P, Bervillé A. 2002. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* **104**: 1353 – 1361.

Boulard M. 1979. Cigales du genre *Cicada* Linné, originaires de Turquie (Homoptera, Cicadidae). *Türkiye Bitki Koruma Dergisi* **3** (2): 67-74.

Boulard M. 1982. Les cigales du Portugal, contribution à leur étude (Homoptera, Cicadidae). *Annales de la Societe Entomologique de France* **18** (2): 181-198.

Buckley TR, Simon C, Chambers GK. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Systematic Biology* **50** (1): 67-86.

Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**: 1487-1495.

Caccone A, Milinkovitch MC, Sbordoni V, Powell JR. 1997. Mithocondrial DNA rates and biogeography in European newts (genus *Euproctus*). *Systematic Biology* **46**: 126-144.

Cameron SA, Williams PH. 2003. Phylogeny of bumble bees in the New World subgenus *Fervidobombus* (Hymenoptera: Apidae): congruence of molecular and morphological data. *Molecular Phylogenetics and Evolution* **28**: 552-563.

Chiba S. 1999. Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. *Evolution*, **53**: 460-471.

Claridge MF. 1985. Acoustic signals in Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology* **30**: 297-317.

Claridge MF. 1995. Species concepts and speciation in insect herbivores: planthopper case studies. *Bollettino di Zoologia* **62**: 53-58.

Claridge MF, De Vriijer P. 1993. Reproductive behaviour: the role of acoustic signals in species recognition and speciation. *In*: R. F. Denno & T. J. Perfect (eds) *Planthoppers their ecology and management*. Chapman & Hall, New York, pp. 216-233.

Cox CB, Moore PD. 2005. *Biogeography: an ecological and evolutionary approach*. 7th Ed. Blackwell Publishing. Oxford, UK.

Crandall KA, Olaf RP, Bininda-Edmonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15** (7): 290-295.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Fu Y. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics* **147**: 915-925.

Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66** (4): 591-600.

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

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

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, **405**: 907-913.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography – or seen genes in space and time. *Molecular Ecol*ogy **10**: 537-549.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B-Biological Sciences* **359**: 183-195.

Hey J. 2001. The mind of the species problem. *Trends in Ecology & Evolution* **16** (7): 326-329.

Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, Penny D. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution* **13** (1): 150-169.

Jordan WC, Foley K, Bruford MW. 1998. Single-strand conformation polymorphism (SSCP) analysis. *In* "*Molecular Tools for Screening Biodiversity, Plants and Animals*" (A. Karp, P. G. Isaac & D. S. Ingram, Eds), Chapman & Hall, pp.152-156.

Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America* **78**: 454-458.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

Kutnik M, Uva P, Brinkworth L, Bagnères A-G. 2004. Phylogeography of two European *Reticulitermes* (Isoptera) species: the Iberium refugium. *Molecular Ecol*ogy **13**: 3099-3113.

Livak KJ. 1984. Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**: 611-634.

Lunt DH, Ibrahim KM, Hewitt GM. 1998. MtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity*. **80**: 633-641.

Mes THM. 2003. Demographic expansion of parasitic nematodes of livestock based on mitochondrial DNA regions that conflict with the infinite-sites model. *Molecular Ecology* **12**: 1555-1566.

Michaux JR, Chevret P, Filippucci M-G, Macholan M. 2002. Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12S rRNA. *Molecular Phylogenetics and Evolution* **23**: 123-136.

Neiman M, Lively CM. 2004. Pleistocene glaciation is implicated in the phylogeographical structure of *Potamopyrgus antipodarum*, a New Zealand snail. *Molecular Ecol*ogy **13**: 3085-3098.

Noor MAF, Larkin, JC. 2000. A re-evaluation of 12S ribosomal RNA variability in *Drosophila pseudoobscura*. *Molecular Biology and Evolution* **17** (6): 938-941.

Page RDM. 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. *Nucleic Acids Research* **28** (20): 3839-3845.

Paterson HEH. 1985. The recognition concept of species. *In* E.S. Vrba, ed. *Species and speciation*. Pretoria, S.A, Transvaal Museum Monograph no. 4, pp.: 21-29.

Pinto GA, Quartau JA, Morgan JC, Hemingway J. 1998. Preliminary data on the sequencing of a fragment of cytochrome b gene of mitochondrial DNA in *Cicada orni* L. (Homoptera: Cicadidae) in Portugal. *Boletim da Sociedade Portuguesa de Entomologia* **VII-6** (188): 57-66.

Pinto-Juma G, Simões PC, Seabra SG, Quartau JA. 2005. Calling song structure and geographical variation in *Cicada orni* Linnaeus (Hemiptera: Cicadidae). *Zoological Studies* **44** (1): 81-94.

Pinto-Juma G, Seabra SG, Quartau JA. Patterns of acoustic variation in *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco. *Bulletin of Entomological Research* - in press.

Pinto-Juma G, Quartau JA, Bruford MW. Population structure of *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco based on mitochondrial DNA analysis. *Bulletin of Entomological Research* - in press.

Pons J, Barraclough TC, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55** (4): 595-609.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics Applications note* **14** (9): 817-818.

Posada D, Buckey TR. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and the Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793-808.

Price BW, Barker NP, Villet MH. 2007. Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology* **16**: 2574-2588.

Quartau JA. 1995. Cigarras esses insectos quase desconhecidos. *Correio da Natureza* **19**: 33-38.

Quartau JA, Rebelo MT. 1994. Sinais acústicos em Cicadidae e Cicadellidae (Homoptera, Auchenorryncha) que ocorrem em Portugal. *In* Vitor Almada e Rui Correia (eds.) *Biologia e Comportamento*, *Actas do I Congresso Nacional de Etologia*, Lisboa 28, 29 e 30 de Novembro de 1991, Publicado por Instituto Superior de Psicologia Aplicada, Lisboa pp: 137-142.

Quartau JA, Ribeiro M, Simões PC, Crespo A. 2000. Taxonomic separation by isozyme electrophoresis of two closely related species of *Cicada* L. (Hemiptera: Cicadoidea) in Portugal. *Journal of Natural History* **34**: 1677-1684.

Quartau JA, Ribeiro M, Simões PC, Coelho MM. 2001. Genetic divergence among populations of two closely related species of *Cicada* Linnaeus (Hemiptera: Cicadoidea) in Portugal. *Insect Systematics & Evolution* 32: 99-106.

Quartau JA, Simões PC. 2005a. *Cicada cretensis* sp. n. (Hemiptera, Cicadidae) from southern Greece. *Biologia Bratislava* **60** (5): 489-494.

Quartau JA, Simões PC. 2005b. Acoustic evolutionary divergence in cicadas: the case study of *Cicada* L. (Hemiptera, Cicadoidea) in southern Europe. Chapter 17. *In*: *Insect Sounds and Communication: Physiology, Behaviour and Evolution*, S.- Drosopoulos & M. Claridge (Editors), Taylor & Francis, USA.

Ramos-Onsins S, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**: 2092-2100.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9** (3): 552-569.

Rohlf FJ. 1973. Algorithm 76. Hierarchical clustering using the minimum spanning tree. *Computer Journal* **16**: 93-95.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Satoh A, Sota T, Uéda T, Enokido Y, Paik JC, Hori M. 2004. Evolutionary history of coastal tiger beetles in Japan based on a comparative phylogeography of four species. *Molecular Ecol*ogy **13**: 3057-3069.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin: A software for population genetics data analysis. Ver. 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Seabra S, Simões PC, Drosopoulos S, Quartau JA. 2000. Genetic variability and differentiation in two allopatric species of the genus *Cicada* L. (Hemiptera, Cicadoidea) in Greece. *Mitteilungen aus dem Museum für Naturkunde in Berlin - Deutsche Entomologische Zeitschrift* **47** (2): 143-145.

Simões PC, Boulard M, Rebelo MT, Drosopoulos S, Claridge MF, Morgan JC, Quartau JA. 2000. Differences in the male calling songs of two sibling species of *Cicada* (Hemiptera: Cicadoidea) in Greece. *European Journal of Entomology* **97** (3): 437–440.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P· 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers· *Annals of the Entomological Society of America* **87** (6): 651-701·

Simon C, Nigro L, Sullivan J, Holsinger K, Martin A, Grapputo A, Franke A, Mcintosh C. 1996. Large differences in substitutional pattern and evolutionary rate of 12S ribosomal RNA genes. *Molecular Biology and Evolution* **13** (7): 923-932.

Sites JW, Marshall JC. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* **18** (9): 462-470.

Sullivan J, Holsinger KE, Simon C. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in Sigmodontine rodents. *Molecular Biology and Evolution* **12** (6): 988-1001.

Swofford DL. 1999. PAUP 4.0 beta version. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, Massachusetts.

Widmer A, Lexer C. 2001. Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution* **16** (7): 267-271.

Wolstenholme DR, Clary DO. 1985. Sequence evolution of *Drosophila* mitochondrial DNA. *Genetics* **109**: 725-744.

Xia X, Xie Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* **92**:371-373.

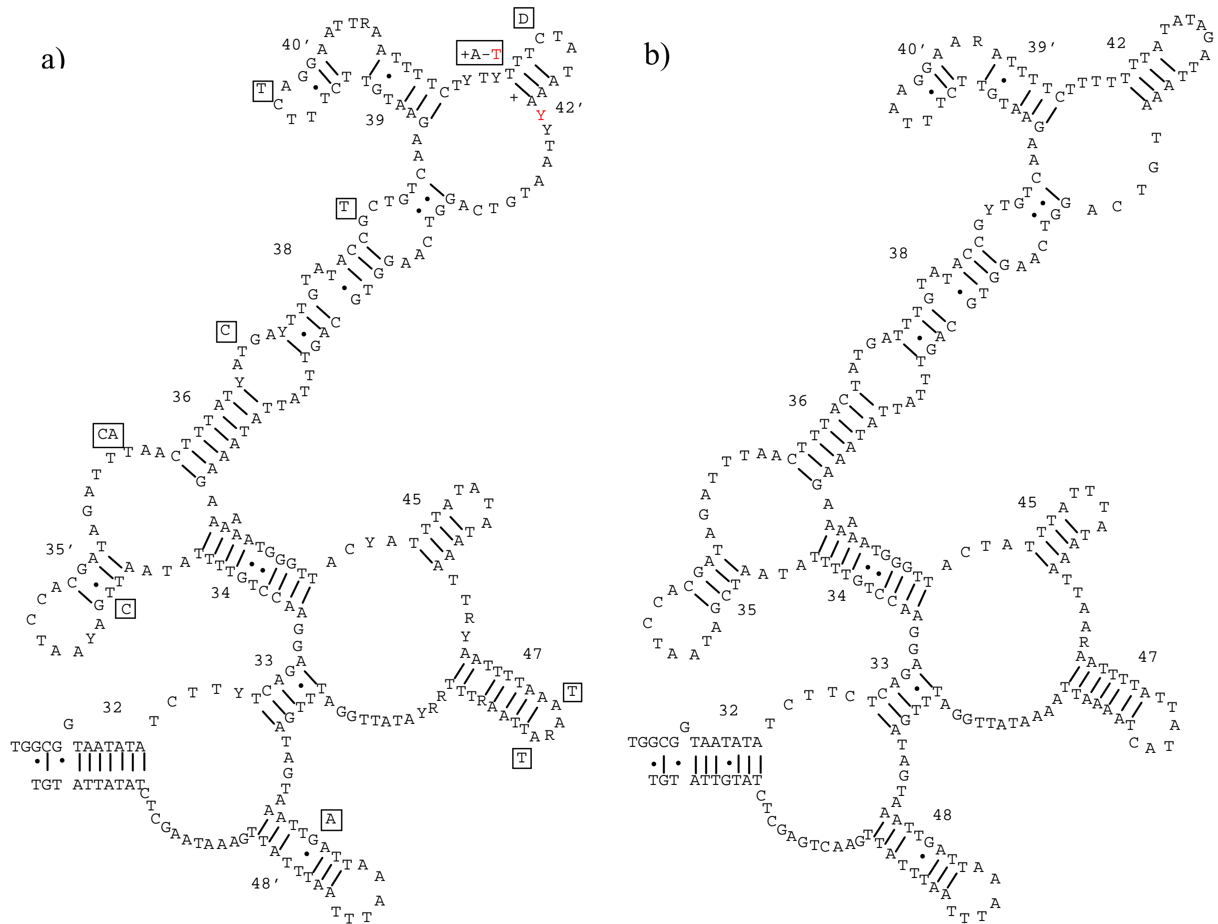
Appendix 1. Distribution of 12S rRNA gene (~390bp) haplotypes among *Cicada* species from the Mediterranean area (* *C. orni* and *C. barbara* sympatry localities; Pt-Portugal, Sp-Spain, Mr-Morocco, Fr-France, Ct- Croatia, Md- Macedonia, Sl-Slovenia, Gr-Greece, Cr-Crete, Tk-Turkey; HCo stands for *C. orni*, HCm for *C. mordoganensis*, HCc for *C. cretensis*, HCb for *C. barbara* and HCl stands for *C. lodosi*)

	HCo1	HCo2	HCo3	HCo4	HCo5	HCo6	HCo7	HCo8	HCo9	HCo10	HCo11	HCo12	HCm1	HCm2
Alcalar Pt														
Alter do Chão Pt	15													
Alvor Pt														
Arrábida Pt*	10													
Caparica Pt		15												
Crato Pt*	14			1										
Foz Côa Pt														
Lisbon Pt		2												
Monforte Pt*	9		1											
Moura Pt														
Piedade Pt	15													
Portel Pt*	14				1									
Sousel Pt														
Tomar Pt	10													
Algeciras Sp	15													
Cordoba Sp														
Jadraque Sp	3													
Sevilla Sp														
Toledo Sp*	2													
Ceuta Sp														
Fès Mr														
Mèknes Mr														
Narbonne Fr	15													
St. Hippolyte Fr	13					1	1							
Isl. Lastovo Ct	1													
Istra Ct	1													
Island Vis Ct	1													
Mount.Zeden Md	1													
Komen Sl	1													
Athens Gr		14												
Evia Gr		5												
Itea Gr		5												
Lesbos Gr								7						
Kithira Gr											4	1		
Naxos Gr										11				
Skyros Gr								4	3					
Kos Gr													7	
Rhodes Gr													9	
Samos Gr													9	1
Heraklion Cr														
Gordes Tk														
TOTAL	**140**	**41**	**1**	**1**	**1**	**1**	**1**	**11**	**3**	**11**	**4**	**1**	**25**	**1**

Appendix 1. (Continuation)

	HCm3	HCm4	HCc1	HCc2	HCb1	HCb2	HCb3	HCb4	HCb5	HCb6	HCb7	HCb8	HCl	N
Alcalar Pt					14						1			15 HCb
Alter do Chão Pt														15 HCo
Alvor Pt					10									10 HCb
Arrábida Pt*					15									10 Co/15HCb
Caparica Pt														15 HCo
Crato Pt*					15									15HCo/15HCb
Foz Côa Pt					14							1		15 HCb
Lisbon Pt														2 HCo
Monforte Pt*					15									10 Co/15HCb
Moura Pt					14					1				15 HCb
Piedade Pt														15 HCo
Portel Pt*					15									15HCo/15HCb
Sousel Pt					15									15 HCb
Tomar Pt														10 HCo
Algeciras Sp														15 HCo
Cordoba Sp					5									5 HCb
Jadraque Sp														3 HCo
Sevilla Sp					2				13					15 HCb
Toledo Sp*					1	1								2HCo/2HCb
Ceuta Sp					15									15 HCb
Fès Mr					14		1							15 HCb
Mèknes Mr					12			3						15 HCb
Narbonne Fr														15 HCo
St. Hippolyte Fr														15 HCo
Isl. Lastovo Ct														1 HCo
Istra Ct														1 HCo
Island Vis Ct														1 HCo
Mout.Zeden Md														1 HCo
Komen Sl														1 HCo
Athens Gr														14 HCo
Evia Gr														5 HCo
Itea Gr														5 HCo
Lesbos Gr														7 HCo
Kithira Gr														5 HCo
Naxos Gr														11 HCo
Skyros Gr														7 HCo
Kos Gr	1	1												9 HCm
Rhodes Gr														9 HCm
Samos Gr														10 HCm
Heraklion Cr			16	1										17 HCc
Gordes Tk													1	1 HCl
TOTAL	**1**	**1**	**16**	**1**	**176**	**1**	**1**	**3**	**13**	**1**	**1**	**1**	**1**	**459**

Appendix 2. Secondary structure of domain III of 12S rRNA gene based on the template described by Hickson *et al.* (1996) obtained for (a) the consensus sequences of the *Cicada orni*, *C. mordoganensis* and *C. cretensis* haplogroups, *C. lodosi* substitutions are shown at the boxes (+ stands for insertion, D stands for deletion), and (b) the consensus sequence of the *C. barbara* haplogroup.



Appendix 3. Kimura 3 parameter (proportion of invariables I=0.8116) distance matrix, obtained by comparison of all pairs of haplotypes analysed in this study, based on an approximately 394 bp fragment of the 12SrRNA gene (HCo stands for *C. orni*, HCm for *C. mordoganensis*, HCc for *C. cretensis* and HCb for *C. barbara* haplotypes).

	Co1	Co2	Co3	Co4	Co5	Co6	Co7	Co8	Co9	Co10	Co11	Co12	Cb1	Cb2	Cb3	Cb4	Cb5	Cb6	Cb7	Cb8	Cm1	Cm2	Cm3	Cm4	Cc1	Cc2
HCo1																										
HCo2	0.3																									
HCo3	0.3	0.5																								
HCo4	0.3	0.5	0.5																							
HCo5	0.3	0.5	0.5	0.5																						
HCo6	0.3	0.5	0.5	0.5	0.5																					
HCo7	0.3	0.5	0.5	0.5	0.5	0.5																				
HCo8	0.5	0.3	0.8	0.8	0.8	0.8	0.8																			
HCo9	1.1	0.8	1.4	0.8	1.4	1.4	1.4	1.1																		
HCo10	1.4	1.1	1.7	1.1	1.7	1.7	1.7	1.4	1.4																	
HCo11	0.8	0.5	1.1	0.5	1.1	1.1	1.1	0.8	0.8	1.1																
HCo12	1.1	0.8	1.4	0.8	1.4	1.4	1.4	1.1	1.1	1.4	0.3															
HCb1	12.5	11.8	13.3	11.8	13.3	13.3	13.3	13.3	11.1	9.9	11.8	11.9														
HCb2	13.2	12.4	14.1	12.4	14.0	14.0	14.1	11.9	11.7	10.4	12.4	12.6	0.0													
HCb3	13.2	12.4	14.0	12.4	14.0	14.0	14.0	11.8	11.7	10.4	12.4	12.5	0.5	0.5												
HCb4	14.2	13.3	15.1	13.3	15.1	15.1	15.1	12.7	12.5	11.1	13.3	13.5	0.5	0.5	0.5											
HCb5	14.2	13.3	15.1	13.3	15.1	15.1	15.1	12.7	12.5	11.1	13.3	13.5	0.5	0.5	0.5	0.5										
HCb6	12.5	11.8	13.3	11.8	13.3	13.3	13.3	11.3	11.1	9.9	11.8	11.9	0.0	0.0	0.5	0.5	0.5									
HCb7	13.3	12.5	14.2	12.5	14.2	14.2	14.2	12.0	11.8	10.4	12.5	12.7	0. 3	0. 3	0.8	0.8	0.8	0. 3								
HCb8	13.3	12.5	14.2	12.5	14.2	14.2	14.2	12.0	11.8	10.5	12.5	12.7	0. 3	0. 3	0.8	0.8	0.8	0. 3	0.5							
HCm1	2.0	1.7	2.3	1.7	2.3	2.3	2.3	2.0	2.0	2.3	1.7	2.0	11.8	12.4	12.4	11.8	13.3	11.8	12.5	12.5						
HCm2	2.3	2.0	2.6	2.0	2.6	2.6	2.6	2.3	1.7	2.6	2.0	2.3	12.5	13.2	13.2	12.5	14.2	12.5	13.3	13.3	0. 3					
HCm3	2.3	2.0	2.6	2.0	2.6	2.6	2.6	2.3	2.3	2.6	2.0	2.3	12.5	13.2	13.2	12.5	14.2	12.5	13.3	13.3	0. 3	0.5				
HCm4	2.3	2.0	2.6	2.0	2.6	2.6	2.6	2.3	2.3	2.6	2.0	2.3	12.5	13.2	13.2	12.5	14.2	12.5	13.3	13.3	0. 3	0.5	0.5			
HCc1	2.6	2.3	3.0	2.3	3.0	3.0	3.0	2.6	2.0	2.3	2.3	2.6	9.2	9.7	9.8	10.3	10.3	9.2	9.7	9.8	2.3	2.6	2.6	2.6		
HCc2	3.0	2.6	3.4	2.6	3.4	3.4	3.4	3.0	2.3	2.6	2.6	3.0	9.8	10.2	10.4	10.9	10.9	9.7	10.3	10.3	2.6	3.0	3.0	3.0	0. 3	
C.lodosi	6.5	7.0	7.0	5.9	7.0	7.0	7.0	7.5	6.0	6.5	7.0	7.0	13.1	14.0	13.6	14.1	13.2	13.1	14.1	14.1	7.6	7.6	8.2	8.2	5.0	5.5

Chapter 3.

Calling song structure and geographic variation in *Cicada orni* Linnaeus (Hemiptera: Cicadoidea)

Published: Pinto-Juma GA, Simões PC, Seabra SG, Quartau JA. 2005. Calling song structure and geographic variation in Cicada orni Linnaeus (Hemiptera: Cicadoidea). *Zoological Studies* **44** (1): 81-94.

Zoological Studies **44**(1): 81-94 (2005)

Zoological Studies

Calling Song Structure and Geographic Variation in *Cicada orni* Linnaeus (Hemiptera: Cicadidae)

Gabriela Pinto-Juma[1], Paula C. Simões[1,2], Sofia G. Seabra[1], and José A. Quartau[1,2,*]

[1]*Centro de Biologia Ambiental C2-3° Piso, Faculdade de Ciências de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*
[2]*Departamento de Biologia Animal, C2-3° Piso, Faculdade de Ciências de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*

(Accepted October 1, 2004)

Gabriela Pinto-Juma, Paula C. Simões, Sofia G. Seabra, and José A. Quartau (2005) Calling song structure and geographic variation in *Cicada orni* Linnaeus (Hemiptera: Cicadidae). *Zoological Studies* **44**(1): 81-94. An analysis of the structure of the calling song of *Cicada orni* Linnaeus, 1758 over a selected part of its distribution range in the Mediterranean area was carried out in order to better understand the pattern of its geographic variation and population history. Calling songs of males from several localities in western and southeastern Europe were recorded from 1995 to 2003, and were analysed in time and frequency domains. The calling songs analyzed constituted a relatively homogeneous group. However, there was some tendency of songs from southeastern Europe (Greece) to group apart from those of western Europe (Iberian Peninsula and France). The inter-echeme interval duration was the variable that contributed most to this separation, but males from Greece showed significant differences in almost every acoustic variable in relation to the remaining studied regions. This acoustic differentiation is in agreement with introductory genetic results still under investigation. Echeme duration proved to be quite constant across the geographic range of this cicada and, as such, is probably one of the most important parameters encoding species-specific information for species recognition. Conversely, the inter-echeme interval was quite variable, and so it is expected that this variable is not an important parameter for species recognition and isolation in *C. orni*. The observed distinctiveness of the populations of the Aegean area may be the result of repeated cycles of isolation in southern refugia through the mountain ranges of the Balkans during the ice ages. Furthermore, it is hypothesized that the Aegean area and West Asia Minor might constitute the main area of origin of *C. orni*. http://www.sinica.edu.tw/zool/zoolstud/44.1/81.pdf

Key words: *Cicada orni*, Calling song, Acoustic signals, Acoustic divergence, Geographic variation.

Cicada orni Linnaeus is one of the most abundant and common cicadas throughout the Mediterranean area, being very familiar for the striking calling songs produced by males during summertime. It is distributed from the Iberian Peninsula in western Europe to Greece and Turkey and some countries in the near East (Nast 1972), as well as around the Black Sea (Popov 1975). This cicada is known to occur in closed high shrubland and woodland (Patterson et al. 1997, Puissant and Sueur 2001, Sueur et al. 2004), more commonly occurring in olive trees, pine trees, oak trees, and also eucalyptus and vineyards.

As in other cicadas, males of *C. orni* produce loud airborne acoustic signals by means of a tymbal mechanism. The tymbals are ribbed cuticle membranes located dorsolaterally in the 1st segment of the abdomen that are distorted by the action of powerful muscles (e.g., Pringle 1954, Popov 1975, Bennet-Clark 1998). Male cicadas can produce different kinds of acoustic signals (e.g., Alexander 1967, Boulard 1995 2000a, Fonseca 1991), with the calling song being the most common one. This is typically regularly patterned and species-specific, and is involved in mate attraction (as shown in *Magicicada* spp. (Cooley and Marshall 2001)) as well as in male

* To whom correspondence and reprint request should be addressed. Tel: 351-21-7500000. Fax: 351-21-7500028. E-mail: jaquartau@fc.ul.pt

Zoological Studies **44**(1): 81-94 (2005)

aggregation and chorusing (as in *C. mordoganensis* (Simões et al. 2000) or in other cicadas as referred to by, for instance, Villet (1992) and Fonseca and Revez (2002). Therefore, according to Paterson's recognition concept of species, the calling song constitutes a distinct specific-mate recognition system (SMRS) that should remain relatively constant by stabilizing selection across the distribution range of the species (Paterson 1985).

Cicada orni belongs to a complex of species which are morphologically very similar but which differ mostly in the calling songs that males produce. Thus, the study of these acoustic signals is very important in the delimitation of close species, as well as to better understand the speciation process.

Several authors have described the calling song of *C. orni* (e.g., Popov 1975, Joermann and Schneider 1987, Fonseca 1991, Boulard 1995) for a few local populations of this species. However, studies on the acoustic variation across the distribution area of cicadas are rare, with the exception of some work on American cicadas (e.g., Moore 1993). For *C. orni*, introductory comparisons of the calling songs in populations from southern France with those from the former USSR showed no obvious geographic variation (Claridge et al. 1979, Claridge 1985). Later, Quartau et al. (1999), when comparing Portuguese with Greek populations, revealed some geographic variation in *C. orni* songs.

The present paper is a more-thorough analysis of the structure of the calling song of *C. orni* over a wider distribution range of the species in the Mediterranean region in order to complement previously known data and to better understand the pattern of its geographic variations.

MATERIAL AND METHODS

Males of *C. orni* from several populations of several different regions along the geographic distribution of the species were recorded during the hottest season (June to Aug.) from 1995 to 2003. Eight populations from the Iberian Peninsula region, three from the south of France, seven from continental Greece, and one specimen each from Corsica and Turkey, totaling 176 specimens, were sampled (Fig. 1, Table 1). Some of the French recordings, as well as the recording from Corsica were provided by Jérôme Sueur (Muséum national d'Histoire Naturelle, Paris, France).

Males were first located by their calling songs, and their songs recorded, followed by collection of specimens by hand or by means of a sweep net. Most of the recordings were carried out using a Sony Dat recorder (TCD-D10 ProII and TCD-D8; at frequency ranges of 20~22,000 and 20~20,000 Hz, respectively, and at a sampling frequency of 44.1 kHz) connected to a dynamic Sony F-780 microphone or a Telinga Pro4PiP microphone (with frequency responses of 50~18,000 and 40~18,000 Hz, respectively). Other recordings were made using a UHER 4200 Report Monitor (with a sampling frequency of 44.1 kHz and a frequency range of 20~25,000 Hz) with an AKG D202 dynamic microphone. The microphone was placed at a distance of at least 30 cm from the calling insect. Acoustic recordings and specimens were kept in the Department of Animal Biology with one of the authors (J.A.Q.).

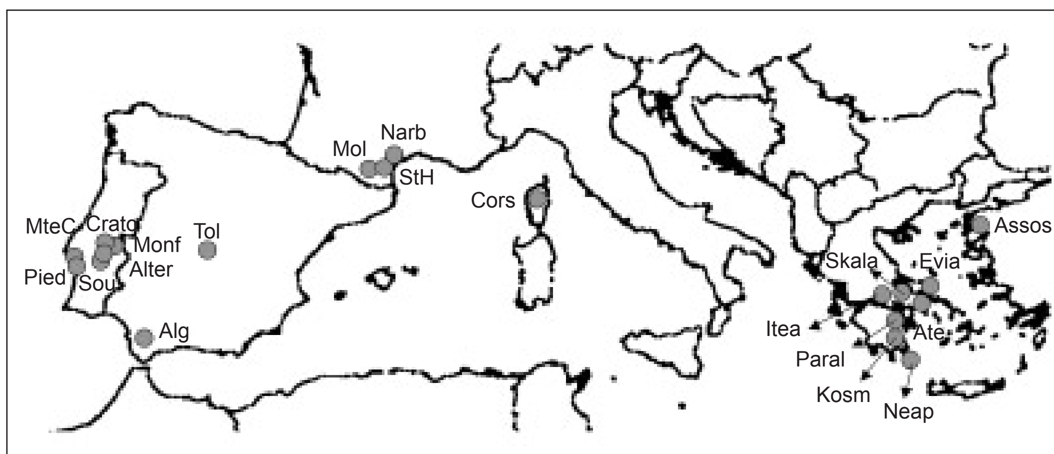


Fig. 1. Map showing sampling localities for *Cicada orni* acoustic signals (for abbreviations see table 1).

Recordings were taken between 09:00 and 18:00 with temperatures ranging from 23 to 38°C.

Sound recordings were digitized using the software Avisoft-SASLab Pro (Specht 2002) at a sampling rate of 44.1 kHz and a resolution of 16 bits, and the time and frequency domains were analyzed. For each specimen, whenever possible, a 1-min recording was used to produce oscillograms, sonagrams (or spectrograms), and mean amplitude spectra (Fig. 2). In the frequency domain, spectra were computed using Fast Fourier transformation with a resolution of 512 points and a Hamming Window.

Avisoft software allows the automatic measurement of time variables derived from the spectrogram, namely the number and duration of acoustic elements (in this case the echemes, which are composed of groups of pulses), and the duration of the interval between them (inter-echeme interval). Spectrum-based variables (peak frequency, minimum frequency, maximum frequency, bandwidths, and quartiles) were also automatically obtained for each echeme (Fig. 2, Table 2). All frequency measurements were calculated from the mean spectrum of each echeme. Time and frequency measurements of the echemes were then averaged, and the mean was taken as the value of the variable for each specimen.

Due to a strong background noise of chorusing cicadas in some samples, which prevented the distinction of pulses, variables like the number of pulses per unit of time were discarded. Moreover, missing data relative to temperature in some populations did not allow us to take this environmental factor into account.

The selected variables were measured for each specimen analyzed. Statistical tests were made using STATISTICA 6.0 software (StatSoft 2001). Variables were not normally distributed, and so nonparametric tests were applied.

The correlation coefficients (Spearman rank order correlations) between all pairs of variables were calculated to determine the relationship

Table 1. Sampled populations of *Cicada orni* with abbreviation names (Abbrev.), number of specimens recorded for sound analyses (*n*), dates of recording, and ambient temperatures

Locality	Abbrev.	*n*	Dates of recording	Temperature range (°C)
IBERIAN PENINSULA				
Algeciras (Andalucía, Spain)	Alg	10	5 Aug. 2001	31~34
Sousel (Alto Alentejo, Portugal)	Sou	11	27 June 2003	27~30
Alter-do-Chão (Alto Alentejo, Portugal)	Alter	7	6~9 Aug. 1997	25~30
Monforte (Alto Alentejo, Portugal)	Monf	16	25 July~7 Aug. 1997	23~38
Crato (Alto Alentejo, Portugal)	Crato	8	27 June 2001	24~26
Piedade (Arrabida, Estremadura, Portugal)	Pied	10	19 July~12 Aug. 1995	-
Monte-da-Caparica (Estremadura, Portugal)	MteC	7	16~22 Sept. 1997	25~30
Toledo (Castilla-La Mancha, Spain)	Tol	2	5 July 2000	-
FRANCE				
Molitg-les-Bains (Languedoc-Roussillon)	Mol	10	17 July 2001	27
St Hippolyte (Languedoc-Roussillon)	StH	11	14 and 17 July 2001	26~33
Narbonne (Languedoc-Roussillon)	Narb	8	16 July 2001	26
CORSICA	Cors	1	-	-
GREECE				
Itea (Athika)	Itea	24	26 and 29 June 2002	-
Skala (Athika)	Skala	4	29 June 2002	-
Evia (Athika)	Evia	12	29 June 2002	-
Athens (Athika)	Ate	18	9~10 July 1997; 15 July 1998	-
Paralio (Peloponnese)	Paral	7	24 June 2002	-
Kosmas (Peloponnese)	Kosm	2	24 June 2002	-
Neapolis (Peloponnese)	Neap	7	25 June 2002	-
TURKEY				
Assos (Aegean coast)	Assos	1	27 June 2003	33
TOTAL		176		

Zoological Studies **44**(1): 81-94 (2005)

between each pair of variables.

To compare the amount of acoustic variation in specimens, populations, or regions, the coefficient of variation (CV) corrected for small samples (Sokal and Rohlf 1981) was applied: $CV = 100 \times (1 + 1/4N) \times SD/Average$; where N is the sample size and SD is the standard deviation. Since the CV is simply the standard deviation expressed as a percentage of the mean, it allows comparisons of variation between sets of data even when the averages greatly differ. The coefficients of variation for each variable were compared at the intra-individual (CVind), intra-population (CVpop), and intra-regional (CVreg) levels.

Altogether, 176 cicadas were analyzed for 13 acoustic variables. In order to reduce the dimen-

Table 2. Description of the acoustic variables analyzed for *Cicada orni* calls

Variable	Description
No. of echemes/s	Number of elements per second
Echeme duration	Duration of each element from start to end (ms)
Inter-echeme interval	Duration between the end of one element and the beginning of the following one (ms)
Echeme period	Duration between the start of one element and the beginning of the following one (ms)
Echeme/inter-echeme interval ratio	Ratio between the echeme duration and the inter-echeme interval
Peak frequency	Frequency of the maximum amplitude on the spectrum
Minimum frequency	Lowest frequency having an amplitude exceeding the threshold (-20 dB)
Maximum frequency	Highest frequency having an amplitude exceeding the threshold (-20 dB)
Bandwidth	Difference between the maximum and minimum frequencies (Hz)
25% Quartile	Frequency below which 25% of the total energy of the spectrum resides
50% Quartile	Frequency below which 50% of the total energy of the spectrum resides (mean frequency of the spectrum)
75% Quartile	Frequency below which 75% of the total energy of the spectrum resides
75% Quartile - 25% Quartile	Difference between the upper (75% quartile) and lower (25% quartile) quartiles of the spectrum (a measure of the pureness of the sound)

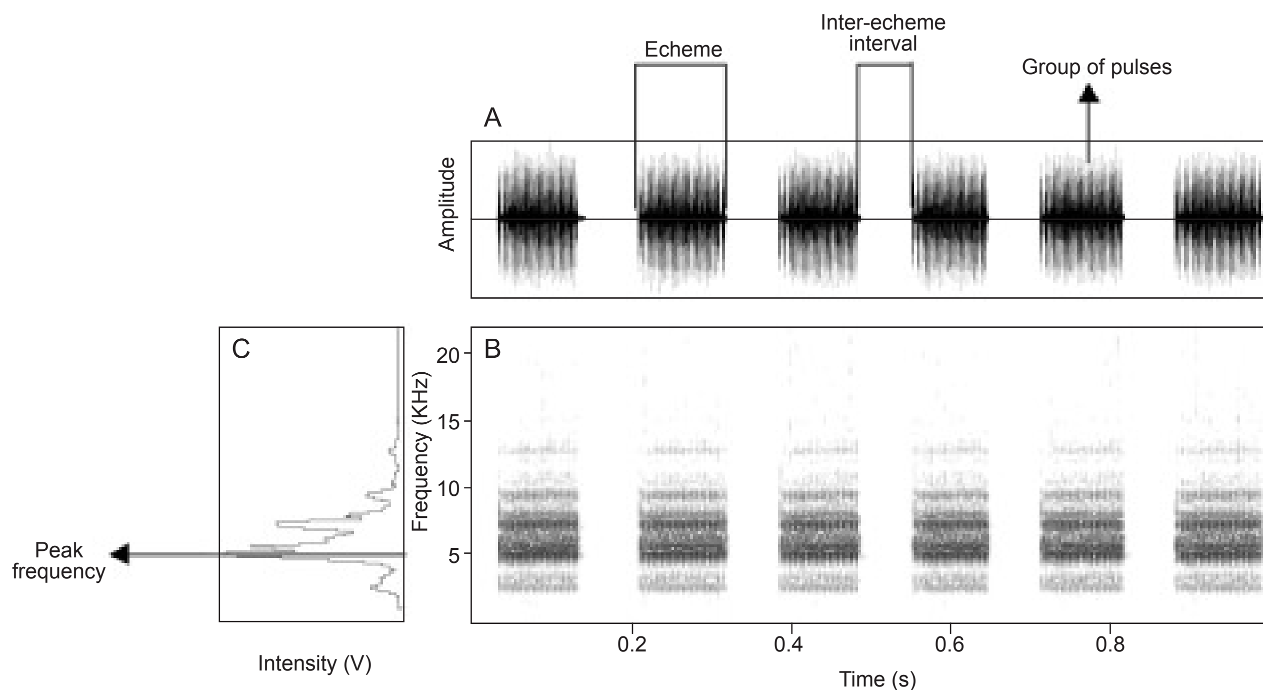


Fig. 2. Calling song of *Cicada orni*. (A) Oscillogram (amplitude vs. time); (B) sonagram or spectrogram (frequency vs. time); and (C) mean amplitude spectrum (frequency vs. amplitude).

sionality of the data matrix and to test for intraspecific acoustic differentiation among these specimens, an ordination through principal component analysis (PCA) was carried out. Data were standardized. The Kaiser criterion was used to retain only components with eigenvalues greater than 1. After this analysis, Kruskal-Wallis (KW) nonparametric tests were used to compare component scores between different regions. When results were significant ($p < 0.05$), Mann-Whitney U tests (MW) were performed to compare pairs of regions (Dytham 2003).

Kruskal-Wallis nonparametric tests were also used to compare several independent samples (regions or populations) for each acoustic variable, followed by Mann-Whitney U tests for 2-sample comparisons when the KW tests were significant (Dytham 2003).

RESULTS

The calling song of *C. orni* is produced by males which can sing continuously from a single site for hours, sometimes chorusing with other males. It is made up of a regular repetition of echemes, which are composed of a variable number of groups of pulses (Fig. 2).

Descriptive statistics of all the acoustic variables analyzed for all specimens studied are shown in table 3. This signal can be described in the time domain as having echemes of 0.08 ± 0.03 (average ± standard deviation) s in duration separated by intervals of 0.15 ± 0.07 s. The spectral characteristics of the signal showed a peak frequency of 4825 ± 486 Hz, and a bandwidth (at -20 dB) of 7233 ± 1437 Hz. Moreover, the mean frequency was 5544 ± 441 Hz, and the frequency difference between the upper and lower quartiles of these species was relatively small (2357 ± 585 Hz) compared with those of other cicada species. The calling songs were more variable in the time (with CVs which ranged from 27.80% to 75.25%) than in frequency domain (with CVs which ranged from 7.60% to 24.84%).

Strong nonparametric correlations between frequency variables were detected (e.g., R_S (peak frequency vs. 25% quartile) = 0.84; R_S (bandwidth vs. maximum frequency) = 0.97; $p < 0.001$). On the other hand, no significant correlations were found between the time and frequency variables. Also, time variables such as echeme duration and the interval between echemes were weakly but significantly correlated ($R_S = -0.35$; $p < 0.001$).

Recordings taken using either the Sony or UHER recorders were compared for the three French populations for which there were specimens recorded with Sony and others with UHER. It was found that there were some high-frequency components in the recordings taken with UHER which did not appear in the recordings taken with Sony. As such, the acoustic variables of maximum frequency, bandwidth, quartile 75%, and the difference between the upper and lower quartiles were significantly higher for the UHER recordings (MW tests, $p < 0.05$). The opposite was found with the minimum frequency (Fig. 3). On account of this, the UHER recordings for these variables (two from

Table 3. Descriptive statistics of the acoustic variables of *Cicada orni* specimens. *n*, number of specimens recorded for sound analyses; SD, standard deviation; CV, coefficient of variation

Variable	*n*	Average ± SD	Minimum	Maximum	CV (%)
No. of echemes/s	176	4.69 ± 1.30	2.23	7.49	27.80
Echeme duration (s)	176	0.08 ± 0.03	0.03	0.21	36.42
Inter-echeme interval (s)	176	0.15 ± 0.07	0.06	0.36	48.28
Echeme period (s)	176	0.23 ± 0.07	0.13	0.45	30.82
Echeme/inter-echeme interval ratio	176	0.71 ± 0.53	0.17	3.31	75.23
Peak frequency (Hz)	176	4824.84 ± 485.61	3851.22	6584.29	10.08
Minimum frequency (Hz)	158	2191.43 ± 390.84	1244.27	4074.77	17.86
Maximum frequency (Hz)	158	9430.15 ± 1397.45	6966.84	13347.91	14.84
Bandwidth (-20 dB) (Hz)	158	7232.92 ± 1437.02	3626.52	11280.30	19.19
25% Quartile (Hz)	176	4658.30 ± 353.53	3603.51	5784.14	7.60
50% Quartile (Hz) (mean freq.)	176	5544.42 ± 440.93	4611.31	6640.51	7.96
75% Quartile (Hz)	158	7038.39 ± 751.82	5405.18	9948.06	10.70
75% Quartile - 25% Quartile (Hz)	158	2356.54 ± 584.83	1242.40	5205.60	24.84

Toledo, five from Molitg-les-Bains, six from St. Hippolyte, and three from Narbonne) were removed from the analysis.

Results of the principal component analysis are shown in figure 4. The 1st four components accounted for 89.06% of the total variation. More than half (64.5%) of the variation in the study was explained by the 1st two components, and more than 3/4 (78.08%) by the 1st three components (C1 = 38.83%, C2 = 25.67%, and C3 = 13.58%). There was no complete separation between the three different regions (Iberian Peninsula, France, and Greece) and/or populations for any combination of the 1st three axes, since specimens mostly appeared as a homogeneous group (Fig. 4).

However, when plotting components 1 and 2, specimens from southeastern Europe (Greece) tended to group on the 3rd and 4th quadrants while those of western Europe (Iberian Peninsula and France) appeared mainly in the 1st and 2nd ones (Fig. 4). Thus, there was a slight separation along the 2nd axis (component 2). In fact, KW nonparametric analysis revealed significant differences for this axis between different regions. MW tests revealed significant differences between Greece and each of the other two regions for component 2, and also for component 1 when comparing Greece with the Iberian Peninsula.

Factor loadings were considerably high (Table 4) for some of the variables and considering that component 2 had the highest loading for interecheme interval duration (-0.92), this was the vari-

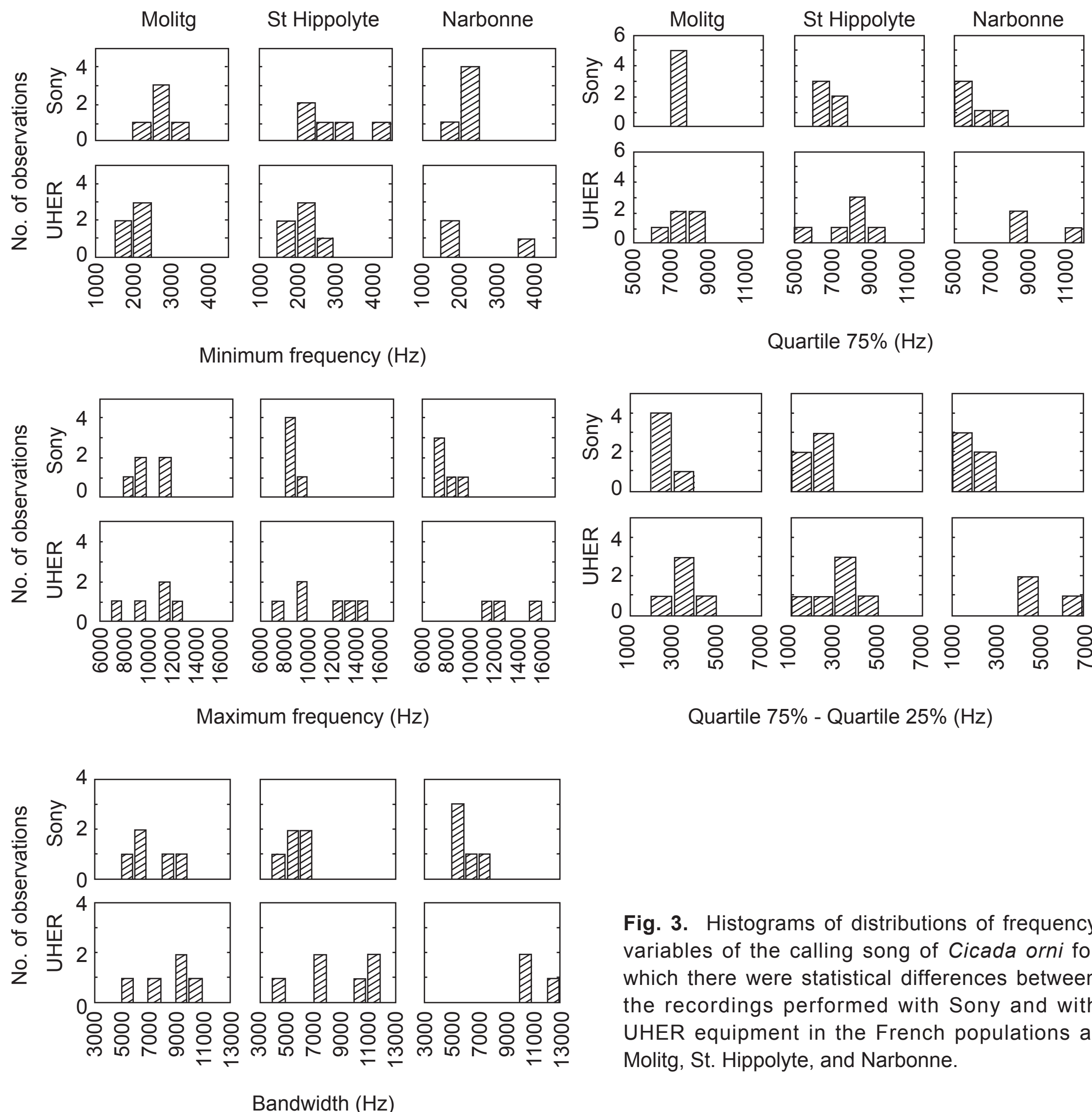


Fig. 3. Histograms of distributions of frequency variables of the calling song of *Cicada orni* for which there were statistical differences between the recordings performed with Sony and with UHER equipment in the French populations at Molitg, St. Hippolyte, and Narbonne.

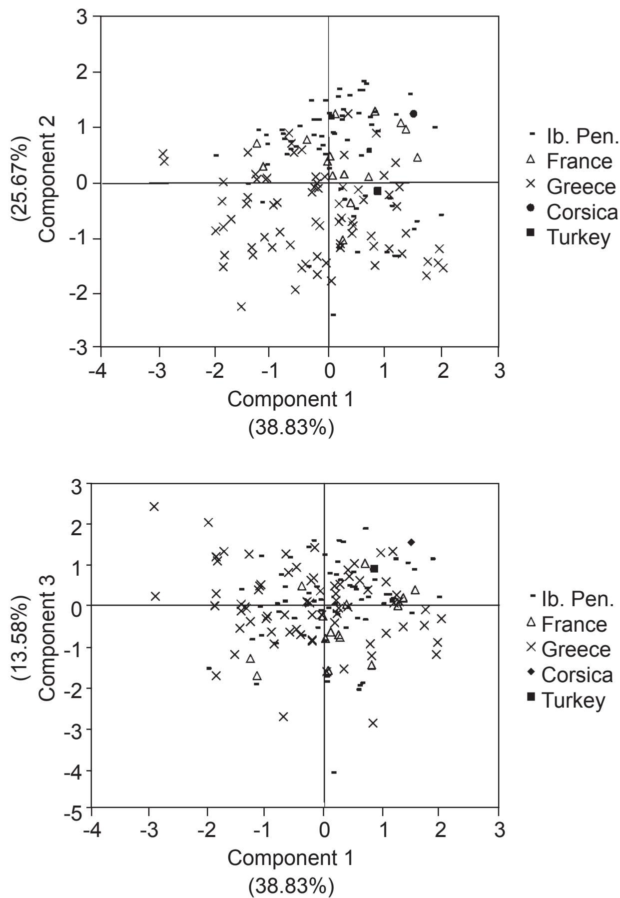


Fig. 4. Bidimensional diagrams of relationships between specimens of *Cicada orni* (176 OTUs) of the principal component analysis based on a correlation matrix between 13 acoustic characters.

able which contributed most to the separation between the populations of Greece and the other two regions. In fact, as shown below, Greek specimens had inter-echeme intervals of longer duration.

There were also significant differences between the populations of France and the Iberian Peninsula, but just for component 3, and which however explained only 13.58% of the total variation.

With respect to the Corsican and Turkish specimens, the Corsican one was grouped with the western European samples, while the Turkish one was grouped with the southeastern ones.

When only uncorrelated variables were used (echeme duration, inter-echeme interval duration, and difference between the upper and lower quartiles), the PCA results were similar (results not shown).

In a comparison of the three regions, the Iberian Peninsula, France, and Greece, the number of echemes per second, inter-echeme interval duration, and echeme period differed for each pair of regions (MW tests, $p < 0.05$). These differences were due to generally longer inter-echeme intervals in Greek cicadas, being intermediate in French cicadas and shorter in the Iberian ones (Fig. 5). The only acoustic variables for which there were no differences between regions (KW tests, $p > 0.05$) were echeme duration, the 75% quartile, and the difference between the upper and lower quartiles. For the remaining frequency vari-

Table 4. Factor loadings of the principal component analysis (PCA) based on a correlation matrix between 13 acoustic variables for *Cicada orni* (using 176 OTUs)

Variable	Component 1	Component 2	Component 3	Component 4
No. of echemes/s	-0.222552	0.851582	0.304946	0.283294
Echeme duration	-0.177807	0.269043	-0.792527	-0.497948
Inter-echeme interval	0.338561	-0.920525	0.026469	-0.086444
Echeme period	0.285371	-0.858498	-0.277505	-0.282083
Echeme/inter-echeme interval ratio	-0.233615	0.672798	-0.584044	-0.323306
Peak frequency	-0.534498	-0.474767	-0.309437	0.388026
Minimum frequency	-0.818619	-0.204684	0.304262	-0.368513
Maximum frequency	-0.747884	-0.314878	-0.298219	0.419953
Bandwidth (-20 dB)	-0.902071	-0.261265	-0.050210	0.145775
25% Quartile	-0.947363	-0.044019	0.075287	-0.044032
50% Quartile	-0.766465	0.136966	0.281331	-0.315707
75% Quartile	-0.258655	0.133888	-0.478907	0.518750
75% Quartile - 25% Quartile	-0.914037	-0.171322	0.175947	-0.230332

Zoological Studies **44**(1): 81-94 (2005)

ables, Greek samples showed significantly higher values than the others, while samples from the Iberian Peninsula and France were similar to each other. The minimum frequency was an exception; French populations differed from the remaining ones in having higher minimum frequencies. With respect to the ratio of echeme/interval, Greek samples showed lower values.

In Greece all of the variables showed some differences among populations (except for the ratio of the echeme/inter-echeme interval) (KW tests, $p < 0.05$). In the Iberian Peninsula and France, every time variable but not every frequency variable showed some differences among populations (KW tests, $p < 0.05$). The frequency variables revealing significant differences among popula-

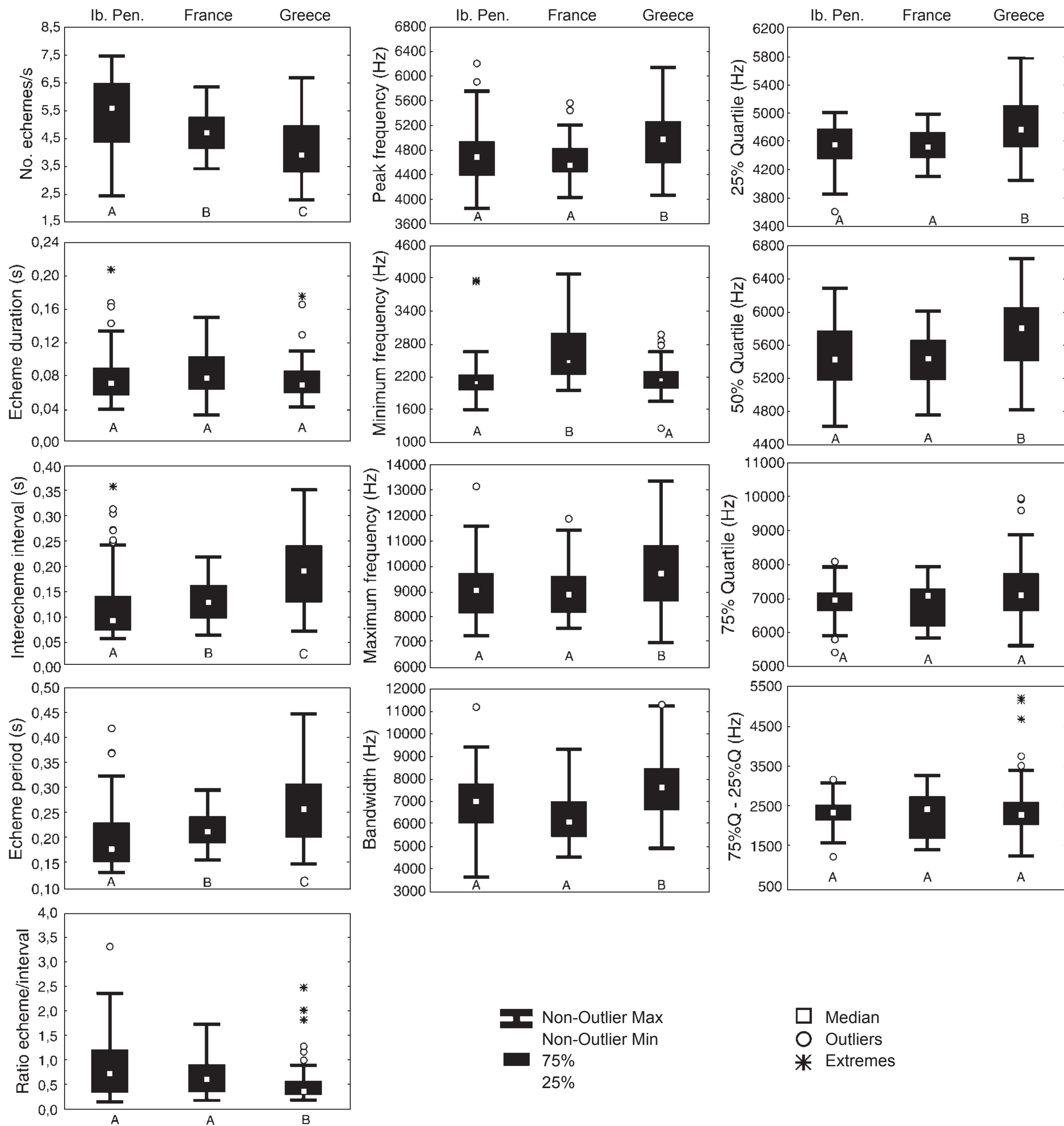


Fig. 5. Boxplots of the 13 acoustic variables of the calling song of *Cicada orni* investigated for the three regions (71 specimens from the Iberian Peninsula, 29 specimens from southern France, and 74 specimens from continental Greece). Different letters indicate that significant differences (Mann-Whitney, $p < 0.05$) exist between those regions.

tions on the Iberian Peninsula were the 25% quartile, the difference between the upper and lower quartiles, and the minimum and maximum frequencies. For the French samples, these variables were the 75% quartile, the difference between the upper and lower quartiles, and the minimum frequency.

Data for the calling song of *C. orni* by Joermann and Schneider (1987) and Popov (1975) for Yugoslavia and southern USSR, respectively, were closer to the values here obtained for the Iberian Peninsula and French cicadas than to the Greek ones, especially for the inter-echeme interval duration. In fact, the range of values of this variable obtained for Yugoslavian samples was 39~127 ms (Joermann and Schneider 1987), while that for southern USSR was 45~87 ms (calculated from Popov (1975) by subtracting the mean echeme duration from the mean period duration of each specimen). These values were closer to the median values obtained for the Iberian Peninsula and France (lower than 150 ms) than to the Greece median value (higher than 150 ms) (Fig. 5).

When comparing populations, some deviated quite obviously from the others for some variables (Fig. 6). The Sousel (Portugal) population specimens had, on average, longer inter-echeme intervals than the remaining Portuguese populations, which was also reflected in longer echeme periods and shorter ratios of echeme/interval (similar to the values found in Greek populations). The Toledo (Spain) population had lower 25% quartile and 50% quartile values, which might have been due to some interference of the recorded material (UHER). In Greece, the Kosmas population showed longer echemes and shorter intervals than the other Greek populations investigated.

In general, the frequency characters were found to be less variable than the time ones, as referred to above. This tendency was true at all three levels of the analysis. The variable presenting the highest coefficient of variation for all levels of comparison was the ratio of echeme/interval (CVind > 28%, CVpop > 48%, and CVreg > 43%). On the other hand, the variables with the lowest coefficients of variation at all levels were the 25%, 50%, and 75% quartiles (CVind < 3% and CVpop and CVreg < 10%) (Table 5). For most variables, the Greek cicadas showed a higher intra-individual coefficient of variation (CVind), a lower intra-population coefficient of variation (CVpop), and a higher intra-region coefficient of variation (CVreg), than did cicadas from the other regions (Table 5).

The mean intra-individual coefficients of variation were always lower than intra-population ones for all variables in all regions (Table 5), which allowed the use of a mean value of each variable measured for each specimen. In fact, the percentage of specimens presenting a CVind higher than the CVpop was always lower than 50% (with the exception of echeme duration for Greece) (Table 6).

In contrast, CVpop values were more similar to CVreg ones for most of the variables than were CVind relative to CVpop values (Table 5). Low percentages of populations with CVpop > CVreg were only found for the inter-echeme interval and echeme period, as well as the minimum frequency in the Iberian Peninsula, while Greece showed low values for echeme duration and the echeme/interval ratio (Table 6). These values were probably due to the atypical populations of Sousel (Portugal) and Kosmas (Greece) which, as referred to before, have very distinct values for these variables (Fig. 6).

DISCUSSION AND CONCLUSIONS

Detailed acoustic analyses of the time and frequency domains revealed some variations in the calling song among specimens and populations of *C. orni* and conform to previously known data. Values obtained in the present study for time and frequency variables of this song are within the range of variation observed in the former Yugoslavia (Joermann and Schneider 1987) and southern USSR (Popov 1975). The results obtained by Boulard (1995 2000b) in Provence (France) also fall within the observed range of the French populations of this study as well as of those reported by Fonseca (1991) for Portugal.

Principal component analysis indicated that in spite of some variations, the calling songs of *C. orni* investigated here constitute a relatively homogeneous group, without clear-cut acoustic differences as expected for conspecifics. However, some significant differences were found among regions, a fact not concordant with the prediction by Paterson (e.g., 1985) that specific mate recognition systems should be invariant within species. Songs from southeastern Europe (Greece) tended to group separately from those of western Europe (Iberian Peninsula and France), and the inter-echeme interval was the variable that contributed most to this separation. In fact, Greek male songs showed longer inter-echeme intervals on average

than the songs of western Europe as found previously by Quartau et al. (1999). An exception is the Portuguese population at Sousel, which generally showed longer inter-echeme intervals. In Greece and Portugal, the influence of temperature is an unlikely explanation for such differences since the recordings were performed within a similar range of temperatures at all sites. Conversely, the population at Kosmas (Greece) showed shorter inter-echeme intervals and longer echemes than the typical Greek populations. As this population was the only one recorded at a high elevation (> 1000 m), more-thorough studies should be carried out at this locality as well as at Sousel to better understand these unusual findings.

In addition to the inter-echeme interval, specimens from Greece showed significant differences in relation to the remaining studied regions for almost every acoustic variable in both the time and frequency domains, as shown in the pairwise comparisons. This acoustic differentiation is in agreement with genetic data, namely, isozymes (Quartau et al. 2001), microsatellites (Seabra in prep.), and mitochondrial DNA (Pinto-Juma in prep.) and is probably due to the considerable isolation of such populations throughout the mountain ranges of the Balkans during the ice ages.

It is interesting to note that the songs of the Corsican specimens as well as the Yugoslavian (Joermann and Schneider 1987) and southern USSR (Popov 1975) ones (for time variables, especially inter-echeme intervals) were closer to those of the Iberian Peninsula and France. On the other hand, and as expected due to reasons of geographic proximity, the single Turkish male investigated here had greater similarities with the Greek populations than to other populations.

The only time variable that showed no significant differences between each pair of regions, i.e. that proved to be quite constant across the geographic range of this cicada, was echeme duration. In a comparison of *C. orni* with its sibling *C. mordoganensis*, echeme duration is the variable which more-obviously readily distinguishes these species, since there is minimal overlap (Simões et al. 2000). Thus, echeme duration is probably one of the most important parameters encoding species-specific information for species recognition. On the other hand and as referred to previously, the inter-echeme interval is quite variable within *C. orni*, and its range of variation considerably overlaps that of *C. mordoganensis*. Therefore, it is expected that the inter-echeme interval alone is not an important parameter for species recognition and isolation in these species.

Characters for the time domain were found to be more variable (i.e., with higher coefficients of variation) than those for the frequency domain as reported for other cicadas (e.g., Sueur and Aubin 2002). This was expected since the spectrum is especially constrained by the physical properties of the sound-producing organ. In fact, as stated for

Table 5. Coefficients of variation for each acoustic variable analyzed within specimens (CVind) within populations (CVpop) (average ± standard deviation) and within regions (CVreg) for the three areas investigated

	Iberian Peninsula			France			Greece		
	CVind	CVpop	CVreg	CVind	CVpop	CVreg	CVind	CVpop	CVreg
No. of echemes/s	-	18.36 ± 8.21	17.48	-	16.08 ± 0.54	10.26	-	20.11 ± 5.00	20.58
Echeme duration	15.36 ± 7.02	30.33 ± 11.90	30.34	16.10 ± 7.29	28.10 ± 3.21	28.93	19.41 ± 6.76	22.4 ± 14.47	41.67
Inter-echeme interval	17.88 ± 8.04	33.39 ± 14.03	48.37	16.42 ± 8.36	27.73 ± 2.64	21.49	27.17 ± 10.21	29.59 ± 4.66	34.00
Echeme period	12.18 ± 7.53	18.98 ± 8.07	25.62	11.70 ± 6.08	16.03 ± 0.97	10.14	18.8 ± 7.44	20.07 ± 4.61	22.43
Echeme/inter-echeme interval ratio	28.56 ± 13.90	54.50 ± 15.81	50.72	30.59 ± 19.18	54.59 ± 17.20	43.97	40.27 ± 19.13	48.72 ± 38.42	79.10
Peak frequency	4.97 ± 4.20	8.34 ± 3.43	3.4	3.27 ± 3.14	6.77 ± 3.40	2.48	5.98 ± 3.14	6.72 ± 3.71	6.17
Minimum frequency	4.85 ± 4.30	12.61 ± 9.87	10.66	11.85 ± 9.71	15.90 ± 9.65	15.12	9.74 ± 8.58	9.85 ± 4.57	6.19
Maximum frequency	6.83 ± 5.99	13.05 ± 4.96	4.54	4.20 ± 4.12	10.26 ± 5.11	12.40	6.02 ± 3.91	11.68 ± 5.74	11.48
Bandwidth	9.62 ± 8.01	18.47 ± 9.93	6.62	9.73 ± 6.92	18.4 ± 4.88	14.69	8.94 ± 5.52	14.82 ± 6.05	14.31
25% Quartile	1.06 ± 0.53	5.54 ± 1.40	2.77	0.77 ± 0.45	4.14 ± 1.32	2.23	1.74 ± 1.36	5.83 ± 2.17	5.17
50% Quartile	1.48 ± 1.08	7.35 ± 2.32	1.78	1.38 ± 0.70	6.2 ± 2.25	5.11	2.06 ± 1.55	6.42 ± 3.54	6.50
75% Quartile	1.79 ± 1.96	7.18 ± 1.88	2.97	2.16 ± 1.95	7.75 ± 2.84	8.71	2.93 ± 2.49	9.97 ± 4.87	9.58
75% Quartile - 25% Quartile	4.97 ± 4.65	14.71 ± 5.30	8.35	6.57 ± 6.35	20.14 ± 7.95	22.07	8.21 ± 6.44	23.58 ± 7.95	23.41

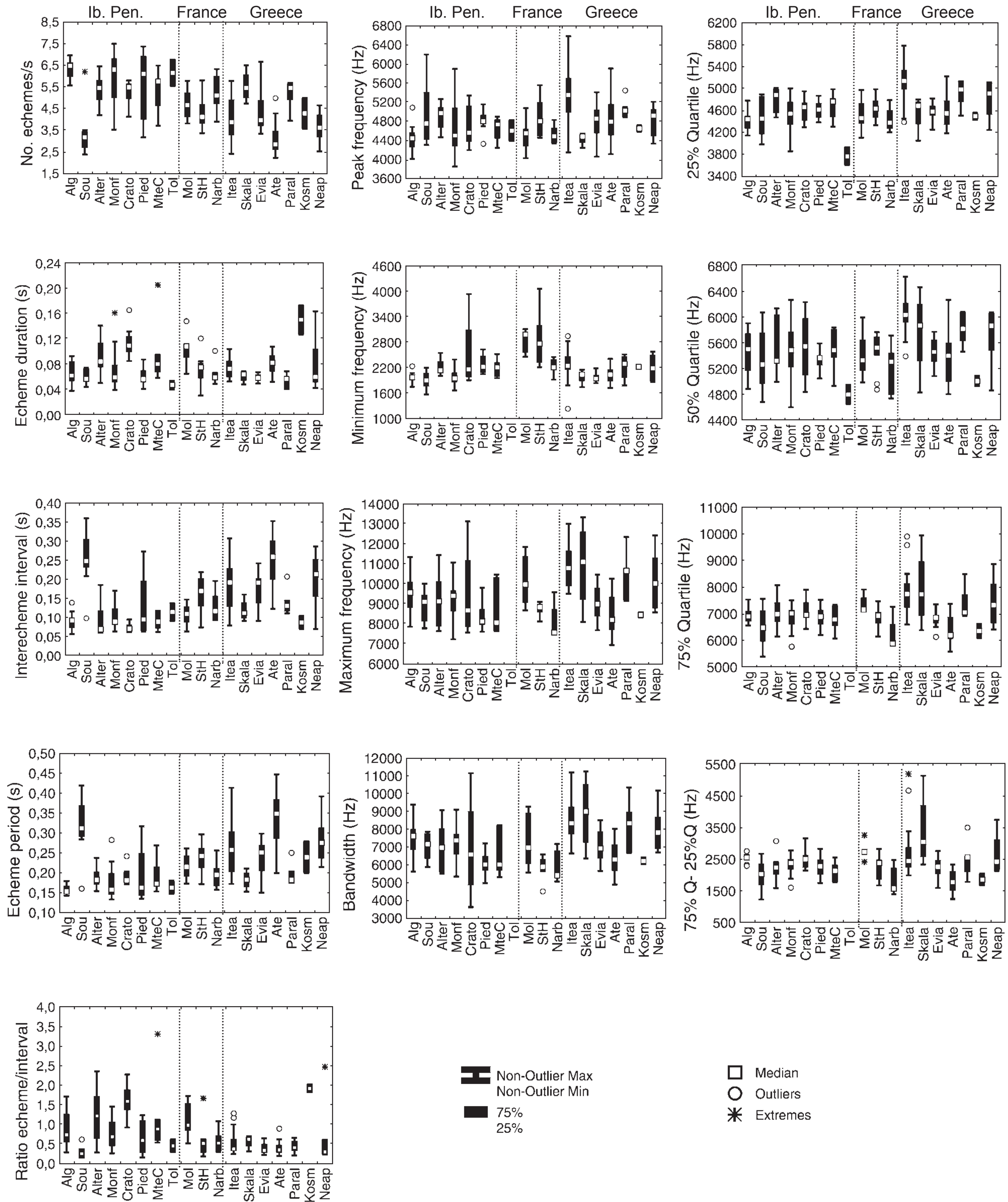


Fig. 6. Boxplots of the 13 acoustic (including five time and eight frequency) variables of the calling song of *Cicada orni* investigated for each population. (for an explanation of the abbreviations see table 1).

Zoological Studies **44**(1): 81-94 (2005)

insects in general by Stumpner and Helversen (2001), the time pattern of the song is usually more important in the recognition of a conspecific signal than its spectrum, since this latter differs much less between related species.

Intra-individual variability was found to be higher in the Greek populations than in those from the other regions investigated. This might be related to a stronger interaction among specimens due to the higher densities of cicadas found in Greece. Consequently, at the intra-population level, the acoustic variability was low. On the other hand, Greek songs also proved to have a higher level of intra-regional variation than songs from the remaining areas. An explanation for this might be the very high density of cicadas in Greece and consequently a more-pronounced chorus effect which is important for attracting females and pair formation (as has also been shown for *Magicicada* spp. by Karban (1981) and Williams and Kimberly (1991)). Therefore, this chorus effect might have promoted a scattered distribution and population isolation and, as such, enhanced divergence between different cicada populations.

It is likely that significant regional differentiation might have occurred during the glacial periods, when *C. orni* was split into isolated populations in at least two main Mediterranean areas – the Iberian Peninsula in the west and the southeastern Aegean area (including the mainland and islands) (Taberlet et al. 1998). Therefore, the extant eastern populations of *C. orni* would have been derived from such isolated Aegean refuge stock, an isolation which would have promoted substantial microevolutionary genetic diversification as in other groups (Avise and Walker 1998). In such a view, the Aegean area and West Asia Minor might correspond to the original area of dispersal and expansion of *C. orni* during Pleistocene climatic cycles, as probably occurred with many widespread species over the entire Mediterranean basin (Oosterbroek and Arntzen 1992, Oosterbroek 1994). In fact and in corroboration with Hewitt (e.g., 1996 1999) and others, the southern European population genomes of many animal species show greater haplotype diversity. As suggested before, this was probably due to the existence of southern ice age refugia, such as the Aegean coastal zones, valleys, and islands, which were subjected to repeated cycles of isolation through the mountain ranges of the Balkans, and from where expansion cycles took place for colonizing more-northerly and western areas during interglacial warmer periods.

In summary, further investigations on *C. orni* are needed not only at the acoustic but also at the genetic and phylogeographic levels to foster a better understanding of the history of the intriguing distinctiveness of its southeastern populations.

Table 6. Percentage of specimens presenting CVind > CVpop and percentage of populations presenting CVpop > CVreg. CVind coefficient of variation within individuals; CVpop coefficient of variation within populations; and CVreg coefficient of variation within regions

	Iberian Peninsula		France		Greece	
	CVind > CVpop	CVpop > CVreg	CVpop > CVind	CVreg > CVpop	CVind > CVpop	CVpop > CVreg
No. of echemes/s	-	50%	-	100%	-	57.10%
Echeme duration	9.86%	37.50%	13.79%	33.30%	52.17%	14.30%
Inter-echeme interval	9.86%	25%	10.35%	100%	34.78%	28.60%
Echeme period	15.49%	12.50%	20.69%	100%	34.78%	28.60%
Echeme/inter-echeme interval ratio	9.86%	62.50%	10.35%	66.70%	27.54%	14.30%
Peak frequency	21.74%	100%	26.67%	100%	26.09%	57.10%
Minimum frequency	11.59%	14.30%	33.33%	33.30%	34.78%	71.40%
Maximum frequency	14.49%	100%	13.33%	33.30%	10.15%	42.90%
Bandwidth	15.94%	100%	13.33%	66.70%	14.49%	57.10%
25% Quartile	0%	100%	0%	100%	2.90%	57.10%
50% Quartile	0%	100%	0%	66.70%	2.90%	42.90%
75% Quartile	2.90%	100%	0%	33.30%	4.35%	42.90%
75% Quartile - 25% Quartile	8.70%	85.70%	0%	33.30%	4.35%	28.60%

Acknowledgments: We are grateful to Jérôme Sueur (Muséum National d'Histoire Naturelle, Paris, France) for recording some of the cicada specimens. We also thank Sakis Drosopoulos (Agricultural University of Athens, Athens, Greece), Stéphane Puissant (Muséum National d'Histoire Naturelle Muséum, Paris, France), and Genage André (Faculdade de Ciências de Lisboa, Portugal), for help in the field, and Mónica Ribeiro and Teresa Fernandes (Faculdade de Ciências de Lisboa, Lisboa, Portugal) for some of the Portuguese recordings. This study was partially financially supported by Fundação para a Ciência e a Tecnologia (FCT, Portugal through PhD grants SFRH/BD/1027/2000 and PRAXIS BD/18229/98), by project no. 537C1 (Portuguese ICCTI and French Embassy in Lisbon), and by the Centre for Environmental Biology (FCT). We are also appreciative of the referees and the English Editor whose comments considerably improved the manuscript.

REFERENCES

Alexander RD. 1967. Acoustical communication in arthropods. Annu. Rev. Entomol. **12:** 495-526.

Avise JC, D Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. Roy. Soc. Lond. B Biol. **265:** 457-463.

Bennet-Clark HC. 1998. How cicadas make their noise. Sci. Am. **5:** 36-39.

Boulard M. 1995. Postures de cymbalisation, cymbalisations et cartes d'identité acoustique de cigales. 1. Généralités et espéces méditerranéennes (Homoptera, Cicadoidea). EPHE Biol. Évol. Insectes **7/8:** 1-72.

Boulard M. 2000a. Appareils, productions et communications sonores chez les insectes en général et chez les cigales en particulier. EPHE Biol. Évol. Insectes **13:** 75-110.

Boulard M. 2000b. Espèce, milieu et comportement. EPHE Biol. Évol. Insectes **13:** 1-14.

Claridge MF. 1985. Acoustic signals in the Homoptera: behavior, taxonomy, and evolution. Annu. Rev. Entomol. **30:** 297-317.

Claridge MF, MR Wilson, JS Singhrao. 1979. The songs and calling sites of two European cicadas. Ecol. Entomol. **4:** 225-229.

Cooley JR, DC Marshall. 2001. Sexual signalling in periodical cicadas, *Magicicada* spp. (Hemiptera: Cicadidae). Behaviour **138:** 827-855.

Dytham C. 2003. Choosing and using statistics, a biologist's guide. 2nd ed. Oxford, UK: Blackwell Science, 248 pp.

Fonseca PJ. 1991. Characteristics of the acoustic signals in nine species of cicadas (Homoptera, Cicadidae). Bioacoustics **3:** 173-192.

Fonseca PJ, MA Revez. 2002. Song discrimination by male cicadas *Cicada barbara lusitanica* (Homoptera, Cicadidae). J. Exp. Biol. **205:** 1285-1292.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. **58:** 247-276.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. **68:** 87-112.

Joermann G, H Schneider. 1987. The songs of four species of cicada in Yugoslavia (Homoptera: Cicadidae). Zool. Anz. **219:** 283-296.

Karban R. 1981. Flight and dispersal of periodical cicadas. Oecologia **49:** 385-390.

Moore TE. 1993. Acoustic signals and speciation in cicadas (Insecta: Homoptera: Cicadidae). *In* DR Lees, D Edwards, eds. Evolutionary patterns and processes. Linnean Society Symposium no. 14, London: Academic Press, pp. 269-284.

Nast J. 1972. Palaearctic Auchenorrhyncha (Homoptera), an annotated checklist. Warszawa, Poland: Polish Scientific Publications, 550 pp.

Oosterbroek P. 1994. Biodiversity of the Mediterranean region. Syst. Conserv. Eval. **50:** 289-307.

Oosterbroek P, JW Arntzen. 1992. Area-cladograms of circum-Mediterranean taxa in relation to Mediterranean paleogeography. J. Biogeogr. **19:** 3-20.

Paterson HEH. 1985. The recognition concept of species. *In* ES Vrba, ed. Species and speciation. Pretoria, SA: Transvaal Museum Monograph no. 4, pp. 21-29.

Patterson IJ, G Massei, P Genov. 1997. The density of cicadas *Cicada orni* in Mediterranean coastal habitats. Ital. J. Zool. **64:** 141-146.

Popov AA. 1975. The structure of the tymbals and the characteristics of the sound signals in singing cicadas (Homoptera, Cicadidae) in the southern regions of the USSR. Entomol. Rev. **54:** 7-35.

Pringle JWS. 1954. A physiological analysis of cicada song. J. Exp. Biol. **31:** 525-556.

Puissant S, J Sueur. 2001. Contribution à l'étude des Cigales de Corse (Hemiptera, Cicadidae). Bull. Soc. Entomol. Fr. **106:** 429-436.

Quartau JA, MT Rebelo, PC Simões, TM Fernandes, MF Claridge, S Drosopoulos, JC Morgan. 1999. Acoustic signals of populations of *Cicada orni* L. in Portugal and Greece (Homoptera: Auchenorrhyncha: Cicadomorpha: Cicadidae). Reichenbachia **33:** 7-80.

Quartau JA, M Ribeiro, PC Simões, MM Coelho. 2001. Genetic divergence among populations of two closely related species of Cicada Linnaeus (Hemiptera: Cicadoidea) in Portugal. Insect Syst. Evol. **32:** 99-106.

Simões PC, MM Boulard, MT Rebelo, S Drosopoulos, MF Claridge, JC Morgan, JA Quartau. 2000. Differences in the male calling songs of two sibling species of *Cicada* (Hemiptera: Cicadoidea) in Greece. Eur. J. Entomol. **97:** 437-440.

Sokal RR, FJ Rholf. 1981. Biometry. The principles and practice of statistics in biological research. New York, NY: WH Freeman, 859 pp.

Specht R. 2002. Avisoft-SASLab Pro-Sound analysis and synthesis laboratory. vers. 4.1d, Berlin.

StatSoft 2001. STATISTICA vers. 6.0, Tulsa.

Stumpner A, D von Helversen. 2001. Evolution and function of auditory systems in insects. Naturwissenschaften **88:** 159-170.

Sueur J, T Aubin. 2002. Acoustic communication in the Palaeartic red cicada, *Tibicina haematodes*: chorus organisation, calling-song structure, and signal recognition. Can. J. Zool. **80:** 126-136.

Sueur J, S Puissant, PC Simões, S Seabra, M Boulard, JA Quartau. 2004. Cicadas from Portugal: revised list of

species with eco-ethological data (Hemiptera: Cicadidae). Insect Syst. Evol. (in press)

Taberlet P, L Fumagali, AG Wust-Saucy, JF Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. **6:** 289-301.

Villet M. 1992. Responses of free-living cicadas (Homoptera: Cicadidae) to broadcasts of cicada songs. J. Entomol. Soc. S. Afr. **55:** 93-97.

Williams KS, GS Kimberly. 1991. Dynamics of periodical cicada chorus centers (Homoptera: Cicadidae: Magicicada). J. Insect Behav. **4:** 275-291.

Chapter 4.

Geographical variation in *Cicada barbara* (Stål)

4.1. Published: Pinto-Juma GA, Quartau JA, Bruford MW. 2008. Population Structure of *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco based on mitochondrial DNA analysis. *Bulletin of Entomological Research* **98**: 1-14.
4.2. Published: Pinto-Juma GA, Seabra SG, Quartau JA. 2008. Patterns of acoustic variation in *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco. *Bulletin of Entomological Research* **98**:15-25.

Bulletin of Entomological Research (2008) **98**, 15–25
doi:10.1017/S0007485307005573
© 2008 Cambridge University Press
Printed in the United Kingdom

Population structure of *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco based on mitochondrial DNA analysis

G.A. Pinto-Juma[1,2]*, J.A. Quartau[1] and M.W. Bruford[2]

[1]Centro de Biologia Ambiental/Departamento de Biologia Animal, Faculdade de Ciências de Lisboa, C2, Campo Grande, 1749-016 Lisboa, Portugal: [2]School of Biosciences, Cardiff University, Cardiff CF10 3US, Wales, UK

Abstract

We assess the genetic history and population structure of *Cicada barbara* in Morocco and the Iberian Peninsula, based on analysis of the mitochondrial cytochrome *b* gene. The divergence between Morocco and the Iberian Peninsula populations was strongly corroborated by the molecular data, suggesting genetically isolated populations with a low level of gene flow. The Ceuta population from Spanish North Africa was more similar to the Iberian populations than the surrounding Moroccan populations, suggesting that the Strait of Gibraltar has not been acting as a strict barrier to dispersal while the Rif Mountains have. The Iberian Peninsula specimens showed a signature of demographic expansion before that which occurred in Morocco, but some of the assumptions related to the demographic parameters should be considered with caution due to the small genetic variation found. The high haplotype diversity found in Morocco implies higher demographic stability than in the Iberian Peninsula populations. These results do not, however, suggest a Moroccan origin for Iberian cicadas; but the most northwest region in Africa, such as Ceuta, might have acted as a southern refuge for Iberian cicadas during the most severe climatic conditions, from where they could expand north when climate improved. The separation of two subspecies within *C. barbara* (*C. barbara lusitanica* and *C. barbara barbara*) finds support with these results.

Keywords: *Cicada barbara*, Hemiptera, mitochondrial cytochrome *b*, subspecies, genetic variation, population structure, demographic expansions, Strait of Gibraltar

(Accepted 28 August 2007)

Introduction

Cicada barbara Stål is usually found associated with olive, pine and, occasionally, eucalyptus trees and occurs in scattered warm habitats in the Iberian Peninsula, whereas it is widely distributed in northwest Africa (Boulard, 1982; Quartau, 1988). Isozyme studies have revealed that *C. barbara* is a genetically homogeneous species in Portugal, especially when compared to *C. orni*, a species with which it occurs in sympatry in some areas of the Iberian Peninsula (Quartau *et al.*, 2000, 2001). One potential explanation for this homogeneity is that *C. barbara* is a relatively recent

*Author for correspondence
Fax: +44 (0)2920874305
E-mail: jumag@cf.ac.uk

immigrant to the Iberian Peninsula from North Africa. A North African origin for this species has previously been proposed by Boulard (1982), who classified specimens from Portugal as a subspecies *C. barbara lusitanica*, distinct from the North African subspecies *C. barbara barbara*, based on differences at the male genitalia and female ovipositor.

Traditional recognition of subspecies based on discontinuities in the geographical distribution of phenotypic traits (Mayr & Ashlock (1991) *in* Phillimore & Owens, 2006) has often been challenged by molecular phylogenetic data, especially when the phylogenetic monophyly of a subspecies is not recovered (e.g. Zink, 2004; Phillimore & Owens, 2006). However, historical introgression or incomplete lineage sorting might reduce the probability of finding reciprocal monophyly in recently divergent populations, inducing an underestimation of the differentiation (Phillimore & Owens, 2006). Nevertheless, patterns of genetic variation found within species at present can reflect differentiation due to geographic isolation, as well as the retreats and expansions of populations due to climatic changes (Hewitt, 1996, 1999, 2000, 2001, 2004a; Taberlet *et al.*, 1998; Besnard *et al.*, 2002a; Gómez *et al.*, 2005). During the ice ages, many species, particularly thermophilic species (such as cicadas), could only survive in favourable regions or refugia, so genetic differentiation found in the present day reflects their survival in different refugia combined with genetic drift and founder effects during re-colonization. The Iberian Peninsula and Maghreb have been frequently recognised as a key southern refugium for many European species during the ice ages (the Atlantic-Mediterranean refugial area; *sensu* Schmitt, 2007), but the survival of exclusively Mediterranean taxa would also have been constrained by these climatic oscillations. According to Jansson & Dynesius (2002), in areas with little climatic change over long periods of time (low ORD, Orbitally Forced Range Dynamics, over an entire period of Milankovitch oscillations, about 100 kyr) such as southern Europe and the tropics, cladogenesis and anagenesis may be induced by a wide range of processes. As a result, high intraspecific genetic variation and finely geographically subdivided species are expected to occur, while, in regions with high ORD, climatic shifts will alter the dynamics of genetic diversity directly by subsampling gene pools and indirectly by selecting for vagility and generalism, consequently decreasing genetic variability.

Information on the influence of the ice ages on Mediterranean taxa has become increasingly available (e.g. Hewitt, 2004b; Schmitt, 2007), and different geological and geophysical conditions have been identified as possible promoters of the present-day distribution and variability of Mediterranean species. The strait of Gibraltar has separated the Iberian Peninsula from North Africa since the end of the Messinian Salinity Crisis about 5.33 Ma (Hsü *et al.*, 1977), dividing the Atlantic-Mediterranean refugial area and representing a barrier for several taxa, resulting in vicariant differentiation (Hewitt, 2004b; Albert *et al.*, 2007; Pinho *et al.*, 2007; Schmitt, 2007). On the other hand, with the increase in the volume of the ice sheets, sea levels dropped and, during the Late Glacial Maximum, were 120–130 m (Yokoyama, *et al.*, 2000) lower than at present. As a consequence, the European and the African continental shelves emerged, the distance between these continents was reduced, and several islands emerged (Collina-Girard, 2001), providing stepping stones for different organisms between the two continents (e.g. snakes (Colubridae) and plants (Asteraceae) (Carranza *et al.*, 2006; Ortiz, *et al.*, 2007). In the Iberian Peninsula, the expansion and contraction of ice sheets have been implicated as generators of intraspecific diversity and substructure within Iberian taxa (e.g. Paulo *et al.*, 2001; Hewitt, 2004b).

Here, we attempt to improve understanding of the demographic history and population structure of *C. barbara* in the Iberian Peninsula and North Africa, based on genetic analysis of a section of the cyt *b* gene. The mtDNA gene cytochrome (cyt) *b* has not been as widely used in insects as other mitochondrial genes, such as cytochrome oxidase I (Caterino *et al.*, 2000), largely because cyt *b* studies have reported a lack of phylogenetic signal for comparisons among distant taxa (e.g. Krzywinski *et al.*, 2001; Simmons & Weller, 2001). However, it has been highlighted that cyt *b* is likely to be more informative within and among very closely related species and populations (Krzywinski *et al.*, 2001; Simmons & Weller, 2001; Drotz, 2003; Monteiro *et al.*, 2004).

We used the cyt *b* gene to explore the genetic history of this species in order to better understand the current patterns of genetic differentiation and geographical distribution. We also aim to determine if the geographical separation of the Iberian Peninsula and North Africa acts as an effective natural barrier between *C. barbara* populations and discuss the concept of the differentiation of two subspecies within *C. barbara*.

Material and methods

Field work was carried out during the summer seasons (July–September) of 1995–2001 in 12 localities from the Iberian Peninsula and three from Morocco (table 1). Male *Cicada barbara* were identified in the field by their calling songs then collected either by hand or with a sweep net and taken to the lab where they were preserved in liquid nitrogen or in 96% ethanol. Mitochondrial DNA sequences from five specimens per locality (except Toledo: $N = 2$) were analysed (table 1).

Molecular analysis

DNA was isolated from a small section of the thoracic muscle (frozen specimens) or from one leg (alcohol preserved) for each cicada, using the methods described in Pinto *et al.* (1998) or using the QIAamp DNA mini kit (QIAgen) following the manufacturer's protocol. Primers to amplify a segment of the mitochondrial cytochrome *b* gene were designed based on sequences obtained directly from purified mtDNA from 30 *C. barbara*; these were CB-J-11004: 5′ CTTACTTAGGGCTCAAC 3′ and CB-N-11351: 5′ TAAAAAGTATCATTCTGG 3′ with a product size of 365 bp (nomenclature based on Simon *et al.* (1994), numbers refer to the mitochondrial *Drosophila yacuba* sequence). PCR reactions comprised 25 µl of 1× *Taq* buffer, 1.25 mM $MgCl_2$, 0.2 mM of dNTPs, 50 pmol of each primer, 1 unit of *Taq* polymerase (Invitrogen) and approximately 50 ng of genomic DNA; or, instead, the puReTaq Ready-To-Go PCR beads (Amersham Biosciences, UK) were used, just adding the same primers and template as before. Cycle conditions were as follows: 3 min at 95°C, then 30 cycles of 95°C for 30 s, 50°C for 45 s and 72°C for 1 min, and a final extension step of 72°C for 7 min. Microcon-100 microconcentrators (Amicon) or the Geneclean turbo for PCR kit (QIAgen, BIO101) were used to clean the PCR products prior to direct sequencing.

Table 1. Sampled populations of *Cicada barbara* with the distribution of cyt *b* gene haplotypes (364 bp) and genetic diversity among localities.

Localities	*N*	Haplotypes						h	π	*d*
		HCb1	HCb2	HCb3	HCb4	HCb5	HCb6			
North Africa										
Fès (Morocco)	5		4	1				0.4	0.0011	0.4
Meknès (Morocco)	5		3			1	1	0.7	0.0022	0.8
Ceuta (Spain)	5	5						0	0	0
Iberian Peninsula										
Sevilla (Andalucía, Spain)	5	4			1			0.4	0.0011	0.4
Cordoba (Andalucía Spain)	5	5						0	0	0
Alcalar (Algarve, Portugal)	5	5						0	0	0
Alvor (Algarve, Portugal)	5	5						0	0	0
Moura (Baixo Alentejo, Portugal)	5	5						0	0	0
Portel (Alto Alentejo, Portugal)	5	5						0	0	0
Sousel (Alto Alentejo, Portugal)	5	5						0	0	0
Monforte (Alto Alentejo, Portugal)	5	5						0	0	0
Crato (Alto Alentejo, Portugal)	5	5						0	0	0
Arrábida (Estremadura, Portugal)	5	5						0	0	0
Toledo (Castilla-La Mancha, Spain)	2	2						0	0	0
Foz Côa (Beira Alta, Portugal)	5	5						0	0	0
Total	**72**	**61**	**7**	**1**	**1**	**1**	**1**			

N, number of specimens used for mtDNA analysis; h, haploptype diversity; π, nucleotide diversity; *d*, mean number of pairwise differences.

Sequencing reactions using the primers described above were performed using the ABI Big Dye v. 2 with better buffer (WebScientific Ltd.) and sequences were electrophoresed in a Perkin Elmer 3100 DNA sequencer. Both DNA strands were always sequenced.

DNA sequences were initially analysed with Chromas software version 1.4 (Conor McCarthy, Griffith University, Queensland, Australia), then edited with Editseq and aligned by the CLUSTAL method with MegAlign (DNAstar applications, Lasergene, USA). Transitions, transversions and synonymous versus nonsynonymous substitutions between sequences were analysed in MEGA version 2.1 (Kumar *et al*., 2001). Modeltest v. 3.5 (Posada & Crandall, 1998) was used to search for the most likely model for DNA substitution. The hierarchical likelihood ratio test (hLRT) and the Akaike information criterion (AIC) revealed HKY 85 (Hasegawa *et al*., 1985) as the most likely model of DNA evolution for these data. Therefore, a HKY distance matrix between the sequences was computed using PAUP (4.0 beta version; Swofford, 1999). Genetic diversity, nucleotide variation, minimum spanning network based on the number of pairwise differences between all pairs of haplotypes and AMOVA (analysis of molecular variance) were calculated using the software Arlequin Ver. 2.000 (Schneider *et al*., 2000). A medium joining network based on the maximum likelihood connections between haplotypes was produced by Network 4.1.1.2 (Fluxus Technology Ltd). Pairwise genetic distances based on Slatkin's linearised F_{st} (Slatkin, 1995) were also determined using Arlequin. AMOVA was used to test different groupings of populations based on the geographical proximity between sampling sites and the pattern of distribution of the haplotypes. The significance of estimates of the variance components for different hierarchical subdivisions (within populations, within groups of populations and among groups) was tested using 1000 non-parametric permutations (Excoffier *et al*., 1992).

Arlequin was also used to estimate the parameters expected under a sudden demographic expansion and to test the goodness-of-fit of mismatch distributions using 1000 bootstrap replicates. DNAsp4 (Rozas *et al*., 2003) was used to plot mismatch distributions among haplotypes and to determine the confidence intervals and significance of the raggedness statistics using coalescent simulations based on the neutral infinite-sites model; and, together with Arlequin, to perform neutrality tests in order to corroborate the mismatch distributions statistics (Tajima's (1989) D-test, Fu & Li's (1993) D* and F*, and Fu's (1997) F_s statistics). Fu's (1997) F_s test statistics have been shown to be more powerful for detecting population growth than the mismatch statistics (Fu, 1997; Ramos-Onsins & Rozas, 2002; Mes, 2003). Fu's F_s assesses population growth (Fu, 1997), recognizing the excess of low frequency alleles in an expanding population compared to the number expected in a stationary one. Tajima's D test assumes that, under neutrality, the number of nucleotide differences between sequences from a random sample should be equal to the number of differences between the polymorphic sites; so, significant negative deviations of Tajima's D from zero might reflect population expansions (Tajima, 1989). Fu & Li's D* and F* statistics were used to compare with F_s tests, as the pattern of significance of these tests may distinguish population growth from the effects of background selection; expansion is expected when F_s is significant and D* and F* are not, while the converse indicates selection (Fu, 1997).

The ratio between the number of variable sites (S) and mean pairwise differences among sequences (d) was also calculated ($S\ d^{-1}$, expansion coefficient). Expanding populations tend to have sequences with higher frequency of unique sites; thus, large $S\ d^{-1}$ values may indicate a recent population expansion, while small values indicate relatively constant long-term population sizes.

Assuming population expansion happened, Arlequin generated the expansion time in mutational units (τ), and

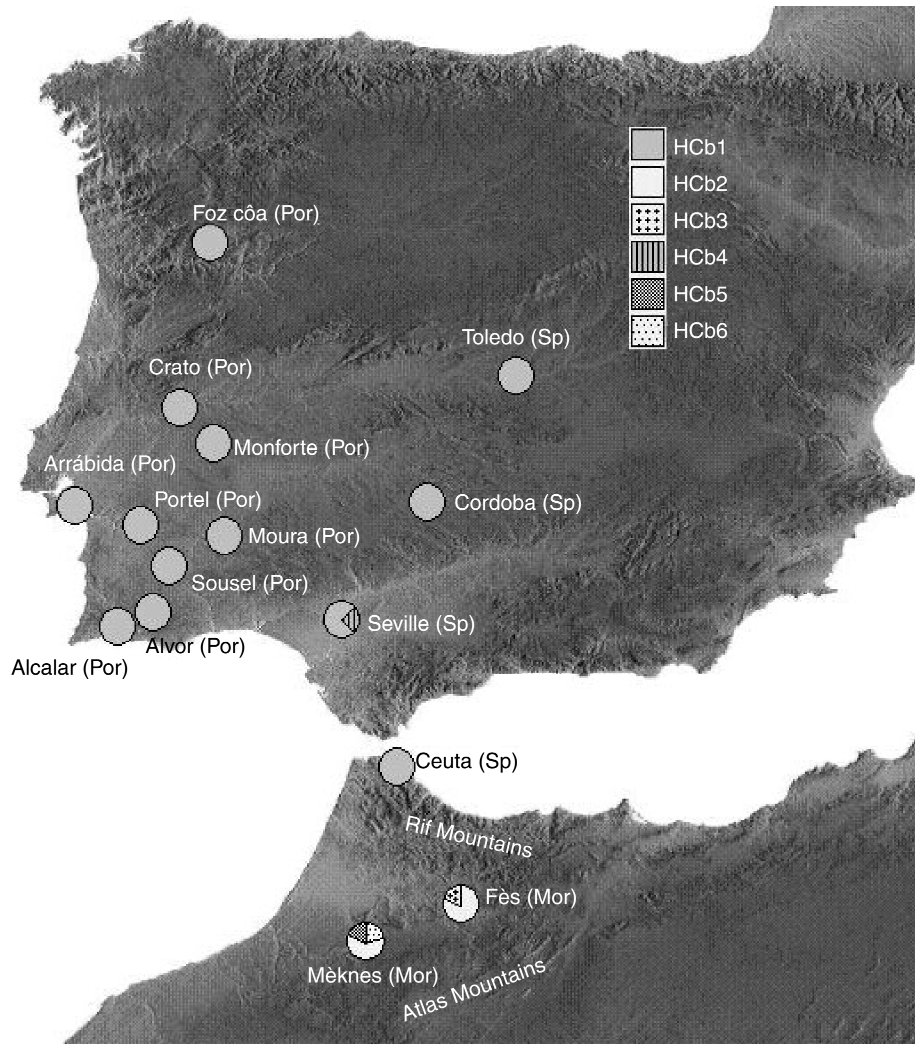


Fig. 1. Distribution and relative frequency of cyt *b* haplotypes among populations of *Cicada barbara* in the Iberia Peninsula and Morocco. The volume of the pie chart representing each haplotype is proportional to the total number of specimens showing that particular haplotype.

the time since expansion was calculated based on the estimators described by Rogers & Harpending (1992), Harpending (1994) and Rogers (1995) ($\tau = 2ut$, where t is the number of generations elapsed between initial population and current population; and $u = 2\mu k$, where μ is the mutation rate per million years and k is the length of the sequence).

Results

Six haplotypes from the 364 bp cyt *b* gene were detected in the 72 cicadas (fig. 1). All localities analysed from the Iberian Peninsula + Ceuta (North Africa) possessed the same haplotype (HCb1); Seville was one exception, here an additional haplotype was found in one individual (HCb4). Neither haplotype was present in Morocco. Both sites in Morocco shared a dominant haplotype (HCb2); but each locality also revealed other exclusive haplotypes, HCb3 was found in Fès and HCb5 and HCb6 in Meknès.

Genetic variation

The mean proportion for each nucleotide was 30.3%: A; 39.5%: T; 12.4%: G; and 17.9%: C – demonstrating the expected AT bias for invertebrate mtDNA (Simon *et al.*, 1994). Six polymorphic sites were found. Substitutions were all transitions, except one A→T transversion, only present in haplotypes found in Morocco. The mean Ti/Tv ratio between all haplotypes was 1.25. Only one non-synonymous substitution occurred in haplotype CBHCb5, where Ile (Isoleucine) was replaced by Thr (Threonine). The sequences encoded 120 amino acids, with leucine the most common (14.2%), followed by phenylalanine (10.8%) as observed in most insects for this gene (Simmons & Weller, 2001).

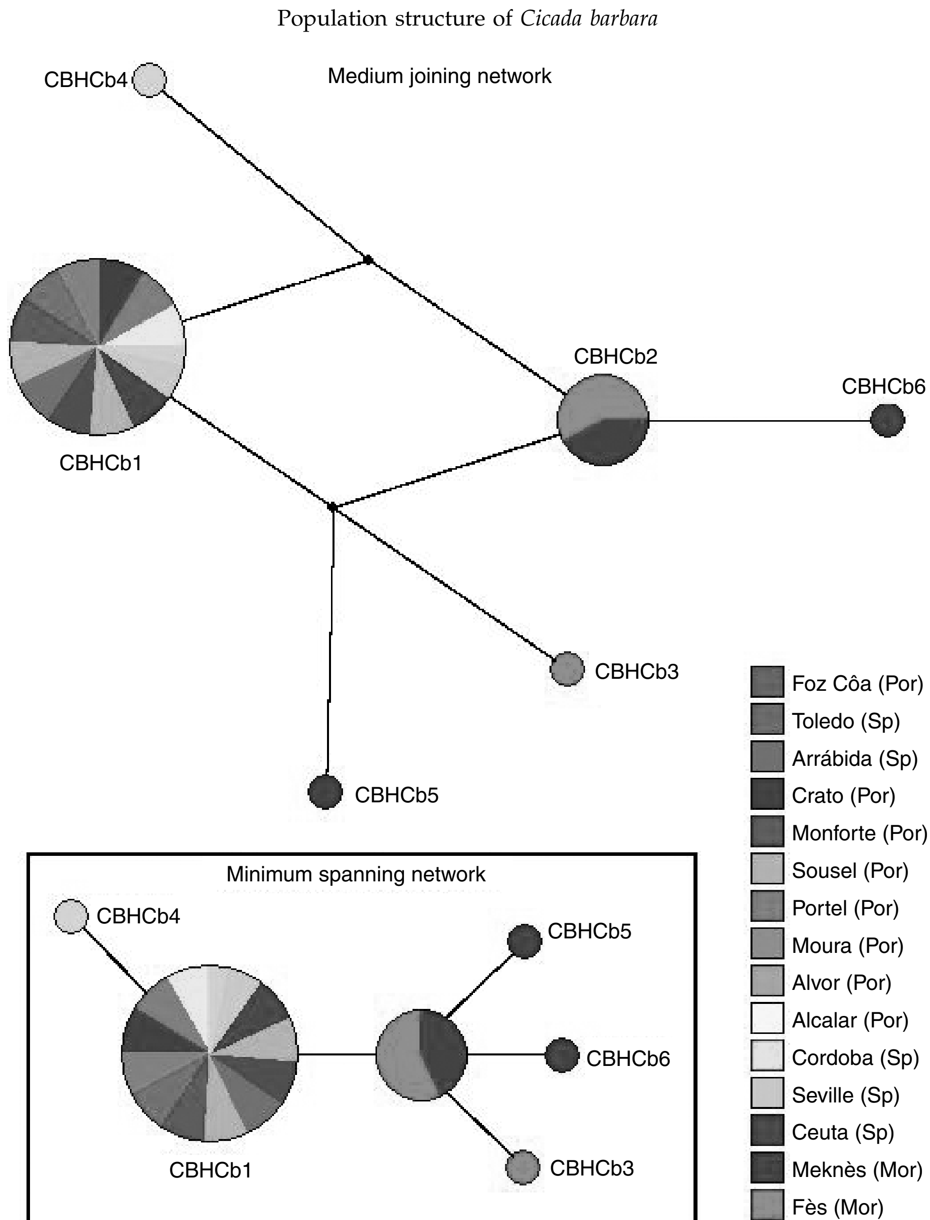


Fig. 2. Medium joining and minimum spanning networks for *Cicada barbara* in the Iberia Peninsula and Morocco based on cyt *b* haplotypes. The volume of the circles representing each haplotype is proportional to the total number of specimens showing that particular haplotype. The colours are representative of the number of specimens per locality. Each connection indicates one mutational step and black dots in the medium joining network represent missing intermediate haplotypes.

The average number of pairwise differences was 0.576 and the average nucleotide diversity, π, was 0.00159. The mean nucleotide diversity among synonymous sites was 0.00423. Haplotype diversity for all data was 0.2758 ± 0.07 (within Morocco, 0.5333 ± 0.18; Iberian Peninsula + Ceuta, 0.0323 ± 0.03).

Meknès had the highest haplotype (h) and nucleotide (π) diversity ($h = 0.7$ and $\pi = 0.0022$; table 1). Both Fès and Seville had two haplotypes and the same values for haplotype and nucleotide diversity ($h = 0.4$ and $\pi = 0.0011$). The mean number of pairwise differences within these localities was 0.4 for Fès and Seville, and 0.8 for Meknès. All remaining localities, including Ceuta, did not show any diversity.

Population structure

A minimum spanning network constructed by hand, based on the number of pairwise nucleotide differences computed using Arlequin and a medium joining network computed by Network based on maximum likelihood, are shown in fig. 2. Even though the number of haplotypes and the connection lengths are low, Moroccan haplotypes group monophyletically.

HKY distances between all pairwise combinations of haplotypes varied from 0.0028 to 0.0084 (table 2). The average genetic distance among haplotypes within Morocco was 0.42%, within the Iberian Peninsula + Ceuta, 0.28% and

Table 2. HKY pairwise distances between all pairs of haplotypes based in a 364 bp fragment of the Cyt *b* gene within *Cicada barbara*.

	CBHCb1	CBHCb2	CBHCb3	CBHCb4	CBHCb5	CBHCb6
CBHCb1	–					
CBHCb2	0.0028	–				
CBHCb3	0.0056	0.0028	–			
CBHCb4	0.0028	0.0056	0.0084	–		
CBHCb5	0.0056	0.0028	0.0056	0.0084	–	
CBHCb6	0.0056	0.0028	0.0056	0.0084	0.0056	–

Table 3. Slatkin's linearised F_{st} values (Slatkin, 1995) between all pairs of samples based in a 364 bp fragment of the Cyt *b* gene within *Cicada barbara*, below diagonal.

	Alcalar	Alvor	Arrábida	Crato	Foz Côa	Monforte	Moura	Portel	Sousel	Ceuta	Cordoba	Fès	Mèknes	Toledo	Sevilla
Alcalar															
Alvor	0														
Arrábida	0	0													
Crato	0	0	0												
Foz Côa	0	0	0	0											
Monforte	0	0	0	0	0										
Moura	0	0	0	0	0	0									
Portel	0	0	0	0	0	0	0								
Sousel	0	0	0	0	0	0	0	0							
Ceuta	0	0	0	0	0	0	0	0	0						
Cordoba	0	0	0	0	0	0	0	0	0	0					
Fès	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*				
Mèknes	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0.7*	0			
Toledo	0	0	0	0	0	0	0	0	0	0	0	0.7*+	0.6*		
Sevilla	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0.7*+	0.6*+	−0.3	

*, significant values ($P < 0.05$).

Table 4. AMOVA results for the two groups of *Cicada barbara* maximizing the Φ_{CT} values analysed for cyt *b* gene (Morocco vs Iberian Peninsula and Ceuta).

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Indices	*P*
Among groups	1	8.872	0.51302 Va	90.77	Φ_{CT} 0.91	+
Among populations within groups	13	0.484	−0.00396 Vb	−0.70	Φ_{SC} −0.08	+
Within populations	57	3.200	0.05614 Vc	9.93	Φ_{ST} 0.90	+
Total	71	12.556	0.56520			

d.f., degrees of freedom; *P*, significance of percentage variation and fixation indices estimated from permutation tests (1000 permutations); Φ_{CT}, correlation of random haplotypes within a group of populations relative to that drawn from the entire species;Φ_{SC}, correlation of random haplotypes within populations relative to that within a regional grouping of populations within a region;Φ_{ST}, correlation of random haplotypes within a population relative to that from the whole species (Excoffier *et al.*, 1992); significant + $P < 0.05$.

between the two groups the average HKY distance was 0.63%.

Slatkin's linearised F_{st} revealed significant genetic differentiation ($P < 0.05$) in all pairwise comparisons between Morocco samples (Fès and Mèknes) versus all other populations (table 3), with values ranging from 0.6 to 0.8. All other sample pairwise comparisons revealed zero or non-significant F_{st} values (samples within the Iberia Peninsula and also between Fès and Meknès).

Several hierarchical analyses of molecular variance were carried out to compare hypotheses of genetic partitioning. The highest Φ_{CT} value, which was statistically significant, was found when samples were separated into two groups (table 4), one including Iberian Peninsula + Ceuta samples and the other including Moroccan samples (Φ_{CT} 0.91). Other partitioning tests with lower Φ_{CT} were also significant. For instance, comparing Iberian Peninsula vs. North Africa (Ceuta and Morocco), a Φ_{CT} value of 0.78 was found. When testing five groups (North Portugal (Foz Côa), South Portugal, South Spain, North Africa (Ceuta) and Morocco), the Φ_{CT} was even lower (0.67).

Demographic analysis

Since there was a significant genetic differentiation between Morocco and the Iberian Peninsula + Ceuta, the historical demography of these two regions was analysed separately. The distribution of pairwise nucleotide differences for the Iberia Peninsula + Ceuta samples and for the Morocco samples alone were computed and plotted (fig. 3). The Iberian Peninsula + Ceuta group showed a unimodal distribution, while Morocco revealed a slightly bimodal

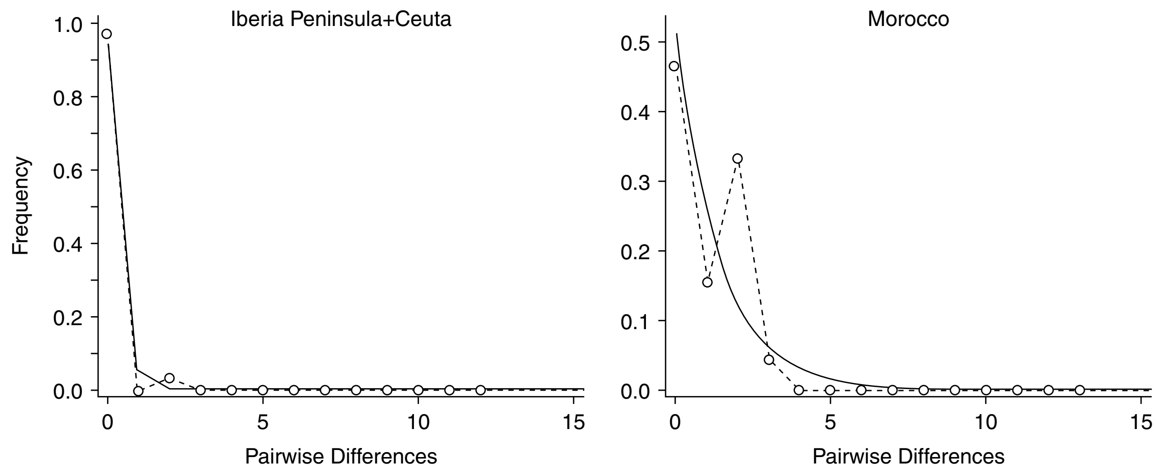


Fig. 3. Mismatch distributions within *Cicada barbara* (—, Exp; ---o---, Obs).

curve, even though both distributions showed the expected pattern for populations that have undergone sudden expansion (table 5). The observed raggedness was high for the Iberia Peninsula + Ceuta and moderate for Morocco, though not significant in both cases. This was most likely due to the small number of haplotypes observed for the first region. The coalescent simulated raggedness, on the other hand, was notably low for the Iberia Peninsula + Ceuta and, again, moderate for Morocco.

Both regions showed significantly negative Fu's F_s values (table 5) indicative of expanding populations. In the Morocco samples, Tajima's D was significantly negative, and Fu and Li's D* and F* were non-significant negative, which, again, reflects population growth and no background selection. Iberia Peninsula + Ceuta also had negative values for these statistics, but Tajima's D was not significant contrary to Fu and Li's D* and F*.

The expansion coefficients estimated for each region revealed high values for the two regions analysed; however, Iberia Peninsula + Ceuta had a much higher expansion coefficient than Morocco.

The τ parameter for the Iberian samples was considerably higher than for the Moroccan samples. Cicada nymphs develop underground and the exact time they spend at this stage is still unknown for most species within genus *Cicada*. In fact, the cicada immature stages may last between two to three years (Quartau, 1995; Boulard & Mondon, 1996; Giralda *et al*., 1998). Considering this, there are two possible scenarios for the expansion time. Considering a mutation rate for the cyt *b* gene of 2.15% (as for *Agabus* species, Coleoptera: Drotz, 2003) and one generation every two years, the time of expansion (t) based in the estimators described by Rogers & Harpending (1992) and Harpending (1994) for Iberian Peninsula cicadas occurred around 191,668 years BP (20,444–191,668 years BP), while Moroccan cicadas are estimated to have expanded about 48,492 years ago (0–113,724 years BP). The estimated time of expansion for the Iberian Peninsula is also at the maximum of its error range, most likely an artefact due to the low number of haplotypes detected. If we consider that these cicadas have a generation time of three years; then, for the Iberian Peninsula, the expansion is estimated to have occurred around 287,502 years BP (30,666–287,502 years BP) and for Morocco about 72,738 years BP (0–170,586 years BP).

Discussion

Population structure and genetic variation

The divergence between Moroccan and the Iberian Peninsula + Ceuta populations is strongly corroborated by the molecular data. The AMOVA results, for the cyt *b* sequence analysed for *C. barbara*, revealed a highly significant Φ_{CT} value of 0.91; and the high mean pairwise Φ_{ST} value (0.90) implies genetically isolated populations with a low level of gene flow.

The Ceuta population from Spanish North Africa was more similar to the Iberian populations than to the surrounding Moroccan populations. This relationship might suggest that the Strait of Gibraltar does not act as a strict barrier to dispersal. The narrowest distance between the southern-most point of Spain and the northwestern-most point of North Africa is 13 km (the strait ranges from 12.9 km to 43 km), possibly within potential flight distances for cicadas. Moreover, the drop in sea level during the last glaciations, revealing the European and the African continental shelves and several islands in between (Collina-Girard, 2001) would have facilitated the movement of cicadas through the Strait of Gibraltar. Wind dispersal over water has been invoked as one of the explanations for dispersal in New Zealand cicadas (Arensburger *et al*., 2004) and sea-surface dispersal through the Strait of Gibraltar by some coleoptera, including wingless insects, has been postulated (e.g. Palmer & Cambefort, 2000; Horn *et al*., 2006). Alternatively, considering that Ceuta is a Spanish territory, a possible human-induced translocation should not be ignored. Trees or shrubs, containing *C. barbara* specimens (eggs or adults), from the Iberian Peninsula might have been introduced. In contrast, the high altitude of the Rif Mountains (highest point is 2455 m) to the south of Ceuta and its climatic extremes, including snow and frost, might provide a strong obstacle to the natural dispersal of these thermophilic cicadas. There is no record of *C. barbara* in the Rif Mountains and, considering the distribution of vegetation and climate along the altitude gradient here (e.g. Moore

Table 5. Diversity, neutrality and expansion parameters for *Cicada barbara* based on a section of cyt *b* gene.

	Iberian Peninsula + Ceuta	Morocco
Number of haplotypes	2	4
Nucleotide diversity	0.00018	0.00264
Haplotype diversity	0.032	0.533
Expansion coefficient (S d^{-1})	30.77 (2/0.065)	4.18 (4/0.956)
Observed raggedness (r)	r = 0.88, $P = 0.89$	r = 0.16, $P = 0.70$
Coalescent simul. raggedness	Average r = 0.05 (95% CI: 0.0–0.14)	Average r = 0.28 (95% CI: 0.05–0.79)
Tajima's D	−1.080	-1.562^{+}
Fu & Li's F*	-2.645^{+}	−1.294
Fu & Li's D*	-2.630^{+}	−1.127
Fu's F_s	-1.817^{+}	-1.964^{+}
τ	3.000 (95% CI: 0.32–3)	0.759 (95% CI: 0–1.78)
Time since expansion (1 generation per 2 years)	191 668 (20,444–191,668) years BP	48 492 (0–113,724) years BP
Time since expansion (1 generation per 3 years)	287 502 (30,666–287,502) years BP	72 738 (0–170,586) years BP

S, polymorphic sites; d, mean number of pairwise nucleotide differences; CI, confidence interval. Significant values for the neutrality estimations ($P < 0.05$) are accentuated with (+) Estimated mean τ parameter was obtained according to the percentile method from the sudden expansion model (confidence level: $\alpha = 0.05$ based on 1000 replicates).

et al., 1998), it is highly unlikely that *C. barbara* occurs above 1000 m. The cultivation of olive trees in these mountains occurs below 500 m (Moore *et al.*, 1998). The presence of a rock glacier site in these mountains is also indicative of the severe past climatic conditions. According to Hughes *et al.* (2006), glacial deposits extend to exceptionally low altitudes for this latitude (*ca.* 36°N), reaching as low as 750 m on some slopes of the massif. The Rif Mountains are known to hold a high rate of endemism, partly as a result of the long isolation of Holartic taxa in the high elevations of these North African mountain ranges. The HKY genetic distances observed between the haplotypes of *C. barbara* were relatively low (0.28–0.84%). For the same gene, Drotz (2003) found genetic distances between 0–3.9% within *Agabus* species (Coleoptera) and Monteiro *et al.* (2004) referred distances of 0.7–1.2% within groups of *Triatoma brasiliensis* (Hemiptera).

Demographic expansions

Overall, the demographic analysis suggests sudden expansions for both regions. Even though the observed raggedness was high, especially for the Iberian Peninsula, coalescent simulations revealed much smaller values, which is probably due to the low number of haplotypes observed in the Iberian Peninsula. All other estimators analysed, including those usually considered as more robust and reliable, strongly supported sudden expansion models. However, the low number of haplotypes and low nucleotide diversity in the Iberian Peninsula could indicate the occurrence of a strong bottleneck at some time in the past.

The Iberian Peninsula and Ceuta specimens showed a more ancient signature of demographic expansion than in Morocco. These results should, however, be treated with caution because the mutation rate used for cyt *b* gene might not be appropriate for *C. barbara*, and the low number of haplotypes detected weakens these tests. Also, the generation time has yet to be accurately determined for this species; and the sample sizes, especially for Morocco, are relatively low.

Cicada barbara can be seen on different kinds of shrubs or trees; but, in the majority of sites, they are found associated with olive and pine. Therefore, it is most likely that the life history of cicadas might be related with the life history of these trees. In fact, Burban & Petit (2003) found a similar pattern of genetic differentiation within maritime pine (*Pinus pinaster*). Three different mitotypes were detected in maritime pine using RFLP analysis within the Mediterranean area: one haplotype was exclusively present in the Iberian Peninsula and in Punta Cires (in the north of Morocco very close to Ceuta), one other was present only in Morocco and the third was present in Eastern Europe and Northeast Africa. On the other hand, the cultivation of olive trees is known to have happened since ancient times, and humans have greatly influenced its dispersal (Besnard *et al.*, 2002a). As such, the genetic structure of olive trees reflects not only the occurrence of refugial zones but also the particular biogeographic conditions of the Mediterranean Basin and the human influence (Besnard *et al.*, 2002a). Nevertheless, strong genetic differentiation between the eastern and western Mediterranean, as well as for Morocco, was shown for olive cpDNA (Besnard *et al.*, 2002a); and high levels of genetic differentiation were revealed by both cpDNA and mtDNA, probably due to the contraction of the distribution during the Quaternary glaciations and subsequent recolonization during Holocene (Besnard *et al.*, 2002b). Even though no eastern samples were available for this study, these data show that in the Iberia Peninsula and Morocco *C. barbara* has a similar pattern of genetic differentiation as both pine and olive trees.

Iberia Peninsula expansions

Whether we consider that cicadas have a two or three year generation time, the expansion time for the Iberian Peninsula cicadas coincides with the Mindel-Riss interglacial period (300,000–180,000 years BP). Furthermore, the previous (Mindel) glaciation (410,000–380,000 years BP) correlates with the strong bottleneck we detected for Iberian cicadas. In fact, it is considered that this glaciation marks the absolute maximum extent of continental ice sheets in the Quaternary (Lindner *et al.*, 2003). During the subsequent glaciations (Riss, 180,000–130,000 years BP; and Würm, approximately 70,000–10,000 years BP), *C. barbara* individuals could have potentially survived in refugia, which persisted in the Iberia Peninsula. Moreover, the last interglacial, the Eemian, which started approximately 126,000 years BP (Sánchez Goñi *et al.*, 2000) allowed Mediterranean

taxa to expand. For instance, the maritime pine in the Iberian Peninsula seems to have been largely favoured in comparison with other Mediterranean conifers due to its wide ecological range (Gómez *et al*., 2005). During glacial periods, *P. pinaster* was found near the coast and mountains, and it has been suggested to have a continuous presence at lower altitudes since 30,000 years BP (Carrión & Van Geel (1999) *in* Gómez *et al*., 2005). Nevertheless, *C. barbara* did not show any geographical structure in Iberia, unlike other Mediterranean species reported (e.g. Paulo *et al*., 2001; Jansson & Dynesius, 2002; Hewitt, 2004b), suggesting high vagility and vulnerability to climate change. The extreme northwest of Africa, such as Ceuta, might, however, have acted as a southern refuge for Iberian cicadas during the most severe climatic conditions. Specimens of *C. barbara* from the Iberian Peninsula and from Ceuta (North Africa) did not show any genetic differentiation; and, as stated before, access between the two margins of the Strait of Gibraltar would have been easier during the Pleistocene.

Moroccan expansions

The estimated expansion time for Moroccan populations, considering a two year generation time, occurred during the last Pleistocene glaciation, which ended around 14,000 years BP. On the other hand, considering a three year generation time, Moroccan cicadas would have expanded after Riss glaciation. Since the end of the Riss glaciation, during the interglacial, the Mediterranean oak forests expanded greatly in Morocco (Hooghiemstra *et al*., 1992), allowing the expansion of other species, e.g. macaques (Modolo *et al*., 2005). Moroccan cicadas might have expanded during this phase and retreated when conditions were not so favourable and finally expanded again at the end of the last ice age, when climatic conditions were improving. In fact, the mismatch distribution of Moroccan specimens shows two peaks, and the signal of the expansion is not as clear as for the Iberian Peninsula populations. This might indicate that populations in Morocco were more stable; and, even though they suffered retreats and expansions, there is no evidence of a very strong bottleneck during these ice ages. Also, the high haplotype variability found in these populations suggests a higher stability than in the Iberian Peninsula populations. These results are in agreement with other studies of Mediterranean taxa suggesting that the most southern populations show higher geographical substructure and stability (e.g. Jansson & Dynesius, 2002; Hewitt, 2004b; Pinho *et al*., 2007).

Origin of Iberian cicadas and subspecies classification

The hypothesis of *C. barbara* being a recent immigrant to the Iberian Peninsula from North Africa (Quartau *et al*., 2001) cannot be confirmed or refuted with these data. Ceuta, one of the most northwestern localities in North Africa, did not clarify this issue, as it showed the same genetic pattern as Iberian populations, but remains significant since it potentially shows that the Strait of Gibraltar does not seem to have prevented dispersal of these cicadas in contrast to the Rif Mountains. Although the low genetic variability of the Iberian populations might suggest a very recent colonization, the complete absence of shared haplotypes with Morocco seems to indicate an ancient separation between these regions. Also, the expansion signal in the Iberian Peninsula is much more ancient than in Morocco. It is possible that a sampling artefact might have hidden shared haplotypes; however, these results do not seem to suggest a Moroccan origin of the Iberian cicadas. Nevertheless, we cannot exclude the hypothesis of a North African origin. In fact, Morocco topography, with the Rif Mountains in the north, the Atlas Mountains in the southeast and the Sahara desert further south, favours the isolation of species (Modolo *et al*., 2005) and consequently their divergence, as also found with the maritime pine (Burban & Petit, 2003) and the olive, *Olea europaea* subsp. *maroccana* (Besnard *et al*., 2002a,b). Also, during the Quaternary climatic fluctuations, the Sahara Desert expanded and retracted repeatedly (Cox & Moore, 2005), separating a northwestern and a northeastern refuge which favoured the differentiation between populations in northwest and northeast Africa, as observed in some coleoptera, snails, lizards, freshwater turtles, shrews and snakes (e.g. Cosson *et al*., 2005; Fritz *et al*., 2005; Horn *et al*., 2006; Carranza *et al*., 2006). It will be important to sample cicadas from other North African regions to clarify this.

The definition of the two subspecies within *C. barbara* (*C. barbara lusitanica* and *C. barbara barbara*) finds some support here, although Moroccan specimens might be also differentiated from the other Northern African individuals as found in other taxa. Species delimitation remains conceptually difficult and for subspecies the subjectivity may be even more pronounced. Davies & Nixon (1992) suggested that evolutionary lineages may become different species if at least one character state is fixed in one group and absent in the other. Templeton (1994) stated that evolutionary lineages must be phylogenetically distinct, show no recent gene flow and reveal ecological or demographic limitations to reproduction. Zink (2004) and Phillimore & Owens (2006) also used monophyly as an indicator of phylogenetic distinctiveness of subspecies. On the other hand, Lee (2004) emphasized the importance of analysing other biological information besides molecular data to define species, and pointed out that it is not possible to associate a specific level of genetic divergence to reproductive isolation. Even so, some of these concepts might suggest that the two groups of cicadas analysed here are differentiating and may be on the way to splitting, as supported by some acoustic differentiation which is involved in mating attraction (see Pinto-Juma *et al*., this journal).

Acknowledgements

We thank Genage André, Sofia Seabra, Paula Simões, Teresa Rebelo, Mónica Ribeiro and Teresa Fernandes (Faculty of Science, University of Lisbon) for their help with field work. We also thank G. Hewitt and an anonymous reviewer for their comments and suggestions on a previous version of this manuscript. This research was supported by the PhD grant PRAXIS BD/18229/98 from the Fundação para a Ciência e Tecnologia (Lisbon, Portugal).

References

Albert, E.M., Zardoya, R. & García-París, M. (2007) Phylogeographical and speciation patterns in subterranean worm lizards of the genus Blanus (Amphisbaenia: Blanidae). *Molecular Ecology* **16**, 1519–1531.

Arensburger, P., Simon, C. & Holsinger, K.E. (2004) Evolution and phylogeny of the New Zealand cicada genus *Kikihia* Dugdale (Homoptera: Auchenorryncha: Cicadidae) with special reference to the origin of the Kermadec and Norfolk Islands' species. *Journal of Biogeography* **31**, 1769–1783.

Besnard, G., Khadari, B., Baradat, P. & Bervillé, A. (2002a) *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* **104**, 1353–1361.

Besnard, G., Khadari, B., Baradat, P. & Bervillé, A. (2002b) Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics* **105**, 139–144.

Boulard, M. (1982) Les cigales du Portugal, contribution à leur étude (Homoptera, Cicadidae). *Annales de la Société Entomologique de France* **18**, 181–198.

Boulard, M. & Mondon, B. (1996) Vies et mémoires de cigales. Equinoxe, Barbentane.

Burban, C. & Petit, R.J. (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**, 1487–1495.

Carranza, S., Arnold, E.N. & Pleguezuelos, J.M. (2006) Phylogeny, biogeography, and evolution of two Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis* (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution* **40**, 532–546.

Caterino, M.S., Cho, S. & Sperling, F.A.H. (2000) The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology* **45**, 1–54.

Collina-Girard, J. (2001) L'Atlantide devant le détroit de Gibraltar? Mythe et géologie. *Sciences de la Terre et des Planètes* **333**, 233–240.

Cosson, J.-F., Hutterer, R., Libois, R., Sarà, M., Taberlet, P. & Vogel, P. (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soridae). *Molecular Ecology* **14**, 1151–1162.

Cox, C.B. & Moore, P.D. (2005) *Biogeography: An Ecological and Evolutionary Approach*. 7th edn. 428 pp. Oxford, UK, Blackwell Publishing.

Davies, J.I. & Nixon, K.C. (1992) Populations, genetic variation and the delimitation of phylogenetic species. *Systematic Biology* **41**, 421–435.

Drotz, M.K. (2003) Speciation and mitochondrial DNA diversification of the diving beetles *Agabus bipustulatus* and *A. wollastoni* (Coleoptera, Dytiscidae) within Macaronesia. *Biological Journal of the Linnean Society* **79**, 653–666.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Fritz, U., Barata, M., Busack, S.D., Fritzsch, G. & Castilho, R. (2005) Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, *Mauremys leprosa* (Reptilia, Testudines, Geoemydidae). *Zoologica Scripta* **35**, 97–108.

Fu, Y. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics* **147**, 915–925.

Fu, Y. & Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Giralda, A.A., Rodriguez, J.M.P. & Pintor, F.G. (1998) La cigarra del olivo o Magrebí-*Cicada barbara* (Stal.). Junta de Extremadura. *Boletin Fitosanitario de Avisos e Informaciones* **1/98**, 1–4.

Gómez, A., Vendramin, G.G., González-Martínez, S.C. & Alía, R. (2005) Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions* **11**, 257–263.

Harpending, H. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**, 591–600.

Hasegawa, M., Kishino, K. & Yano, T. (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160–174.

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

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87–112.

Hewitt, G.M. (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.

Hewitt, G.M. (2001) Speciation, hybrid zones and phylogeography – or seen genes in space and time. *Molecular Ecology* **10**, 537–549.

Hewitt, G.M. (2004a) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **359**, 183–195.

Hewitt, G.M. (2004b) The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology* **1**, 1–16.

Hooghiemstra, H., Stalling, H., Agwu, C.O.C. & Dupont, L.M. (1992) Vegetational and climatic changes at the northern fringe of the Sahara 250,000–5,000 years BP: evidence from 4 marine pollen records located between Portugal and the Canary Islands. *Review of Palaeobotany and Palynology* **74**, 1–53.

Horn, A., Roux-Morabito, G., Lieutier, F. & Kerdelhue, C. (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera: Scolytinae). *Molecular Ecology* **15**, 1603–1615.

Hsü, K.J., Montadert, L., Bernoulli, D., Bianca, C.M., Erickson, A., Garrison, R.E., Kidd, R.B., Mèliéres, F., Müller, C. & Wright, R. (1977) History of the Mediterranean salinity crisis. *Nature* **267**, 399–403.

Hughes, P.D., Woodward, J.C. & Gibbard, P.L. (2006) Quaternary glacial history of the Mediterranean mountains. *Progress in Physical Geography* **30**, 334–364.

Jansson, R. & Dynesius, M. (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics* **33**, 741–777.

Krzywinski, J., Wilkerson, R.C. & Besansky, N.J. (2001) Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. *Molecular Phylogenetics and Evolution* **18**, 479–487.

Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

Lee, M.S.Y. (2004) The molecularisation of taxonomy. *Invertebrate Systematics* **18**, 1–6.

Lindner, L., Dzierzek, J., Marciniak, B. & Nitychoruk, J. (2003) Outline of Quaternary glaciations in the Tatra Mts.: their development, age and limits. *Geological Quarterly* **47**, 269–280.

Mes, T.H.M. (2003) Demographic expansion of parasitic nematodes of livestock based on mitochondrial DNA regions that conflict with the infinite-sites model. *Molecular Ecology* **12**, 1555–1566.

Modolo, L., Salzburger, W. & Martin, R.D. (2005) Phylogeography of Barbary macaques (*Macaca sylvanus*) and the origin of the Gibraltar colony. *The Proceedings of the National Academy of Sciences Online* **102**, 7392–7397.

Monteiro, F.A., Donnelly, M.J., Beard, C.B. & Costa, J. (2004) Nested clade and phylogeographic analyses of the chagas disease vector *Triatoma brasiliensis* in northeast Brazil. *Molecular Phylogenetics and Evolution* **32**, 46–56.

Moore, H.M., Fox, H.R., Harrouni, M.C. & El Alami, A. (1998) Environmental challenges in the Rif mountains, northern Morocco. *Environmental Conservation* **25**, 354–365.

Ortiz, M.Á., Tremetsberger, K., Talavera, S., Stuessy, T. & García-Castaño, J.L. (2007) Population structure of Hypochaeris salzmanniana DC. (Asteraceae), an endemic species to the Atlantic coast on both sides of the Strait of Gibraltar, in relation to Quaternary sea level changes. *Molecular Ecology* **16**, 541–552.

Palmer, M. & Cambefort, Y. (2000) Evidence for reticulate palaeogeography: beetle diversity linked to connection-disjunction cycles of the Gibraltar strait. *Journal of Biogeography* **27**, 403–416.

Paulo, O.S., Dias, C., Bruford, M.W., Jordan, W.C. & Nichols, R.A. (2001) The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 1625–1630.

Phillimore, A.B. & Owens, I.P.F. (2006) Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences* **273**, 1049–1053.

Pinho, C., Harris, D.J. & Ferrand, N. (2007) Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology* **16**, 1191–1205.

Pinto, G.A., Quartau, J.A., Morgan, J.C. & Hemingway, J. (1998) Preliminary data on the sequencing of a fragment of cytochrome b gene of mitochondrial DNA in *Cicada orni* L. (Homoptera: Cicadidae) in Portugal. *Boletim da Sociedade Portuguesa de Entomologia* **7**(6), 57–66.

Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics Applications Note* **14**, 817–818.

Quartau, J.A. (1988) A numerical analysis of interspecific morphological differences in two closely related species of *Cicada* (Homoptera, Cicadoidea) in Portugal. *Great Basin Naturalist Memoirs* **12**, 171–181.

Quartau, J.A. (1995) Cigarras esses insectos quase desconhecidos. *Correio da Natureza* **19**, 33–38.

Quartau, J.A., Ribeiro, M., Simões, P.C. & Crespo, A. (2000) Taxonomic separation by isozyme electrophoresis of two closely related species of *Cicada* L. (Hemiptera: Cicadoidea) in Portugal. *Journal of Natural History* **34**, 1677–1684.

Quartau, J.A., Ribeiro, M., Simões, P.C. & Coelho, M.M. (2001) Genetic divergence among populations of two closely related species of *Cicada* Linnaeus (Hemiptera: Cicadoidea) in Portugal. *Insect Systematics and Evolution* **32**, 99–106.

Ramos-Onsins, S. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092–2100.

Rogers, A.R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615.

Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552–569.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Sánchez Goñi, M.F., Turon, J.-L., Eynaud, F., Schackleton, N.J. & Cayre, O. (2000) Direct land/sea correlation of the Eemian and its comparison with the Holocene: a high-resolution palynological record of the Iberian Margin. *Netherlands Journal of Geosciences* **79**, 345–354.

Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* **4**(11), 1–13.

Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver. 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Simmons, R.B. & Weller, S.J. (2001) Utility and evolution of cytochrome *b* in insects. *Molecular Phylogenetics and Evolution* **20**, 196–210.

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

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Swofford, D.L. (1999) PAUP 4.0 beta version. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, Massachusetts.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G. & Cosson, J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453–464.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Templeton, A.R. (1994) The role of molecular genetics in speciation studies. pp. 455–477 *in* Schierwater, B., Streit, B., Wagner, G.P. & DeSalle, R. (*Eds*) *Molecular Ecology and Evolution: Approaches and Applications*. Basel, Switzerland, Birkhauser Verlag.

Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L.K. (2000) Timing of the Last Glacial Maximum from observed sea-level minima. *Nature* **406**, 713–716.

Zink, R.M. (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society B: Biological Sciences* **271**, 561–564.

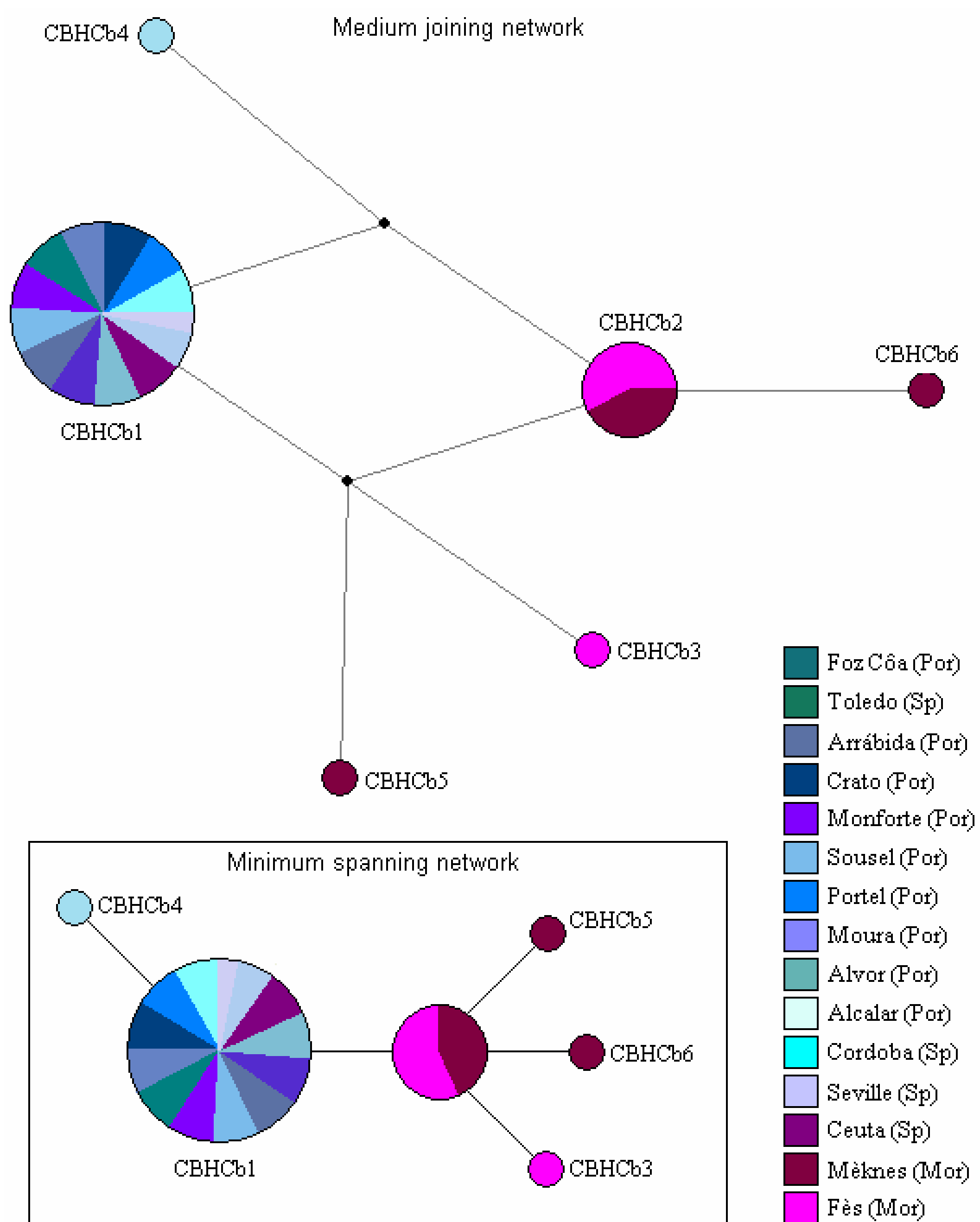


Fig. 2. Medium joining and minimum spanning networks for *Cicada barbara* in the Iberia Peninsula and Morocco based on cyt *b* haplotypes. The volume of the circles representing each haplotype is proportional to the total number of specimens showing that particular haplotype. The colours are representative of the number of specimens per locality. Each connection indicates one mutational step and black dots in the medium joining network represent missing intermediate haplotypes.

Bulletin of Entomological Research (2008) **98**, 1–14
doi:10.1017/S0007485307005561
© 2008 Cambridge University Press
Printed in the United Kingdom

Patterns of acoustic variation in *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco

G.A. Pinto-Juma[1,2]*, S.G. Seabra[1] and J.A. Quartau[1]

[1]Centro de Biologia Ambiental/Departamento de Biologia Animal, Faculdade de Ciências de Lisboa, C2, Campo Grande, 1749-016 Lisboa, Portugal: [2]School of Biosciences, Cardiff University, Cardiff CF10 3US, Wales, UK

Abstract

Field recordings of the calling song and of an amplitude modulated signal produced by males of *Cicada barbara* from North Africa and the Iberian Peninsula were analysed in order to assess the geographical acoustic variation and the potential usefulness of acoustic data in the discrimination of subspecies and populations. Sound recordings were digitized and the frequency and temporal properties of the calls of each cicada were analysed. In all regions studied, peak frequency, quartiles 25, 50 and 75% and syllable rate showed low coefficients of variation suggesting inherent static properties. All frequency variables were correlated with the latitude, decreasing from south to north. In addition, most acoustic variables of the calling song showed significant differences between regions, and PCA and DFA analyses supported a partitioning within this species between Iberian Peninsula + Ceuta and Morocco, corroborating mtDNA data on the same species. Therefore, the subspecific division of *C. barbara* into *C. barbara barbara* from Morocco and *C. barbara lusitanica* from Portugal, Spain and Ceuta finds support from the present acoustic analyses, a result which is also reinforced by molecular markers.

Keywords: acoustic signals, *Cicada barbara*, subspecies, acoustic divergence, calling song

(Accepted 29 January 2007)

Introduction

Cicadas are Hemipteran insects mainly characterized by the presence in males of cuticle membranes (tymbals) placed dorsolaterally in the first and second segments of the abdomen. Tymbals can be distorted by the action of powerful muscles producing loud airborne acoustic signals (e.g., Pringle, 1954; Popov, 1975; Young & Bennet-Clark, 1995; Bennet-Clark, 1997, 1998a, 1999). Males produce different kinds of acoustic signals, but the calling song involved in mate attraction is typically species-specific and females are only attracted to the calls of the conspecific males (e.g. *Magicicada* spp.: Cooley & Marshall, 2001).

Paterson (1985) pointed out that species are primarily characterized by unique specific mate recognition systems (SMRSs). These serve to ensure the mating of conspecifics and result secondarily in reproductive isolation (Claridge & de Vriijer, 1994). It is common to find that the most striking differences between closely related sympatric species concern characters used in mate recognition (*e.g*. Maynard-Smith, 1989); thus, SMRSs should allow the discrimination of

*Author for correspondence
Fax: +44 (0)2920874305
E-mail: jumag@cf.ac.uk

species. As such, calling songs constitute distinct specific-mate recognition systems (SMRSs) and have been commonly used to recognize cicada species in both natural and farmed ecosystems, especially when no other so distinctive characters are available, such as in the genera *Tosena*, *Pomponia* and *Lyristes* (e.g. Boulard, 1988, 2006), *Cicadetta* (e.g. Gogala & Trilar, 2004) or in *Tettigetta* (e.g. Boulard, 1982; Quartau & Boulard, 1995; Sueur *et al*., 2004). Paterson's (1985) recognition concept of species emphasizes that SMRSs and, therefore, the calling song should be maintained relatively constant by stabilizing selection across the distribution range of the species.

Cicada barbara Stål is known to occur in the Iberian Peninsula, Italy and in North Africa (Nast, 1972; Boulard, 1982; Quartau, 1988) in scattered high temperature environments and is usually associated with Mediterranean trees and shrubs, and frequently in farmed environments as olive tree orchards. This cicada produces a typical sound composed of a continuous series of pulses without pauses (Boulard, 1982, 1995; Fonseca, 1991; Quartau & Rebelo, 1994). Boulard (1982), based on small differences in the male genitalia and the female ovipositor, created a new subspecific taxon for *C. barbara* from Portugal, i.e. *C. barbara lusitanica* Boulard, in comparison with the nominal subspecies from North Africa. An investigation of the genetic differentiation also supports the splitting of this species into *C. barbara lusitanica* and *C. barbara barbara* Stål (see Pinto-Juma *et al*., this journal); however, their geographical boundaries may be questionable since a population in Ceuta (North Africa) has been found to cluster with the Iberian Peninsula specimens.

It should be pointed out that the discrimination of subspecies has been a source of controversy (e.g. Phillimore & Owens, 2006); and, indeed, biological information from different sources should be used when defining species (Lee, 2004). At the subspecific level, the use of complementary data should also assist on classification.

Intraspecific variation in the calling song can occur at a variety of scales and geographical population divergence can lead to reproductive isolation and subsequent speciation (Zuk *et al*., 2001). Otte (1994, *in* Shaw & Herlihy, 2000) suggested that acoustic features of the calling song of swordtail crickets in Hawaii are among the earliest characters to change in the speciation process. The same was also referred to *Drosophila willistoni* complex (Gleason & Ritchie, 1998). Boulard (1995) also referred to the potential importance of comparing the calling songs of *C. barbara barbara* and *C. barbara lusitanica*, but no such comparison has been made yet. On the other hand, the courtship song of both subspecies has been compared by Boulard (1995), who stated that *C. barbara lusitanica* from Portugal has shorter phrases with higher periodicity in comparison with *C. barbara barbara* from North Africa. However, no statistical significance tests have ever been accomplished. The courtship song has been described as a continuous amplitude modulated sound, presenting also modulation in frequency and in syllable period (Fonseca, 1991; Boulard, 1995).

In this paper, the acoustic variation of the calling song, as well as of an amplitude modulated signal of *C. barbara* are assessed and compared in both the Iberian Peninsula and North Africa populations. Moreover, the usefulness of the present acoustic data in the splitting of *C. barbara* into subspecies, *C. barbara barbara* and *C. barbara lusitanica*, is evaluated.

Fig. 1. Recording sites in the Iberian Peninsula and North Africa of the *Cicada barbara* specimens studied.

Material and methods

Specimens of *Cicada barbara* Stål were sampled in the summer (July–September) over several years (1996–2002) from different localities of the Iberian Peninsula (*C. barbara lusitanica*) and North Africa (*C. barbara barbara*) (fig. 1, table 1). Males were identified in the field by their calling songs, located and recorded on a digital Sony DAT recorder (TCD-D10 ProII and TCD-D8 at frequency ranges of 20–22,000 Hz and 20–20,000 Hz, respectively and at a sampling frequency of 44.1 kHz) connected to a dynamic Sony F-780 microphone with frequency responses of 50–18,000 Hz. All recordings were made approximately at the same distance from the cicada (within at least 30 cm to 1 m), to avoid distance interference on the frequency measurements. Air temperature was measured at most collecting sites at the place where cicadas were singing.

Acoustic analysis

The software Avisoft-SASLab Pro (Specht, 2002) was used to digitize sound recordings as in Pinto-Juma *et al*. (2005). A sound fragment of about 60 s was used to produce oscillograms, sonagrams (or spectrograms) and mean amplitude spectra (using Fast Fourier transformation with a resolution of 512 points and a Hamming Window and 50% overlap for temporal resolution), allowing temporal and spectral analyses. The frequency variables measured included: peak frequency; minimum and maximum frequencies; bandwidth; quartiles 25, 50, 75% and quartile 75%–quartile 25%. The number of syllables (a syllable being a unitarian group of pulses, e.g. Fonseca, 1991) was counted per unit of time – 30 fragments of about 0.1 s were analysed and the average number of syllables was calculated for each specimen.

Table 1. Sampled populations of *Cicada barbara*.

Localitites	Date	Habitat	Temperature (°C)	*N*
North Africa (*C. barbara barbara*)				
Ceuta (Spain)	21/7 and 22/7/1999	Eucalyptus	29–35	12
Fès South (Morocco)	3/8/2001	Olive tree orchard	34	6
Fès (Morocco)	2/8/2001	Olive tree orchard	31–35	10
Meknès (Morocco)	4/8/2001	Olive tree orchard	31–37	9
Iberian Peninsula (*C. barbara lusitanica*)				
Alcalar (Algarve, Portugal)	23/8/1995		–	10
Alvor (Algarve, Portugal)	28/8/1995		–	3
Arrábida (Estremadura, Portugal)	27/7/1995	Pine trees	–	5
Cordoba (Andalucía Spain)	6/9/2000	Olive tree orchard	34	5
Crato (Alto Alentejo, Portugal)	6–8/7/1999; 15/7/1999; 1–3/8/1999	Olive tree orchard	26–41	14
Monforte (Alto Alentejo, Portugal)	22–24/7/1995	Olive tree orchard	–	6
Moura (Baixo Alentejo, Portugal)	28/8/2001	Olive tree orchard	31–34	11
Portel (Alto Alentejo, Portugal)	24/7 and 10/8/2001	Olive tree orchard	31–35	10
Sevilla (Andalucía, Spain)	6/8/2001	Olive tree orchard	38–41	7
Sousel (Alto Alentejo, Portugal)	8/9/2001	Olive tree orchard	33–35	11
TOTAL				**119**

Temperature intervals refer to the range of temperature within each collecting site at the place and time where cicadas were singing; *N*, number of males acoustically recorded.

Statistical analyses were performed using MINITAB version 14 software (Minitab Inc., 2004). Anderson-Darling Normality tests were performed to verify the distribution for each variable; and, since the null hypothesis of a normal distribution was always rejected, all the subsequent analyses followed non-parametric statistics.

Following Boulard's (1982) subspecific classification, the new mtDNA data (see Pinto-Juma *et al*., this journal) and the fact that no apparent barriers exist between the populations from Portugal and Spain the Iberian Peninsula specimens were considered as *C. barbara lusitanica*. Similarly, North African specimens were considered as *C. barbara barbara*. However, since the specimens from Ceuta in northern Africa revealed the same genetic pattern as the Iberian ones (see Pinto-Juma *et al*., this journal), several tests were carried out grouping Ceuta with the remaining North African localities and, on the other hand, with the Iberian ones.

Spearman's rank correlations were analysed between time and frequency variables and the temperature in the field. Descriptive statistics of each variable for each region were also analysed, including the coefficient of variation. The coefficients of variation (CV) for each variable were calculated within populations (CVpop) and within regions (CVreg). Since these coefficients are the standard deviation expressed as a percentage of the mean of each variable, it was possible to compare the variation between different sets of data. Mann-Whitney U (MW) and Kruskal-Wallis (KW) non-parametric tests were used to compare regions and populations within each region, respectively. The significance of multiple tests was adjusted according to Dunn-Sidák method (Dytham, 2003), by reducing the critical *P* value from 0.05 to $1–(0.95)^{1/k}$, where k is the number of tests performed. The correlation between each variable and the latitude was also assessed in order to test any variation gradient within the geographical distribution of the specimens.

Frequency measurements and the syllable rate were used to assess differences between specimens and regions using multivariate methods. A Principal Component Analysis (PCA) was performed in order to reduce the variables to a small number of components, as well as to assess the correlation between the variables and those components (component loadings). Kruskall-Wallis and Mann-Whitney tests were then used to compare the component scores obtained for the individuals between regions. Discriminant Function Analysis (DFA), with two and three regions defined *a priori*, was also performed to determine statistical significant discriminant functions that might separate these groups.

Besides the more generally known calling song, amplitude modulated signals were analysed in five males of *C. barbara lusitanica* from Portugal (one from Moura, two from Portel and one from Sousel, all recorded in 2001; and one from Crato, recorded in 1997) and in eight males of *C. barbara barbara* from North Africa (three from Meknès, three from Fès and one from Fès South, all recorded in 2001; and one from Ceuta, recorded in 1999). Recordings and digitizing were performed as described above. Ten phrases per male, each composed of one high amplitude section (Section I) and one low amplitude section (Section II) were analysed whenever possible. The peak frequency and duration of each section in each phrase were obtained using Avisoft and the number of syllables in three fragments of about 0.1 s was counted for each section of each phrase and an average was obtained for each section per specimen.

Non-parametric tests were used to compare the two phrasal sections (Wilcoxon test for two related samples) and to compare African and Iberian samples (Mann-Whitney U test for two independent samples). Correlations between the acoustic variables and the temperature were calculated using the Spearman correlation coefficient.

Results

Acoustic analyses were based on the recordings of a total of 119 specimens from 14 localities, four from North Africa and ten from the Iberian Peninsula (fig. 1, table 1).

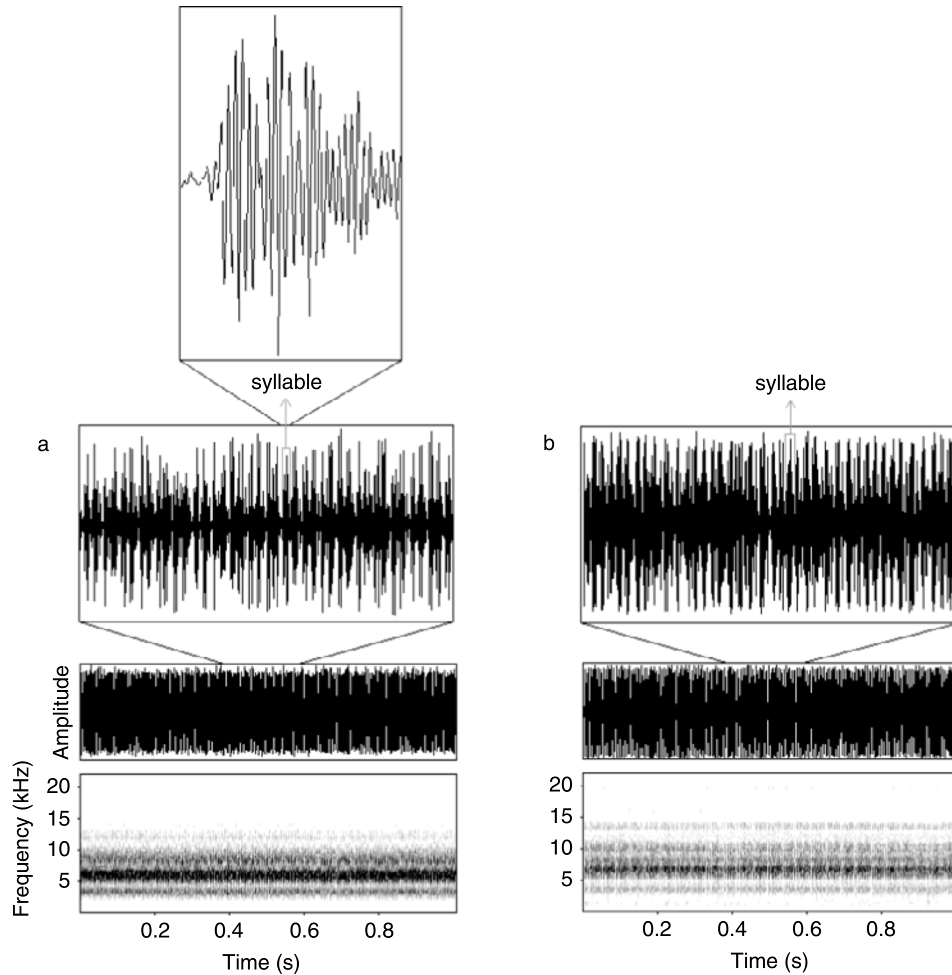


Fig. 2. Oscillograms (time vs. amplitude) and sonagrams (time vs. frequency) of the calling songs of (a) one male of *Cicada barbara lusitanica* from Crato, Portugal and (b) one male of *C. barbara barbara* from Fès, Morocco. A fragment of 0.2 seconds is shown, with an individual syllable pointed out, which is further magnified in (a) to show the pulses.

Calling song

The calling song of *Cicada barbara* consists of a continuous emission of pulses produced by the tymbals (fig. 2). Each tymbal is responsible for the repetitive production of a group of pulses which we call syllable. The tymbals alternate their sound production, but slightly overlap (Fonseca, 1991). In the calling song, each syllable lasts about 5 ms (*ca.* 200 syllables s^{-1}; table 2). In the males analysed ($N = 119$), two were apparently either using only one tymbal or there was a synchronous action of both tymbals, as revealed by the oscillograms with half of the number of syllables produced

Table 2. Descriptive statistics of the acoustic variables of *Cicada barbara*.

	N	Mean ± SD	Min.	Max.	CV (%)
Interval of time analysed (s)	119	57.6 ± 9	10.2	68.4	
Peak frequency (Hz)	119	6330 ± 437.3	5160	7660	6.9
Minimum frequency (Hz)	119	2845.2 ± 346.3	940	4730	12.5
Maximum frequency (Hz)	119	11291 ± 1791	7920	17910	15.9
Bandwidth (Hz)	119	8442 ± 1800	5160	15330	21.3
Quartile 25% (Hz)	119	5924.2 ± 363.2	4990	6710	6.1
Quartile 50% (Hz)	119	6740.5 ± 384.3	5770	7490	5.7
Quartile 75% (Hz)	119	8708 ± 701.2	6460	10240	8.1
Quart75%–Quart25% (Hz)	119	2783.9 ± 453	1300	4040	16.3
Syllables s^{-1}	116	200.9 ± 18.7	137.6	252.3	9.3

SD, standard deviation; CV, coefficient of variation.

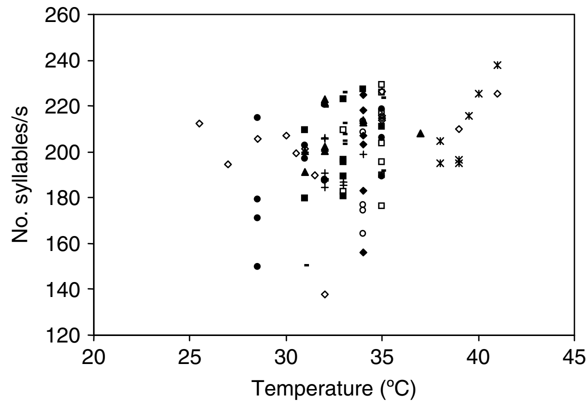


Fig. 3. Variation of the syllable rate with the temperature in specimens of *Cicada barbara*. ◆, Fès south; ■, Fès; ▲, Meknès; ●, Ceuta; ⚹, Sevilla; ○, Cordoba; +, Moura; −, Portel; □, Sousel; ◇, Crato.

per unit of time. In this case, the frequency values were not significantly different from when both tymbals were functioning (MW test, $P > 0.05$), but the variable 'syllable rate' had to be discarded.

The peak frequency for all the specimens analysed varied from 5160 to 7660 Hz. The bandwidth ranged between 5160 and 15,330 Hz, and the difference between upper and lower quartiles ranged between 1300 and 4040 Hz (table 2).

No correlation between frequency variables and temperature was detected. However, the syllable rate was slightly correlated with temperature: $r_S = 0.34$ ($P < 0.001$) when considering all specimens (see fig. 3).

The calling songs differed significantly between populations for all variables (KW tests, $P < 0.001$), except for the mean syllable rate ($P = 0.06$). For all population pairwise comparisons within the Iberian Peninsula, MW tests ($P < 0.05$) showed that Moura and Sousel had higher maximum frequencies (see appendix I). Within North African populations, including Ceuta samples, the variables peak frequency, minimum frequency, and quartiles 25, 50 and 75% differed significantly. Ceuta specimens seem responsible for most of these differences, since significantly lower values for peak frequency and quartiles 25% and 50% were found in Ceuta when compared to the Morocco samples (MW tests in table 3, boxplots in fig. 4).

When temperature was controlled by performing ANCOVA (temperature as covariate), the differences between North African populations were still present for most variables, except for minimum frequency. This does not necessarily contradict the non-correlation between frequency variables and temperature. In fact, ANCOVA is the only test which allows controlling the temperature, but it is a parametric test, so the minimum frequency result might be a simple statistical artefact due, for instance, to a violation in the test criteria.

The influence of temperature could not be tested to explain the low syllable rate values found in Alvor, as this information was missing. However, when excluding specimens for which there was no temperature data, there was no significant differences between populations in syllable rate (KW test, $P > 0.05$).

A negative significant correlation between each frequency variable of all populations and their latitude was found ($P < 0.05$), indicating that there is a decreasing gradient on frequency parameters from south to north (table 4, appendix I). The syllable rate did not show a significant correlation with latitude when all populations were analysed.

Calling song and the differentiation between regions

Corroborating molecular analyses, the Ceuta population revealed considerable acoustic differences in comparison with the remaining North African populations, so multiple MW tests (with the significance level corrected via Dunn-Sidák method) were performed in order to evaluate different subregions (table 3). The mean syllable rate and the minimum frequency were not significantly different for any of the subregions compared, but all other frequency variables showed differences depending on the regions.

Ceuta revealed more highly significant differences when compared to Morocco (peak frequency, quartile 25% and quartile 50%) than when compared to the Iberian Peninsula (only bandwidth). MW tests between Morocco vs. Iberian Peninsula + Ceuta also revealed more significant differences than North Africa (Morocco and Ceuta) vs. Iberian Peninsula (see table 3). All comparisons where Ceuta was included suggested less differentiation between this population and the Iberian Peninsula than with Moroccan populations.

Table 3. Mann-Whitney U tests comparing two regions for each acoustic variable in *Cicada barbara*.

Variables tested	Morocco + Ceuta vs. Iberian Peninsula	Morocco vs. Iberian Peninsula + Ceuta	Morocco vs. Iberian Peninsula	Morocco vs. Ceuta	Iberian Peninsula vs. Ceuta
	P	*P*	*P*	*P*	*P*
Peak frequency	0.264	0.000	0.003	0.000	0.016
Minimum frequency	0.513	0.022	0.052	0.004	0.089
Maximum frequency	0.000	0.000	0.000	0.620	0.003
Bandwidth	0.000	0.001	0.000	0.987	0.000
Quartile 25%	0.050	0.000	0.001	0.000	0.247
Quartile 50%	0.031	0.000	0.001	0.000	0.343
Quartile 75%	0.000	0.000	0.000	0.013	0.341
Quart75%–Quart25%	0.001	0.003	0.001	0.962	0.003
Syllables s^{-1}	0.938	0.379	0.497	0.177	0.385

Significant differences highlighted in grey are according to the method of Dunn-Sidák (Dytham, 2003) ($P < 1–(0.95)^{1/k}$, where k = number of tests performed, in this case k = 45, so $P < 0.001$).

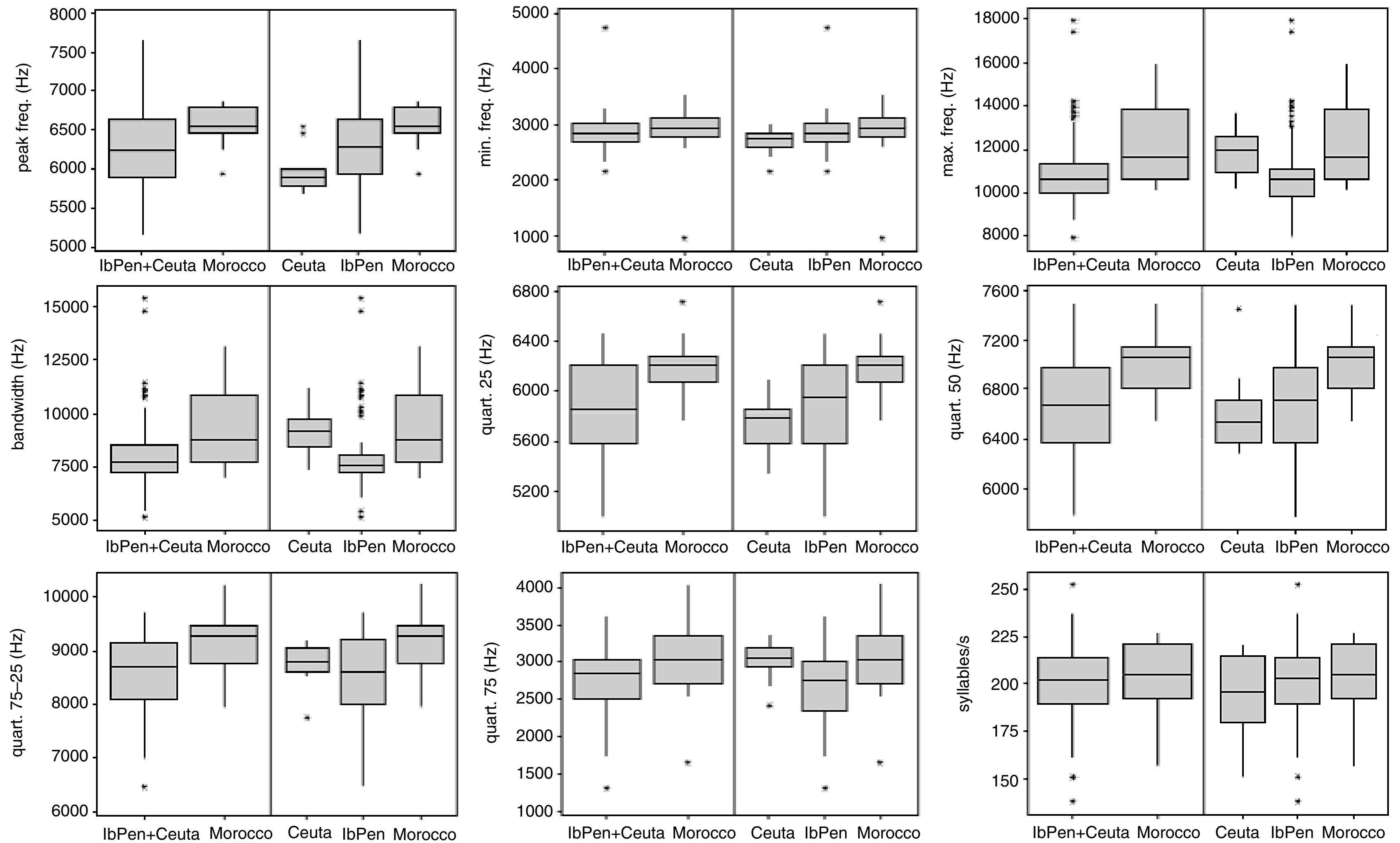


Fig. 4. Boxplots of the acoustic variables analysed for *Cicada barbara* at two regions (Iberian Peninsula + Ceuta and Morocco) and three regions (Ceuta, Iberian Peninsula and Morocco).

Boxplots represent the sample distribution: the rectangular box corresponds to approximately the middle 50% of the data, the top of the box is the third quartile (75%), the bottom is the first quartile (25%) and the horizontal line is the median; vertical lines extending to either side indicate the general extent of the data; *, outlier.

Table 4. Spearman correlation (r_S) between latitude and acoustic variables analysed for *Cicada barbara*.

	Peak frequency (Hz)	Minimum frequency (Hz)	Maximum frequency (Hz)	Bandwidth (Hz)	Quartile 25% (Hz)	Quartile 50% (Hz)	Quartile 75% (Hz)	Quart75% – Quart25% (Hz)	Syllables s^{-1}
r_S	−0.267	−0.219	−0.431	−0.407	−0.351	−0.391	−0.385	−0.317	0.102
P	0.003	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.276

P is the level of significance; for $P<0.05$, the correlation is significant. Significant differences are highlighted in grey.

Table 5. Mean, standard deviation (SD) and coefficient of variation (CV) for each acoustic variable of *Cicada barbara* on each region.

	Morocco				Iberian Peninsula + Ceuta			
	N	Mean ± SD	CVpop (%)	CVreg (%)	N	Mean ± SD	CVpop (%)	CVreg (%)
Peak frequency (Hz)	25	6586.7 ± 241.6	3.83	3.67	94	6261.7 ± 453	4.94	7.23
Minimum frequency (Hz)	25	2904.9 ± 470.4	15.81	16.19	94	2829.3 ± 320.6	8.72	11.33
Maximum frequency (Hz)	25	12371 ± 1799	14.65	14.54	94	11004 ± 1684	12.73	15.3
Bandwidth (Hz)	25	9461 ± 1941	21.53	20.51	94	8171 ± 1669	17.10	20.42
Quartile 25% (Hz)	25	6174.5 ± 208.1	3.40	3.37	94	5857.6 ± 367.3	4.26	6.27
Quartile 50% (Hz)	25	6985.2 ± 232.4	3.25	3.33	94	6675.5 ± 391.3	3.77	5.86
Quartile 75% (Hz)	25	9188 ± 527	5.75	5.74	94	8580.5 ± 688.5	6.12	8.02
Quart75%–Quart25% (Hz)	25	3013 ± 451.8	14.63	15	94	2722.9 ± 435.6	13.52	16
Syllables s^{-1}	24	203.52 ± 17.58	8.94	8.64	92	200.18 ± 19.06	8.13	9.52

CVpop (%), average of the CV between populations in each region; CVreg (%), CV between individuals within the region.

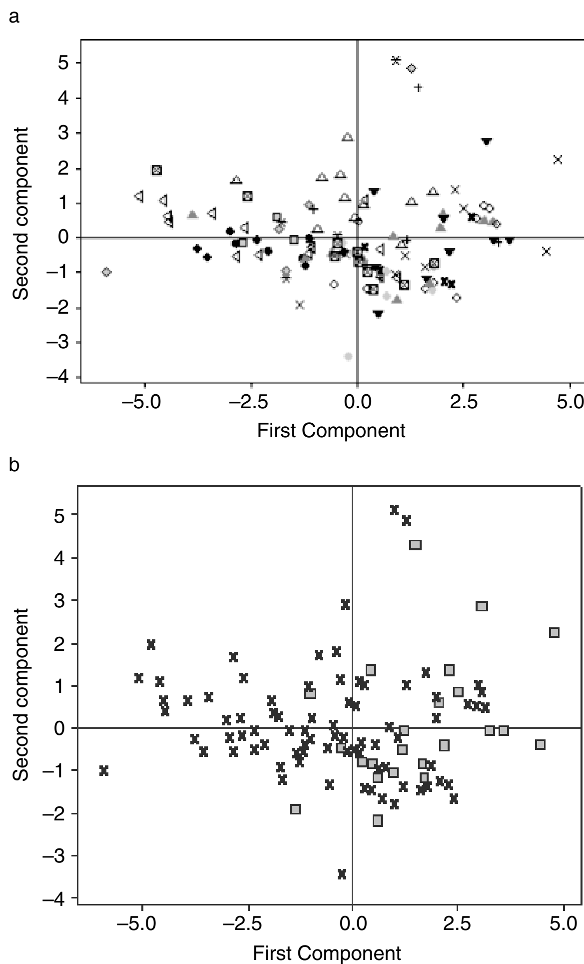


Fig. 5. Component scores obtained from the Principal Component Analysis (PCA) based on a correlation matrix between all nine acoustic variables analysed for *Cicada barbara*: (a) ◆, Alc; ▩, Alv; ◈, Arr; △, Ceu; ✖, Cor; ◁, Cra; ▼, Fès; +, FèsS; ×, Mek; ⋇, Mon; ◇, Mou; ⊠, Por; ◆, Sev; ▲, Sou population level and (b) ✖, Iberian Peninsula + Ceuta; ▩, Morocco region level.

Similar levels of coefficients of variation (CV) were found within populations in the Iberian Peninsula + Ceuta and within populations in Morocco (table 5). However, within-region CV were generally higher in Iberian Peninsula + Ceuta than in Morocco; only the CV of the minimum frequency and of the bandwidth for Morocco was higher than the CV of Iberian Peninsula + Ceuta. The variable presenting the highest CV of all was bandwidth, while the lowest CV was found for quartile 50%, for both regions.

PCA did not reveal non-overlapping groups between regions, however Moroccan specimens predominantly aggregated in the upper and lower right quadrants of the diagram, while Iberian specimens were mostly spread along the upper and lower left quadrants (fig. 5). KW tests showed significant differences for the first two components between the scores of specimens of Morocco, Iberian Peninsula and Ceuta ($P<0.001$). MW tests between Morocco and Iberian Peninsula + Ceuta also revealed highly significant differences for the first two components (same level of significance, $P<0.001$). Comparing Morocco and the Iberian Peninsula alone, MW tests revealed even stronger significant differences for component 1 with $P<0.0001$; Morocco compared to Ceuta also had significant differences ($P<0.005$), while the Iberian Peninsula compared to Ceuta did not show significant differences ($P=0.650$).

The proportion of the total variance (table 6a) accounted for by the first component was 52%, the first two combined

Table 6. Multivariate analyses based on acoustic variables for *Cicada barbara*.

a) Principal Component Analysis (PCA) based on nine acoustic variables: eigen analysis and component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigen analysis							
Eigen value	4.679	1.714	1.033	0.701	0.521	0.261	0.093
Proportion	52%	19%	11.5%	7.8%	5.8%	2.9%	1%
Cumulative	52%	71%	82.5%	90.3%	96.1%	99%	100%
Component loadings per variable							
Peak frequency	0.319	−0.309	0.116	−0.562	0.096	0.672	
Minimum frequency	0.196	−0.472	−0.23	0.489	0.641	0.064	
Maximum frequency	0.343	0.452	−0.046	−0.069	0.423	−0.065	
Bandwidth	0.302	0.544	0.001	−0.167	0.293	−0.077	
Quartile 25%	0.395	−0.298	−0.011	−0.139	−0.096	−0.516	
Quartile 50%	0.411	−0.206	0.103	−0.19	−0.191	−0.311	
Quartile 75%	0.439	−0.012	−0.028	0.248	−0.302	0	
Quart75%–Quart25%	0.364	0.219	−0.035	0.494	−0.39	0.413	
Syllables s^{-1}	−0.011	0.028	−0.959	−0.226	−0.159	0.026	

PC, principal component.

b) Discriminant Function Analysis (DFA) results based on seven acoustic variables.

	Two Regions		Three regions		
	Morocco	Iberian Peninsula + Ceuta	Morocco	Iberian Peninsula	Ceuta
Total *N*	24	95	24	83	12
Correct classification	17	61	16	43	9
Misplaced	7	34	8	40	3
Proportion of correct classification	70.8%	63.0%	66.7%	49.4%	63.6%
Overall correct classification	64.7% (with cross validation 60.3%)		54.3% (with cross validation 45.7%)		

The frequency variables Q75%–Q25% and bandwidth were highly correlated with other predictors and had to be discarded from this analysis.

components accounted for 71%, and 82.5% for the first three components. The cumulative variance accounted by consecutive components decreased rapidly, which is indicative of significant inter-correlation between the variables. Most variables (excluding the syllable rate which had the lowest loading attributed, 0.011) were almost equivalently responsible for the first component, the loadings varying from 0.196 (minimum frequency) to 0.439 (quartile 75%). Even though most frequency variables were cross-correlated, the score plot of the PCA showed a better separation of the regions than other plots using only non-correlated variables (data not shown).

For DFA, two different groupings were tested: one separating Morocco from the Iberian Peninsula + Ceuta and the other separating three regions, Morocco, Iberian Peninsula and Ceuta (table 6b). The frequency variables Q75–Q25 % and bandwidth had to be discarded from this analysis, as they were highly correlated with other variables (predictors).

DFA separating Morocco from the Iberian Peninsula + Ceuta had a global proportion of correct classification of 64.7%; after cross validation, the proportion was 60.3%. Morocco specimens were less misclassified than the ones from Iberian Peninsula + Ceuta since Morocco alone had 70.8 % of correct classification, while the Iberian Peninsula + Ceuta had 63%. The overall correct categorization of DFA for the three regions, considering Ceuta as a separate region, was 54.3% and 45.7% after cross validation. For each region separately, Morocco had 66.7% of correct categorization, the Iberian Peninsula had the lowest proportion of correct categorization (49.4%) and Ceuta had 63.6%. The lower overall proportion of correct classification observed suggests that the two regions defined in the first test, Morocco and the Iberian Peninsula together with Ceuta, are more adequate to explain the differentiation between the specimens.

Amplitude modulated signal

The amplitude modulated signal analysed here is similar in pattern to the one referred by Fonseca (1991) and Boulard (1995) as a courtship song. However, from our field observations, we cannot corroborate that it is truly a courtship song, being instead generally associated with the disturbance caused by the approaching of the operator during the field recordings. However, this is an ethological issue that requires further clarification.

In the amplitude modulated signal, Section I of the phrase had higher amplitude (fig. 6) and also significantly lower duration and higher syllable rate than Section II (Wilcoxon test, $P = 0.001$ for both) (table 7). On the other hand, peak frequency was not significantly different between Sections I and II (Wilcoxon test, $P = 0.701$). Temperature had no significant correlation with any of the acoustic variables measured in this modulated signal (Spearman correlation, $P > 0.1$). There was no significant difference between African and Iberian samples for each variable or between Morocco samples and Iberian and Ceuta samples taken together (MW

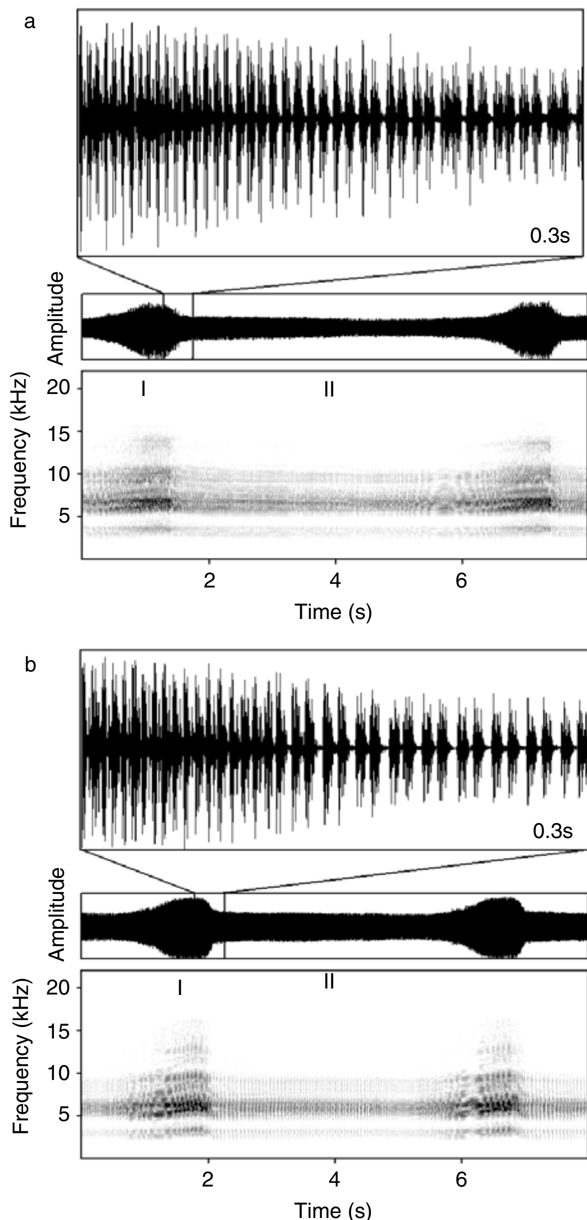


Fig. 6. Oscillogram and sonagram of the amplitude modulated signal of (a) one male of *Cicada barbara lusitanica* from Moura and (b) one male of *C. barbara barbara* from Fès. Sections I and II of the signal are signalled; and a fragment of 0.3 seconds of the transition between sections is shown enlarged on top, depicting the clear difference in syllable rates between Section I and II.

tests, $P > 0.1$). The acoustic variable with the highest coefficients of variation was the temporal variable duration (CV from 21.3% to 32.7%). On the other hand, the peak frequency and the syllable rate were more constant (CV from 5.2% to 10.8%) (table 7). Also, the intra-individual coefficients of variation of duration were, on average, 14.5% for Section I and 29.4% for Section II and were lower for peak frequencies and syllable rate (2% and 3.7% for the first; 4.4% and 8% for the latter).

Discussion

Acoustic variation

The values found for the acoustic variables of the calling song in the present study are within the range found by other authors. Fonseca (1991) reported syllable rates from $162\,s^{-1}$ to $238\,s^{-1}$ on specimens of *C. barbara lusitanica* (recorded in the field with free tethered animals, but unfortunately without reference to the locality) and a peak frequency of 5.5–6.5 kHz, a minimum frequency of 2–3 kHz and a maximum of 9.5–11 kHz. Boulard (1995) described the calling song of a male of *C. barbara barbara* from North Africa with a peak frequency at 6.2 kHz. If we consider that the two regions analysed in this study, Iberian Peninsula + Ceuta and Morocco, correspond to the subspecies *C. barbara lusitanica* and *C. barbara barbara*, respectively, the two subspecies showed different means but with overlapping values for these acoustic variables: *C. barbara lusitanica* had a mean ($\pm$SD) peak frequency of 6.3 ± 0.4 kHz, a minimum frequency of 2.8 ± 0.3 kHz, a maximum frequency of 11 ± 1.7 kHz and a syllable rate of $200 \pm 19.1\,s^{-1}$; while *C. barbara barbara* had a mean peak frequency of 6.6 ± 0.2 kHz, a minimum frequency of 2.9 ± 0.5 kHz, a maximum frequency of 12.4 ± 1.8 kHz and a syllable rate of $203 \pm 17.6\,s^{-1}$.

Temperature had a significant effect on the syllable rate, which increased with air temperature, as observed for this species previously (Fonseca, 1991) and for other cicadas (e.g. *Tettigeta argentata*, *T. josei* and *Tympanistalna gastrica*, in Fonseca & Revez (2002a)).

The syllable rate did not show any significant differences between populations and regions in our study, although most other acoustic characters revealed significant differences. However, variability between individuals (within populations) was relatively high (see appendix II), increasing the complexity of the results. According to Gerhardt (1991), the properties of the calling songs of frogs and insects can be categorized as static, responsible for the quality of calls, mainly concerning species recognition; or dynamic, determining the quantity of signalling, and more related with mating success. Static characters should be highly stereotyped within and between males, whereas dynamic characters can vary considerably (Gerhardt, 1991). We did not assess within-male variability as we would need several calling songs from the same male in the field, but variability between males can also show this pattern. In fact, for *C. barbara* in both regions studied, the peak frequency, quartiles 25, 50 and 75% seem to be static characters, according to Gerhardt (1991) classification, usually with less than 5% variability. The minimum and maximum frequency, bandwidth and quartiles 75–25% varied in most cases more than 12%, corresponding, therefore, to the Gerhardt dynamic characters. However, these particular characters depend considerably on recording settings, like distance, and the variability found might also reflect some recording heterogeneity. The syllable rate showed coefficients of variation between 5–8%, slightly higher than static properties but lower than most dynamic characters. Nonetheless, Fonseca & Revez (2002b) proved that the temporal pattern in *C. barbara lusitanica* can influence long-range communication; so, this character might be constrained by specific selective

Table 7. Descriptive statistics for the acoustic variables measured in the phrases of the amplitude modulated signal of *Cicada barbara barbara* and *C. barbara lusitanica*.

		C. barbara barbara				*C. barbara lusitanica*				Average intra-individual CV (%)
		Average ± SD	Minimum	Maximum	CV (%)	Average ± SD	Minimum	Maximum	CV (%)	
Section I	Peak frequency (Hz)	6334.1 ± 322.6	5758	6854	5.2	5971.2 ± 627.9	5137.1	6771.7	10.8	2
	Duration (s)	0.9 ± 0.3	0.7	1.6	29.9	0.9 ± 0.2	0.7	1.1	21.3	14.5
	Syllable rate (s^{-1})	217.8 ± 17.8	175.1	229.1	8.4	225.1 ± 14.3	209.9	242.6	6.5	4.4
Section II	Peak frequency (Hz)	6415.6 ± 343.8	5791	6860	5.5	6137.1 ± 506.6	5287.1	6534	8.5	3.7
	Duration (s)	2.8 ± 0.9	1.9	4.4	32.7	2.7 ± 0.9	2.1	4.1	32.3	29.4
	Syllable rate (s^{-1})	123.1 ± 9.9	101.4	133.6	8.2	132.4 ± 7.3	120.5	140.6	5.7	8
	Phrase Duration (Section I + Section II)	3.8 ± 1.1	2.8	5.6	30.4	3.6 ± 1	2.9	5.3	29.2	22.3
	Duration Section II/ Duration Section I	3.1 ± 0.6	1.9	4.1	20.3	3.1 ± 0.5	2.6	3.7	15.3	35

SD, standard deviation; CV, coefficient of variation.

pressures and perform as static properties. Cicadas are difficult to keep in a laboratory, and females are particularly difficult to observe in the field. Therefore, female preference' studies are hard to develop and, for *C. barbara*, there is no literature available. Fonseca & Revez (2002b) studied the males' response to natural and modified songs and managed to successfully determine the relative importance of song parameters in song discrimination. Besides the importance of the temporal pattern, they also recognized that the most attractive peak frequency was 6 kHz and that *C. barbara* males can discriminate peak frequencies differing by 1–2 kHz. The amplitude modulated signal analysed in our study had similar low coefficients of variation for the peak frequency and the syllable rate of the calling song, corroborating the significance of these parameters on within species communication. However, future female choice experiments are needed to elucidate this significance.

The generally higher acoustic variability among populations in the Iberian Peninsula compared to Morocco might have been due to sampling effort, since more populations were analysed in Iberia. Habitat heterogeneity might also have had some influence, since topographic and latitudinal differences between sampling sites in the Iberian Peninsula were higher than in Morocco. Furthermore, all populations studied from Morocco occur in olive tree orchards, while in the Iberian Peninsula (and Ceuta) some of the populations analysed were also found in pine and even in eucalyptus trees (table 1).

All populations analysed here revealed significant differences for most acoustic variables. This differentiation might be partially due to a 'chorusing' effect; cicadas tend to sing in groups apparently to increase the chance of attracting conspecific females and to avoid predators (Villet, 1992; Cooley & Marshall, 2001; Fonseca & Revez, 2002b) or as a result of inter-male competition (Greenfield *et al.*, 1997; Sueur & Aubin, 2002), which might have supported a scattered distribution favouring divergence between different populations, as also hypothesised for *C. orni* populations in Europe (Pinto-Juma *et al.*, 2005). Moreover, sound propagation depends not only on the insects' anatomy and physiology but also on the physical environmental conditions (Bennet-Clark, 1998b), so cicadas may need to adjust some qualities of the calling songs to local environmental conditions in order to maximise sound communication.

Regional variation and subspecies

Although most acoustic characters revealed significant differences between populations and regions, the statistical analyses performed did not completely separate any region or subspecies and an overlap of specimens from different regions was often observed. In addition, the amplitude modulated signal on which Boulard (1995) found support for the splitting of the species into *C. barbara barbara* and *C. barbara lusitanica* did not show any significant differences between these subspecies in our study.

The low but significant correlation between most acoustic variables and latitude showed that there is a slightly gradient, with all frequency variables decreasing from south to north. This gradient and some high coefficients of variation for some variables probably complicated the effective separation of regions. Nonetheless, the significant differences of most acoustic variables between regions do suggest a partitioning within this species between Iberian Peninsula + Ceuta and Morocco. The higher level of variation in the Iberian Peninsula compared to Morocco certainly influenced the results obtained for the DFA, as shown by the high proportion of misclassified samples in the Iberian Peninsula. The Ceuta population tended to separate from the other regions, but the split of this population from the Iberian Peninsula did not have statistical support. On the other hand, the separation between Morocco and the remaining populations was supported by all analyses performed. Because these calling songs are likely to be constrained by species-specific morphological/physiological characteristics, significant differences at higher levels, such as at the regional level, might indicate partitioning within species; and, as such, the present data support the subspecific division of *C. barbara* of Boulard (1982) with *C. barbara lusitanica* present in the Iberian Peninsula and

Ceuta and *C. barbara barbara* present in Morocco. The fact that all static properties of the calling song, excluding the syllable rate, had significant differences between these two regions reinforces this supposition since these variables are probably more associated to species recognition (Gerhardt, 1991).

The Ceuta population, in spite of being located in North Africa, seems to be more similar to the Iberian Peninsula populations than to those located in the Moroccan mainland south of the Rif Mountains. This might be expected since the Rif Mountains might have caused the segregation of Ceuta and the Moroccan populations analysed and, thus, increased their divergence. On the other hand, the strait of Gibraltar, separating the Iberian Peninsula from Ceuta, is relatively short and does not appear to have acted as an effective barrier. These results corroborate and complement the analyses of the populations at the molecular level (Pinto-Juma *et al*., this journal) since the same separation between regions was achieved.

Even though the acoustic signals analysed are variable at the inter-individual and population levels, the acoustic divergence observed at the regional level seems more consistent, as observed also in *Cicada orni* (Pinto-Juma *et al*., 2005), and may be useful to distinguish between closely related taxa. However, determining the level of differences between two groups of populations of a given species in order to split it into independent subspecies is highly complex.

In the present study, *C. barbara* presented a highly stereotyped calling song, which, notwithstanding this, allowed some discrimination between regions and, therefore, supported its splitting into two subspecies. But it should be noted that had this marker been applied alone it would not be sufficient to sustain such subspecific differentiation. However, the combination of the acoustic data with the molecular analyses for the same populations have reinforced and strengthened the delimitation of the two different subspecies within *C. barbara*.

Acknowledgements

We are very grateful to Genage André, Paula Simões, Teresa Rebelo, Mónica Ribeiro and Teresa Fernandes (Faculty of Science, University of Lisbon) for their help with field work. Moreover, we thank Paula Simões, Octávio Paulo and Michael Bruford for having discussed some topics analysed here with us. This research was supported by two PhD grants (PRAXIS BD/18229/98 and SFRH/BD/1027/2000) from the Fundação para a Ciência e Tecnologia (Lisbon, Portugal).

References

Bennet-Clark, H.C. (1997) Tymbal mechanics and the control of song frequency in the cicada *Cyclochila australasiae*. *Journal of Experimental Biology* **200**, 1681–1694.

Bennet-Clark, H.C. (1998a) How cicadas make their noise. *Scientific American* **5**, 36–39.

Bennet-Clark, H.C. (1998b) Size and scale effects as constraints in insect sound communication. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **353**, 407–419.

Bennet-Clark, H.C. (1999) Resonators in insect sound production: How insects produce loud pure-tone songs. *Journal of Experimental Biology* **202**, 3347–3357.

Boulard, M. (1982) Les cigales du Portugal, contribution à leur étude (Homoptera, Cicadidae). *Annales de la Société Entomologique de France* **18**, 181–198.

Boulard, M. (1988) Les *Lyristes* d'Asie Mineure (Homoptera, Cicadidae). I. Sur deux formes éthospécifiques syntopiques et description de deux espèces nouvelles. *L'Entomologiste* **44**, 153–167.

Boulard, M. (1995) Postures de cymbalisation, cymbalisations et cartes d'identité acoustique de cigales. 1. Généralités et espéces méditerranéennes (Homoptera, Cicadoidea). *EPHE, Biologie et Insectes* **7**, 1–72.

Boulard, M. (2006) Acoustic signals, diversity and behaviour of cicadas (Cicadidae, Hemiptera). pp. 331–349 *in* Drosopoulos, S. & Claridge, M.F. (*Eds*) *Insect Sounds and Communication*. New York, NY, CRC Press, Taylor & Francis.

Claridge, M.F. & de Vrijer, P. (1994) Reproductive behaviour: the role of acoustic signals in species recognition and speciation. pp. 216–233 *in* Denno, R.F. & Perfect, T.J. (*Eds*) *Planthoppers their Ecology and Management*. New York, NY, Chapman & Hall.

Cooley, J.R. & Marshall, D.C. (2001) Sexual signalling in periodical cicadas, *Magicicada* spp. (Hemiptera: Cicadidae). *Behaviour* **138**, 827–855.

Dytham, C. (2003) *Choosing and Using Statistics, a Biologist's Guide*. 2nd edn. 248 pp. Oxford, UK, Blackwell, Science.

Fonseca, P.J. (1991) Characteristics of the acoustic signals in nine species of cicadas (Homoptera, Cicadidae). *Bioacoustics* **3**, 173–192.

Fonseca, P.J. & Revez, M.A. (2002a) Temperature dependence of cicada songs (Homoptera, Cicadoidea). *Journal of Comparative Physiology A* **187**, 971–976.

Fonseca, P.J. & Revez, M.A. (2002b) Song discrimination by male cicadas *Cicada barbara lusitanica* (Homoptera, Cicadidae). *Journal of Experimental Biology* **205**, 1285–1292.

Gerhardt, H.C. (1991) Female choice in treefrogs: static and dynamic acoustic criteria. *Animal Behaviour* **42**, 615–635.

Gleason, J.M. & Ritchie, M.G. (1998) Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* complex: do sexual signals diverge the most quickly? *Evolution* **52**, 1493–1500.

Gogala, M. & Trilar, T. (2004) Bioacoustic investigations and taxonomic considerations on the *Cicadetta montana* species complex (Homoptera: Cicadoidea: Tibicinidae). *Annals of the Brazilian Academy of Sciences* **76**, 316–324.

Greenfield, M.D., Tourtellot, M.K. & Snedden, W.A. (1997) Precedence effects and the evolution of chorusing. *Proceedings of the Royal Society B: Biological Sciences* **264**, 1355–1361.

Lee, M.S.Y. (2004) The molecularisation of taxonomy. *Invertebrate Systematics* **18**, 1–6.

Maynard-Smith, M.J. (1989) *Evolutionary Genetics*. Oxford, UK, Oxford University Press.

Minitab, Inc. (2004) MINITAB Release 14, Statistical Software.

Nast, J. (1972) *Palaearctic Auchenorrhyncha* (*Homoptera*): *an annotated check list*. 550 pp. Warsaw, Poland, Polish Academy of Sciences, Institute of Zoology.

Paterson, H.E.H. (1985) The recognition concept of species. pp. 21–29 *in* Vrba, E.S. (*Ed.*) *Species and Speciation*. Pretoria, South Africa, Transvaal Museum Monograph no. 4.

Phillimore, A.B. & Owens, I.P.F. (2006) Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences* **273**, 1049–1053.

Pinto-Juma, G., Simões, P.C., Seabra, S.G. & Quartau, J.A. (2005) Calling song structure and geographical variation in *Cicada orni* Linnaeus (Hemiptera: Cicadidae). *Zoological Studies* **44**, 81–94.

Popov, A.V. (1975) The structure of the tymbals and the characteristics of the sound signals in singing cicadas (Homoptera, Cicadidae) in the Southern regions of the USSR. *Entomological Review* **54**, 7–35.

Pringle, J.W.S. (1954) A physiological analysis of cicada song. *Journal of Experimental Biology* **31**, 525–556.

Quartau, J.A. (1988) A numerical analysis of interspecific morphological differences in two closely related species of *Cicada* (Homoptera, Cicadoidea) in Portugal. *Great Basin Naturalist Memoirs* **12**, 171–181.

Quartau, J.A. & Boulard, M. (1995) *Tettigetta mariae* n. sp., nouvelle Cigale lusitanienne (Homoptera, Cicadoidea, Tibicinidae). *EPHE, Biologie et Évolution des Insectes* **7**, 105–110.

Quartau, J.A. & Rebelo, M.T. (1994) Sinais acústicos em Cicadidae e Cicadellidae (Homoptera, Auchenorryncha) que ocorrem em Portugal. pp. 137–142 *in* Almada, V. & Correia, R. (*Eds*) *Biologia e comportamento. Actas do I Congresso Nacional de Etologia*, Publicado por Instituto Superior de Psicologia Aplicada, 28–30 Novembro 1991, Lisboa.

Shaw, K.L. & Herlihy, D.P. (2000) Acoustic preference functions and song variability in the Hawaiian cricket *Laupala cerasina*. *Proceedings of the Royal Society B: Biological Sciences* **267**, 577–584.

Specht, R. (2002) Avisoft-SASLab Pro-sound analysis and synthesis laboratory. Version 4.1d, Berlin, Germany.

Sueur, J. & Aubin, T. (2002) Acoustic communication in the Palaearctic red cicada *Tibicina haematodes*: chorus organization, calling song structure, and signal recognition. *Canadian Journal of Zoology* **80**, 126–136.

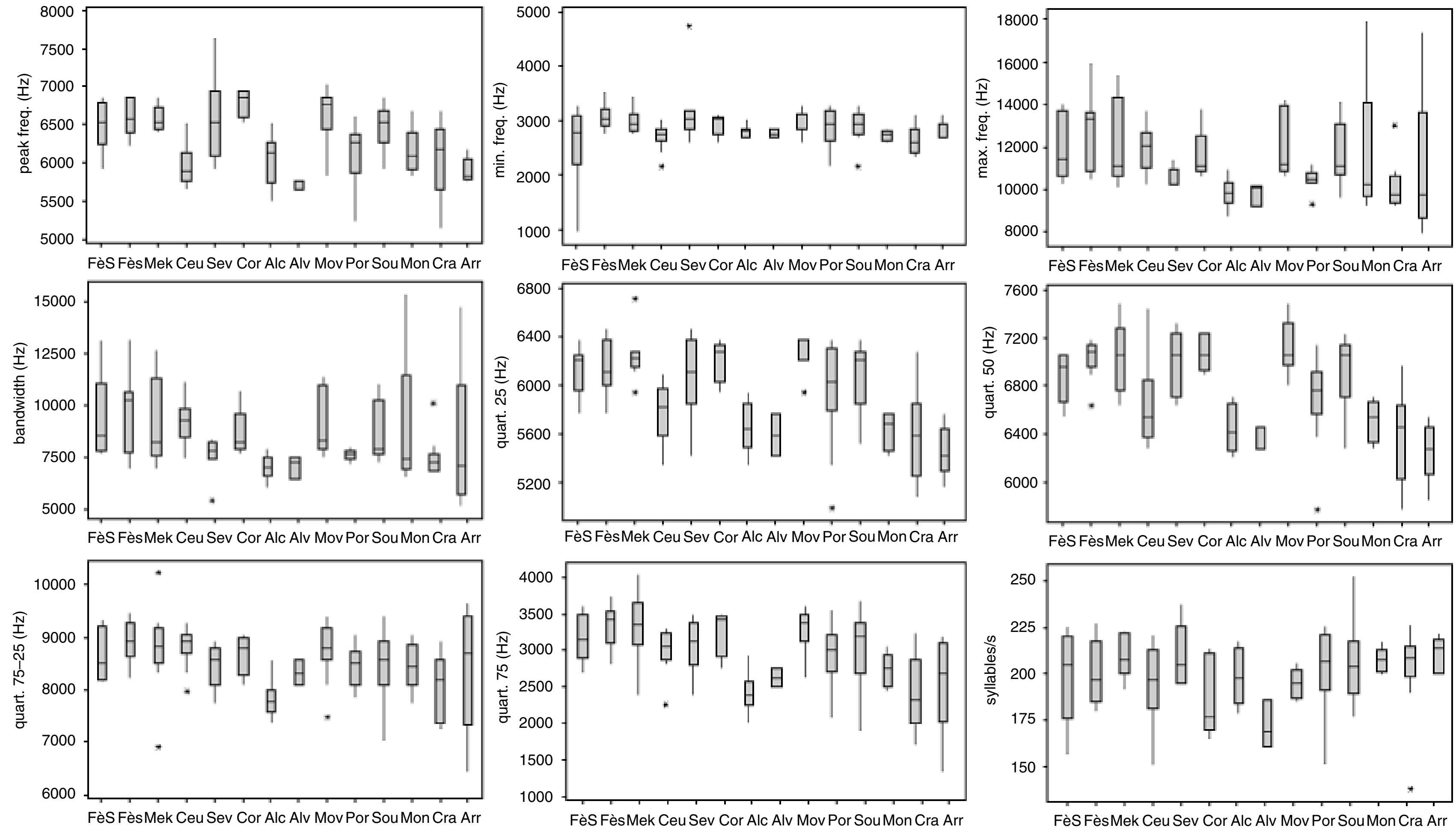
Sueur, J., Puissant, S., Simões, P.C., Seabra, S.G., Boulard, M. & Quartau, J.A. (2004) Cicadas from Portugal: revised list of species with eco-ethological data (Hemiptera: Cicadidae). *Insect Systematics & Evolution* **35**, 177–187.

Villet, M. (1992) Responses of free-living cicadas (Homoptera: Cicadidae) to broadcasts of cicada songs. *Journal of the Entomological Society of Southern Africa* **55**, 93–97.

Young, D. & Bennet-Clark, H.C. (1995) The role of the tymbal in cicada sound production. *Journal of Experimental Biology* **198**, 1001–1019.

Zuk, M., Rotenberry, J.T. & Simmons, L.W. (2001) Geographical variation in calling song of the field cricket *Teleogryllus oceanicus*: the importance of spatial scale. *Journal of Evolutionary Biology* **14**, 731–741.

Appendix I. Boxplots of the acoustic variables analysed for *Cicada barbara* at the sampled sites.



Boxplots represent the sample distribution: the rectangular box corresponds to approximately the middle 50% of the data, the top of the box is the third quartile (75%), the bottom is the first quartile (25%) and the horizontal line is the median; vertical lines extending to either side indicate the general extent of the data; *, outlier; FèS, Fès south; Mek, Meknès; Ceu, Ceuta; Sev, Sevilla; Cor, Cordoba; Alc, Alcalar; Alv, Alvor; Mou, Moura; Por, Portel; Sou, Sousel; Mon, Monforte; Cra, Crato; Arr, Arrábida.

Appendix II. Values of average ± standard deviation and coefficient of variation for each acoustic variable per population.

		Peak frequency (Hz)	Minimum frequency (Hz)	Maximum frequency (Hz)	Bandwidth (Hz)	Quartile 25% (Hz)	Quartile 50% (Hz)	Quartile 75% (Hz)	Quart75%–Quart25% (Hz)	Syllables s^{-1}
Morocco	FesS	6515 ± 343 5.3	2558 ± 830 32.4	11894 ± 1553 13.1	9334 ± 2072 22.2	6127 ± 207 3.4	6878 ± 210 3.0	9022 ± 482 5.3	2896 ± 381 13.2	199 ± 26 12.8
	Fes	6612 ± 249 3.8	3062 ± 231 7.5	12688 ± 1757 13.8	9619 ± 1880 19.5	6142 ± 217 3.5	7023 ± 168 2.4	9247 ± 429 4.6	3105 ± 296 9.5	201 ± 17 8.6
	Mek	6606 ± 163 2.5	2962 ± 221 7.5	12339 ± 2105 17.1	9371 ± 2142 22.9	6243 ± 205 3.3	7014 ± 303 4.3	9232 ± 674 7.3	2990 ± 634 21.2	209 ± 11 5.4
Iberian Península + Ceuta	Ceu	5991 ± 272 4.5	2700 ± 240 8.9	11914 ± 1084 9.1	9211 ± 1047 11.4	5765 ± 237 4.1	6607 ± 340 5.1	8787 ± 406 4.6	3022 ± 262 8.7	193 ± 21 10.6
	Sev	6590 ± 593 9.0	3194 ± 705 22.1	10710 ± 443 4.1	7513 ± 984 13.1	6061 ± 359 5.9	6997 ± 262 3.7	8817 ± 534 6.1	2756 ± 305 11.1	210 ± 17 8.0
	Cord	6816 ± 187 2.7	2924 ± 203 6.9	11556 ± 1266 11.0	8626 ± 1188 13.8	6196 ± 171 2.8	7076 ± 153 2.2	9108 ± 442 4.9	2912 ± 281 9.6	188 ± 22 11.8
	Alc	6051 ± 319 5.3	2788 ± 108 3.9	9838 ± 689 7.0	7047 ± 593 8.4	5645 ± 195 3.5	6438 ± 201 3.1	7970 ± 399 5.0	2325 ± 274 11.8	198 ± 14 7.3
	Alv	5740 ± 52 0.9	2753 ± 85 3.1	9813 ± 524 5.3	7060 ± 536 7.6	5593 ± 175 3.1	6400 ± 104 1.6	8263 ± 175 2.1	2670 ± 170 6.4	172 ± 13 7.4
	Mou	6676 ± 342 5.1	2981 ± 236 7.9	12266 ± 1580 12.9	9282 ± 1637 17.6	6237 ± 123 2.0	7090 ± 223 3.1	9186 ± 404 4.4	2949 ± 387 13.1	195 ± 9 4.4
	Port	6146 ± 398 6.5	2855 ± 357 12.5	10469 ± 517 4.9	7610 ± 275 3.6	5946 ± 447 7.5	6671 ± 380 5.7	8703 ± 610 7.0	2757 ± 274 10.0	203 ± 22 11.0
	Sou	6518 ± 283 4.3	2886 ± 304 10.5	11639 ± 1477 12.7	8747 ± 1374 15.7	6079 ± 267 4.4	6909 ± 308 4.5	8851 ± 694 7.8	2772 ± 462 16.7	206 ± 22 10.4
	Mon	6238 ± 349 5.6	2695 ± 104 3.9	11333 ± 3248 28.7	8638 ± 3303 38.2	5610 ± 150 2.7	6527 ± 168 2.6	8438 ± 308 3.6	2828 ± 338 12.0	207 ± 7 3.3
	Cra	6081 ± 461 7.6	2666 ± 255 9.5	10067 ± 994 9.9	7397 ± 854 11.6	5636 ± 390 6.9	6364 ± 365 5.7	8023 ± 753 9.4	2386 ± 404 16.9	204 ± 22 10.7
	Arr	5922 ± 167 2.8	2772 ± 187 6.7	10846 ± 3737 34.5	8072 ± 3801 47.1	5456 ± 219 4.0	6264 ± 255 4.1	8196 ± 1014 12.4	2740 ± 893 32.6	210 ± 10 4.6

FèS, Fès south; Mek, Meknès; Ceu, Ceuta; Sev, Sevilla; Cord, Cordoba; Alc, Alcalar; Alv, Alvor; Mou, Moura; Port, Portel; Sou, Sousel; Mon, Monforte; Cra, Crato; Arr, Arrábida.

Chapter 5.

Unprecedented number of copies and insertion events of mitochondrial cytochrome *b Numts* in *Cicada* L. (Hemiptera, Cicadoidea)

Submitted

Chapter 5.

Unprecedented number of copies and insertion events of mitochondrial cytochrome *b Numts* in *Cicada* L. (Hemiptera, Cicadoidea)

G. A. PINTO-JUMA[1,2]*, J. A. QUARTAU[1] & M.W.BRUFORD[2]

[1]Centro de Biologia Ambiental/Departamento de Biologia Animal, Faculdade de Ciências de Lisboa, C2, Campo Grande, 1749-016 Lisboa, Portugal: [2]School of Biosciences, Cardiff University, Cardiff CF10 3TL, Wales, UK

Abstract

Background

Mitochondrial (mt) DNA copies in the nuclear genome (*Numt*s) are known to occur in most eukaryotic organisms. However the number of copies and integration events can vary widely even between closely related species and their occurrence seems to be governed by species-specific mechanisms that control the accumulation and loss of nuclear DNA. *Numts* have a wide relevance for the understanding of genome evolution but due to their similarity with mtDNA, *Numts* are often accidentally co-amplified during PCR in mtDNA phylogenetic studies which can lead to incorrect conclusions. Here we describe and analyse the occurrence of a large number of nuclear copies in cicadas which were originally detected by PCR co-amplification of mtDNA in a related analysis.

Results

Cloned PCR products of a fragment of putative mt cytochrome (cyt) *b* sequence from five individuals of *Cicada barbara* and two *C. orni* (20 clones per specimen) revealed the presence of 87 and 26 haplotypes, respectively. Within these, several divergent mtDNA-like sequences were detected. The identification of authentic mtDNA in these species was only possible when purified mtDNA extracted from each species was sequenced. Between 45-56% of the *Numts* in these cicadas are originated from mtDNA translocated to the nuclear genome over a recent evolutionary timescale.

Conclusions

The use of universal primers for PCR of cyt *b* mtDNA in *C. barbara* exclusively amplifies *Numts* in most specimens. The frequency of *Numts* in this species appears to be the highest reported so far for a small mtDNA fragment. The integration of mtDNA fragments into the cicada nuclear genome is frequent and ongoing as suggested by the presence of *Numts* with different levels of divergence from authentic mtDNA. Further study of those species with a high number of *Numts* can potentially improve our knowledge of the processes involved in the assimilation of mitochondrial DNA into the nuclear genome.

*Corresponding author

Background

Knowledge of nuclear copies of mitochondrial genes, or as commonly known, *Numt*s [1], has expanded considerably since late 1990s. Numerous studies have reported the occurrence of *Numt*s and they are thought to be present in most eukaryotic organisms [2, 3, 4]. The incorporation of *Numts* in the yeast genome has been shown to be related to the repair of double strand breakage (DBS) by means of non-homologous end-joining (NHEJ) [5, 6]. This phenomenon seems to occur in all eukaryotes and may be implicated in the ongoing transfer of *Numts* observed in some organisms [4]. MtDNA can be released from mitochondria when disruption of the organelle membranes occurs during gametogenesis [7], or during autophagy, organelle fusion or division or cell stress [8]. Once integrated into the nuclear genome *Numts* lose their function and evolve as noncoding sequences at a slower evolutionary rate than in the mitochondrion [9], for this reason in some studies *Numts* have been used as molecular 'fossil' sequences, for example in humans where the nuclear mutation rate is considerably lower than the mitochondrial rate [2, 10]. Using this approach, Mishmar *et al.* [11] confirmed an African origin for the human ancestors by comparison with a *Numt* sequence and Hay *et al.* [12] were able to use a *Sphenodon Numt* as an 'outgroup' sequence for a group with no close evolutionary relatives.

The genome sequences of an increasing number of taxa allow a new understanding of *Numts* in genomes, their diversity and implications and the mechanisms underlying their evolution. In humans, several hundred *Numt* copies have been recognised [3, 7, 13, 14, 15] and their study has already enabled an estimate of the rate of mtDNA integration in the nucleus [15, 16], clarified the causes of some genetic diseases in humans [e.g. 17] and suggested mechanisms underlying their evolution [e.g. 18]. Although some examples of insects demonstrating an absence of, or just a few *Numts* such as *Anopheles* and *Drosophila* (Diptera) [3, 4] have been described, almost all eukaryotes studied possess *Numts* [2, 3, 4]. The genome sequence of some species has also raised the number of records of *Numts*, even in species previously reported as having none or just a few, such as the honeybee and some fish [19, 20, 21]. The number of copies found so far varies considerably from just a few as in *Drosophila* and chicken, to nearly 2000 in rice [3, 4, 22] and in the bee genome [19, 20]. The number of *Numts* in an organism depends on species-specific mechanisms that control the accumulation and loss of nuclear DNA [4, 23, 24], but the size of the nuclear and mitochondrial genomes does not correlate with their abundance [3, 4, 22].

Due to their similarity to mtDNA, *Numts* are often accidentally co-amplified during PCR in mtDNA studies, especially when universal primers are used [2, 4]. This is particularly frequent in phylogenetic studies. Guidelines have been suggested to avoid the contamination of mtDNA sequences with nuclear insertions [2, 4, 25] but it is likely that many studies contain data including undetected *Numt* contamination. Harrison *et al*. [26] predicted that up to 22% of human deposited sequences could be *Numts* and Thalmann *et al*. [27] highlighted the great similarity between some *Numts* and mtDNA in humans and apes. We report here a case of co-amplification of *Numts* in two closely related species from the genus *Cicada* and a range of analyses to analyse and differentiate *Numts* from authentic mtDNA.

Results

Detection of Numts

Attempts to sequence a section of the cyt *b* gene in the two sibling species *Cicada orni* and *C. barbara* resulted in unambiguous sequences for *C. orni* but not always for *C. barbara*. None of the sequences obtained had insertions, deletions or stop codons typical of nuclear copies of mtDNA, but some of the sequences obtained for *C. barbara* revealed ambiguities. Furthermore, one specimen (reference Cb 234, from Crato) exhibited very high divergence from other *C. barbara* specimens and repeated sequencing revealed ambiguities and differences from the first attempts. Additional specimens revealed similar discrepancies, implying that more than one fragment was being sequenced implying either multiple divergent mitochondrial copies or *Numts*. Further attempts to optimise PCR were carried out but for most *C. barbara* specimens, PCR products gave two bands of very similar size. DNA extracted directly from one of the bands was clearly identified as a *Numt* (Cb 821) since it included a 3bp insertion, whilst the second band always produced ambiguous sequences.

Assessment of Numts

PCR products from five *C. barbara* specimens and two *C. orni* were cloned (Table 1). This allowed the isolation and sequencing of fragments from individual PCR products. Sequences differing by one base pair were conservatively considered identical. From 111 *C. barbara* clones, 87 different sequences were produced and 26 *C. orni* sequences were detected in 40 clones. These sequences were very diverse and identifying *Numts* was complex as a result. All sequences presenting deletions, insertions or stop codons were categorised as

Numts and these varied from 258bp to the 486bp (Appendix 1). Sequences lacking these characteristics were analysed with respect to the cyt *b* gene structure and amino-acid models.

Table 1 - *Cicada barbara* and *C. orni* specimens used in this study

sp	Specimens Reference	Locality	Collection date	Observations
C. barbara	Cb 50	Arrábida (Portugal)	27-07-1995	Cloned
	Cb 106	Alcalar (Portugal)	23-08-1995	cloned
	Cb 234	Crato (Portugal)	24-07-1996	cloned
	Cb 737	Foz Côa (Portugal)	11-07-1999	Model sequence for specific mtDNA primers
	Cb 739	Foz Côa (Portugal)	11-07-1999	cloned
	Cb 775	Ceuta (Spain)	20-07-1999	cloned
	Cb 821	Ceuta (Spain)	21-07-1999	Model sequence for specific numt primers
	Cb 1816-Cb 1828	Moura (Portugal)	22-08-2001	mtDNA extraction
	Cb 1832	Moura (Portugal)	28-08-2001	mtDNA extraction
	Cb 1840-Cb 1850	Moura (Portugal)	28-08-2001	mtDNA extraction
	Cb 1551-Cb 1552	Sousel (Portugal)	28-07-2001	mtDNA enriching
C. orni	Co 509	Alter do Chão (Portugal)	05-07-1998	cloned
	Co1085	Lisbon (Portugal)	20-09-2000	cloned
	Co 1181-Co 1186	Estremoz (Portugal)	26-06-2001	mtDNA extraction
	Co 1546-Co 1550	Sousel (Portugal)	28-07-2001	mtDNA extraction
	Co 1180	Estremoz (Portugal)	26-06-2001	mtDNA enriching

We used criteria to analyse amino-acid substitutions including the physicochemical distances between amino acid replacements [28], amino-acid exchangeability using Argyle's method [29] and using conservative/non-conservative sites in the cyt *b* gene (comparing with known cyt *b* sequences from *Drosophila*, *Apis* and mammals). The amino-acid substitutions among putative mtDNA sequences are shown in Table 2 and the details of this assessment are in Appendix 2. This analysis recognized a further 27 *Numts* within *C. barbara* and 12 within *C. orni*. However, five *C. barbara* and eight *C. orni* clones remained classified as putative mtDNA. Most of these sequences considered of potentially mitochondrial origin translated similar and structurally appropriate amino-acids (Table 2, Appendix 2).Therefore, simple analysis of amino-acid substitutions could not efficiently categorise all sequences.

Table 2 - Amino-acid substitutions within mtDNA-like clones (no deletions, insertions or stop codons) from a PCR fragment of cyt *b* gene (477 bp translated 159 aa) for *Cicada barbara* and *C. orni* specimens. Only variable positions are shown

Reference	Freq	Sequence
		1111111111111111111111111111111111
		1222222333333444444455555666666777778888888889999990000001111122222333333334444444555
		12598123579346789123578914589345679046890123456890256780136780135624789012567890135678123
Reference	Freq	YFWSNSAIYGNWIGGFADNALRFYLIMVITIIHLFLGNPLGLMSNIDIPFPSIKDLGLMMMMLIMEDDIPANPLPNHIQPEYLFAYLRS
~~234-7~~	1	..........N...............................E....V.........V..MFV....T...L................
~~234-9~~	1	.......F..S................................L..................MF......L.D..............H.
~~234-10~~	1	..........S...............................TTS..V.....S........MF.......................H.
~~234-15~~	1	.....L....S..................K................................MF.TG..T....................
~~234-17~~	1	....Y..L.SS....L.Y...L.MF..D.......................
~~234-18~~	1	..........N...........................T.....D...............TMLV..........R.........T....
234-24	1	..........N...............................E....V..............MFV....T....................
~~234-25~~	1	.......V..N...........................A.....D.................MF.............Y............
~~234-30~~	1	..C.Y....SS................A..................................MFV.........................
~~106-3/14~~	2	DV........N.....................R........S..G.E..I...........TMLV..........R.......V.T....
~~106-16~~	1	..........N...........L.........................S......S.....MLV....................E....
106-18	1	..........S..L.MF..D.......................
~~106-19~~	1	..........N.....................R...........G....I...........TMLV...G......R.......V.T....
~~106-23~~	1	..........N................I..............S..T.V.....S.......VMF..........................
50-1	1	.......V..N...........................A.....D.......T.........MF..........................
~~50-2~~	1	......E...N.................V.............D....V..............MF..........................
50-3	1	..........S...........L.................S................T.L.MF.....T.V............L.....
~~50-8~~	1	..........S..................K................................MF......................H...
~~50-11/12~~	2	..........S..D.S...............V.F............................MF.....M....................
~~50-18~~	1	....K.....NC..AS...G........V........K..S...D.....L.......V..MF...............L.......I..
~~50-19~~	1	....K.....NC..A....G........V........K..S...D.....L.......V..MF............K..........I..
739-1	15	..........N.................V.............E....V..............MFV....T....................
~~739-5~~	1	........C.N.............F....M..............D....L............MF..G...S.......T.S.........
739-9	1	..........N.................V.............E....V..............MFV....T....................
~~739-15~~	1	........C.N............N.....M..............D.................MF......S...................
739-18	1	..........N.L...............V.............E....V..............MFV....T....................
739-19	1	..........N.................V.............E...GV..............MFV....T....................
739-20	1	..........N........V........V.............E....V..............MFV....T....................
~~775-2~~	1	....K.....NC.......G........V.....L..K..S...D.....L.......V..MF.........................L.

Sequence	n	Alignment
~~775-3~~	1	..........N...Q......Q....................TL...V.....S......MF......T.....H..............
~~775-9~~	1	.......V..N........................S..A.....D.......T.Y.....MF..........L................
~~775-11/18~~	2	.......V..N...........................A.....D..V....T.......MF..........L................
~~775-12~~	1	..........S...MF..............Y............
~~775-14/19~~	2	..........S....................................V...F........MFV.....T....................
~~775-15~~	1	..........N..........Q....................TL...V.....S......MF......T.....H..............
~~775-17~~	1	...L......N.................V.......V.....E....V............MF...........................
~~1085-1~~	1	..........S.......................L......M..................LL..................G........
1085-2	1	..........S...ML...........................
~~1085-4~~	1	..........S...LL...................H.......
1085-6	6	..........S...LL...........................
~~1085-7~~	1	..........S..M......LL...........................
1085-8	1	..........S...LL...........................
~~1085-9~~	1	..........S.............P...............................R...LL...........................
~~1085-10~~	1	..........S.........................E.......................LF..........................P
1085-12	1	..........S.....T...S.......................................LL...........................
~~1085-13~~	1	..........S...LL...........................
1085-14	1	..........N.............W...................................MF..................D........
~~1085-16~~	1	..........S.T.....D...LL...........................
1085-17	1	..........S........T...............M........................LL...........................
~~509-1~~	1	..........S...LL.V.....................H...
509-2	7	..........S...LL...........................
509-3	1	..........S...LL..D........................
~~509-5~~	1	..........S..................S.........W.......T............LL...........M...............
~~509-6~~	1	..........S....S..........T................................VLL...........................
~~509-7/8~~	2	..........N...................T........F............T.......MF....E..................C...
~~509-10~~	1	..........S...LL.................L.........
509-13	1	..........S..............F..................................LL...........................
509-17	1	..........S...ML...........................
		YFWSNSAIYGNWIGGFADNALRFYLIMVITIIHLFLGNPLGLMSNIDIPFPSIKDLGLMMMMLIMEDDIPANPLPNHIQPEYLFAYLRS

References strikethrough are sequences identified as *Numts* after amino-acids substitution analysis.

Highly frequent substitutions: 29 S/N 110 M/L 111 L/F

Cells in red represent conservative positions in the cytochrome *b* gene (of mammals [48], honeybee [47] and Drosophila [46])

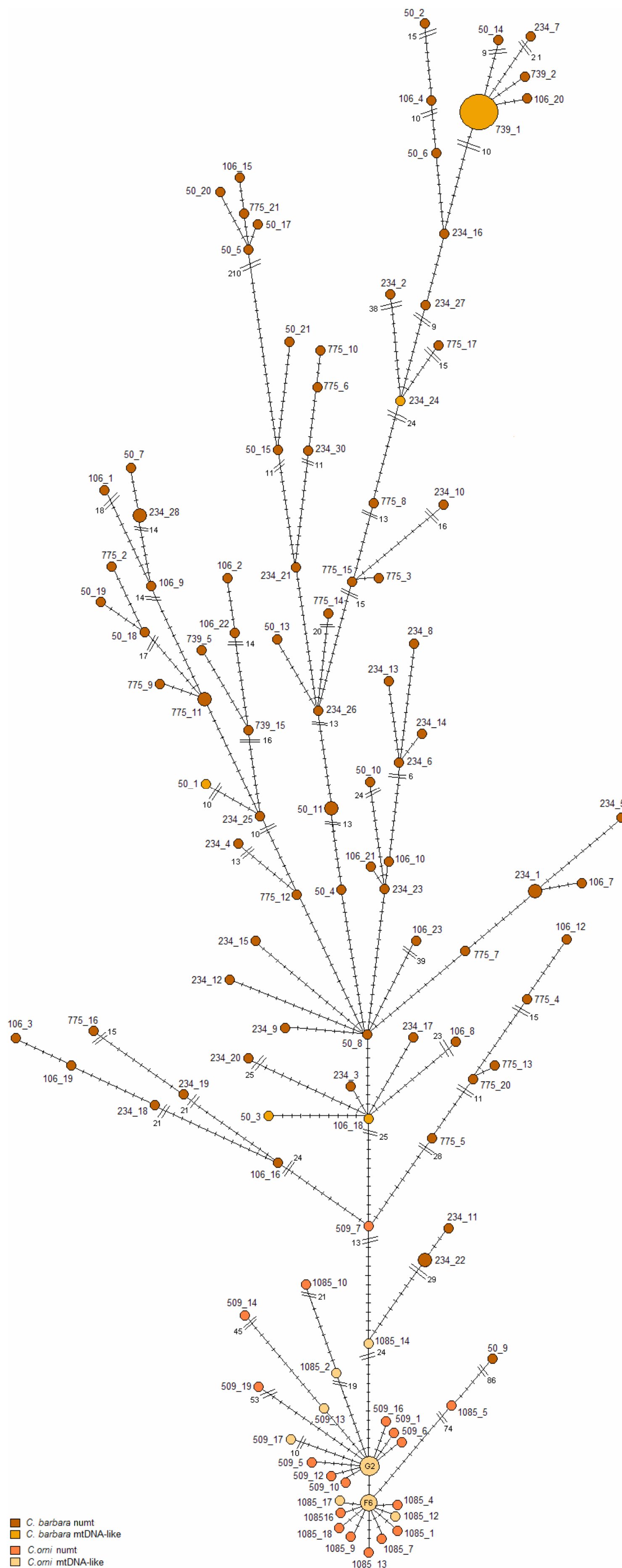


Figure 1 - Minimum spanning network between the clones from a PCR fragment of cyt *b* gene of five specimens of *Cicada barbara* and two of *C. orni*.

Haplotype names = specimen number and number of the clone. Circle area for each haplotype reflects the number of clones showing that haplotype, the colours are representative of the species and the classification of the haplotypes as *Numt* or as mtDNA-like. Connection lengths > 10 were reduced and the number of steps is shown at that position.

A minimum spanning network (Fig. 1) revealed the complexity of the connections between all sequences obtained through cloning. This divergence is evident even between clones from the same species. Also, the putative mtDNA haplotypes, highlighted in Fig. 1, were highly divergent (*p* distances of 2.1% - 11.3% for *C. barbara*). One candidate mtDNA haplotype was found in multiple clones (739-1) which came, as expected, from a single specimen. All other specimens lacked a candidate mtDNA sequence found in more than one clone. Within *C. orni*, *p* distances between mtDNA-like ranged from 0.4% - 7.9%. Two mtDNA candidate sequences were present in multiple clones. Between the two species, putative mtDNA sequences diverged between 7.9% and 13.4%. However, the presence of many divergent candidate mtDNA sequences prevented conclusive identification of mtDNA based on sequence analysis alone.

Authentic mtDNA

MtDNA from 11 (*C. orni*) and 25 (*C. barbara*) individuals was extracted using a standard CsCl gradient method (adapted from [43] and [44]). Surprisingly, we were not able to amplify the cyt *b* segment from *C. barbara* mtDNA via PCR using universal primers and only specific primers designed based on a solitary unambiguous presumed-mtDNA PCR generated sequence from Cb737 from Foz Côa resulted in amplification of the cyt *b* gene in the *C. barbara* mtDNA isolate, resulting in a identical sequence to the model specimen. For *C. orni* the amplification of the cyt *b* fragment with universal primers was straightforward, and mtDNA sequence similar (*p* = 0.44%) to clones ref. 509-2/1085-6 was obtained. We cloned the cyt *b* PCRs from purified mtDNA and only one mtDNA sequence was found in each case (from five clones per species).

Similar Numts vs. divergent mtDNA

Table 3 summarizes the results obtained for all clones analysed using authentic mtDNA data and Appendix 3 details variability statistics. Within *C. barbara* only one of the haplotypes (739-1) may truly be considered of mtDNA origin. The other mtDNA-like sequences were highly divergent, much higher than the overall mtDNA-cyt *b* intraspecific divergence found within *C. barbara* from the Iberian Peninsula and Morocco (less than 1%) ([30]; Table 4), so they could not be mistaken for mtDNA. The closest haplotype to 739-1 mtDNA (234-24) was 4.4% divergent from *C. barbara* mtDNA, exceeding interspecific sequence divergence levels between two species within the same genus, e.g. *C. orni* vs. *C. mordoganensis* (3.8%, Table 4). Thus, the clones in this species revealed 87 haplotypes, from

which 86 are likely *Numts* and only one is considered to be of mtDNA origin. The proportion of *Numts* in different specimens varied considerably, from 15% in the Foz Côa specimen to 100% in other localities (Crato, Arrábida and Ceuta). The proportion of *Numts* also varied between species: 84% of the clones obtained in *C. barbara* were likely *Numts* in *C. orni* this value was 57.5%.

Table 3 - Summary of Numts and mtDNA clones of a cyt *b* PCR fragment from five specimens *Cicada barbara* and two specimens *C. orni*

Cb = *Cicada barbara* and Co = *C. orni*, H = number of haplotypes. Between parentheses is the number of sequences if different from the number of haplotypes. Sequences differing in a single base pair were considered the same haplotype.

Specimen Ref[a]	Nº of clones	H	specimens with shared Haplotypes	*Numts*[1]	*Numts* (aa criteria)	mtDNA-like[2]	**truly mtDNA[3]**	*Numts*
Cb 234	30	29	1 Numt 234/106 1 Numt 234/50	20 (21)	8	1	**-**	100%
Cb 106	20	19	1 Numt 106/234 1 mtDNA 106/739	13	4 (5)	2	**1**	95%
Cb 50	21	20	1 Numt 50/234	13	5(6)	2	**-**	100%
Cb 739	20	4	1 mtDNA 739/106	1	2	1 (17)	**1 (17)**	15%
Cb 775	20	18	-	10	8 (10)	-	**-**	100%
Total Cb	**111**	**87**		**55 (58)**	**27 (31)**	**5 (22)**	**1 (18)**	**84%**
Co 1085	20	15	2 mtDNA 1085/509	2	7	6 (11)	**4(9)**	55%
Co 509	20	13	same	4	5 (6)	4 (10)	**2 (8)**	60%
Total Co	**40**	**26**		**6**	**12 (13)**	**8 (21)**	**4 (17)**	**57.5%**

1 Straight forward *Numts* (with frameshift and stop codons); 2 Sequences without any of the usual *Numts* characteristics and with credible amino acids substitutions; 3 Based on mtDNA extraction

In *C. orni* three of the eight mtDNA-like haplotypes were divergent from authentic mtDNA, differing in more than 16 substitutions and with genetic distances above 3.8% (Table 5). Considering that the mtDNA-cyt *b* intraspecific divergence within this species across Europe reaches up to 2% (Table 5), these clones are most likely to be *Numts*. However, the remaining mtDNA-like clones differ by 0.4-1.9% from authentic mtDNA. We cannot fully exclude the possibility of PCR artefacts for the clones more similar to authentic mtDNA but these clones and certainly the less similar ones might be a case of a very recent nuclear translocation or of different mtDNA copies present in one single individual (heteroplasmy).

Table 4 - Mean, minimum and maximum *p* distances between all groups of haplotypes obtained from cyt *b* PCR clones from five *Cicada barbara* specimens and two *C. orni*, and mtDNA sequences from *C. barbara*, *C. orni*, *C. mordoganensis*, *C. cretensis* and *Tettigetta josei*

	Cb *Numts*	Cb mtDNA	Co *Numts*	Co mtDNA	*C. mordoganen*	*C. cretensis*	*C. orni*	*C. barbara*
Cb *Numts*	0.100 [0.003-0.185]							
Cb mtDNA	0.118 [0.003-0.176]	-						
Co *Numts*	0.125 [0.070-0.188]	0.151 [0.120-0.166]	0.05 [0.006-0.137]					
Co mtDNA	0.124 [0.072-0.163]	0.149 [0.147-0.154]	0.032 [0.003-0.127]	0.010 [0.003-0.020]				
C. mordoganensis	0.117 [0.077-0.154]	0.145 [0.143-0.146]	0.051 [0.028-0.127]	0.038 [0.033-0.045]	0.004 [0.003-0.006]			
C. cretensis	0.120 [0.056-0.157]	0.149 0.146-0.152]	0.07 [0.047-0.133]	0.059 [0.055-0.063]	0.061 [0.058-0.063]	0.006 [0.003-0.008]		
C. orni	0.122 [0.077-0.157]	0.147 [0.146-0.149]	0.03 [0.003-0.125]	0.008 [[0.003-0.014]	0.037 [0.033-0.039]	0.059 [0.055-0.061]	0.006 [0.003-0.008]	
C. barbara	0.116 [0.011-0.177]	0.011 [0.008-0.014]	0.15 [0.125-0.167]	0.149 [0.141-0.157]	0.143 [0.139-0.147]	0.144 [0.138-0.150]	0.147 [0.141-0.152]	0.005 [0.003-0.008]
T. josei	0.203 [0.174-0.246]	0.224 -	0.187 [0.175-0.206]	0.183 [0.181-0.188]	0.184 [0.182-0.185]	0.196 [0.193-0.202]	0.184 [0.182-0.185]	0.226 [0.224-0.227]

Numts evolution and integration events

The Neighbour Joining tree (Fig. 2), reveals several clades comprising *Numts*, however most of them lack significant statistical support. We included mtDNA sequences of other species from the genus *Cicada* in this analysis and highlighted the position of these in the tree, as well as mtDNA-like sequences. *Numts* are present in all branches including that containing authentic mtDNA sequences. The level of divergence between *Numts* and authentic mtDNA is remarkably wide; *C. barbara Numts* diverge from mtDNA by as little as 0.3% and as much as 17.6% (Table 5). *Numts* from *C. orni* possess slightly lower levels of divergence (up to 12.7% from mtDNA), but the number of clones sampled was smaller.

Such a wide range of divergence between *Numts* raises questions about their origin. *Numts* are known to originate from mitochondrial insertions in the nuclear genome but also from secondary duplications [15, 19, 31, 32]. To distinguish these, we estimated the minimum number of mtDNA nuclear insertions folllowing Bensasson *et al* [15, 31]. The incomplete knowledge of the dimension of *Numts* in cicadas and the limited size of the fragment analysed here prevented the use of other methods such as in Hazkani-Covo and Graur [32]. In order to explain our data, within the 100 *Numts* analysed, a minimum of 54 independent mitochondrial integrations in the nuclear genome are inferred. Within *C. barbara*, a minimum of 45 separate integration events are inferred. In *C. orni* 45% of cyt *b Numts* are inferred to have originated from independent integration events (a minimum of nine mtDNA transfers producing 20 *Numts*). We did not find any evidence for a mtDNA transfer event responsible for *Numts* from both species simultaneously.

Table 5 - Pairwise *P* distances between mtDNA-like clone sequences from a PCR fragment of cyt *b* gene

	234-24	106-18	50-1	50-3	739-1	1085-2	1085-6	1085-12	1085-14	1085-17	509-2	509-13
234-24	-											
106-18	0.094	-										
50-1	0.088	0.071	-									
50-3	0.102	0.021	0.079	-								
739-1	0.044	0.104	0.094	0.113	-							
1085-2	0.115	0.086	0.084	0.094	0.115	-						
1085-6	0.129	0.096	0.094	0.104	0.127	0.054	-					
1085-12	0.129	0.100	0.098	0.109	0.127	0.058	0.004	-				
1085-14	0.117	0.084	0.079	0.092	0.115	0.075	0.073	0.077	-			
1085-17	0.134	0.100	0.098	0.109	0.132	0.058	0.004	0.008	0.077	-		
509-2	0.127	0.094	0.092	0.102	0.125	0.054	0.004	0.008	0.071	0.008	-	
509-13	0.127	0.096	0.102	0.104	0.129	0.069	0.019	0.019	0.079	0.023	0.015	-
509-17	0.117	0.088	0.096	0.098	0.127	0.058	0.038	0.042	0.079	0.042	0.033	0.035

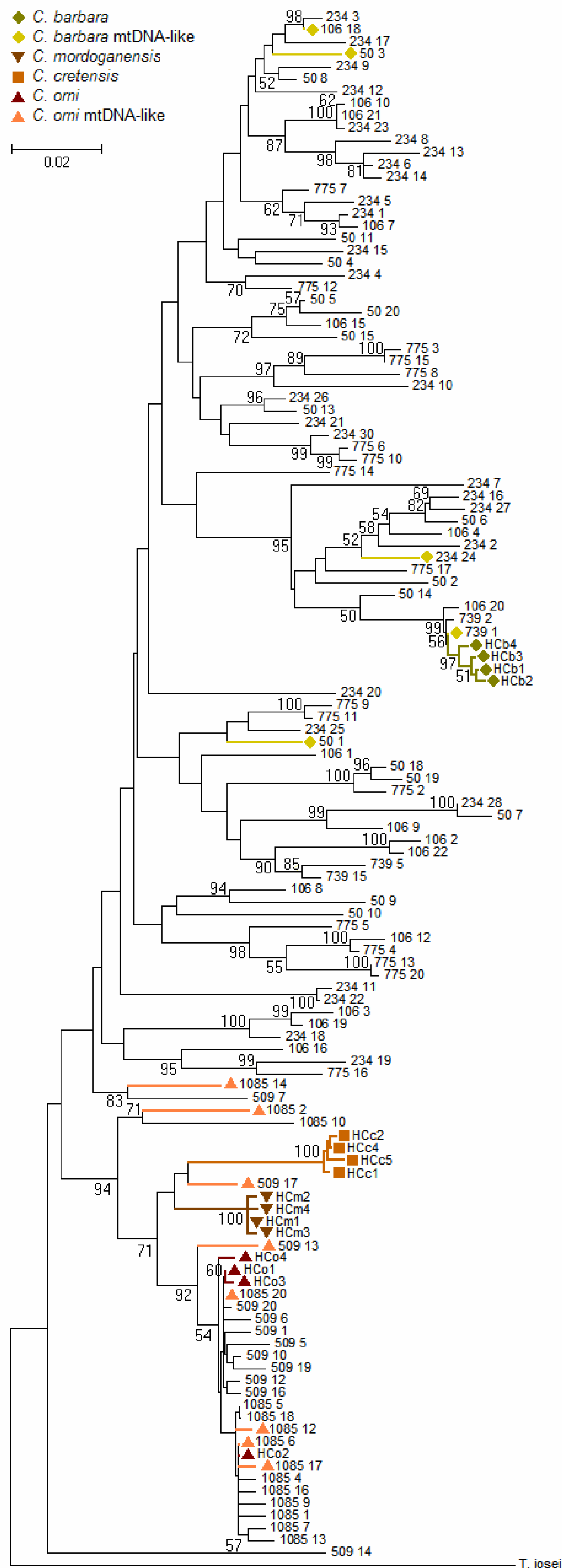


Figure 2 - Neighbour-joining tree of clones from a PCR fragment of cyt *b* for five *Cicada barbara* and two *C. orni* and authentic mtDNA for *C. barbara*, *C. orni*, *C. mordoganensis* and *C. cretensis*.

Clone names = specimen number and number of the clone; H = haplotype, Cb, Co, Cm and Cc = *C. barbara*, *C. orni*, *C. mordoganensis* and *C. cretensis* and haplotype reference *Tettigetta josei* (Tibicinidae) was used as outgroup. 1000 Bootstrap replicates were performed (only bootstrap values > 50% are shown).

Discussion

Numt frequency

The number of *Numts* reported here is to our knowledge unprecedented in the literature for a small DNA fragment. However, since the mitochondrial genome of cicadas has not been completely sequenced we cannot assume that the pattern observed in the cyt *b* fragment will hold for the remainder of the genome. However, to our knowledge, the number of nuclear copies from a single mtDNA fragment found in *C. barbara* is the highest ever reported. The density of nuclear copies recently reported for the whole genome of the honeybee [19] originating from separate sites of mitochondrial genes (using 50 bp intervals), never exceeded 54, whilst we found 86 nuclear copies derived from a single fragment of cyt *b* gene in *C. barbara* and the *Numts* detected are likely a small proportion of the total. We rarely found multiple *Numt* sequences and their number increased linearly with the number of clones analysed (Fig. 3).

However, these results are not completely unexpected. Bensasson *et al.* [31] and Sunnucks and Hales [23] reported multiple copies in other hemimetabolous insects, grasshoppers and aphids respectively, and both studies predicted a higher number of undetected *Numts*. Pons and Vogler [33] found evidence of high DNA loss within tiger beetles and suggested that *Numts* might be rare in holometabolous insects contrasting to hemimetabolous species. New findings for the honeybee and the flour beetle seem to contradict the rarity of *Numts* in holometabolous insects [19, 20] but our data reinforce the assumption of their abundance in hemimetabolous species.

The proportion of *Numt* mtDNA cloned between the two species analysed here was rather different: 84% in *C. barbara* and 57.5% in *C. orni* and if we exclude clones from the specimen of *C. orni* from Foz Côa, this difference would be even greater. This does not necessarily indicate a smaller number of *Numts* in *C. orni*. The disparity may be due to the inability to amplify mtDNA using universal primers in most *C. barbara* specimens, so only *Numts* were cloned as opposed to authentic sequences. However, we cannot exclude the possibility that the frequency of mtDNA integrations in the nucleus and the rate of nuclear DNA loss differs considerably between these species as is the case in other taxa [3, 23, 24]. Also, population/species demographic events such as bottlenecks and range expansions can potentially increase the rate of fixation of *Numts* [21, 34], so species with different

demographic histories may show different rates. Only the sequencing the genome of these cicadas can clarify these questions.

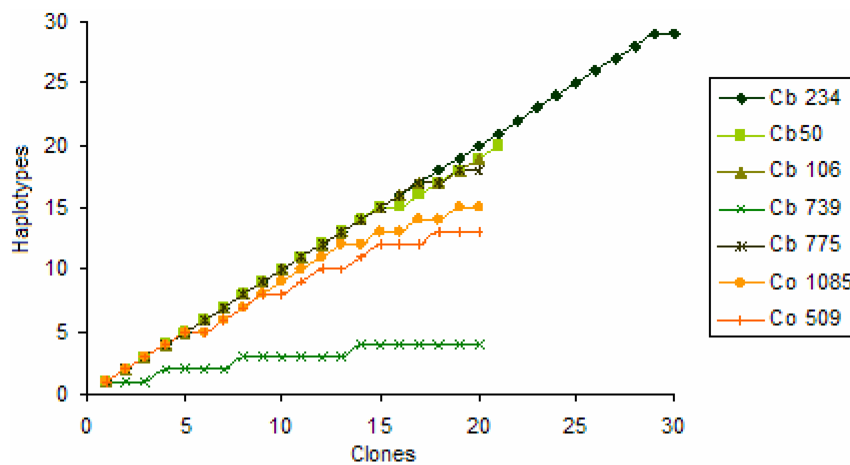


Figure 3 - Frequency of haplotypes per number of clones.

An ongoing process

In cicadas, the integration of mtDNA fragments into the nuclear genome seems to be a frequent and ongoing process as observed in humans and other organisms [6, 7, 13, 15, 16, 34]. We found *Numts* with many levels of divergence from authentic mtDNA suggesting that in cicadas, the integration of mtDNA fragments into the nuclear genome is a continuous process. The rate of mtDNA transfer to the nuclear genome found here of 45-56% is within the same range proposed for humans (32-85%) [15, 4, 16] but is the highest reported for insects. So, in these cicadas approximately half of the *Numts* found originated from new integrations of mtDNA into the nucleus, whilst the other half are probably due to duplications of previous *Numts*.

The presence of some *Numts* with very small genetic divergence from authentic mtDNA implies that very recent transfers have occurred. Very recent *Numts* in the human genome have also been reported, these have been found more frequently in nuclear genes associated with inherited disease in humans [4, 17, 34]. The high rate of accumulation of mtDNA in the nuclear genome found in humans is about the same order as the rate of point substitutions estimated for noncoding DNA, so it should be possible to distinguish individuals based on the presence/absence of *Numts* [4, 15, 34].

Numt contamination in phylogenetic studies

The detection in our study of several sequences from the same specimen not possessing any obvious *Numt* signature highlights some of the inherent problems when

confusing *Numt*s with mtDNA. Whatever the divergence of *Numt*s in relation to the corresponding mtDNA, these can certainly contaminate phylogenetic studies if ignored. In our preliminary analysis using the DNA sequences obtained via PCR with universal primers, *C. orni* and *C. barbara* divergence was estimated in about 7% (unpublished) but when considering authentic mtDNA this divergence increases to approximately 15%.

Many mtDNA phylogenetic studies have reported the inadvertent co-amplification of *Numts* [2, 4, 13] and recently diverse instances in crayfish [35], humans and gorillas [27, 36], in acanthocephalans [37] and in muskoxen [38] were reported. Martin [25] commented that confusing *Numts* with real mtDNA may incorrectly suggest horizontal gene transfer, and Parr *et al*. [39] raised the possibility of *Numt* contamination of standard heteroplasmy disease-associated analysis. Most studies suggest the standardization of preventive methods to avoid *Numt* amplification. DNA extraction methods enriching for mtDNA or the amplification by long-range PCR or reverse transcriptase PCR are methods which can reduced the probability of *Numt* co-amplification with mtDNA [2, 4, 25]. These methods should be particularly employed when new taxa lacking mtDNA information are analysed.

Conclusions

Our results provide additional evidence of the extremely high number of nuclear copies in hemimetabolous insects expanding other findings in grasshoppers and aphids. The number of *Numts* found here is highly likely to comprise a small proportion of the true number in these species, and is to our knowledge the highest reported so far for a small fragment of mtDNA. The integration of mtDNA into the nuclear genome of cicadas appears to have occurred very frequently and is likely an ongoing process judging by the extremely recent integrations found in this study. The study of organisms with such a high rate of *Numts* can potentially improve our knowledge of the processes involved in the assimilation of mitochondrial DNA in the nucleus.

We showed that universal cyt *b* primers almost exclusively amplify *Numts* in *C. barbara* and that it was effectively impossible to identify mtDNA based only on sequence analysis of clones. The implementation of methods already described by several authors in order to avoid *Numt* amplification should be a priority when initiating any phylogenetic mtDNA based study.

Methods

Specimens of *Cicada orni* and *C. barbara* used in this study (Table 1) were collected at localities in Portugal and in Ceuta (Spanish territory in North Africa) during the summers of 1995-2000 and were freshly frozen into liquid nitrogen and moved later to -80ºC (Table 1).

Standard DNA extraction, PCR and sequencing

DNA extraction of individual specimens is described elsewhere [30] and the amplification of approximately 453 bp of the cytochrome *b* gene (cyt *b*) was carried out via PCR using the universal primers CB-J-10933 and CB-N-11367 [40] in a 25µl reaction (1x *Taq* buffer, 1.25mM $MgCl_2$, 0.2mM of dNTPs, 50pmol of each primer, 1unit of *Taq* polymerase (Invitrogen) and approximately 50ng of genomic DNA) under cycling conditions of 3 min at 95ºC, followed by 30 cycles of 94ºC for 45 s, 56ºC for 45 s, and 72ºC for 1.15 min, with a final extension of 72ºC for 7 min. PCR optimization varied $MgCl_2$ concentration and the annealing temperature. PCR products were purified using the Microcon-100 microconcentrators (Amicon) following the manufacturer's protocols. Products showing double bands were purified by electrophoresis in an agarose gel: each band was extracted from the gel, immersed in water and used for sequencing.

DNAstar (Lasergene, USA) was used to edit, align through the CLUSTAL method and translate the sequence into amino-acid data. Due to the persistence of ambiguities in the DNA sequence data, several pairs of primers were design based on an assumed mtDNA sequence and a numt sequence using Primerselect from DNAstar (Lasergene, USA). The most effective pair of primers for the amplification of mtDNA were CB-J-11004: 5'CTTACTTAGGGCTC AAC3' and CB-N-11351: 5'TAAAAAGTATCATTCTGG3' (name adapted from [40]), and for the nuclear copies the following pair of primers were used NC-95: 5'ATTGAATTTGAGGTGGAT3' and NC-397: 5'GTGGATTATTAGCAGGGATG AAG3'.

Cloning

PCR products of five *C. barbara* specimens and two *C. orni* were cloned using a pGEM®-T easy vector system kit (Promega) following the manufacturer's instructions. Positive colonies were grown in LB media and the isolation of recombinant plasmid DNA was done using the QIAprep Miniprep kit. At least 20 positive colonies from each specimen were sequenced (forward and reverse) as described previously but using the universal M13 primers.

MtDNA extraction methods

A DNA extraction method enriching for mtDNA was used as in Zhang and Hewitt [41], adapted from Lansman *et al*. [42]. Additionally, a CsCl/ethidium bromide gradient mtDNA extraction method was used in cicadas preserved from fresh in liquid nitrogen and later transferred into a freezer at -80ºC. The methods used to extract mtDNA from *C. orni* and *C. barbara* specimens were slightly different due to different amounts of tissue available. For *C. barbara* approximately 10.5g of tissue (29 specimens) was ground into a fine powder in liquid nitrogen and homogenised on ice in 30ml of MSB-Ca^{2+} with 0.05x volume of Dis-EDTA. This mixture was centrifuged at 4ºC for 5min at 2,500g to pellet cell debris and supernatant was transferred to a new tube and spun again for four times under the same conditions. The resulting supernatant was centrifuged in a Beckman 12-21 ultracentrifuge (rotor JA-20) at 20,000g for 15min at 4ºC to pellet mitochondria. The supernatant was discarded and the pellet was ressuspended in 0.25M Sucrose-TE and spun at 10,000g for 15min at 4ºC. The resulting supernatant was discarded and the pellet was ressuspended in 4ml TNE and incubated at room temperature for 5min, 800µl 20% SDS was added and incubated 2min before adding 800µl TNE. Exactly 5.4g of CsCl was dissolved in 5.2ml of this mixture to precipitate lipids and proteins debris, incubated at 4ºC for 15min and centrifuged for 20min at 6,000rpm. The protein-SDS complex was removed with a spatula and the liquid was transferred into an ultracentrifugation tube containing 200µl of Etidium Bromide (10mg/ml) and the tube was filled to 1cm from the top with 1g/ml of CsCl in TNE prior to ultracentrifugation. The process was repeated using the DNA from 11 *C. orni* specimens (4g of cicada tissue), adapting from Moritz and Hillis, [43] and White *et al*. [44].

Each sample from *C. orni* and *C. barbara* was centrifuged in a Beckman L7 ultracentrifuge (rotor SCO40TI) at 20,000g for 40h. The purified mtDNA was removed with a syringe from the middle layer visible under UV light into new tubes. The mtDNA from these experiments was amplified, cloned (5 clones per mitochondrial purified DNA sample) and sequenced as described before.

Sequence analysis

All sequences were analysed in detail and translated according to the invertebrate mitochondrial DNA code (using EditSeq and MegAlign from Lasergene). Arlequin ver. 3.0 [45] was used to assess molecular diversity and to calculate a minimum spanning tree of all DNA sequences which was manually drawn based on the connection length between OTUs.

Due to the number and variety of DNA sequences found, we used the following criteria in order to identify putative mtDNA sequences for the cyt *b* gene:

- Neither insertions nor deletions over the entire studied fragment, comparing to known mtDNA sequences and the structure of this gene and absence of stop codons.
- Physicochemical distances [28] between amino-acids replacements usually lower than 65 (comparisons between the analysed sequences and other known cyt *b* sequences such as *Drosophila* [46], honeybee [47] and mammals [48]).
- High probability of amino-acid exchangeability according to Argyle's method [29] was also taken into account.
- Lower frequency of replacements on conservative positions.
- High frequency of amino-acid replacement among the analysed sequences increases the probability of replacements.

Phylogenetic analysis and numt integration events

A Neighbour Joining tree was computed using *MEGA* version 3.1 [49] based on nucleotide *p*-distance model with pairwise deletion option and a bootstrap resampling phylogeny test was assessed using 1000 replicates. DNA sequences for the same cyt *b* fragment from the species *C. orni* and *C. barbara* and also from the closely related *C. mordoganensis* and *C. cretensis* species were also included in this analysis. Tree was rooted using the cyt *b* sequence of *Tettigetta josei*.

We used the method described by Bensasson *et al* [15, 31] to determine the minimum number of separate integrations of mtDNA into the nuclear genome in our data. This approach is based in the assumption that *Numts* lack their original mitochondrial function [3, 31, 50] and as such, *Numts* originated from the same mtDNA integration should evolve evenly without a significant bias at their nucleotide sites contrasting with the mtDNA sequences which favour the accumulation of changes at third codon positions.

We excluded from our analysis *Numt* sequences containing large deletions (106-15, 50-5, 50-17, 50-20, 775-21, 50-9, 1085-5 and 509-19) because these would distort the general codon position bias (would not show any bias compared to most other sequences due to so many differences in all codon positions). All other 100 *Numts* were compared with each other. The number of pairwise nucleotide differences for each codon position was determined using *MEGA* version 3.1. Each pairwise comparison was tested for significant codon position bias

applying a $\chi 2$ test (df 2, *P* 0.05). *Numt* sequences with significant codon position bias in their pairwise comparison are likely to originate from separate mitochondrial transfers so from this data we can infer the minimum number of separate integration events.

Acknowledgements

We are very grateful to Stephen Funk for providing the facilities in the Zoological Institute of London to perform the mtDNA extraction (CsCl gradient method) and for his technical advice. We sincerely appreciate the helpful discussions with Douda Bensasson. We also thank Genage André, Sofia Seabra, Paula Simões, Teresa Rebelo, Mónica Ribeiro and Teresa Fernandes (Faculty of Science, University of Lisbon) for helping collect cicadas in the field. This research was supported by a PhD grant (PRAXIS BD/18229/98) from the Fundação para a Ciência e Tecnologia (Lisbon, Portugal).

References

1. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. **Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat.** *J Mol Evol* 1994, 39: 174-190
2. Bensasson D, Zhang D-X, Hartl DL, Hewitt GM. **Mitochondrial pseudogenes: evolution's misplaced witnesses.** *Trends Ecol Evol* 2001, 16: 314-321
3. Richly E, Leister D. **NUMTs in sequenced eukaryotic genomes.** *Mol Biol Evol* 2004, 21: 1081-1084
4. Leister D. **Origin, evolution and genetic effects of nuclear insertions of organelle DNA.** *Trends in Genetics* 2005, 21 (12): 655-663
5. Yu X, Gabriel A. **Patching broken chromosomes with extranuclear cellular DNA.** *Mol Cell* 1999, 4: 873-881
6. Ricchetti M, Fairhead C, Dujon B. **Mitochondrial DNA repairs double-strand breaks in yeast chromosomes.** *Nature* 1999, 402: 96-100
7. Woischnik M, Moraes CT. **Patterns of organization of human mitochondrial pseudogenes in the nuclear genome.** *Genome Res* 2002, 12: 885-893
8. Thorsness PE, Weber ER. **Escape and migration of nucleic acids between cloroplasts, mitochondria, and the nucleus.** *Int Rev Cytol* 1996, 165: 207-234

9. Perna NT, Kocher TD. **Mitochondrial DNA: molecular fossils in the nucleus.** *Curr Biol* 1996, 6: 128-129

10. Fukuda M, Wakasugi S, Tsuzuki T, Nomiyama H, Shimada K, Miyata T. **Mitochondrial DNA-like sequences in the human nuclear genome.** *J Mol Biol* 1985, 186: 257-266

11. Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC. **Mitochondrial DNA-like sequences in the nucleus (numts): insights into our African origins and the mechanism of foreign DNA integration.** *Hum Mutat* 2004, 23:125-133

12. Hay JM, Sarre SD, Daugherty CH. **Nuclear mitochondrial pseudogenes as molecular outgroups for phylogenetically isolated taxa: a case study in *Sphenodon*.** *Heredity* 2004, 93: 468-475

13. Mourier T, Hansen AJ, Willerslev E, Arctander P. **The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus.** *Mol Evol Biol* 2001,18:1833-1837

14. Tourmen Y, Baris O, Dessen P, Jacques C, Malthiery Y, Reynier P. **Structure and chromosomal distribution of human mitochondrial pseudogenes.** *Genomics* 2002, 80: 71-77

15. Bensasson D, Feldman MW, Petrov DA. **Rates of DNA duplication and mitochondrial DNA insertion in the human genome.** *J Mol Evol* 2003, 57: 343–354

16. Hazkani-Covo E, Sorek R, Graur D. **Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications.** *J Mol Evol* 2003, 56:169-174

17. Turner C, Killoran C, Thomas NST, Rosenberg M, Chuzhanova NA, Johnston J, Kemel Y, Cooper DN, Biesecker LG. **Human genetic disease caused by *de novo* mitochondrial nuclear transfer.** *Hum Genet* 2003, 112: 303-309

18. Keller I, Bensasson D, Nichols RA. **Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes.** *PLOS Genetics*, 2007, 3 (2):e22

19. Pamilo P, Viljakainen L, Vihavainen A. **Exceptionally high density of NUMTs in the honeybee genome.** *Mol Biol Evol* 2007, 24 (6): 1340-1346

20. Behura SK. **Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome.** *Mol Biol Evol* 2007, 24 (7): 1492-1505

21. Antunes A, Ramos MJ. **Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes.** *Genomics* 2005, 86: 708-717

22. Pereira SL, Baker AJ. **Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics.** *BMC Evol Biol* 2004, 4: 17

23. Sunnucks P, Hales DF. **Numerous transposed sequences of mitochondrial cytochrome oxidase I_II in aphids of the genus *Sitobion* (Hemiptera: Aphididae).** *Mol Biol Evol* 1996, 13: 510-524

24. Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM. **Genomic gigantism: DNA loss is slow in mountain grasshoppers.** *Mol Biol Evol* 2001, 18: 246-253

25. Martin W. **Lateral gene transfer and other possibilities.** *Heredity* 2005, 94: 565-566

26. Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M. **Molecular Fossils in the Human Genome: Identification and Analysis of the Pseudogenes in Chromosomes 21 and 22.** *Genome Res* 2002, 2: 272-280

27. Thalmann O, Hebler J, Poinar HN, Paabo S, Vigilant L. **Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes.** *Mol Ecol* 2004, 13: 321-335

28. Grantham R. **Amino-acid difference formula to help explain protein evolution**. *Science* 1974, 185 (4154): 862-864

29. Argyle E. **A similarity ring for amino-acids based on their evolutionary substitution rates.** *Orig Life Evo Biosph* 1980, 10 (4): 357-360

30. Pinto-Juma GA, Quartau JA, Bruford MW. **Population structure of *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco based on mitochondrial DNA analysis.** *B Entomol Res* 2008, 98: 15-25

31. Bensasson D, Zhang D-X, Hewitt GM. **Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes.** *Mol Biol Evol* 2000, 17: 406-415

32. Hazkani-Covo E, Graur D. **A comparative analysis of numt evolution in human and chimpanzee.** *Mol Biol Evol* 2007, 24 (1): 13-18

33. Pons J, Vogler AP. **Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation if tiger beetles.** *Mol Biol Evol* 2005, 22 (4): 991-1000

34. Ricchetti M, Tekaia F, Dujon B. **Continued colonization of the human genome by mitochondrial DNA.** *PLOS Biol* 2004, 2: E273

35. Nguyen TTT, Murphy NP, Austin CM. **Amplification of multiple copies of mitochondrial cytochrome b gene fragments in the Australian freshwater crayfish, *Cherax destructor* Clark (Parastacidae: Decapoda).** *Anim Genet* 2002, 33: 304-308

36. Anthony NM, Clifford SL, Bawe-Johnson M, Abernethy KA, Bruford MW, Wickings EJ. **Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: guidelines for their diagnosis in complex sequence databases.** *Mol Phylogenet Evol* 2007, 43: 553-566

37. Benesh DP, Hasu T, Suomalainen L-R, Valtonen ET, Tiirola M. **Reliability of mitochondrial DNA in an acanthocephalan: the problem of pseudogenes.** *Int J Parasit* 2006, 36: 247-254

38. Kolokotronis S-O, MacPhee RD, Greenwood AD. **Detection of mitochondrial insertions in the nucleus (numts) of Pleistocene and modern muskoxen.** *BMC Evol Biol* 2007, 7: 67

39. Parr RL, Maki J, Reguly B, Dakubo GD, Aguirre A, Wittock R, Robinson K, Jakupciak JP, Thayer RE. **The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation.** *BMC Genomics* 2006, 7: 185

40. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. **Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers.** *Ann Entomol Soc Am* 1994, 87 (6): 651-701

41. Zhang D-X; Hewitt GM. **Nuclear integrations: challenges for mitochondrial DNA markers.** *Trends Ecol Evol* 1996, 11 (6): 247-251

42. Lansman RA, Shade RO, Shapira JF, Avise JC. **The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications.** *J Mol Evol* 1981, 17: 214-226

43. Moritz C, Hillis DM. ***Molecular Systematics.*** Sinauer Associates Inc 1996

44. White PS, Tatum OL, Tegelström H, Densmore LD. **Mitochondrial DNA isolation, separation, and detection of fragments.** In: *Molecular Genetic analysis of populations.* Edited by AR Hoelzel. Oxford University Press 1998, 65-99

45. Excoffier L, Laval G, Schneider S. **Arlequin ver. 3.0: An integrated software package for population genetics data analysis.** *Evolutionary Bioinformatics Online* 2005, 1: 47-50

46. Clary DO, Wolstenholme DR. **The mitochondrial molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code.** *J Mol Evol* 1985, 22: 252-271

47. Crozier RH, Crozier YC. **The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization.** *Genetics* 1993, 135: 97-117

48. Irwin, DM, Kocher TD, Wilson AC. **Evolution of cytochrome *b* gene of mammals.** *J Mol Evol* 1991, 32: 128-144

49. Kumar S, Tamura K, Nei M: ***MEGA*3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, 5: 150-163

50. Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W. **The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants.** *FEBS Lett* 1993, 325 (1,2): 140-145

51. Nei M, Kumar S. **Evolutionary Change of Amino Acid Sequences.** *Molecular Evolution and Phylogenetics.* Oxford University Press US 2000, 17-32

Appendix 1 - Description of clones from a PCR fragment of cyt b gene of five specimens *Cicada barbara* and two specimens *C. orni*

Reference	Freq.	Total bp	Transl. bp	Nº aa	Classif.	Deletions/Position		Insertions/Position		Stop codons	Frame shift
Cb 234: 30 sequences / 29 haplotypes [haplotypes shared with other specimens: 234-1 (106-17), 234-II (50-16)]											
234-1/106-17	2	478	477	150	Numt	1bp/209bp				9	yes
234-2	1	456	456	151	Numt	32bp/0-32bp				0	No
234-3	1	478	477	145	Numt	1bp/80bp				14	yes
234-4	1	476	474	158	Numt	3bp/354-356bp				0	No
234-5	1	478	477	151	Numt	1bp/209bp				8	yes
234-6	1	478	477	153	Numt	1bp/337bp				6	yes
234-7	1	479	477	159	Numt					0	
234-8	1	478	477	153	Numt	1bp/337bp				6	yes
234-9	1	479	477	159	Numt					0	
234-10	1	479	477	159	Numt					0	
234-11	1	477	477	144	Numt	1bp/387bp	1bp/39bp			15	yes
234-12	1	478	477	145	Numt	1bp/111bp				14	yes
234-13	1	477	477	147	Numt	1bp/337bp	1bp/125bp			12	yes
234-14	1	478	477	154	Numt	1bp/337bp				5	yes
234-15	1	479	477	159	Numt					0	
234-16	1	478	477	144	Numt	1bp/50bp				15	yes
234-17	1	479	477	159	Numt					0	
234-18	1	479	477	159	Numt					0	
234-19	1	480	480	150	Numt	1bp/308bp		1bp/64bp	1bp/209	10	yes
234-20	1	478	477	149	Numt	1bp/210bp				10	yes
234-21	1	477	477	148	Numt	1bp/89bp	1bp/308bp			11	yes
234-22/50-16	2	478	477	157	Numt	1bp/387bp				2	yes
234-23	1	478	477	147	Numt	1bp/160bp				12	yes
234-24	1	479	477	159	Mt-like					0	
234-25	1	479	477	159	Numt					0	
234-26	1	478	477	145	Numt	1bp/89bp				14	yes
234-27	1	478	477	144	Numt	1bp/50bp				15	yes
234-28/234-29	2	479	477	157	Numt					2	No

234-30	1	479	477	159	Numt				0	
Cb 106: 20 sequences / 19 haplotypes [haplotypes shared with other specimens: 106-17 (234-1), 106-6 (739-1)]										
106-1	1	478	477	141	Numt	1bp/38bp			18	yes
106-2	1	478	477	154	Numt	1bp/333bp			5	yes
106-3/106-14	2	479	477	159	Numt				0	
106-4	1	478	477	149	Numt	1bp/192bp			10	yes
106-6 (1 ≠ 739-1)	1	479	477	159					0	
106-7	1	477	477	152	Numt	1bp/209bp	1bp/309bp		7	yes
106-8	1	485	483	161	Numt			6bp/137bp	0	No
106-9	1	479	477	157	Numt				2	No
106-10	1	478	477	147	Numt	1bp/160bp			12	yes
106-12	1	481	480	157	Numt	1bp/364bp		3bp/391bp	3	yes
106-15	1	258	258	84	Numt	221bp/214-434			2	yes
106-16	1	479	477	159	Numt				0	
106-17 (= 234-1)	1	478	477	150		1bp/209bp			9	yes
106-18	1	479	477	159	Mt-like				0	
106-19	1	479	477	159	Numt				0	
106-20	1	480	480	158	Numt			1bp/465bp	2	yes
106-21	1	477	477	146	Numt	1bp/160bp	1bp/460bp		13	yes
106-22	1	478	477	154	Numt	1bp/333bp			5	yes
106-23	1	479	477	159	Numt				0	
Cb 50: 21 sequences / 20 haplotypes [haplotypes shared with other specimens: 50-16 (234-II)]										
50-1	1	479	477	159	Mt-like				0	
50-2	1	479	477	159	Numt				0	
50-3	1	479	477	159	Mt-like				0	
50-4	1	477	477	156	Numt	2bp/356-357bp			3	yes
50-5	1	257	255	80	Numt	221bp/214-434	1bp/95bp		5	yes
50-6	1	477	477	150	Numt	1bp/50bp	1bp/162bp		9	yes
50-7	1	479	477	157	Numt				2	No
50-8	1	479	477	159	Numt				0	
50-9	1	344	342	113	Numt	135bp/0-135bp			1	No
50-10	1	479	477	150	Numt	1bp/337bp		1bp/160bp	9	yes

50-11/50-12	2	479	477	159	Numt					0	
50-13	1	480	480	153	Numt	1bp/218				7	yes
50-14	1	479	477	158	Numt					1	No
50-15	1	478	477	153	Numt	1bp/270bp				6	yes
50-16 (= 234-22)	1	478	477	157		1bp/387bp				2	yes
50-17	1	258	258	84	Numt	221bp/214-434				2	yes
50-18	1	479	477	159	Numt					0	
50-19	1	479	477	159	Numt					0	
50-20	1	479	255	85	Numt	221bp/214-434bp	1bp/460bp			0	yes
50-21	1	475	474	150	Numt	1bp/270bp	2bp/112-113bp	1bp/148bp		8	yes
Cb 739: 20 sequences / 4 haplotypes [haplotypes shared with other specimens: 739-1 (106-6)]											
739-1/739-7/739-10/739-12/739-13/739-14/739-16/739-3/739-4/739-6/739-8/739-9/739-11/739-17/73918/739-19/739--20/106-6	17	479	477	159	Mt-like					0	
739-2	1	476	474	149	Numt	1bp/209bp	1bp/12bp	1bp/299bp		9	yes
739-3 (1 ≠ 739-1)	1	479	477	159						0	
739-4 (1 ≠ 739-1)	1	479	477	159						0	
739-5	1	479	477	159	Numt					0	
739-6 (1 ≠ 739-1)	1	479	477	159						0	
739-8 (1 ≠ 739-1)	1	479	477	159						0	
739-9 (1 ≠ 739-1)	1	479	477	159						0	
739-11 (1 ≠ 739-1)	1	478	477	150		1bp/209bp				9	yes
739-15	1	479	477	159	Numt					0	
739-17 (1 ≠ 739-1)	1	480	480	151					1bp/9bp	9	yes
739-18 (1 ≠ 739-1)	1	479	477	159						0	
739-19 (1 ≠ 739-1)	1	479	477	159						0	
739-20 (1 ≠ 739-1)	1	479	477	159						0	
Cb 775: 20 sequences / 18 haplotypes [haplotypes shared with other specimens: none]											
775-2	1	479	477	159	Numt					0	
775-3	1	479	477	159	Numt					0	
775-4	1	481	480	159	Numt	1bp/471bp			3bp/391bp	1	yes

775-5	1	480	480	152	Numt			1bp/246bp	8	yes
775-6	1	479	477	157	Numt				2	No
775-7	1	478	477	149	Numt	1bp/209bp			10	yes
775-8	1	479	477	158	Numt				1	No
775-9	1	479	477	159	Numt				0	
775-10	1	479	477	156	Numt				3	No
775-11/775-18	2	479	477	159	Numt				0	
775-12	1	479	477	159	Numt				0	
775-13	1	479	477	156	Numt	1bp/336bp		1bp/246bp	3	yes
775-14/775-19	2	479	477	159	Numt				0	
775-15	1	479	477	159	Numt				0	
775-16	1	473	471	153	Numt	1bp/35bp	5bp/152-156bp		4	yes
775-17	1	479	477	159	Numt				0	
775-18(= 775-11)	1	479	477	159					0	
775-19 (= 775-14)	1	479	477	159					0	
775-20	1	480	480	153	Numt			1bp/246bp	7	yes
775-21	1	258	258	84	Numt	221bp/214-434			2	yes
Co 1085: 20 sequences / 15 haplotypes [haplotypes shared with other specimens: 1085-20 (509-2)]										
1085-1	1	479	477	159	Numt				0	
1085-2	1	479	477	159	Mt-like				0	
1085-3 (1 ≠ 1085-6)	1	479	477	159					0	
1085-4	1	479	477	159	Numt				0	
1085-5	1	395	393	131	Numt	84bp/0-84bp			0	No
1085-6/1085-11/1085-19/1085-3/1085-15/1085-8/509-15	6	479	477	159	Mt-like				0	
1085-7	1	479	477	159	Numt				0	
1085-8 (1 ≠ 1085-6)	1	479	477	159					0	
1085-9	1	479	477	159	Numt				0	
1085-10	1	479	477	159	Numt				0	
1085-12	1	479	477	159	Mt-like				0	
1085-13	1	479	477	159	Numt				0	
1085-14	1	479	477	159	Mt-like				0	

1085-15(1 ≠ 1085-6)	1	479	477	159						0	
1085-16	1	479	477	159	Numt					0	
1085-17	1	479	477	159	Mt-like					0	
1085-18	1	477	477	149	Numt	1bp/11bp	1bp/299bp			10	yes
1085-20 (= 509-2)	1	479	477	159						0	
Co 509: 20 sequences / 13 haplotypes [haplotypes shared with other specimens: 509-2 (1085-20), 509-15 (1085-6)]											
509-1	1	479	477	159	Numt					0	
509-2/509-18/509-9/1085-20/509-3/ 509-4/509-11/509-20	7	479	477	159	Mt-like					0	
509-3 (1 ≠ 509-2)	1	479	477	159						0	
509-4 (1 ≠ 509-2)	1	479	477	159						0	
509-5	1	479	477	159	Numt					0	
509-6	1	479	477	159	Numt					0	
509-7 509-8	2	479	477	159	Numt					0	
509-10	1	479	477	159	Numt					0	
509-11 (1 ≠ 509-2)	1	479	477	159						0	
509-12	1	478	477	146	Numt	1bp/62bp				13	yes
509-13	1	479	477	159	Mt-like					0	
509-14	1	486	486	156	Numt			3bp/230bp	4bp/234bp	6	yes
509-15 (1 ≠ 1085-6)	1	479	477	159						0	
509-16	1	478	477	158	Numt	1bp/378bp				1	yes
509-17	1	479	477	159	Mt-like					0	
509-19	1	418	417	139	Numt	61bp/419-479				0	No
509-20 (1 ≠ 509-2)	1	479	477	159						0	

Grey lines are haplotypes included in another main haplotype (shows only one nucleotide difference). Identification of *Numts* based on amino-acid substitutions is in grey.

Appendix 2 - Amino-acid substitutions within mtDNA-like clones from a section of cytochrome *b* gene (477 bp, 159 amino-acids translated) of *Cicada barbara* and *C. orni*

Reference	Freq	Amino-acid substitutions														
~~234-7~~	1	29N (46)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	106V(M) 4/17 (21)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89)	131L(P) 3/2 (98)						
~~234-9~~	1	23F(I) 8/7 (21)	29S (46)	83L(S) 1 (145)	110M (15)	111F (22)	128L(P) 3/2 (98)	130D(N) 36/42 (23)	152H(R) 10/8 (29)							
~~234-10~~	1	29S (46)	82T(M) 2/1 (81)	83T(S) 38/32 (58)	84S(N) 20/34 (46)	88V(I) 33/57 (29)	97S(K) 8/7 (121)	110M (15)	111F (22)	152H(R) 10/8 (29)						
~~234-15~~	1	21L(S) 1 (145)	29S (46)	63K(T) 11/8 (78)	110M (15)	111F (22)	115T(M) 2/6 (81)	116G(E) 4/7 (98)	127T(I) 7/11 (89)							
~~234-17~~	1	18Y(N) 4/3 (143)	23L(I) 9/22 (5)	27S(G) 21/16 (56)	29S (46)	38L(F) 13/6 (22)	107L(M) 45/8 (15)	110M (15)	111F (22)	116D(E) 53/56 (45)						
~~234-18~~	1	29N (46)	41Y(D) 0 (160)	78T(P) 4/5 (38)	84D(N) 36/42 (23)	108T(M) 2/6 (81)	110M (15)	111L (22)	113V(I) 33/57 (29)	135R(P) 4/5 (103)	147T(A) 32/22 (58)					
234-24	1	29N (46)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89)								
~~234-25~~	1	23V(I) 33/57 (29)	29N (46)	78A(P) 22/13 (27)	84D(N) 36/42 (23)	110M (15)	111F (22)	137Y(H) 4 (83)								
~~234-30~~	1	5C(W) 0 (215)	18Y(N) 4/3 (143)	27S(G) 21/16 (56)	29S (46)	58A(V) 18/13 (64)	110M (15)	111F (22)	113V(I) 33/57 (29)							
~~106-3/14~~	2	1D(Y) 0 (160)	2F(V) 1/0 (177+)	29N (46)	66R(H) 10/8 (29)	81S(L) 1 (145+)	84G(N) 6/12 (80+)	86E(D) 53/56 (45+)	90I(F) 7/8 (153)	108T(M) 2/6 (81+)	110M (15)	111L (22)	113V(I) 33/57 (29)	135R(P) 4/5 (103)	145V(L) 15/11 (32)	147T(A) 32/22 (58)
~~106-16~~	1	29N (46)	48L(F) 13/6 (22)	89S(P) 1 (74+)	100S(L) 1 (145+)	110M (15)	111L (22)	113V(I) 33/57 (29)	147E(A) 17/10 (107+)							
106-18	1	29S (46)	107L(M) 45/8 (15)	110M (15)	111F (22)	116D(E) 53/56 (45)										
~~106-19~~	1	29N (46)	66R(H) 10/8 (29)	84G(N) 6/12 (80)	90I(F) 7/8 (153)	108T(M) 2/6 (81)	110M (15)	111L (22)	113V(I) 33/57 (29)	124G(D) 6/11 (94)	135R(P) 4/5 (103)	145V(L) 15/11 (32)	147T(A) 32/22 58)			

~~106-23~~	1	29N (46)	58I(V) 33/57 (29)	82S(M) 1/4 (135)	85T(I) 3/11 (89)	88V(I) 33/57 (29)	97S(K) 8/7 (121)	108V(M) 4/17 (21)	110M (15)	111F (22)							
50-1	1	23V(I) 33/57 (29)	29N (46)	78A(P) 22/13 (27)	84D(N) 36/42 (23)	96T(I) 7/11 (89+)	110M (15)	111F (22)	127T(I) 7/11 (89+)								
~~50-2~~	1	22E(A) 17/10 (107+)	29N (46)	59V(I) 33/57 (29)	82D(M) 0 (160)	88V(I) 33/57 (29)	110M (15)	111F (22)									
50-3	1	29S (46)	48L(F) 13/6 (22)	81S(L) 1 (145+)	103T(L) 3/2 (92)	107L(M) 45/8 (15)	110M (15)	111F (22)	129V(A) 18/13 (64)	146L(F) 3/6 (22)							
~~50-8~~	1	29S (46)	63K(T) 11/8 (78)	110M (15)	111F (22)	148H(Y) 4 (83+)											
~~50-11/12~~	2	29S (46)	36D(G) 6/11 (94+)	38S(F) 2/3 (155)	65V(I) 33/57 (29)	67F(L) 6/3 (22)	110M (15)	111F (22)	127M(I) 12/5 (10)								
~~50-18~~	1	18K(N) 13/25 (94+)	29N (46)	33C(W) 0 (215)	37A(G) 21 (60)	38S(F) 2/3 (155)	43G(A) 21 (60)	59V(I) 33/57 (29)	76K(N) 13/25 (94)	80S(G) 21/16 (56)	84D(N) 36/42 (23)	92L(P) 2/3 (98+)	106V(M) 4/17 (21)	110M (15)	111F (22)	139L(Q) 3/6 (113)	151I(L) 9/22 (5)
~~50-19~~	1	18K(N) 13/25 (94+)	29N (46)	33C(W) 0 (215)	37A(G) 21 (60)	43G(A) 21 (60)	59V(I) 33/57 (29)	76K(N) 13/25 (94+)	80S(G) 21/16 (56)	84D(N) 36/42 (23)	92L(P) 2/3 (98+)	106V(M) 4/17 (21)	110M (15)	111F (22)	136K(N) 13/25 (94+)	151I(L) 9/22 (5)	
739-1	15	29N (46)	59V(I) 33/57 (29)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89+)								
~~739-5~~	1	25C(Y) 3 (194+)	29N (46)	51F(L) 13/6 (22)	63M(T) 2/6 (81+)	84D(N) 36/42 (23)	90L(F) 6/13 (22)	110M (15)	111F (22)	116G(E) 4/7 (98+)	128S(P) 12/17 (74+)	138T(I) 7/11 (89+)	140S(P) 12/17 (74+)				
739-9	1	29N (46)	59V(I) 33/57 (29)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89+)								
~~739-15~~	1	25C(Y) 3 (194+)	29N (46)	49N(Y) 4/3 (143)	63M(T) 2/6 (81+)	84D(N) 36/42 (23)	110M (15)	111F (22)	128S(P) 12/17 (74+)								
739-18	1	29N (46)	34L(I) 9/22 (5)	59V(I) 33/57 (29)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89+)							
739-19	1	29N (46)	59V(I) 33/57 (29)	82E(M) 1/0 (126)	86G(D) 6/11 (94+)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89+)							
739-20	1	29N (46)	43V(A) 18/13 (64)	59V(I) 33/57 (29)	82E(M) 1/0 (126)	88V(I) 33/57 (29)	110M (15)	111F (22)	113V(I) 33/57 (29)	127T(I) 7/11 (89+)							

~~775-2~~	1	18K(N) 13/25 (94+)	29N (46)	33C(W) 0 (215)	37A(G) 21 (60)	43G(A) 21 (60)	59V(I) 33/57 (29)	69L(F) 6/13 (22+)	76K(N) 13/25 (94+)	80S(G) 21/16 (56+)	84D(N) 36/42 (23)	92L(P) 2/3 (98+)	106V(M) 4/17 (21+)	110M (15)	111F (22)	152L(R) 1 (102)
~~775-3~~	1	29N (46)	47Q(R) 10/9 (43+)	82T(M) 2/6 (81+)	83L(S) 1 (145+)	88V(I) 33/57 (29+)	97S(K) 8/7 (121+)	110M (15)	111F (22)	127T(I) 7/11 (89+)	135H(P) 3/5 (77)					
~~775-9~~	1	23V(I) 33/57 (29+)	29N (46)	70S(L) 1 (145+)	78A(P) 22/13 (27)	84D(N) 36/42 (23)	96T(I) 7/11 (89+)	98Y(D) 0 (160)	110M (15)	111F (22)	131L(P) 3/2 (98+)					
~~775-11/18~~	2	23V(I) 33/57 (29+)	29N (46)	78A(P) 22/13 (27)	84D(N) 36/42 (23)	88V(I) 33/57 (29+)	96T(I) 7/11 (89+)	110M (15)	111F (22)	131L(P) 3/2 (98+)						
~~775-12~~	1	29S (46)	110M (15)	111F (22)	137Y(H) 4 (83+)											
~~775-14/19~~	2	29S (46)	88V(I) 33/57 (29+)	95F(S) 1/3 (155)	110M (15)	111F (22)	113V(I) 33/57 (29+)	127T(I) 7/11 (89+)								
~~775-15~~	1	29N (46)	47Q(R) 10/9 (43+)	82T(M) 2/6 (81+)	83L(S) 1 (145+)	88V(I) 33/57 (29+)	97S(K) 8/7 (121+)	110M (15)	111F (22)	127T(I) 7/11 (89+)	135H(P) 3/5 (77)					
~~775-17~~	1	9L(S) 1 (145+)	29N	59V(I) 33/57 (29+)	74V(G) 5/3 (109)	82E(M) 1/0 (126)	88V(I) 33/57 (29+)	110M (15)	111F (22)							
~~1085-1~~	1	29S (46)	69L(F) 6/13 (22+)	81M(L) 45/8 (15+)	110L (15)	111L (22)	141G(E) 4/7 (98+)									
1085-2	1	29S (46)	110M (15)	111L (22)												
~~1085-4~~	1	29S (46)	110L (15)	111L (22)	143H(Y) 4 (83+)											
1085-6	6	29S (46)	110L (15)	111L (22)												
~~1085-7~~	1	29S (46)	97M(K) 20/4 (95)	110L (15)	111L (22)											
1085-8	1	29S (46)	110L (15)	111L (22)												
~~1085-9~~	1	29S (46)	51P(L) 3/2 (98+)	101R(G) 0/1 (125)	110L (15)	111L (22)										
~~1085-10~~	1	29S (46)	74E(G) 4/7 (98+)	110L (15)	111F (22)	153P(S) 12/17 (74+)										

1085-12	1	29S (46)	39T(A) 32/22 (58+)	110L (15)	111L (22)				
~~1085-13~~	1	29S (46)	45S(L) 1 (145+)	110L (15)	111L (22)				
1085-14	1	29N (46)	51W(L) 4/0 (61)	110M (15)	111F (22)	141D(E) 53/56 (45+)			
~~1085-16~~	1	29S (46)	34T(I) 7/11 (89+)	42D(N) 36/42 (23)	110L (15)	111L (22)			
1085-17	1	29S (46)	43T(A) 32/22 (58+)	70M(L) 45/8 (15+)	110L (15)	111L (22)			
~~509-1~~	1	29S (46)	110L (15)	111L (22)	115V(M) 4/17 (21+)	148H(Y) 4 (+83)			
509-2	7	29S (46)	110L (15)	111L (22)					
509-3	1	29S (46)	110L (15)	111L (22)	116D(E) 53/56 (45+)				
~~509-5~~	1	29S (46)	63S(T) 38/32 (58)	79W(L) 0 (61)	88T(I) 7/11 (89+)	110L (15)	111L (22)	132M(L) 45/8 (15+)	
~~509-6~~	1	29S (46)	38S(F) 2/3 (155)	55T(M) 2/6 (81+)	108V(M) 4/17 (21+)	110L (15)	111L (22)		
~~509-7/8~~	2	29N (46)	64T(I) 7/11 (89+)	79F(L) 13/6 (22+)	96T(I) 7/11 (89+)	110M (15)	111F (22)	122E(D) 53/56 (45+)	148C(Y) 3 (194+)
~~509-10~~	1	29S (46)	110L (15)	111L (22)	140L(P) 2/3 (98+)				
509-13	1	29S (46)	54F(I) 7/8 (21)	110L (15)	111L (22)				
509-17	1	29S (46)	110M (15)	111L (22)					

Amino-acids are represented by a single letter code preceded by their position in the DNA fragment analysed here and followed by the expected amino-acid for that site. Numbers below correspond to the amino acid substitution probabilities for one PAM (accepted point mutations: represents one amino acid substitution per 100 amino acid sites) based on the amino acid substitution matrix for the evolutionary distance of 1 PAM (x 10,000) [51]. Within parenthesis is the physico/chemical distance between amino-acids substitutions. References strikethrough are sequences identified as *Numts* after amino-acids substitution analysis. Cells in red represent conservative positions in the cytochrome *b* gene

Appendix 3 - Summary statistics of clones from a cyt *b* gene PCR fragment per specimen cloned

	Cb 234			Cb 106			Cb 50			Cb 739			Cb 775			Co 1085			Co 509		
	Total	mtDNA	numt	Total	mtDNA	numt	Total	mtDNA	numt	Total	mtDNA	numt	Total	mtDNA	numt	Total	mtDNA	numt	Total	mtDNA	numt
Gene copies	30	-	»	20	1	19	21	-	»	20	17	3	20	-	»	20	11	9	20	10	10
Haplotypes	29	-	»	19	1	18	20	-	»	4	1	3	18	-	»	16	6	9	14	4	9
Usable nucleotide sites	464	-	»	472	462	472	464	-	»	462	462	462	466	-	»	462	462	462	469	462	469
Polymorphic sites	249	-	»	337	-	337	407	-	»	64	-	64	307	-	»	158	52	130	160	22	152
Gene diversity (+/-)	0.99 (0.01)	-	»	0.99 (0.02)	-	0.99 (0.02)	0.99 (0.02)	-	»	0.28 (0.13)	-	1.00 (0.28)	0.99 (0.02)	-	»	0.95 (0.04)	0.73 (0.14)	1.00 (0.05)	0.92 (0.06)	0.53 (0.18)	0.98 (0.05)
Transitions	189	-	»	160	-	160	155	-	»	53	-	53	144	-	»	67	42	38	73	18	66
Transversions	56	-	»	56	-	56	59	-	»	8	-	8	48	-	»	16	11	9	34	4	31
TS/TV	3.37	-	»	2.86	-	2.86	2.63	-	»	6.62	-	6.63	3.00	-	»	4.19	3.82	4.22	2.15	4.50	2.13
Substitutions	245	-	»	216	-	216	214	-	»	61	-	61	192	-	»	83	53	47	107	22	97
Indels	49	-	»	237	-	237	362	-	»	3	-	3	233	-	»	85	0	85	70	0	70
sites with TS	181	-	»	157	-	157	150	-	»	53	-	53	141	-	»	66	42	37	72	18	65
sites with TV	56	-	»	56	-	56	59	-	»	8	-	8	48	-	»	16	11	9	34	4	31
sites with subst.	212	-	»	194	-	194	196	-	»	61	-	61	172	-	»	80	52	46	99	22	91
sites with indels	49	-	»	237	-	237	362	-	»	3	-	3	233	-	»	85	0	85	70	0	70
Nucleotide composition % C	15.66	-	»	15.68	17.75	15.57	14.99	-	»	17.39	17.75	15.33	14.97	-	»	14.59	14.46	14.73	14.56	14.50	14.59
T	39.79	-	»	39.69	38.31	39.77	40.35	-	»	38.63	38.31	40.42	40.51	-	»	41.97	42.03	41.92	42.01	42.21	41.81
A	31.93	-	»	31.83	30.52	31.90	31.46	-	»	30.58	30.52	30.95	31.44	-	»	29.90	30.03	29.74	30.16	29.87	30.48
G	12.62	-	»	12.80	13.42	12.76	13.20	-	»	13.40	13.42	13.30	13.08	-	»	13.54	13.48	13.61	13.27	13.42	13.12
Mean number of pairwise dif.	45.81 (20.42)	-	»	68.20 (30.70)	-	68.77 (31.03)	106.88 (47.82)	-	»	10.96 (5.20)	-	42.67 (25.85)	63.95 (28.81)	-	»	19.76 (9.12)	11.64 (5.72)	29.78 (14.40)	23.41 (10.75)	4.87 (2.59)	40.51 (19.26)
Nucleotide diversity	0.10 (0.05)	-	»	0.14 (0.07)	-	0.15 (0.07)	0.23 (0.12)	-	»	0.02 (0.01)	-	0.09 (0.07)	0.14 (0.07)	-	»	0.04 (0.02)	0.03 (0.01)	0.06 (0.04)	0.05 (0.03)	0.01 (0.01)	0.09 (0.05)

Chapter 6.

General Discussion

Doubt is not a pleasant condition, but certainty is absurd.
Voltaire

Chapter 6.

GENERAL DISCUSSION

The current study represents the first step towards understand the evolutionary history of the genus *Cicada* in the Mediterranean region. Cicadas are well known for their intraspecific acoustic communication systems, but are rarely studied in a phylogenetic context, particularly Palaearctic species (Suer *et al*. 2007). Furthermore, studies integrating the application of Paterson's recognition concept (based on Specific Mate Recognition Systems - SMRS), such as the acoustic signals, and genetic differentiation in Mediterranean cicada species in order to better understand the evolutionary patterns of species are rare. Exceptions are the work developed in parallel by some colleagues (e.g. Ribeiro 1998; Seabra 2007) using isozyme and microsatellite molecular markers. Here, the phylogenetic relationship of *Cicada* species from the Mediterranean area was assessed and an in depth analysis of population differentiation in *Cicada orni* and *C. barbara* was based on acoustic and mtDNA markers.

MtDNA and NUMTS

The detection of nuclear copies of mtDNA (*Numts*) obscuring routine analysis of sequences of the cytochrome *b* gene (cyt *b*) was crucial for subsequent data analyses in the current study. *Numts* are often accidentally co-amplified during PCR in mtDNA phylogenetic studies and if undetected can lead to incorrect conclusions, for instance, *Numts* have even been mistaken for dinosaur mtDNA and ancient monkey sequences (Bensasson *et al*. 2001). In this study, the number of *Numts* reported is to our knowledge unprecedented in the literature for such a small DNA fragment analysed (further discussed in Chapter 5), and thus, cicadas might be a useful model for studying the processes involved in the assimilation of mitochondrial DNA into the nucleus.

In order to avoid co-amplification of *Numts,* specific cyt *b* primers were designed for *C. barbara*, and the cyt *b* and 12SrRNA sequences analysed were confirmed to be exclusively mitochondrial following amplification and sequencing of the same fragments directly from purified *C. orni* or *C. barbara* mtDNA. In addition, the secondary structure and well-conserved sequence motifs of the 12SrRNA section was checked to confirm agreement with previous descriptions for other insects (discussed in Chapter 2) and the cyt *b* sequence was analysed to verify the expected pattern of amino-acid substitutions.

The fragment of the mitochondrial 12SrRNA gene analysed in the current study was useful for species discrimination and to determine the phylogeny of the genus *Cicada* in the Mediterranean area. However, the population structure and demography studies of *C. orni* and *C. barbara* were not completely resolved. For *C. barbara*, the 12SrRNA gene did not reveal any population structure, whilst cyt *b* gene analysis allowed the separation of populations in two regions. This is not surprising, since 12SrRNA gene is more conserved than the cyt *b* gene. Nevertheless, the use of an mtDNA approach in this study was unquestionably appropriate, but in order to answer some of the questions that arose from this study the use of more variable nuclear markers would be necessary.

Acoustic and molecular signal evolution in Cicada species

This study confirmed that within the genus *Cicada,* species which have been discriminated mainly on the basis of acoustic characters constitute independent evolutionary lineages. Thus, Paterson's species concept (1985), stressing species unique specific mate recognition systems (SMRS), such as acoustic communication in cicadas, does constitute a practical operational concept for this group, as also found in frogs (Angulo & Reichle 2008).

Speciation in the genus *Cicada* has not occurred with small genetic divergences. Contrary to preliminary allozyme data (see Quartau *et al.* 2000*,* 2001; Seabra *et al.* 2000), the level of divergence found at the mtDNA level between *C. barbara* and the other *Cicada* species was high (10-13.7% in 12SrRNA – see Chapter 2; and 14.3-14.9% in cyt *b* – see Chapter 5). Also, the closest pair of species analysed in this study, *C. orni* and *C. mordoganensis,* was moderately divergent (2.3% in 12SrRNA – see Chapter 2; and 3.8% in cyt *b* – see Chapter 5), whilst allozyme analysis failed to distinguish this species pair (see Seabra *et al.* 2000). Furthermore, microsatellite data (Seabra 2007) corroborates these mtDNA findings. The most basal *Cicada* species, *C. barbara* and *C. lodosi*, as denoted by 12S rRNA, have calling songs characterised by a continuous signal which might be a plesiomorphic character and the discontinuous state present in most *Cicada* species may be a derived evolutionary innovation within the genus, as also suggested by Quartau & Simões (2006).

The presence of the three most closely related *Cicada* species in the Aegean area and the fact that *C. orni* is the most recent diverged species within this genus suggests that the Aegean area may be the region of origin of *C. orni*. The occurrence of various paleogeographic events resulting in the separation of islands from the mainland (from before

10MYA up to the early Pleistocene) would have favoured vicariant processes probably leading to isolation and divergence between populations and, consequently, the formation of new species (Kasapidis *et al*. 2005). Conversely, the Messinian salinity crisis (5.9-5.3MYA) and the Pleistocene ice ages affected the sea level, possibly revealing land-bridge connections between islands and mainland that would have supported dispersal (Kasapidis *et al*. 2005). In fact, the distribution of species on Aegean islands has been generally attributed to vicariance, but also to dispersal, both natural and human mediated (e.g. Douris *et al*. 1998; Kasapidis *et al*. 2005).

Considering the efficiency of acoustic and mtDNA markers for discriminating cicada species, these signals seem to have evolved in parallel at the species level. However, although the closest pairs of species present in the Aegean area, *C. orni*, *C. mordoganensis* and *C. cretensis*, can be generally distinguished acoustically, the intraspecific range of most acoustic variables overlaps between species (see Quartau & Simões 2005). Since these species have never been found in sympatry, genetically isolated populations may retain the ancestral advertisement calling song through stabilizing selection. Genetically differentiated units with poor calling song differentiation are common in allopatric systems (Heyer & Reid 2003), as found for instance in cicadas of the *Platypleura stridula* species complex in South Africa (Price *et al*. 2007). In this species complex, six clades were identified based on mtDNA analysis, each with specific host-plant associations and distinct geographical ranges whilst at the acoustic level only four clades could be identified.

Population differentiation in Cicada orni and C. barbara

Cicada orni and *C. barbara* showed considerable intraspecific acoustic variation, with populations from each species revealing differences for most acoustic variables (see Chapters 3 and 4). Such population diversification might be explained by the tendency of males to form choruses in order to increase the chance of attracting conspecific females and avoiding predators (Villet 1992; Cooley & Marshall 2001; Fonseca & Revez 2002a), inter-male competition (Greenfield *et al*. 1997; Sueur & Aubin 2002), or even environmental micro-adaptations in order to maximise sound propagation and communication (Bennet-Clark 1998). Acoustic responses to environmental factors are evident in cicadas. In this study, temporal acoustic variables, particularly syllable rate, correlated with environmental temperature. This was expected, because temporal variables are dependent on the contraction/expansion frequency of tymbal muscles which is influenced by body temperature (Sanborn 2006).

However, despite this intraspecific acoustic variation, the calling song profiles observed at the regional level were consistent and corroborated the main genetic structure observed for *C. orni* and *C. barbara*.

Acoustic and cytochrome *b* gene analysis within *C. barbara* specimens from the Iberian Peninsula and North Africa (Spanish Ceuta and Morocco) suggests a divergence between the Iberian Peninsula (plus Ceuta in the North Africa) and Morocco populations. There was no evidence of geographical structure within the Iberian Peninsula. However, the presence of unique 12SrRNA haplotypes in some localities, in addition to the common haplotype, suggested that colonization was not recent, or that populations have undergone considerable bottlenecks and lineage sorting since the colonization. Also, the complete absence of shared cyt *b* haplotypes with Morocco seems to indicate an ancient separation between these regions (see Chapters 2 and 4 for further discussion). These findings support the subspecific division of *C. barbara* of Boulard (1982), with *C. barbara lusitanica* present in the Iberian Peninsula and *C. barbara barbara* present in Morocco.

The Ceuta population was more closely related to the Iberian Peninsula populations than to the Morocco ones. However, microsatellite data (Seabra 2007) showed a higher differentiation from Iberian populations (*C. barbara lusitanica*) than from the Moroccan populations (*C. barbara barbara*). These contrasting results might indicate that mitochondrial and nuclear markers in these cicadas have different evolutionary rates, or might be the result of asymmetric introgression as a consequence of different behaviour (migration, reproduction) of the males and females from the different subspecies (e.g. mayfly, Hughes *et al*. 2003; honeybees, Kraus *et al*. 2007). In fact, differentiation in nuclear gene markers reflects both male and female dispersal while differentiation at mitochondrial DNA markers reflects only female dispersal. An alternative explanation is that mitochondrial DNA may be subjected to selective forces (William *et al*. 1995).

Although no population structure was found for *C. barbara* within the Iberian Peninsula, Seville specimens possessed the most common haplotype plus one unique 12SrRNA haplotype, and this locality showed also the single different haplotype found in the Iberian Peninsula for cyt *b*. Also, the population from Foz Côa might be differentiated in view of *Numts* data (see Chapter 5). Specimens from Foz Côa were the only ones to amplify mtDNA using universal primers suggesting that the flanking regions of the cyt *b* section amplified may be differentiated from other populations. This differentiation would not be surprising, considering that Foz Côa is the most known northern point of *C. barbara*'s

distribution in Portugal and is considerably isolated surrounded by mountains with apparently no other *C. barbara* populations nearby.

In *C. orni*, acoustic and genetic differentiation was consistent in separating continental Greek cicadas from Western Europe. Cicadas from the Aegean islands were also found genetically differentiated, but no acoustic analysis was performed for these populations in this study.

Contrary to *C. barbara*, which did not show differentiation based in 12SrRNA sequences alone, *C. orni* specimens revealed a significant genetic structure based on this gene. Since the 12SrRNA gene is more conserved than the cyt *b* gene it is likely that *C. orni* populations went through processes of differentiation earlier than *C. barbara*, or they have been demographically more stable. Such levels of differentiation found in continental *C. orni*, both at acoustic and genetic levels (including microsatellite differentiation, Seabra 2007), may be indicative of a subspecific diversification, especially taking into account the subspecific denomination of the Iberian and North African specimens of *C. barbara*, as also suggested by Seabra (2007).

Greek samples alone (continental Greece, and each of the islands Naxos, Kithira, Lesbos and Skyros), also showed a highly significant genetic structure and the genetic distances between *C. orni* haplotypes from the Greek islands and the Western Europe reached up to 1.7%, which is equivalent to the minimum distance found between *C. mordoganensis* and *C. orni* haplotypes (range of 1.7- 2.6% divergence). However, acoustic analysis of *C. orni* from Lesbos (Simões *et al*. 2006) could not support the designation of a subspecies status of this population since its calling songs appear within the limits of variation found in several populations in the Aegean islands and the Greek and Turkish mainlands. However, as discussed previously, allopatric genetically isolated populations may retain the ancestral advertisement calling song through stabilizing selection, or, alternatively, the sounds from other islands and mainlands might be also differentiated, thus the range of acoustic variation for this species might be inflated (see Chapter 4 for further discussion of the complex definitions of subspecies).

Dispersal barriers and postglacial re-colonization

Cicada orni and *C. barbara* show signatures of demographic expansion in the Mediterranean area related with climate changes during the Pleistocene (discussed in detail in Chapters 2 and 4). *Cicada orni* in continental Greece and *C. barbara* in Morocco, seem to

have been more stable than the Iberian and Western Europe cicadas, suggesting that these populations did not suffer severe demographic contractions as previously found in other organisms from southern populations (e.g. Jansson & Dynesius 2002; Hewitt 2004; Pinho *et al*. 2007).

The differentiation between *C. orni* from Greece and the rest of Europe suggests that the Greek populations remained isolated, probably confined by the Balkan Mountains, and did not contribute greatly to the re-colonization of the other European habitats. Similarly, *C. barbara* from Morocco might have been confined, with the Rif Mountains in the North, the Atlas Mountains in the east-south and the Sahara desert further south. In fact, it has been suggested that Morocco topography favours the isolation of species (e.g. Modolo *et al*. 2005) and consequently their divergence. Analysis of more North African sites would be very important to test for the isolation of Morocco populations and the true origin of the Iberian populations.

The mtDNA data from the *C. barbara* population from Ceuta in North Africa, seems to corroborate the notion of isolation of Moroccan cicadas by revealing the same haplotype as the Iberian cicadas. The same pattern was found in pine trees from a locality in the North of Morocco very close to Ceuta (Punta Cires), which showed the same haplotype as the Iberian Peninsula. In fact, the genetic differentiation found in *C. orni* and in *C. barbara* shows a similar pattern as the one found for both olive, *Olea europaea*, and pine trees, *Pinus pinaster* (e.g. Besnard & Bervillé 2000; Besnard *et al*. 2002a, 2002b; Burban & Petit 2003). This is not surprising, considering that adult cicadas from these species have a strong association with these trees.

The Strait of Gibraltar does not seem to constitute a strict barrier to dispersal of cicadas, as evidenced by the similarity of *C. barbara* population from Ceuta with the Iberian ones. However, the conflicting mtDNA and microsatellite data (the latter indicating that Ceuta population is more similar to Moroccan cicadas than to the Iberian ones; Seabra 2007), might suggest that the Strait of Gibraltar, at least in recent times, does constitute a dispersal barrier. In fact, the paleogeographic events favouring the dispersal through the Strait of Gibraltar are ancient (further discussed in Chapter 4), so since then, any other kind of dispersal through the Mediterranean Sea must have been sporadic. Preliminary studies concerning the dispersal ability in *C. orni* showed that males have a low pre-mating dispersal (100-150m) (Simões & Quartau 2007) probably due to the male aggregation behaviour of forming singing chorus, however, there is still no information regarding female dispersal.

In Europe, the range of Mountains of Pyrenees does not appear to constitute a severe constraint to dispersal of *C. orni* since cicadas from France are not differentiated from the Iberian (data corroborated by microsatellites, Seabra 2007). However, in view of the high impact of other Mountain ranges on the dispersal of cicadas, it seems likely that dispersal may have occurred along the Mediterranean coastal areas of Spain and France, instead of dispersal through the mountain ranges.

The hypothesis of an Aegean area origin of *C. orni* species and the current segregation of this species in Greece might indicate considerable differences, either in topography and/or in climate, at the Balkan Mountains along the evolutionary history of these cicadas.

Sympatric populations of C. orni and C. barbara

The analysis of calling songs from sympatric populations of *C. orni* and *C. barbara* in Portugal did not reveal the presence of any hybrids nor any character displacement in the acoustic variables (see Appendix I: Seabra *et al*. 2006). Genetically, microsatellite data also did not show any hybrids (Seabra 2007), and no misidentified specimens were found in mtDNA analysis, which seems to confirm that these two species were already highly differentiated when they came into contact.

The calling songs of these two cicada species are easily distinguished by gross-temporal variables: *C. barbara* produces a continuous shrill while *C. orni* emits successive series of short shrills alternated with short pauses. However, although some acoustic variables, such as, peak frequency and the syllable rate are usually distinct between these two species, their range overlaps. Also, in the field, *C. barbara* continuous calling song will probably cover up the short pauses from the *C. orni* calling song, especially when other characters of the song overlap. The acoustic interference when these species occur in sympatry may constitute a disadvantage for both species, since males and females may waste more energy to meet their conspecifics, but *C. orni* individuals may be more disadvantaged because the temporal characteristics of their calling songs may become more difficult to recognize by receivers. Fonseca & Revez (2002b) showed that manipulation of the gross temporal patterns of *C. barbara* calling song by introducing pauses was significantly less attractive to other males than the natural calling song, while a modified *C. orni* song, without pauses, proved as attractive as the *C. barbara* song. So acoustic interferences in the field may greatly influence individual responses, however, female preferences have not been tested as it is difficult to implement such playback/response studies in the field.

Several behaviour adjustments have been observed in areas where both *C. barbara* and *C. orni* occur simultaneously which might be indicative of stress induced by acoustic competition (examples in Appendix I), but one of the most striking instances is the fact that in late Summer *C. orni* individuals either disappear or are very rare in sympatric areas, while they remain abundant in allopatric localities. Thus, when both species occur in sympatry, *C. orni* might be at a further disadvantage and this might explain why *C. orni* is absent from North Africa whereas *C. barbara* is present in both North Africa and Iberian Peninsula.

Further work

To complete the phylogeographic analysis of *C. orni* it will be necessary to extend the mtDNA analysis to specimens from a wider geographical range, particularly, from Italy, the eastern Mediterranean Europe and western Asia and Middle East. This may reveal different refugia on the Balkan Peninsula, as well as, clarifying the migration routes taken by cicadas in the Eastern Europe. Also, a more extensive sampling of *C. barbara* specimens in North Africa and within the western Mediterranean islands where it also occurs, should provide the answers to some of the hypotheses raised here, including the possible origin of *C. barbara*, the identification of other potential refugia and the expansion routes along the North Africa.

An estimation of divergence times among cicada species would allow a better understanding of the mechanisms involved in speciation. Divergence times may be estimated, for instance, from DNA data using a Bayesian MCMC (Markov Chain Monte Carlo) approach (Thorne & Kishino 2002) or a maximum likelihood based method (Sanderson 2002). However, an accurate divergence dating would require information on the exact generation time of these cicadas, which is not currently available.

Information on the life cycle of the cicadas from genus *Cicada*, their effective dispersal capability, their morphological variation and female preferences to the acoustic traits is still deficient. Analysis of acoustic female preferences would allow identifying which characteristics of the calling songs are under sexual selection, and also testing the efficiency of these premating isolating barriers between species and populations.

A more detailed analysis using a more variable unlinked nuclear marker group, such as AFLPs (Amplified Fragment Length Polymorphisms) could reveal a clearer evolutionary history of the genus *Cicada* and its population structure. Particularly, when considering the difficulties with the application of microsatellites to these cicada species (e.g. nulls alleles, Seabra 2007), AFLPs markers could provide valuable answers. This technique, firstly

developed by Vos *et al*. (1995), allows the whole genome to be screened and it has higher reproducibility, resolution and sensitivity than other techniques, plus, no prior sequence information is needed for amplification. AFLPs can address problems, such as, identifying hybridization or introgression, species phylogenies, population or species level genetic structure and genetic diversity estimates, assignment of migratory individuals to populations, and identifying loci affecting phenotypes or under selection (Bensch and Akesson 2005). By changing the restriction enzymes adaptors or primers, AFLP can also be adapted to address more specific questions and to identify simple single locus markers, for instance SNPs (single nucleotide polymorphisms (Meksem *et al*. 2001; Nicod & Largiader 2003), which might be also a useful tool for population and conservation genetic studies within the genus *Cicada*.

Data gathered so far for these cicada species, including the mtDNA and the acoustic study presented here, isozyme (Ribeiro 1998; Seabra *et al*. 2000; Quartau *et al*. 2000, 2001) and microsatellite data (Seabra *et al*. 2002; Seabra 2007), and the work in progress by J. A. Quartau, P. Simões and other collaborators regarding mostly the biogeography, ecology and acoustic differentiation between populations from the Aegean area (Quartau & Simões 2005, 2006; Simões *et al*. 2000, 2006), provide a powerful insight into the patterns of speciation and evolution within cicadas. It is expected that some of the uncertainties and new hypotheses derived from the present study may constitute a solid basis for stimulating further research.

References

Angulo A, Reichle S. 2008. Acoustic signals, species diagnosis, and species concepts: the case of a new cryptic species of *Leptodactylus* (Amphibia, Anura, Leptodactylidae) from the Chapare region, Bolivia. *Zoological Journal of the Linnean Society* **152**: 59-77.

Bennet-Clark HC. 1998. Size and scale effects as constraints in insect sound communication. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **353**: 407-419.

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

Bensch S, Akesson M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**: 2899-2914.

Besnard G, Bervillé A. 2000. Multiple origins for Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. Compte Rendu del `Académie des Sciences, Paris*, Sciences de la Vie/ Life Sciences* **323**: 173-181.

Besnard G, Khadari B, Baradat P, Bervillé A. 2002a. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* **104**: 1353-1361.

Besnard G, Khadari B, Baradat P, Bervillé A. 2002b. Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics* **105**: 139-144.

Boulard M. 1982. Les cigales du Portugal, contribution à leur étude (Homoptera, Cicadidae). *Annales de la Société Entomologique de France* (N.S.) **18**(2): 181-198.

Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**: 1487-1495.

Cooley JR, Marshall DC. 2001. Sexual signalling in periodical cicadas, *Magicicada* spp. (Hemiptera: Cicadidae). *Behaviour* **138**: 827-855.

Douris V, Cameron RAD, Rodakis GC, Lecanidou R. 1998. Mitochondrial phylogeography of the land snail *Albinaria* in Crete: long-term geological and short-term vicariance effects. *Evolution* **52** (1): 116-125.

Fonseca PJ, Revez MA. 2002a. Temperature dependence of cicada songs (Homoptera, Cicadoidea). *Journal of Comparative Physiology A*, **187**, 971-976.

Fonseca PJ, Revez MA. 2002b. Song discrimination by male cicadas *Cicada barbara lusitanica* (Homoptera, Cicadidae). *Journal of Experimental Biology.* **205**: 1285-1292.

Greenfield MD, Tourtellot MK, Snedden WA. 1997. Precedence effects and the evolution of chorusing. *Proceedings of the Royal Society B: Biological Sciences* **264**: 1355-1361.

Hewitt GM. 2004. The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology* **1**: 1-16.

Heyer WR, Reid YR. 2003. Does advertisement call variation coincide with genetic variation in the genetically diverse frog taxon currently known as *Leptodactylus fuscus* (Amphibia: Leptodactylidae)? *Anais da Academia Brasileira de Ciências* **75**: 39-54.

Hughes JM, Mather PB, Hillyer MJ, Cleary C & Peckarsky B. 2003. Genetic structure in a montane mayfly *Baetis bicaudatus* (Ephemeroptera: Baetidae), from the Rocky Mountains, Colorado. *Freshwater Biology* **48**: 2149-2162.

Jansson, R. & Dynesius M. (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics* **33**, 741-777.

Kasapidis P, Magoulas A, Mylonas M, Zouros E. 2005. The phylogeography of the gecko *Cyrtopodion kotschyi* (Reptilia: Gekkonidae) in the Aegean archipelago. *Molecular Phylogenetics and Evolution* **35**(3): 612-623.

Kraus FB, Franck P, Vandame R. 2007. Asymmetric introgression of African genes in honeybee populations (*Apis mellifera* L.) in Central Mexico. *Heredity* **99**: 233-240.

Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot DA. 2001. Conversion of AFLP bands into high-throughput DNA markers. *Molecular Genetics and Genomics* **265**: 207-214.

Modolo L, Salzburger W, Martin RD. 2005. Phylogeography of Barbary macaques (*Macaca sylvanus*) and the origin of the Gibraltar colony. *The Proceedings of the National Academy of Sciences Online (US)* **102**: 7392-7397.

Nicod JC, Largiader CR. 2003. SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Research* **31**: art. nº -e19.

Paterson HEH. 1985. The recognition concept of species. *In* ES Vrba (Ed.) *Species and Speciation*. Transvaal Museum, Pretoria, pp. 21–29.

Pinho C, Harris DJ, Ferrand N. 2007. Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology* **16**: 1191-1205.

Price BW, Barker NP, Villet MH. 2007. Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology*, **16**: 2574-2588.

Quartau JA, Simões PC. 2005. *Cicada cretensis* sp. n. (Hemiptera, Cicadidae) from southern Greece. *Biologia Bratislava* **60**(5): 489-494.

Quartau JA, Simões PC. 2006. Acoustic evolutionary divergence in cicadas: the species of *Cicada* L. in Southern Europe. *In*: S Drosopoulos & MF Claridge (Eds). *Insect sounds and communication*. *Physiology, Behaviour, Ecology and Evolution*. Taylor & Francis Group, Boca Raton, Chapter 17, pp. 227–237.

Quartau JA, Ribeiro M, Simões PC, Crespo A. 2000. Taxonomic separation by isozyme electrophoresis of two closely related species of *Cicada* L. (Hemiptera: Cicadoidea) in Portugal. *Journal of Natural History* **34**: 1677-1684.

Quartau JA, Ribeiro M, Simões PC, Coelho MM. 2001. Genetic divergence among populations of two closely related species of *Cicada* Linnaeus (Hemiptera: Cicadoidea) in Portugal. *Insect Systematics & Evolution* **32**: 99-106.

Ribeiro M. 1998. *Diferenciação em duas espécies gémeas do género Cicada L. (Insecta, Homoptera, Cicadoidea): Cicada orni L. e Cicada barbara (Stal)*. Tese de Mestrado, Faculdade de Ciências da Universidade de Lisboa.

Sanborn AF. 2006. Acoustic Signals and Temperature. *In*: S Drosopoulos & MF Claridge (Eds). *Insect sounds and communication*. *Physiology, Behaviour, Ecology and Evolution*. Taylor & Francis Group, Boca Raton, Chapter 7, pp. 111–125.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**:101-109.

Seabra SG. 2007. *Patterns of divergence in populations of two Mediterranean species of genus Cicada L. (Hemiptera, Cicadidae) based on microsatellite genetic markers and acoustic data*. *PhD Thesis*, Faculdade de Ciências da Universidade de Lisboa.

Seabra SG, Pinto-Juma GA, Quartau JA. 2006. Calling songs of sympatric and allopatric populations of *Cicada barbara* and *C. orni* (Hemiptera: Cicadidae) on the Iberian Peninsula. *European Journal of Entomology* **103**: 843-852.

Seabra S, Simões PC, Drosopoulos S, Quartau JA. 2000. Genetic variability and differentiation in two allopatric species of the genus *Cicada* L. (Hemiptera, Cicadoidea) in Greece. *Mitteilungen aus dem Museum für Naturkunde in Berlin - Deutsche Entomologische Zeitschrift* **47**(2): 143-145.

Seabra SG, Wilcock HR, Quartau, JA, Bruford MW. 2002. Microsatellite loci isolated from the Mediterranean species *Cicada barbara* (Stal) and *C. orni* L. (Hemiptera, Cicadoidea). *Molecular Ecology Notes*, 2: 173-175.

Simões PC, Quartau JA. 2007. On the dispersal of males of *Cicada orni* in Portugal (Hemiptera: Cicadidae). *Entomologia Generalis* **30**(3): 8pp.

Simões PC, Boulard M, Quartau JA. 2006. On the taxonomic status of *Cicada orni* Linnaeus (Hemiptera, Cicadidae) from Lesbos island in Greece. *Zootaxa* **1105**: 17-25.

Simões PC, Boulard M, Rebelo MT, Drosopoulos S, Claridge MF, Morgan JC, Quartau JA. 2000. Differences in the male calling songs of two sibling species of *Cicada* (Hemiptera: Cicadoidea) in Greece. *European Journal of Entomology* **97**(3): 437-440.

Sueur J, Aubin T. 2002. Acoustic communication in the Palaearctic red cicada *Tibicina haematodes*: chorus organization, calling song structure, and signal recognition. *Canadian Journal of Zoology* **80**: 126-136.

Suer J, Vanderpool D, Simon C, Ouvrard D, Bourgoin T. 2007. Molecular phylogeny of the genus *Tibicina* (Hemiptera, Cicadidae): rapid radiation and acoustic behaviour. *Biological Journal of the Linnean Society* **91**: 611-626.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51**(5): 689-702.

Villet M. 1992. Responses of free-living cicadas (Homoptera: Cicadidae) to broadcasts of cicada songs. *Journal of the Entomological Society of Southern Africa*, **55**: 93-97.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP - a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407-4414.

William J, Ballard O, Kreitman M. 1995. Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution* **10**: 485-488.

Appendix I

Calling songs of sympatric and allopatric populations of *Cicada barbara* and *C. orni* (Hemiptera: Cicadidae)

Published: Seabra SG, Pinto-Juma GA, Quartau JA. 2006. Calling songs of sympatric and allopatric populations of *Cicada barbara* and *C. orni* (Hemiptera: Cicadidae). *European Journal of Entomology* **103**: 843-852

Eur. J. Entomol. **103**: 843–852, 2006
ISSN 1210-5759

Calling songs of sympatric and allopatric populations of *Cicada barbara* and *C. orni* (Hemiptera: Cicadidae) on the Iberian Peninsula

SOFIA G. SEABRA, GABRIELA PINTO-JUMA and JOSÉ A. QUARTAU

Centro de Biologia Ambiental & Departamento de Biologia Animal, Faculdade de Ciências de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal; e-mail: sgseabra@fc.ul.pt

Key words. Cicadidae, *Cicada barbara*, *C. orni*, cicadas, acoustics, calling song, sympatry, allopatry

Abstract. Calling songs of the sibling species *Cicada barbara* and *C. orni* were studied in sympatric and allopatric populations on the Iberian Peninsula, where the distribution ranges of both species overlap. No difference was found in any acoustic property for the sympatric and allopatric populations of *C. barbara* studied and only one variable (minimum frequency) was significantly different between sympatric and allopatric populations of *C. orni*. No hybrids with intermediate songs were found and no character displacement in the calling song was detected. It is very likely that these species were already considerably differentiated when they met on the Iberian Peninsula. Particularly, premating (or even postmating) isolating mechanisms (according to Mayr's Biological Species Concept) or different specific-mate recognition systems (in the view of the Paterson's Recognition Concept of Species) were most likely already present, which prevented hybridization between this pair of species. It is assumed that the calling songs are the most important premating isolating mechanism corresponding to the specific-mate recognition systems of these species of cicadas.

INTRODUCTION

Cicada barbara Stål and *C. orni* Linnaeus (Hemiptera, Cicadidae) are a pair of sibling cicada species, very similar in morphology, with only slight differences in external characteristics. Examples are differences in the wing spots (Quartau, 1988; Ribeiro, 1998) and in the male genitalia, namely in the lengths of the pygofer (and its dorsal spine), the tenth abdominal segment and its appendages (which are shorter in *C. barbara*) and in the width of the shaft of the aedeagus (thinner in *C. orni*) (Quartau, 1988). In contrast, the acoustic signals produced by the males of these two species are distinct, since *C. barbara* produces a continuous series of pulses without pauses (Boulard, 1982, 1995; Fonseca, 1991; Quartau & Rebelo, 1994) and *C. orni* a series of pulses forming echemes alternating with intervals of silence (Popov, 1975; Boulard, 1982; Fonseca, 1991; Quartau et al., 1999, 2000; Pinto-Juma et al., 2005). In addition, sequencing of the 12S rRNA mitochondrial gene indicates that these two species constitute distinct evolutionary units (Pinto-Juma et al., unpubl.).

Both *C. barbara* and *C. orni* occur in Mediterranean woodland or shrubland, mainly in olive, oak or pine tree groves, but also on other trees and shrubs and even on scattered vegetation in cities (Quartau, 1995). In Corsica *C. orni* was found by Puissant & Sueur (2001) on two types of vegetation: either it is 0.5 to 2 m high with ligneous plants covering more than 60% of the area or more than 2 m high with ligneous plants covering more than 40% of the area. Only *C. barbara* is present both in North Africa and the Iberian Peninsula. *C. orni* is distributed in south west, central and eastern Europe, western Asia and the Middle East (Popov, 1975; Quartau & Fonseca, 1988; Schedl, 1973, 1999). Moreover the Iberian populations of *C. barbara* are considered by Boulard (1982) to be a subspecies, *C. barbara lusitanica* Boulard, different from the type subspecies, which only occurs in north Africa (*C. barbara barbara* Stål).

The distribution ranges of *C. barbara* and *C. orni* overlap on the Iberian Peninsula, and males of both species in a few localities can be seen singing on the same trees at the same time. However, there is some seasonal displacement as *C. orni* emerges earlier in summer (June) than *C. barbara* (July/August). In allopatric populations, the adults of *C. orni* disappear in September/October, whereas in sympatric populations they usually disappear much earlier (August). *C. barbara* shows less variation: it appears somewhat earlier where it occurs allopatrically than sympatrically but disappears at the same time, in September/October (Ribeiro, 1998; and pers. obs.).

Despite the clear differences in the calling songs of males of *C. barbara* and *C. orni*, the structure of the tymbals and mechanism of sound production in these species is similar, as expected in closely related species of the same genus (Fonseca, 1991, 1994). In both species the inward distortion of one of the tymbals (membranes located in the first segment of the abdomen, which are distorted by the tymbal muscles) produces usually three pulses of sound and the outward distortion another pulse, which is usually superimposed on the inward pulses of the other tymbal (Fonseca, 1991). The tymbals alternate in the production of sound.

Males of *C. barbara* and *C. orni* sing continuously for hours, usually without changing their location. This pattern of behaviour is described in Cooley & Marshall (2001), as one extreme in the range of systems found in acoustically signalling insects, and is observed in other species, such as the Neotropical cicada *Fidicina mannifera* (Cocroft & Pogue, 1996), the Palaearctic *Tibicina* species (Quartau & Simões, 2003; Sueur & Aubin, 2004)

and the Australian *Cystosoma saundersii* (Doolan, 1981). Males of both species of *Cicada* studied sing during the hottest part of the day, sometimes even at night if temperatures are well above 30°C.

The calling songs of cicada males are known to attract conspecific females in other species, for example in *Pycna semiclara* and *Azanicada zuluensis* (Villet, 1992), in *Cystosoma saundersii* and *Cyclochila australasiae* (Daws et al., 1997) and in *Magicicada* spp. (Cooley & Marshall, 2001), and are also known to result in the aggregation of males, for example in *Azanicada zuluensis* (Villet, 1992) and *Magicicada* spp. (Cooley & Marshall, 2001).

Experiments on the behaviour of cicadas in general, and *C. barbara* and *C. orni* in particular, are difficult in captivity (Fonseca & Revez, 2002a; Simões & Quartau, unpubl.). In the field, the males of both species often aggregate on the same trees, with a single male song often eliciting other males to sing. Moreover, females are often seen flying to the trunks or branches of trees where males are singing (pers. obs.). Playback experiments carried out with males of *C. barbara* (Fonseca & Revez, 2002a) has shown that females of this species can discriminate conspecific song from that of *C. orni*, as they respond (by singing) more quickly to the conspecific than heterospecific song. Also, altering the temporal pattern of the calling song of *C. barbara* reduce the response of the males, which do not respond to songs with pauses longer than 30 ms, which is more characteristic *C. orni* song, but respond to a modified *C. orni* song without pauses. No experiments were done yet about species discrimination by females.

Where sibling species occur sympatrically it is possible to address questions about isolation mechanisms and recognition processes. According to the Biological Species Concept, species are isolated from each other by "isolating mechanisms" (Dobzhansky, 1951), which include premating or postmating mechanisms (Mayr, 1948, 1963). According to Dobzhansky (1951), if two incompletely reproductively isolated species meet, after a period of allopatry, the reinforcement of isolating mechanisms might evolve by natural selection to prevent heterospecific matings. Under this scenario, sympatric populations would be expected to show more marked reproductive character displacement in mating mechanisms in comparison to allopatric populations. However, Paterson (1985) criticizes this concept since he considers the premating and postmating isolating mechanisms as effects, not evolving to serve the function of preventing hybridization between species. In his Recognition Concept of Species, species are seen as "that most inclusive population of individual biparental organisms which share a common fertilization system" (Paterson, 1985 p. 25). Speciation is seen by Paterson as an incidental effect resulting from the adaptation of the characters of the fertilization system (or Specific-Mate Recognition System – SMRS), among others, to a new habitat. According to the same author, if two populations have interpopulation "sterility" (heterozygote disadvantage) but share a common SMRS, they cannot coexist, since natural selection will act to eliminate the cause of hybrid disadvantage, in which case, reproductive character displacement is not expected (Paterson, 1985).

Evidence of reproductive character displacement is difficult to find. Marshall & Cooley (2000) report a case in one pair of north-American cicada species, *Magicicada tredecim* and *M. neotredecim*. These authors found that *M. neotredecim* produces higher dominant frequency calls when it occurs sympatrically than allopatrically with *M. tredecim*, but the latter species maintains the same frequency throughout their distribution range. This is the expected pattern if there is a process of reinforcement driven by selection against wasteful heterospecific matings (Marshall & Cooley, 2000). Alternatively, this character displacement could be viewed as an adaptation of *M. neotredecim* to the acoustic environment created by *M. tredecim*. In fact, the background noise created by the calling activity of another species may affect the receiver's ability to perceive signals and promote directional selection of the signal produced (Villet, 1995).

In this paper the evidence for hybrids with intermediate characteristics in calling songs between *C. barbara* and *C. orni* was assessed. Differences between the calling songs of sympatric and allopatric populations of each species on the Iberian Peninsula, particularly, whether the songs were more similar or different when species occurred sympatrically, were also investigated.

MATERIAL AND METHODS

Field procedures

Eleven populations of *C. barbara* and eight of *C. orni* were sampled on the Iberian Peninsula where these species occurred allopatrically (six for *C. barbara* and four for *C. orni*) and sympatrically (five for *C. barbara* and four for *C. orni*) (Fig. 1 and Table 1).

Male cicadas were located by their calling songs. Recordings of the acoustic signals were made using a Sony Dat recorder (TCD-D10 ProII; frequency range 20–22000 Hz; sampling frequency 44.1 kHz) connected to a dynamic Sony F-780 microphone (frequency responses 50–18000 Hz). Ambient temperature was taken in the shade at the time of each recording at all localities except Alcalar, Alvor, Monforte, Casalinho and Piedade.

Sound analysis

Sound recordings were digitized using the software Avisoft-SASLab Pro (Specht, 2002) at a sampling rate of 44.1 kHz and a resolution of 16 bits. Time and frequency analysis were performed on one-minute recordings of each individual (whenever possible). Fast Fourier transformation with a resolution of 512 points and a Hamming Window was applied to compute the frequency spectra (Fig. 2). The gross-temporal variables analysed in *C. orni* were: number and duration of echemes and duration of the interval between them (interecheme interval) (Fig. 2). Echeme rates, periods and ratios echeme/interval were then calculated. Spectrum-based variables analysed for both species were: peak frequency, minimum frequency, maximum frequency, bandwidths, and quartiles (described in Pinto-Juma et al., 2005). In *C. orni*, all frequency measurements were calculated from the mean spectrum of each echeme. Time and frequency measurements of the echemes were then averaged, and the mean was taken as the value of the variable for each speci-

Fig. 1. Allopatric populations of *Cicada barbara* (dark circles) and *C. orni* (empty circles), and sympatric populations of both species (triangles) sampled on the Iberian Peninsula

men. A fine-temporal property of the signals, the syllable rate, was also calculated for both species. The number of groups of pulses (syllables) was counted in 30 echemes for *C. orni* (first and last syllables in each echeme were discarded due to their different characteristics – see Fig. 2) and 30 fragments of c. 0.1 s for *C. barbara*. The average number of syllables per unit of time was calculated for each specimen. These syllable rates correspond to the production of pulses by the two tymbals.

Statistical analysis

Nonparametric Mann-Whitney tests were used to compare each acoustic variable between species and between sympatric and allopatric populations of each species. Coefficients of variation for each acoustic variable were compared between species using the nonparametric Wilcoxon signed rank test. Spearman nonparametric correlations between ambient temperature and each acoustic variable were also computed for each species. The significance of multiple tests was assessed by reducing the critical P value according to the Dunn-Sidák method (Dytham, 2003), from 0.05 to 1–(0.951/k), where k is the number of tests performed.

Multivariate analysis was applied to the data matrix of 9 acoustic variables (the variables common to both species) measured for 158 individuals, namely a Principal Components Analysis (PCA) and a Discriminant Function Analysis (DFA). PCA is used to reduce a large number of variables to a smaller number of factors (or components), with no need to specify a dependent variable. The percentage of variance explained by each of the components is given, as well as the correlation coefficients between the variables and the components (component loadings). The component scores obtained for the individuals can be used to compare groups (in this case there are four groups: allopatric *C. barbara*, sympatric *C. barbara*, sympatric *C. orni* and allopatric *C. orni*) with Kruskall-Wallis and Mann-Whitney tests. In DFA, on the other hand, groups are determined beforehand. It derives discriminant functions that best separate the groups. The statistical significances of the functions in discriminating the groups are given, as well as the correla-

TABLE 1. Populations of *Cicada barbara* and *C. orni* sampled on the Iberian Peninsula in areas where they occurred allopatrically and sympatrically.

Localitites	N	Dates of recording	Temperatures (°C)
C. barbara occurs allopatrically			
Sevilla (Andalucía, Spain)	7	6/8/2001	38–41
Córdoba (Andalucía Spain)	5	6/9/2000	34
Alcalar (Algarve, Portugal)	10	23/8/1995	–
Alvor (Algarve, Portugal)	3	28/8/1995	–
Moura (Baixo Alentejo, Portugal)	11	28/8/2001	31–34
Foz Côa (Beira Alta, Portugal)	3	11/7/1999	34
C. barbara occurs sympatrically			
Portel (Alto Alentejo, Portugal)	10	24/7 and 10/8/2001	31–35
Sousel (Alto Alentejo, Portugal)	11	8/9/2001	33–35
Monforte (Alto Alentejo, Portugal)	6	22/7 to 24/7/1995	–
Crato (Alto Alentejo, Portugal)	14	6/7 to 8/7/1999; 15/7/1999; 1/8 to 3/8/1999	26–41
Casalinho (Estremadura, Portugal)	5	27/7/1995	–
C. orni occurs allopatrically			
Algeciras (Andalucía, Spain)	10	5/8/2001	31–34
Alter-do-Chão (Alto Alentejo, Portugal)	7	6 to 9/8/1997	25–30
Piedade (Arrábida, Estremadura, Portugal)	10	19/7 to 12/8/1995	–
Monte-da-Caparica (Estremadura, Portugal)	7	16 to 22/9/1997	25–30
C. orni occurs sympatrically			
Sousel (Alto Alentejo, Portugal)	11	27/6/2003	27–30
Monforte (Alto Alentejo, Portugal)	16	25/7 to 7/8/1997	23–38
Crato (Alto Alentejo, Portugal)	8	27/6/2001	24–26
Arrábida (Estremadura, Portugal)	4	18/8 and 10/9/1997	30
TOTAL	158		

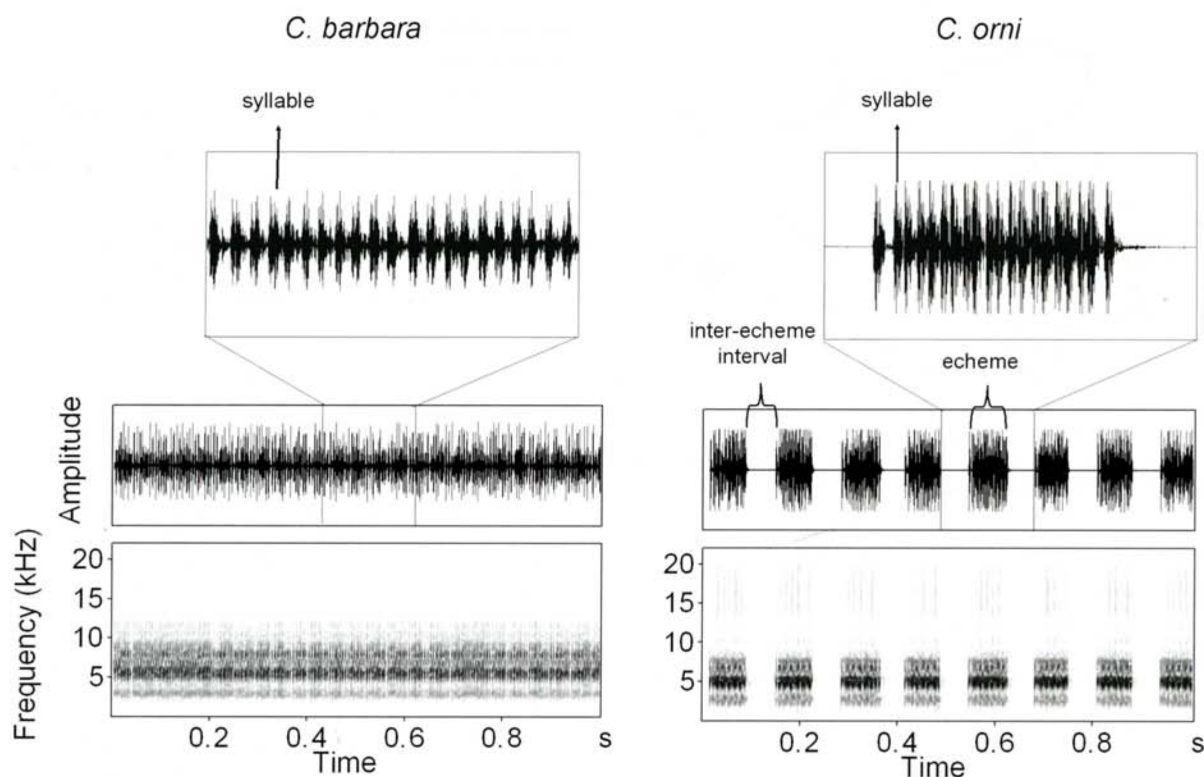


Fig. 2. Oscillograms (amplitude vs. time) and sonagrams (frequency vs. time) of the calling songs of *Cicada barbara* and *C. orni*.

tions of each variable to each discrimination function. All statistical analyses were carried out in SPSS Version 10.0.

RESULTS

Cicada barbara and *C. orni* differed significantly in every acoustic variable compared (Mann-Whitney, $p < 0.001$). The average values of frequency variables of *C. orni* were always lower than those of *C. barbara*. The peak frequency was 4709.8 Hz ± 452.44 (average ± standard deviation) in *C. orni* and 6283.5 ± 476.40 in *C. barbara*. On the other hand, the syllable rate was on average higher in *C. orni* (224.3 ± 25.00) than, in *C. barbara* (201.4 ± 18.78) (Table 2, Fig. 2). The syllable periods, derived from the syllable rates, ranged from 3.96 to 7.27 ms in *C. barbara* and from 3.84 to 6.72 ms in *C. orni*. The syllable rates for only one of the tymbals are half and the syllable periods double of these values.

The inter-individual coefficients of variation of the frequency variables varied from 6.24% to 20.85% in *C. barbara* and from 6.15% to 18.79% in *C. orni*. The syllable rate has a coefficient of variation of 9.35% in *C. barbara* and of 11.20% in *C. orni*. No significant difference between species was found in the coefficients of variation for the acoustic variables (Wilcoxon test, $p = 0.859$). The inter-individual coefficients of variation for the temporal variables in *C. orni* were higher than the frequency variables and ranged from 24.90% to 68.38%.

No significant differences in any of the acoustic variables were found between sympatric and allopatric populations of *C. barbara* (Mann-Whitney tests). In *C. orni*, only the minimum frequency differed significantly between sympatric and allopatric populations (Mann-Whitney test, $p = 0.002$), with lower values when it occurred sympatrically than allopatrically (values for sympatric populations deviating more from those of *C. barbara*) (Fig. 3). This difference was not due to differences in temperature as no frequency variable significantly correlated with temperature. On the other hand, ambient temperature was found to be positively correlated with the syllable rate in both *C. barbara* and *C. orni* (Fig. 4) ($r_S = 0.360$, $p = 0.006$ for *C. barbara* and $r_S = 0.714$, $p < 0.001$ for *C. orni*). The correlation for *C. barbara* was only marginally significant after applying the Dunn-Sidák correction (critical P value = 0.0057), while that for *C. orni* was still strongly significant (critical P value = 0.0037). In *C. orni*, the echeme duration was negatively correlated ($r_S = -0.631$, $p < 0.001$), the echeme rate was positively correlated ($r_S = 0.574$, $p < 0.001$) and the echeme period negatively correlated ($r_S = -0.574$, $p < 0.001$) with temperature. The correlations between inter-echeme interval and temperature, and between the ratio echeme/interval and temperature were not significant ($r_S = 0.030$, $p = 0.827$ and $r_S = -0.329$, $p = 0.012$, respectively) after applying the Dunn-Sidák correction.

In the PCA, the percentage of variation explained by the first four components was 95.3% (64.1% by the first component alone, 80.0% by the first two and 89.2% by the first three). The component loadings were high for all variables in the first component (from 0.642 for band-

TABLE 2. Average ± standard deviation (SD), minimum (Min.), maximum (Max.) and coefficient of variation (CV, corrected for small size samples as in Sokal & Rohlf (1981)) of the acoustic variables of the calling songs of *Cicada barbara* and *C. orni* recorded on the Iberian Peninsula.

Acoustic variables	*C. barbara*					*C. orni*				
	N	Average ± SD	Min.	Max.	CV(%)	N	Average ± SD	Min.	Max.	CV(%)
Peak frequency (Hz)	85	6283.5 ± 476.40	5080	7660	7.60	73	4709.8 ± 452.44	3851	6199	9.64
Minimum frequency (Hz)	85	1830.7 ± 337.53	2150	4730	11.96	73	2125.5 ± 372.79	1583	3952	17.60
Maximum frequency (Hz)	85	10840.5 ± 1699.39	7920	17910	15.72	73	9100.2 ± 1185.43	7239	13144	13.07
Bandwidth (Hz)	85	8006.2 ± 1664.69	5160	15330	20.85	73	6970.2 ± 1305.07	3626	11174	18.79
Quartile 25% (Hz)	85	5839.1 ± 421.82	4300	6460	7.24	73	4572.4 ± 280.27	3854	5011	6.15
Quartile 50% (Hz)	85	6665.2 ± 414.62	5680	7490	6.24	73	5472.0. ± 420.26	4611	6280	7.71
Quartile 75% (Hz)	85	8543.2 ± 715.16	6460	9730	8.40	73	6894.6 ± 517.50	5405	8085	7.53
Quart75%–Quart25% (Hz)	85	2704.1 ± 452.18	1300	3620	16.77	73	2321.9 ± 363.30	1242	3170	15.70
No. syllables/s	82	201.4 ± 18.78	137.6	252.3	9.35	53	224.3 ± 25.00	148.7	260.3	11.20
No. Echemes/s						73	5.37 ± 1.333	2.38	7.49	24.90
Echeme duration (s)						73	0.078 ± 0.0308	0.039	0.207	39.58
Inter-echeme interval (s)						73	0.123 ± 0.0720	0.058	0.362	58.53
Echeme period (s)						73	0.201 ± 0.0649	0.134	0.420	32.34
Ratio echeme/interval						73	0.894 ± 0.6093	0.166	3.314	68.38

width to 0.969 for quartile 75% in frequency variables and –0.484 for syllable rate). All the variables were significantly correlated, except for minimum frequency *vs.* bandwidth. There were no significant differences in the component scores of Component 1 and 2 between sympatrically and allopatrically occurring specimens of the same species (Mann-Whitney test, $p > 0.05$), but the sympatrically occurring specimens had a greater dispersion in the scatterplot (Fig. 5) than the allopatrically occurring ones in both species on both axis. This greater dispersion is probably the reason for the slight proximity observed between some sympatric specimens of one species and those of the other species.

DFA gave an overall correct classification rate of 62.96%. Three functions were computed but only one was significant (Wilk's $\lambda = 0.178$, $p < 0.0001$) and accounted for 97.6% of the variation. The structure matrix indicated that the frequency variables quartile 25%, peak frequency, quartile 50%, quartile 75% and minimum frequency were the most important in determining Function 1. The classification table (Table 3) showed that almost every *C. barbara* and *C. orni* individual where correctly classified to species, with the exception of three individuals (one sympatric *C. barbara* was classified as allopatric *C. orni*; one sympatric *C. orni* was classified as sympatric *C. barbara*; and one allopatric *C. orni* was classified as sympatric *C. barbara*). However, a substantial number of individuals of both allopatric and sympatric groups were misclassified. Individuals of *C. orni* that occurred sympatrically were mostly classified in the allopatric group. When cross-validating, only 48.9% of the grouped samples were correctly classified, which indicates that this analysis did not discriminate between these groups.

When performing the Discriminant Analysis separately for each species, the discriminant function separating sympatric and allopatric *C. barbara* was significant (Wilk's $\lambda = 0.828$, $p = 0.044$) and correctly classified 65.9% of the original samples and 53.7% of cross-validated samples. The variables that were more highly correlated with the discriminant function were quartile 50% (0.621), peak frequency (0.531), syllable rate (–0.504) and minimum frequency (0.493). When the Discriminant Analysis was applied to allopatric and sympatric *C. orni*, the discriminant function was not significant (Wilks' $\lambda = 0.872$, $p = 0.479$, when the nine variables common to both species were used; and Wilks' $\lambda = 0.743$, $p = 0.260$, when all 14 variables were used).

DISCUSSION

The calling songs of *Cicada barbara* and *C. orni* are easily distinguished by the human ear since the first spe-

TABLE 3. Classification of groups using the Discriminant Function Analysis. Values are percentages of samples (individuals) in the actual group predicted to belong to each group. Number of samples (individuals) in parentheses.

Actual Group	Predicted group membership			
	allopatric Cb	sympatric Cb	sympatric Co	allopatric Co
allopatric *C. barbara*	56.8% (21)	43.2% (16)	0.0% (0)	0.0% (0)
sympatric *C. barbara*	26.7% (12)	71.1% (32)	0.0% (0)	2.2% (1)
sympatric *C. orni*	0.0% (0)	3.3% (1)	76.7% (23)	20.0% (6)
allopatric *C. orni*	0.0% (0)	4.3% (1)	56.5% (13)	39.1% (9)

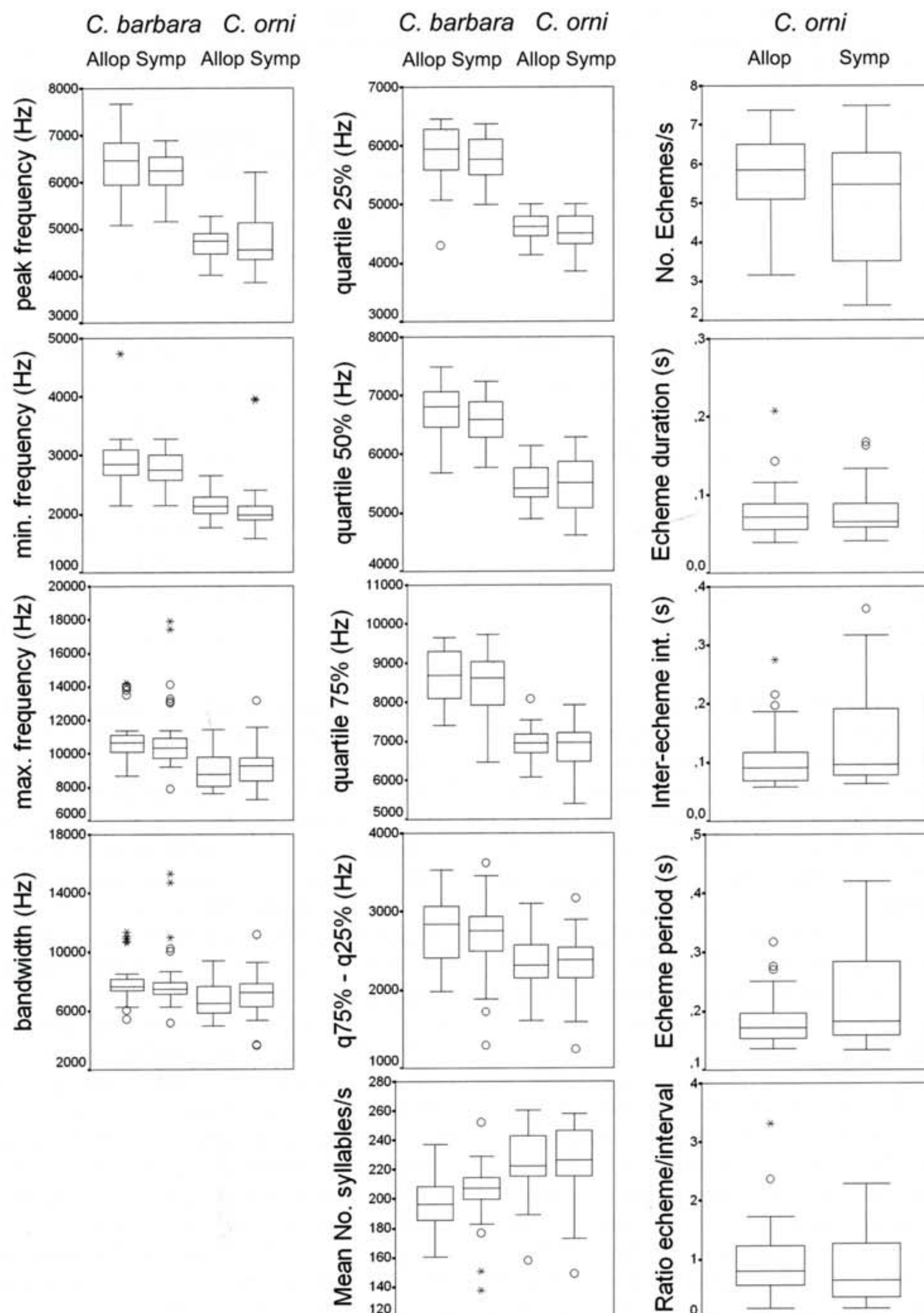


Fig. 3. Boxplots of the acoustic variables of the songs of *Cicada barbara* and *C. orni* occurring allopatrically (Allop) and sympatrically (Symp).

cies produces a continuous shrill and the latter a successive series of short shrills alternating with short pauses. At the acoustic frequency level these species are also generally distinct, *C. barbara* produces a higher peak frequency (average of 6.3 kHz) than *C. orni* (average of 4.7 kHz), with a difference of more than 1 kHz between the averages for each species. However, this is not a totally diagnostic characteristic because there is an overlap between species (see Fig. 3). In fact, some *C. orni* males produce a sound with a peak frequency above 6 kHz and some *C. barbara* males a peak frequency as low as 5 kHz. Also the syllable rates of the two species overlap substantially, despite the average being significantly higher in *C. orni* than in *C. barbara*. In fact, the presence of a few specimens that, in some song variables, were similar to those of the other species caused the Principal Components Analysis and the Discriminant Function Analysis to indicate an incomplete separation between the species.

Differences in the frequency of the sounds are generally related to the size of the sound producing or resonator

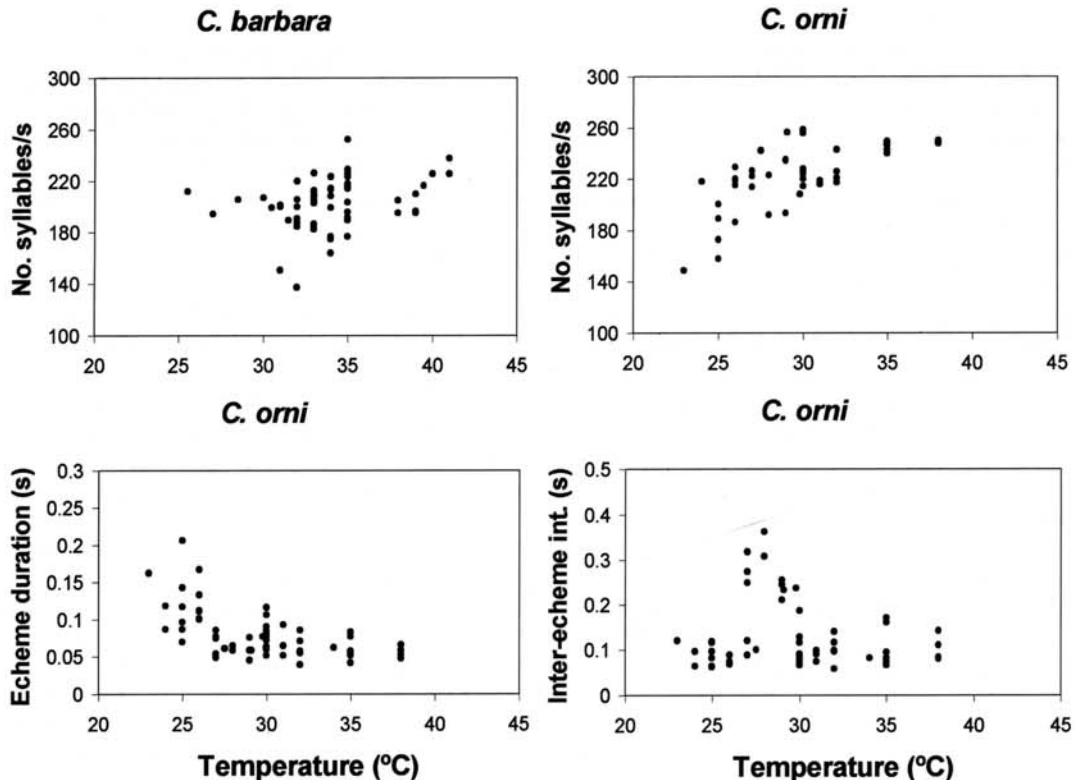


Fig. 4. Scatterplots of the number of syllables per second relative to temperature in *Cicada barbara* and *C. orni* and of both echeme duration and inter-echeme interval relative to temperature in *C. orni*.

organ (Young & Josephson, 1983). A significant negative correlation is usually found between body length and the dominant song frequency in cicada species, with larger species producing lower dominant frequency songs (Bennet-Clark & Young, 1994). This also appears to be the case in the pair of species analysed since *C. barbara* is usually smaller than *C. orni* (Ribeiro, 1998).

Variability in the frequency characteristics of calling songs of cicadas of these species was relatively low (coefficient of variation usually below 20%), as expected since these characteristics are constrained by physical properties of the sound-producing organ. On the other hand, the variability of the gross-temporal characteristics in *C. orni* was high, as also found by Pinto-Juma et al. (2005). However, the fine-temporal characteristic of the songs of both species, the syllable rate, had similar coefficients of variation to the ones of the frequency variables. According to Gerhardt (1994), many female insects choose males whose signals have species-typical values of fine-temporal properties, such as pulse rate or duration. This would lead to stabilizing selection on these characteristics of the song.

However, the variability of the characteristics of each song should be analysed with caution, since the estimate of the variability is dependent on our ability to measure it, which might be different in scaling from the perceptual systems of the intended receivers (McGregor, 1991).

No cicadas with songs intermediate between *C. barbara* and *C. orni* were found in this study. All male cicadas were clearly identified by the gross-temporal differences in their songs (*C. barbara* song is continuous and that of *C. orni* discontinuous). The overlap between species in some song variables is believed to be due to natural variation and not hybridization. Moreover, there was no evidence of character displacement in the acoustic variables studied. Only one acoustic variable (minimum frequency) was significantly different in allopatric and

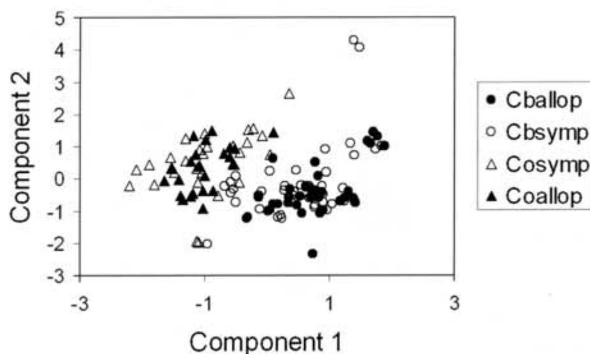


Fig. 5. Scores of the first two PCA components extracted from a data matrix, which was composed of 9 acoustic variables measured for 158 individuals of *Cicada barbara* (Cb) and *C. orni* (Co) occurring allopatrically (allop) and sympatrically (symp).

sympatric populations of *C. orni* showing a deviation from the *C. barbara* values. The multivariate analyses did not indicate any consistent differentiation between sympatric and allopatric populations at the acoustic level.

Also, previous morphometric work, based on the head, wings and male genitalia of *C. barbara* and *C. orni*, did not reveal differences between sympatric and allopatric populations (Ribeiro, 1998).

It is likely that the calling songs, which are most certainly part of the Specific-Mate Recognition Systems in these species, were already sufficiently differentiated before these species came into contact, and so, no hybridization occurred. In fact, at the mtDNA level, *C. barbara* and *C. orni* are highly divergent distinct lineages (Pinto-Juma et al., unpubl.). In terms of isolating mechanisms, the calling song is a prezygotic isolating mechanism that prevents hybridization between species. Since this mechanism was already differentiated before the species came into contact, acoustic character displacement was not necessary to ensure the correct selection by conspecific females. According to Gerhardt (1994), if the signals are already differentiated before the species come into contact, selection would only need to sharpen the selectivity of females, an aspect that should also be studied in these cicadas. *C. barbara* males can discriminate the frequency and temporal characteristics of their song and react preferentially to conspecific songs than to *C. orni* song (Fonseca & Revez, 2002a), but the preferences of females are unknown.

The overlap in song frequencies between the two species could interfere with the communication in these species, particularly, *C. barbara* song could "mask" *C. orni* song since it is continuous and make the temporal characteristics of *C. orni* song more difficult to perceive by receivers. The fact that late in the summer season, *C. orni* is not or rarely found where they occur sympatrically while it remains abundant where it occurs on its own, might be a result of "acoustic competition". Different ecological adaptations could eventually reduce this competition. For instance, different singing positions in the vegetation of cicada species that occur sympatrically is described for *C. orni/Lyristes* (Claridge et al., 1979) and for *C. orni/Tibicina haematodes* (Sueur & Aubin, 2003). Such segregation may also occur between *C. orni* and *C. barbara*, as observed qualitatively by Ribeiro (1998). Boulard (1982) records in Arrábida *C. barbara* singing on the seaward side and *C. orni* on the inland side of the mountain. However, in olive orchards at some sampling sites (e.g. Crato, Portel and Sousel) it was common to see males of both species singing on the same trunks or branches. A study of the micro-habitats occupied by these species (including their singing position on the trees) throughout the summer season (when only *C. orni* is present, when both species are present and when only *C. barbara* is present) is needed to test for ecological adaptations.

In order to exclude any potential effect geographic isolation, only Iberian Peninsula populations were analysed in this study. In fact, as detected previously, populations of these species from other areas differ from the Iberian Peninsula populations (Quartau et al., 1999, Pinto-Juma et al., 2005; Pinto-Juma et al., unpub.). The inter-echeme interval in *C. orni* song is longer in Greek than in Iberian Peninsula populations and the peak frequency is on average higher in Greece (Pinto-Juma et al., 2005). All frequency variables were higher in *C. barbara* from Morocco than from the Iberian Peninsula. It is most likely that these differences in *C. orni* populations are due to geographic distance, since the calling songs in France are similar to those in the Iberian Peninsula. In *C. barbara,* the differences could also be due to geographic isolation between North African and Iberian cicadas. Ecological factors, such as habitat and climatic conditions could be responsible for the direct or indirect selection of certain sound frequencies, specifically the lower frequencies found in both species on the Iberian Peninsula.

In the genus *Cicada* this is the only pair of species known to occur sympatrically. Other species like *C. mordoganensis* Boulard and *C. cretensis* Quartau & Simões, which are present in the Aegean area, have a calling song very similar in temporal and frequency pattern to that of *C. orni*, which is also present in the same region (Simões et al., 2000; Quartau & Simões, 2005), but they are not known to occur sympatrically with *C. orni*. It would be interesting to know if *C. lodosi* Boulard, which has a continuous song like *C. barbara*, but occurs in Turkey (Boulard, 1979), occurs sympatrically with any of the species with discontinuous songs.

The syllable rate increased significantly with temperature in *C. orni*, but only marginally significantly in *C. barbara*. In another study on more populations of *C. barbara* this correlation is significant (Pinto-Juma et al., in prep.). Also, Fonseca (1991) records a temperature dependent syllable period in *C. barbara* and an increase in syllable rate with temperature is described for other cicada species, such as *Tettigeta argentata*, *Tettigeta josei* and *Tympanistalna gastrica* (Fonseca & Revez, 2002b).

In *C. orni* the echeme duration decreased significantly with increasing temperature and there was a very low non-significant correlation between the inter-echeme interval and temperature. In a previous study of one population of *C. orni* at Crato, several recordings per individual at different temperatures showed the same variation with temperature in echeme duration and inter-echeme interval as recorded in this study, though the significance values are marginal for echeme duration and below the critical 0.05 value for inter-echeme interval (echeme duration: $r_S = -0.248$, $p = 0.079$; inter-echeme interval: $r_S = 0.339$, $p = 0.015$) (Quartau et al., 2000). These differences could be due to a sampling effect or other factors. In fact, as reported before by Pinto-Juma et al. (2005), the population at Sousel had longer inter-echeme intervals than other *C. orni* populations, which is not a temperature dependence outcome, since longer intervals would be expected at higher temperatures (Quartau et al., 2000) and Sousel individuals were sampled at temperatures ranging from only 27°C to 30°C. Excluding this population, the results in terms of correla-

tions remained the same (inter-echeme interval positively but not significantly correlated with temperature). The conditions under which *C. orni* population at Sousel were recorded might have been unusual and not detected by the researchers. Nevertheless, Sousel is an area where *C. orni* and *C. barbara* occur sympatrically, and thus, a character displacement might be occurring in which longer silent pauses occur between echemes in the song of *C. orni*, making it more different from the song of *C. barbara*, which has no silent pauses. However, this hypothesis is not consistent with the findings for other areas where these species occur sympatrically. Furthermore, some cicadas from Piedade, where *C. orni* occurs allopatrically have also long inter-echeme intervals compared to other populations (Pinto-Juma et al., 2005). However, this is also inconclusive since the environment temperature was not taken at the time of the recording in this population.

ACKNOWLEDGEMENTS. We are grateful to G. André and P. Simões (Faculdade de Ciências de Lisboa, Portugal) for help in the field. We also thank M. Ribeiro, T. Fernandes and J. Sueur for some of the recordings. J. Sueur gave important contribution in the discussion of results. We thank one referee and the language editor for their important contributions in improving the manuscript. This study was financially supported by Fundação para a Ciência e a Tecnologia (F.C.T., Portugal; PhD grants SFRH/BD/1027/2000 and PRAXIS BD/18229/98).

REFERENCES

Bennet-Clark H.C. & Young D. 1994: The scaling of song frequency in cicadas. *J. Exp. Biol.* **191**: 291–294.

Boulard M. 1979: Cigales du genre Cicada Linne, originaires de Turquie (Hom. Cicadidae). *Türk. Bit. Kor. Derg.* **3**(2): 67–74.

Boulard M. 1982: Les cigales du Portugal, contribution à leur étude (Hom. Cicadidae). *Annls. Soc. Entomol. Fr.* **18**: 181–198.

Boulard M. 1995: Postures de cymbalisation et cartes d'identité acoustique de cigales. 1. Généralités et espéces méditerranéennes (Homoptera, Cicadoidea). *EPHE Biol. Evol. Insects* **7/8**: 14–25.

Claridge M.F., Wilson M.R. & Singhrao J.S. 1979: The songs and calling sites of two European cicadas. *Ecol. Entomol.* **4**: 225–229.

Cocroft R.B. & Pogue M. 1996: Social behavior and communication in the Neotropical cicada Fidicina mannifera (Fabricius) (Homoptera: Cicadidae). *J. Kansas Entomol. Soc.* **69**(4): 85–97.

Cooley J.R. & Marshall D.C. 2001: Sexual signaling in periodical cicadas, Magicicada spp. (Hemiptera: Cicadidae). *Behaviour* **138**: 827–855.

Daws A.G., Hennig R.M. & Young D. 1997: Phonotaxis in the cicadas Cystosoma saundersii and Cyclochila australasiae. *Bioacoustics* **7**: 173–188.

Dobzhansky T. 1951: *Genetics and the Origin of Species. 3rd ed.* Columbia University Press, New York.

Doolan J.M. 1981: Male spacing and the influence of female courtship behaviour in the bladder cicada, Cystosoma saundersii Westwood. *Behav. Ecol. Sociobiol.* **9**: 269–276.

Dytham C. 2003: *Choosing and Using Statistics, a Biologist's Guide. 2nd ed.* Blackwell, Science, Oxford, 248 pp.

Fonseca P.J. 1991: Characteristics of the acoustic signals in nine species of cicadas (Homoptera, Cicadidae). *Bioacoustics* **3**: 173–192.

Fonseca P.J. 1994: *Acoustic Communication in Cicadas (Homoptera; Cicadoidea): Sound Production and Sound Reception.* Thesis, Faculty of Sciences, University of Lisbon, 282 pp.

Fonseca P.J. & Revez M.A. 2002a: Song discrimination by male cicadas Cicada barbara lusitanica (Homoptera, Cicadidae). *J. Exp. Biol.* **205**: 1285–1292.

Fonseca P.J. & Revez M.A. 2002b: Temperature dependence of cicada songs (Homoptera, Cicadoidea). *J. Comp. Physiol. (A)* **187**: 971–976.

Gerhardt H.C. 1994: Selective responsiveness to long-range acoustic signals in insects and anurans. *Am. Zool.* **34**: 706–714.

Marshall D.C. & Cooley J.R. 2000: Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year Magicicada neotredecim. *Evolution* **54**: 1313–1325.

Mayr E. 1948: The bearing of the new systematics on genetical problems. The nature of species. *Adv. Genet.* **2**: 205–237.

Mayr E. 1963: *Animal Species and Evolution.* Harvard University Press, Cambridge, Massachussetts, 797 pp.

McGregor P.K. 1991: The singer and the song: on the receiving end of bird song. *Biol. Rev.* **66**: 57–81.

Paterson H.E.H. 1985: The recognition concept of species. In Vrba E.S. (ed.): *Species and Speciation.* Transvaal Museum Monograph No. 4, Transvaal Museum, Pretoria, pp. 21–29.

Pinto-Juma G., Simões P.C., Seabra S.G. & Quartau J.A. 2005: Calling song structure and geographic variation in Cicada orni Linnaeus (Hemiptera: Cicadidae). *Zool. Stud.* **44**(1): 81–94.

Popov A.V. 1975: The structure of the tymbals and the characteristics of the sound signals in singing cicadas (Homoptera, Cicadidae) in the southern regions of the USSR. *Entomol. Rev.* **54**(2): 7–35.

Puissant S. & Sueur J. 2001: Contribution à l'étude des cigales de Corse (Hemiptera, Cicadidae). *Bull. Soc. Entomol. Fr.* **106**: 429–436.

Quartau J.A. 1988: A numerical taxonomic analysis of interspecific morphological differences in two closely related species of Cicada (Homoptera, Cicadidae) in Portugal. *Great Basin Nat. Mem.* **12**: 171–181.

Quartau J.A. 1995: Cigarras-esses insectos quase desconhecidos. *Correio Natur.* **19**: 33–38.

Quartau J.A. & Fonseca P.J. 1988: An annotated check-list of the species of cicadas known to occur in Portugal (Homoptera: Cicadoidea). In Vidano C. & Arzone A. (eds): *Proceedings of the 6th Auchenorrhyncha Meeting, Turin, Italy, September 7–11, 1987.* Consiglio Nazionale delle Ricerche, Turin, pp. 367–375.

Quartau J.A. & Rebelo M.T. 1994: Sinais acústicos em Cicadidae e Cicadellidae (Homoptera, Auchenorryncha) que ocorrem em Portugal. In Almada V. & Correia R. (eds): *Actas do I Congresso Nacional de Etologia, Lisboa 28, 29 e 30 de Novembro de 1991.* Instituto Superior de Psicologia Aplicada, Lisboa, pp. 137–142.

Quartau J.A. & Simões P.C. 2003: Bioacoustic and morphological differentiation in two allopatric species of the genus Tibicina Amyot (Hemiptera: Cicadoidea) in Portugal. *D. Entomol. Z.* **50**: 113–119.

Quartau J.A. & Simões P.C. 2005: Cicada cretensis sp. n. (Hemiptera, Cicadidae) from southern Greece. *Biologia* **60**: 489–494.

Quartau J.A., Rebelo M.T., Simões P.C., Fernandes T.M., Claridge M.F., Drosopoulos S. & Morgan J.C. 1999: Acoustic signals of populations of Cicada orni L. in Portugal and Greece (Homoptera: Auchenorrhyncha: Cicadomorpha: Cicadidae). *Reichenbachia* **33**(8): 71–80.

QUARTAU J.A., SEABRA S. & SANBORN A. 2000: Effect of ambient air temperature on the calling song of Cicada orni Linnaeus, 1758 (Hemiptera: Cicadidae) in Portugal. *Acta Zool. Cracov.* **43**(3–4): 193–198.

RIBEIRO M. 1998: *Diferenciação em duas espécies gémeas do género Cicada L. (Insecta, Homoptera, Cicadoidea): Cicada orni L. e Cicada barbara (Stal).* Thesis, Faculty of Sciences, University of Lisbon, 70 pp.

SCHEDL W. 1973: Zur Verbreitung, Bionomie und Ökologie der Singzikaden (Homoptera: Auchenorrhyncha, Cicadidae) der Ostalpen und ihrer benachbarten Gebiete. *Ber. Nat. Med. Ver. Innsbruck* **60**: 79–94.

SCHEDL W. 1999: Contribution to the singing cicadas of Israel and adjacent countries (Homoptera, Auchenorrhyncha: Cicadidae et Tibicinidae). *Linzer Biol. Beitr.* **31**: 823–837.

SIMÕES P.C., BOULARD M.M., REBELO M.T., DROSOPOULOS S., CLARIDGE M.F., MORGAN J.C. & QUARTAU J.A. 2000: Differences in the male calling songs of two sibling species of Cicada (Hemiptera: Cicadoidea) in Greece. *Eur. J. Entomol.* **97**: 437–440.

SOKAL R.R. & RHOLF F.J. 1981: *Biometry. The Principles and Practice of Statistics in Biological Research.* W.H. Freeman, New York, 859 pp.

SPECHT R. 2002: *Avisoft-SASLab Pro- Sound analysis and synthesis laboratory.* Version 4.1d, Berlin.

SUEUR J. & AUBIN T. 2003: Is microhabitat segregation between two cicada species (Tibicina haematodes and Cicada orni) due to calling song propagation constraints? *Naturwissenschaften* **90**: 322–326.

SUEUR J. & AUBIN T. 2004: Acoustic signals in cicada courtship behaviour (order Hemiptera, genus Tibicina). *J. Zool. (Lond.)* **262**: 217–224.

VILLET M. 1992: Responses of free-living cicadas (Homoptera: Cicadidae) to broadcasts of cicada songs. *J. Entomol. Soc. Sth. Afr.* **55**(1): 93–97.

VILLET M. 1995: Intraspecific variability in SMRS signals: some causes and implications in acoustic signaling systems. In Lambert D.M. & Spencer H.G. (eds): *Speciation and the Recognition Concept. Theory and Applications.* The John Hopkins University Press, Baltimore and London, pp. 422–439.

YOUNG D. & JOSEPHSON R.K. 1983: Pure-tone songs in cicadas with special reference to the genus Magicicada. *J. Comp. Physiol.* **152**: 197–207.

Received January 19, 2006; revised and accepted May 10, 2006