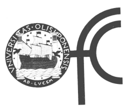
UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

PATTERNS OF DIVERGENCE IN POPULATIONS OF TWO MEDITERRANEAN SPECIES OF GENUS *CICADA* L. (HEMIPTERA, CICADIDAE) BASED ON MICROSATELLITE GENETIC MARKERS AND ACOUSTIC DATA

Sofia Gonçalves Seabra

Tese orientada por:
Professor Doutor José Alberto Quartau e Professor Doutor Michael W. Bruford

DOUTORAMENTO EM BIOLOGIA
(BIOLOGIA EVOLUTIVA)
2007



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

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DO ENSINO SUPERIOR Portugal

Financial support was provided by Fundação para a Ciência e a Tecnologia, Portugal (III Quadro Comunitário de Apoio co-financed by Fundo Social Europeu and by national funds from MCES; PhD grant SFRH/BD/1027/2000).

À minha mãe

Ao Tiago e à Inês

Carlos Picado, 2005

Nota Prévia

Nos termos do disposto no n.º 1 do Artigo 40, Capítulo V da Deliberação n.º 961/2003 do Regulamento de Estudo Pós-Graduados da Universidade de Lisboa, publicado no Diário da República - II Série, n.º 153, de 5 de Julho de 2003, a presente tese foi elaborada com o aproveitamento do resultado de alguns trabalhos já publicados e realizados em colaboração com outros investigadores, tendo eu participado plenamente em todas as fases do trabalho: planeamento, obtenção, análise crítica e discussão dos resultados bem como na redacção. Esclareço que dois dos co-autores dos artigos científicos resultantes desta tese, José A. Quartau e Michael W. Bruford, são meus orientadores. Helen R. Wilcock, foi quem me introduziu às técnicas laboratoriais e me guiou no processo de isolamento dos microssatélites. Na análise acústica colaborei com duas alunas de doutoramento que preparam as suas teses tendo como alvo de estudo o mesmo género *Cicada*: Gabriela Pinto-Juma estuda a filogenia e a biogeografia do género *Cicada* na área Mediterrânica e Paula Simões realiza trabalho sobre a divergência acústica e a biogeografia insular de espécies de *Cicada* existentes na área do Mar Egeu.

Os trabalhos já publicados ou aceites para publicação no âmbito desta tese são:

Seabra SG, Wilcock HR, Quartau JA & Bruford MW. 2002. Microsatellite loci isolated from the Mediterranean species *Cicada barbara* (Stål) and *C. orni* L. (Hemiptera, Cicadoidea). *Molecular Ecology Notes*, 2: 173–175.

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

Seabra SG, Pinto-Juma G & 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.

Pinto-Juma GA, **Seabra SG** & Quartau JA. 2007. Patterns of acoustic variation in *Cicada barbara* Stål (Hemiptera, Cicadoidea) from the Iberian Peninsula and Morocco. *Bulletin of Entomological Research* (accepted).

Os trabalhos em fase de preparação para submissão são:

Seabra SG, Quartau JA & Bruford MW. Microsatellite variation in Mediterranean *Cicada* species (Hemiptera, Cicadoidea): high frequency null alleles and their impact on population genetic data. (Manuscript).

Seabra SG, Quartau JA & Bruford MW. Spatio-temporal genetic variation in sympatric and allopatric populations of *Cicada barbara* and *C. orni* (Hemiptera: Cicadoidea). (Manuscript).

Seabra SG & Quartau JA. Temporal variation of the acoustic characters of the calling songs of *Cicada barbara* and *C. orni* (Hemiptera: Cicadidae) at the individual and population levels. (Manuscript).

11 de Abril de 2007

Sofia Gonçalves Seabra

Agradecimentos / Acknowledgements

Gostaria de agradecer a todos os que contribuíram para que a realização deste trabalho fosse possível ou que tornaram de uma maneira ou de outra este percurso menos penoso.

Ao Professor Doutor José Alberto Quartau por me ter sugerido o tema da tese e pelo apoio no planeamento e realização deste trabalho. Por me ter desperto para o fascinante mundo dos insectos ainda na sua cadeira de Entomologia e por partilhar os seus conhecimento sobre as cigarras e sobre os aspectos evolutivos dos insectos. Foram muito importantes também o seu entusiasmo e incentivo no decorrer do trabalho, bem como as suas críticas pertinentes.

To Professor Michael Bruford for accepting to be my supervisor and giving me access to the Genetics lab in Cardiff School of Biosciences, where I have learnt so much and where I found wonderful colleagues and friends. I am grateful to Mike for allowing me this important experience in my scientific life and for his guidance during the course of this work.

Ao financiamento da Fundação para a Ciência e a Tecnologia por ter permitido a realização do trabalho (bolsa SFRH/BD/1027/2000).

A todos os que colaboraram e me ajudaram no trabalho de campo: Genage André, Gabriela Juma, Paula Simões, Jérôme Sueur, Stéphane Puissant, Ruben André, Imo Juma, Diana Marques, Tiago Marques, Veríssimo Dias, Maria João Gonçalves e Tiago Picado. À Paula Simões pelas cigarras da Grécia. Às pessoas que contribuiram em anos anteriores para a colecção de cigarras do grupo: Mónica Ribeiro, Teresa Fernandes, Teresa Rebelo. À Teresa Fernandes agradeço ter-me acompanhado na minha primeira saída de campo das cigarras.

À Gabriela Juma e à Paula Simões por terem sido companheiras sempre presentes nesta viagem e pela colaboração em muitas das fases do trabalho.

To everyone that helped me in the laboratory work and genetic analysis at Cardiff School of Biosciences: Helen Wilcock, Carlos Fernandes, Elisabete Pires, Evren Koban, Kathryn Jeffery, Patrícia Faria, Trinidad Perez and Yoshan Moodley.

To Helen Wilcock that I met in the lab at Cardiff and that was a wonderful colleague and friend. Tragically she left us too soon. She was a crucial help in the beginning of my lab work and I owe very much to her patience and helpfulness. She will always be in my memory as an example of a very professional researcher and an intelligent and wise person.

À Jérôme Sueur pour avoir partagé ses connaissances sur l'acoustique et les cigales, pour son aide au travail de terrain et pour avoir capturé et enregistré quelques cigales utilisées dans cet étude. Aussi pour m'avoir permit l'accès à beaucoup de bibliographie au Muséum national d'Histoire naturelle, Paris. Et pour avoir discuté les résultats acoustiques. À lui et à Stéphane Puissant je veux remercier leur gentillesse et amitié et aussi leur accueil à Perpignan (Languedoc-Roussillon, France) pour le travail de terrain. J'ai bien appréciée leur enthousiasme avec l'étude des cigales et leur connaissance des espèces et du terrain. Je veux aussi remercier à Stéphane et à Nadine Fille leur hospitalité chez eux.

Ao financiamento do Programa de Cooperação Luso-Francês 2001 do ICCTI & Embaixada Francesa em Lisboa (projecto n.º 537 C1) que permitiu a execução de trabalho de campo no Sul de França em 2001.

À Professeur Michel Boulard pour avoir partagé ses connaissances sur les espèces de cigales.

À David Ouvrard, Laurent Fauvre et à Prof. Jean-Marc Thibaud pour leur hospitalité au Muséum national d'Histoire naturelle, Paris.

To Alex Wilson, Paul Sunnucks and Chris Simon for giving me access to the unpublished primer sequences they developed for North-American cicadas.

Ao Tiago Marques pela ajuda em partes da análise estatística, e por ter realizado a regressão logística no R.

À Paula Simões, ao Octávio Paulo, ao José Pedro Amaral e à Sofia Morais pela leitura cuidadosa dos manuscriptos.

To the anonymous referees that commented on the submitted papers.

To all the colleagues and friends in Cardiff that made my stay there more pleasant and really special: Evren Koban, Trinidad Perez, Elisabete Pires, Gabriela Juma (and her sweet baby girl Madalena), Elena Mavraki, Carlos Fernandes, Helen Wilcock, Simon Shayler, Patrícia Faria, Erika Baus, Rhys Jones, Benoît Goossens, Yoshan Moodley, Sheena James, Fairus Jalil, Mireille Johnson, Nuria Agusti, Lounès Chikhi, Nicky Anthony, Jo Cable, Ciara Dodds, Ceri Williams, Mererid Howells, Kathryn Jeffery, Angela Formia, Noel, Janice Blumenthal, Iris Pereira, Simon Rees, Katherine Dunbar, Michele Drès, Massoud Madjdzadeh, Steve Casey, Georgina Harper, Amy, Nicky, Jorge, Tiago, Carlos, Adelaida, Vera, Barry, Marco, Lucile… It was really a wonderful experience to meet people from all around the world and to make some true friendships.

To Helen Dawah who always welcomed me at her home in Cardiff and gave me so much support and friendship. Her cosy home with her lovely cats (first Morgan and then Frankie) and her beautiful garden were really enjoyable. My thanks also to Hassan Dawah for his kindness.

À Celeste Jacinto por nos ter recebido tão calorosamente nas visitas a Birmingham.

Aos meus colegas e amigos em Lisboa que me acompanharam e com quem partilhei as minhas venturas e desventuras neste percurso, agradeço o apoio e em muitos dos casos a empatia: Filipe Ribeiro, Sofia Morais, Sheila Dias, Maria João Magalhães, Rodrigo Almeida, Luísa Chaves, Tiago Marques, Ana Filipa Filipe, Tânia Nobre, Patrícia Tiago, Carla Rego, Mário Boieiro, Anabela, Sónia, Israel, Nuno, Marta Silva, Patrícia Salgueiro, Carina, Cristiane Silveira, Cristina Luís, Natacha Mesquita, Ana Luísa Rego, Luísa Camacho, Luísa Branquinho, Cecília Rodrigues, Luís Costa, Nuno Costa, Luís Teixeira, Ana Teresa, Luzia, Sónia, Ana...

À Leonor e à D. Lurdes da biblioteca de Biologia Animal pelo profissionalismo e simpatia. À Leonor agradeço também a sua amizade ao longo destes anos.

À Rosinha do Crato que foi uma boa amiga nos tempos que passei naquela simpática vila alentejana.

Ao meu tio Wolfgang Lind por ter traduzido do alemão alguns artigos sobre a distribuição das cigarras.

Ao meu sogro Carlos Picado, pela ilustração da cigarra; a ele e à Gracelina Leitão agradeço o incentivo para terminar a tese.

À Paula Simões por me ter "apresentado" às cigarras ainda no laboratório. Pelo seu apoio ao longo destes anos e por partilhar comigo as experiências tanto de trabalho como de maternidade!

À Gabriela Juma por me ter acompanhado em muitas das fases do trabalho, tanto em Cardiff como em Lisboa e nalgumas viagens de trabalho de campo. Pela sua amizade e generosidade. Tive e tenho a felicidade de acompanhar ao longo destes anos o nascimento e crescimento da sua doce Madalena. Recordarei sempre o Imo pela sua simpatia e grande humanidade, e pelo seu empenho e paciência na viagem a Marrocos.

Ao Genage André pela sua imprescindível ajuda no trabalho de campo, desde o início. Pelos seus ensinamentos das técnicas de captura de cigarras, pelo seu empenho, pela sua boa disposição e palavras de incentivo e de esperança. Mas principalmente por ser um bom amigo.

Ao Filipe Ribeiro pela amizade e boa disposição e pelo pragmatismo e desembaraço com que enfrenta as situações.

À Sofia Morais pelas palavras amigas de incentivo mesmo quando as coisas corriam menos bem. Apesar de longe (às vezes muito longe, nestes últimos anos!) sinto sempre que estamos bem perto.

Às aulas de yoga por me terem permitido manter a mente e o corpo mais ou menos sãos nesta última fase do trabalho.

À minha família que sempre me deu todo o apoio e incentivo. Às minhas queridas avós, ao meu irmão, à minha cunhada, aos meus tio(a)s e primo(a)s. Ao meu pai que eu gostaria que estivesse mais perto. Ao meu avô Amadeu, com muita saudade.

À minha sogra Mariana Garrido pelo carinho e amizade e pelo apoio constantes ao longo destes anos; por me ter recebido no seu acolhedor apartamento em Paris na minha ida ao Museu.

À minha mãe, a quem dedico esta tese e a quem agradeço por estar sempre presente. Inclusivamente para apanhar cigarras no campo!

Ao Tiago por existir. E por me ter ajudado na produção desta tese, mas principalmente na produção de algo muito precioso, a nossa Inês. Também aos dois dedico esta tese.

"Minha filha tem um adormecer custoso. Ninguém sabe os medos que o sono acorda nela. Cada noite sou chamado pai e invento-lhe um embalo. Desse encargo me saio sempre mal. Já vou pontuando fim na história quando ela me pede mais: "E depois?" O que a Rita quer é que o mundo inteiro seja adormecido. E ela sempre argumenta um sonho de encontro ao sono: quer ser lua. A menina quer luarejar e, os dois, faz contarmo-nos assim, eu terra, ela lua."

Mia Couto, *Contos do Nascer da Terra* (1997)

Abstract

Cicada barbara (Stål) and *C. orni* L. are two Mediterranean cicada species, very similar in morphology, that produce distinct acoustic mating signals and that have overlapping distribution ranges in the Iberian Peninsula, occurring in sympatry in several locations. In this study six microsatellite loci were isolated and applied in *C. barbara*, four of which were also applied in *C. orni*. For some loci high frequencies of null alleles were found and data were analysed taking into account their presence. Male calling song characteristics were also analysed and included frequency and fine-temporal variables in both species, as well as gross-temporal variables in *C. orni*. Using both genetic and acoustic data, the geographical and temporal variation in these species was studied. No evidence of hybridization using either marker was found, enabling us to infer that the isolating barriers between these species are efficient. Also, comparing sympatric and allopatric populations, no substantial evidence of "character displacement" of the calling songs was found. Partitioning of geographic variation, both genetic and acoustic, in each species, revealed the main following patterns: Iberian Peninsula and Northwestern Africa populations of *C. barbara* showed higher differentiation between than within each region, supporting the *C. barbara* subspecific divisions (*C. barbara lusitanica* in Iberian Peninsula and *C. barbara barbara* in Northwestern Africa) and highlighting the isolation coincident with the presence of physical barriers to gene-flow; differentiation between populations of *C. orni* from both sides the Pyrenees was very low, and probably this mountain range does not constitute a significant barrier for the dispersal of this species; Greek populations of *C. orni* were found to be highly differentiated from Western European populations at both levels; within the Iberian Peninsula, *C. orni* showed a pattern of isolation-by-distance which *C. barbara* did not. No evidence of reproductively isolated broods was found in either genetic or acoustic data.

Key-words: *Cicada*; microsatellites; acoustics; geographic variation; temporal variation

Sumário

Cicada barbara (Stål) e *C. orni* L. são duas espécies de cigarras mediterrânicas, muito semelhantes morfologicamente, que produzem sinais acústicos de acasalamento distintos, e cujas áreas de distribuição se sobrepõem na Península Ibérica, existindo em simpatria nalguns locais. Neste trabalho seis *loci* microssatélites foram isolados e amplificados em *C. barbara*, quatro dos quais foram também amplificados em *C. orni*. Nalguns *loci* detectaram-se frequências de alelos nulos elevadas e os dados foram analisados tendo em conta a sua presença. Foram também analisadas características do som de chamamento dos machos, incluindo variáveis de frequência e de tempo. Os dados genéticos e acústicos foram utilizados para estudar a variação geográfica e temporal nestas espécies. Nenhum dos marcadores demonstrou evidência de hibridação, permitindo inferir que as barreiras de isolamento entre espécies são eficientes. Adicionalmente, não foi encontrada evidência substancial de "desvio de caracteres" no sinal acústico de acasalamento, comparando populações simpátricas e alopátricas. A caracterização da variação geográfica, tanto a nível genético como acústico, em cada espécie, revelou os seguintes padrões: maior diferenciação entre as populações de *C. barbara* da Península Ibérica e do Noroeste de África do que dentro de cada região, apoiando as divisões subespecíficas (*C. barbara lusitanica* na Península Ibérica e *C. barbara barbara* no Noroeste de África) e realçando o isolamento coincidente com a presença de barreiras físicas ao fluxo genético; por outro lado, a diferenciação entre as populações de *C. orni* dos dois lados dos Pirinéus revelou-se muito reduzida, provavelmente por esta cadeia montanhosa não constituir uma barreira significativa para a dispersão desta espécie; as populações de *C. orni* da Grécia revelaram-se substancialmente diferenciadas das populações da Europa Ocidental a ambos os níveis; na Península Ibérica, *C. orni* revelou um padrão de isolamento por distância, o que não aconteceu com *C. barbara*. Dos dados genéticos e acústicos não houve evidência de existência de isolamento reprodutor entre diferentes anos de emergência das cigarras.

Palavras-chave: *Cicada*; microssatélites; acústica; variação geográfica; variação temporal

Resumo

Padrões de divergência em populações de duas espécies mediterrânicas do género *Cicada* L. (Hemiptera, Cicadidae) baseados em marcadores genéticos de microssatélites e em dados acústicos.

Cicada barbara e *C. orni* são duas espécies-gémeas (semelhantes em termos de morfologia) de cigarras cujos machos produzem sinais acústicos de acasalamento (sons de chamamento) distintos. Os sons, de elevada intensidade, são produzidos por acção dos tímbalos, estruturas membranosas existentes dorsolateralmente no primeiro segmento abdominal. Os machos adultos de ambas as espécies formam agregados, cantando em simultâneo em grandes coros, permanecendo cada macho muitas vezes na mesma posição de canto durante várias horas seguidas. *C. barbara* produz um sinal de chamamento contínuo, composto por uma sequência de pulsos de som. *C. orni* possui um canto descontínuo, produzindo um série repetitiva de elementos acústicos (os equemas), compostos pelos pulsos de sons, alternados com intervalos de silêncio. *C. barbara* encontra-se distribuída no Norte de África, em algumas ilhas mediterrâneas e na Península Ibérica. Duas subspécies encontram-se descritas com base na morfologia, *C. barbara barbara* (a partir de material da Tunísia) e *C. barbara lusitanica* (a partir de material de Portugal). *C. orni* é uma das cigarras mais abundantes e comuns no Sul e Centro da Europa (incluindo nas Penínsulas Ibérica, Itálica e Balcânica), estando também descrita para a Ásia ocidental e Médio Oriente. A Península Ibérica constitui uma zona de sobreposição das áreas de distribuição das duas espécies, havendo alguns locais de simpatria onde machos de ambas as espécies emitem o som de chamamento, muitas vezes simultaneamente a partir dos mesmos troncos ou ramos das árvores. Existe um desvio sazonal na presença de adultos entre ambas espécies com *C. orni* emergindo mais cedo (Junho) do que *C. barbara* (Julho/Agosto) nos dois tipos de áreas, e também entre situações de alopatria e de simpatria. Em áreas de alopatria os adultos de *C. orni* desaparecem em Setembro/Outubro e, quando em simpatria com *C. barbara*, desaparecem muito mais cedo (Agosto). Os adultos de *C. barbara* aparecem um pouco mais cedo em alopatria do que em simpatria e desaparecem em Setembro/Outubro em ambas as áreas. Estas áreas de simpatria são ideais para testar a eficiência das barreiras de isolamento/sistema de reconhecimento das espécies. Na presente tese foram usados marcadores genéticos altamente variáveis, os microssatélites, com o objectivo de detectar hibridação e também com o objectivo de estudar a variação geográfica e temporal (entre diferentes anos) nestas espécies. Adicionalmente, os sons de chamamento dos machos foram analisados com os mesmo objectivos, procurando sinais acústicos de características intermédias (esperados em híbridos de insectos acústicos), e também descrevendo a variação acústica a nível geográfico e temporal. Foram comparadas áreas de simpatria e áreas de alopatria na Península Ibérica de modo a testar a eficiência das barreiras reprodutoras, assim como a existência de “desvio de caracteres” no sinal de chamamento nas áreas de simpatria.

Dado que não existiam *loci* microssatélites isolados para estas espécies nem para espécies próximas, procedeu-se à construção de uma biblioteca genómica enriquecida para a presença de microssatélites, tendo sido isolados seis *loci* para *C. barbara*, quatro dos quais também amplificaram produtos polimórficos em *C. orni* (tendo os restantes dois *loci* apresentado apenas produtos monomórficos). Na análise de microssatélites foram também utilizadas duas populações de ilhas do Mar Egeu (parte leste do Mediterrâneo) de outras duas espécies do mesmo género, *C. cretensis* e *C. mordoganensis*. Verificou-se uma elevada frequência de alelos nulos (alelos que não amplificam durante a PCR devido a mutações no local de emparelhamento dos *primers*) nalguns *loci*, os quais poderão ser causadores de problemas na intepretação dos resultados de estrutura genética das populações. Foi realizada uma abordagem estatística ao problema, utilizando um ajustamento dos dados tendo em conta a presença de alelos nulos. Os resultados em termos de variabilidade e diferenciação genéticas, para cada um dos *loci*, baseados quer nos dados originais quer nos dados ajustados foram então comparados. Os resultados de variabilidade genética (tanto no número de alelelos como na heterozigotia esperada) foram muito influenciados pela presença de alelos nulos, o que seria de esperar para o número de alelos mas não necessariamente para a heterozigotia esperada (H_e). Por outro lado, a diferenciação genética, estimada com base em F_{ST}, foi muito similar nas comparações entre dados originais e corrigidos para a presença de alelos nulos. Adicionalmente, uma análise considerando os genótipos *multi-locus*, utilizando todos os *loci* ou apenas aqueles sem evidência de alelos nulos, permitiu verficar que os resultados quer em termos de variabilidade relativa quer em termos de padrões de diferenciação entre populações ou regiões geográficas foram semelhantes para as duas séries de dados.

A gravação dos sons de chamamento dos machos das duas espécies foi realizada utilizando um gravador digital e um microfone uni-direccional, tendo sido medida a temperatura ambiente aquando de cada gravação. Os sons foram analisados em software apropriado em termos de frequência e de tempo. Em ambas as espécies foram utilizadas as seguintes variáveis: frequência de máxima amplitude e taxa de produção de sílabas (sendo uma sílaba um conjunto de pulsos de som produzidos por um dos tímbalos). Para *C. orni* foram analisadas adicionalmente as seguintes variáveis: duração dos equemas, duração dos intervalos entre equemas, taxa de produção de equemas, período dos equemas e razão duração do equema/duração do intervalo entre equemas.

Os principais resultados obtidos foram:

i) *Cicada barbara* e *C. orni* revelaram alelos e frequências alélicas distintos que permitiram separar as duas espécies numa análise multivariada (Análise Factorial de Correspondências). No entanto, é possível que haja hibridação ocasional dado que dois indivíduos (um de cada espécie) de áreas simpátricas possuíam um alelo típico da outra espécie (no *locus* Cib03). A existência de alelos nulos poderá também ter impedido que se revelassem alelos resultantes de hibridação. As espécies divergiram em todas as variáveis acústicas analisadas, não tendo sido encontrados híbridos com propriedades acústicas intermédias. Deste modo, não foram encontradas evidências sólidas,

genéticas nem acústicas, de hibridação entre *Cicada barbara* e *C. orni*, pelo que se pode concluir que as barreiras de isolamento são eficazes.

ii) Na Península Ibérica não houve diferenças genéticas nem acústicas entre populações simpátricas e alopátricas. Uma excepção ocorreu numa população simpátrica de *C. orni* (Sousel) que apresentou divergência numa variável acústica em relação às restantes populações. Este padrão poderá favorecer a hipótese de desvio de caracteres por interacção com outra espécie.

iii) A divisão de *C. barbara* em duas subespécies, uma presente no Noroeste de África e outra na Península Ibérica, é apoiada pela maior diferenciação encontrada entre subespécies do que dentro de cada subspécie, quer em termos genéticos quer em termos acústicos.

iv) As populações da Península Ibérica e do Sul de França apresentaram fraca diferenciação entre si e dentro de cada região. Por outro lado, as populações de ambas as regiões evidenciaram elevada divergência genética e acústica em relação às populações gregas.

v) Algumas populações exibiram diferenças significativas nas variáveis acústicas em relação às outras da mesma região, o que poderá reflectir condições ambientais ou competitivas divergentes.

vi) Tanto *C. barbara* como *C. orni* apresentaram baixos valores de diferenciação genética entre populações na Península Ibérica. Um padrão de isolamento por distância foi observado em *C. orni* mas não em *C. barbara*.

vii) A diferenciação genética e acústica entre diferentes anos de amostragem foi em geral muito reduzida, não revelando nenhum indício de isolamento reprodutor entre diferentes anos de emergência das cigarras.

viii) As características acústicas analisadas apresentaram valores de variabilidade muito diferentes entre si. A frequência de máxima amplitude e a taxa de produção de sílabas apresentaram variabilidades baixas em ambas as espécies, enquanto as características temporais de *C. orni* duração dos equemas e intervalo entre equemas apresentaram grande variabilidade intra- e inter-individual. Adicionalmente, a duração dos equemas foi a única característica que não variou significativamente entre regiões geográficas.

Contents

List of tables xxv

List of figures xxvii

CHAPTER 1 – GENERAL INTRODUCTION 1

1.1. SPECIES AND SPECIATION 3

1.1.1. Acoustic mating signals 5

1.2. GENERAL OVERVIEW OF CICADA BIOLOGY, ECOLOGY AND BEHAVIOUR 7

1.2.1. Genus Cicada L. 1758 9

1.2.1.1. Reproductive isolation between Cicada barbara and C. orni 12

1.3. OBJECTIVES AND STRUCTURE OF THE THESIS 15

CHAPTER 2 – GENETIC ANALYSIS 17

2.1. INTRODUCTION TO MICROSATELLITES 19

2.1.1. CHARACTERISTICS, APPLICATIONS AND LIMITATIONS OF MICROSATELLITES 19

2.1.2. MICROSATELLITE ISOLATION AND GENOTYPING 21

2.1.3. MICROSATELLITE DATA ANALYSIS 23

2.1.3.1. Hardy-Weinberg equilibrium 23

2.1.3.2. Null alleles 24

2.1.3.3. Population genetic analysis 25

2.2. ISOLATION OF MICROSATELLITES 27

2.2.1. PREPARING DNA FOR ENRICHMENT 27

2.2.2. ENRICHMENT FOR MICROSATELLITE REPEATS 28

2.2.3. CLONING 28

2.2.4. SCREENING FOR COLONIES WITH MICROSATELLITES 28

2.2.5. PRIMER DESIGN AND TESTING 28

2.3. MICROSATELLITE ANALYSIS 30

2.3.1. MATERIAL AND METHODS 30

2.3.1.1. Sampling sites 30

2.3.1.2. Sample processing 30

2.3.1.3. Data analysis 33

2.3.1.3.a) Single-locus analysis 33

2.3.1.3.b) Multi-locus analysis 35

2.3.2. Results **38**
a) Single-locus analysis 38
(i) *Hardy-Weinberg analysis* 38
(ii) *Molecular characteristics* 41
(iii) *Microsatellite variability* 43
(iv) *Population differentiation* 43
b) Multi-locus analysis 46
i) Microsatellite variability 46
ii) Population structure 49
2.3.3. Discussion **56**
a) Single-locus analysis 56
b) Multi-locus analysis 59

CHAPTER 3 – ACOUSTIC ANALYSIS **65**

3.1. INTRODUCTION TO THE ACOUSTIC ANALYSIS **67**
3.1.1. Acoustic communication – sound production, transmission and reception in cicadas 67
3.1.2. Sound recording, processing and analysis 69
3.2. MATERIAL AND METHODS **71**
3.2.1. Field procedures and sound recording 71
3.2.2. Sound analysis 74
3.2.3. Statistical analysis 78
3.2.3.1. Geographic variation in the calling song of C. barbara and C. orni 78
3.2.3.2. Amplitude modulated signal in C. barbara 79
3.2.3.3. Comparisons between species in allopatry and sympatry 79
3.2.3.4. Temporal variability at the individual and population level 79
3.3. RESULTS **80**
3.3.1. Description of the calling songs of *Cicada barbara* and *C. orni* 80
3.3.2. Geographic variation of the calling song in *Cicada barbara* 83
3.3.3. Amplitude modulated signal in *Cicada barbara* 84
3.3.4 Geographic variation of the calling song in *Cicada orni* 86
3.3.5. Comparisons between species 90
3.3.6. Individual variation 91
3.3.7. Temporal variation at the population level 96

3.4. DISCUSSION......97
3.4.1. TEMPERATURE EFFECT 97
3.4.2. GEOGRAPHIC VARIATION OF THE CALLING SONG AND OF THE AMPLITUDE MODULATED SIGNAL OF *CICADA BARBARA*......98
3.4.3. GEOGRAPHIC VARIATION OF THE CALLING SONG OF *CICADA ORNI* 99
3.4.4. CALLING SONGS IN SYMPATRIC AND ALLOPATRIC POPULATIONS OF *CICADA BARBARA* AND *C. ORNI* IN THE IBERIAN PENINSULA......101
3.4.5. TEMPORAL VARIATION OF THE CALLING SONG CHARACTERS OF *CICADA BARBARA* AND *C. ORNI* AT THE INDIVIDUAL AND POPULATION LEVELS......103

CHAPTER 4 - GENERAL DISCUSSION 105

4.1. FINAL REMARKS 110

CHAPTER 5 - REFERENCES 113

APPENDICES 135

Appendix I – Values of F_{IS} calculated for each locus, population and dataset. 137

Appendix II – Number of alleles and heterozygosity in each population per locus per dataset. 139

Appendix III – F_{ST} average values ± standard deviations for each locus and each dataset calculated between *C. barbara* populations and between *C. orni* populations – for all populations, within each region and among regions. 141

Appendix IV – Multilocus F_{ST} values (loci in Hardy-Weinberg equilibrium) for each pair of populations of *C. barbara* (a) and of *C. orni*, *C. cretensis* and *C. mordoganensis* (b)...... 143

Appendix V – Total body length of cicadas (from tip of head to tip of the right wing) from Iberian Peninsula (N=248), Ceuta (N=25) and Fès (Morocco) (N=47)...... 145

List of tables

Page

Table 2.1. Characterization of *Cicada* L. microsatellite loci. The motif, primer sequences, annealing temperatures (T_a) and $MgCl_2$ concentrations are given for each locus. 29

Table 2.2. Number of cicadas in each sampled population of *Cicada barbara*, *C. orni*, *C. cretensis* and *C. mordoganensis*. Percentage of samples of each population stored with each type of method, of each type of tissue, and extracted with each type of extraction method. 32

Table 2.3. Values of F_{IS} calculated for each locus/population combination. Shaded values are statistically significant: *p<0.05; **p<0.01; ***p<0.001. N – evidence of null alleles detected by the program MICRO-CHECKER; S – evidence of stuttering detected by the same program. 39

Table 2.4. Estimated coefficients β and significance values of the logistic regression models of the binary variable amplification success/failure of the locus onto the variables type of storage, type of tissue, type of extraction and species (coded as dummy variables). Significance levels: *p<0.05; **p<0.01; ***p<0.001. 40

Table 2.5. Estimated null allele frequencies using the Brookfield 2 algorithm in MICRO-CHECKER (Null) and frequencies of the most common allele in the original data (MCA). 42

Table 2.6. P-values of nonparametric tests comparing the measures of variability (number of alleles – Na, and expected heterozygosity – H_E) among species (Cb – *C. barbara* and Co – *C. orni*) and among regions for each locus for each dataset (Mann-Whitney tests for two-sample comparisons and Kruskal-Wallis tests for multiple-sample comparisons). 44

Table 2.7. Percentage of variation among regions (Groups), among populations within regions (A-pop) and within populations (W-pop) obtained from the AMOVA for each dataset (ORIG, MIN. MED and MAX). For *C. barbara* two groups were defined (Iberian Peninsula and North Africa) and for *C. orni* two separate AMOVA were carried out, one comparing Iberian Peninsula and France and another comparing Western Europe (Iberian Peninsula and France populations) and Greece. 45

Table 2.8. Average number of alleles (bootstrapped) and average expected heterozygosity (He) across loci for each population of *Cicada barbara*, *C. orni, C. cretensis* and *C. mordoganensis* using all loci and using only the loci in Hardy-Weinberg equilibrium (for *C. barbara*: loci Cib01, Cib06 and Cio08; and for the other species: loci Cib03 and Cib07). * populations not included in the bootstrapping; presented values for these populations are not bootstrapped. 47

Table 2.9. Percentage of variation in AMOVA found among populations (Pop.), among years within populations (Years) and within years in *C. barbara* (Crato, Portel and Sousel), in *C. orni* from Iberian Peninsula (Portel and Algeciras) and in *C. orni* from France (Narbonne and St. Hippolyte) for single-locus and multilocus analysis. 54

Table 3.1. Number of males (N) of *Cicada barbara* (Cb) and *C. orni* (Co) recorded for the calling songs, with range of ambient temperatures at recording (T) and indication of populations in allopatry (A) or sympatry (S) relative to the congeneric species. 73

Table 3.2. Number of recordings (N) of the calling song of males of *C. barbara* (Cb) and *C. orni* (Co) obtained at different times of day and temperatures in the locality of Crato in 1999. 75

Table 3.3. Number of males (N) of *C. barbara* (Cb) and *C. orni* (Co) recorded in each locality in each year and range of temperatures at recording (T). 76

Table 3.4. Descriptive statistics of the acoustic variables of the calling song of *Cicada barbara*: N – number of individuals analysed; SD – standard deviation; CV – coefficient of variation; at the bottom of the table between parenthesis is the range (minimum–maximum). For abbreviations of population names see Table 3.1. 81

Table 3.5. Descriptive statistics of the acoustic variables of the calling song of *C. orni*. N – number of individuals analysed; SD – standard deviation; Range – minimum and maximum; CV – coefficient of variation. 82

Page

Table 3.6. Descriptive statistics for the acoustic variables measured in the phrases of the amplitude modulated signal of *Cicada barbara barbara* (North Africa) and *C. barbara lusitanica* (Iberian). SD – standard deviation; CV – coefficient of variation. 86

Table 3.7. Correlations (factor loadings) between each variable and each component extracted from the PCA carried out in the matrix of the six acoustic variables (syllable rate excluded) for each *C. orni* individual. 89

Table 3.8. Descriptive statistics of the acoustic variables in *C. barbara* and *C. orni* from the Iberian Peninsula. N – number of individuals analysed; SD – standard deviation; Range – minimum and maximum; CV – coefficient of variation. 90

Table 3.9. Spearman correlation coefficients of each acoustic variable with the ambient temperature for each individual cicada and for the total of the observations of *Cicada barbara* (a) and of *C. orni* (b) from Crato in 1999. Partial correlations (parametric) of the acoustic variables and temperature controlled for hour of day are also given. *Significant after Dunn-Sidák correction. 94

Table 3.10. P-values obtained from nonparametric tests (two-samples – Mann-Whitney test, and more than two samples – Kruskal-Wallis test) comparing the years of sampling within each locality and comparing the localities for each species, *C. barbara* (a) and *C. orni* (b). ANCOVA results are also given, with covariates being temperature and hour of day. *Significant after Dunn-Sidák correction. 96

List of figures

Page

Figure 1.1. a) Lateral view of a male cicada with location of the tymbal and tympanal organs; b) Schematic cross section at base of abdomen (Adapted from McGavin, 2001); c) Lateral view of the left tymbal showing the sclerotized elements (gray) and the unsclerotized membrane containing resilin (white) (Adapted from Young & Bennet-Clark, 1995). 8

Figure 2.1. Sampled populations of each species (● – allopatric populations of *Cicada barbara*; ○ – allopatric populations of *C. orni*; ▲ – sympatric populations of *C. barbara* and *C. orni*; ■ – population of *C. cretensis*; □ – population of *C. mordoganensis*). 31

Figure 2.2. Allele frequencies observed in *C. barbara* (Cb), *C. orni* (Co), *C. cretensis* (Cc) and *C. mordoganensis* (Cm) for each microsatellite locus. 48

Figure 2.3. Scatterplots of $F_{ST}/(1-F_{ST})$ and natural logarithm of the geographic distance among *C. barbara* populations from Iberian Peninsula and North-West Africa (a) and among *C. orni* populations from Iberian Peninsula and Southwest France (b). 50

Figure 2.4. Correspondence Factorial Analysis of individual cicadas *C. barbara* from the Iberian Peninsula and North Africa (Ceuta and Morocco). 51

Figure 2.5. Correspondence Factorial Analysis of individual cicadas *C. orni* (from the Iberian Peninsula, France, and from the Greek populations of Athens, Naxos and Skyros) , *C. cretensis* and *C. mordoganensis*. In a) all the individuals are shown and in b) the Iberian and French individuals were excluded from the scatterplot. 52

Figure 2.6. Estimated membership coefficients (q) for each individual *C. barbara* (each single vertical line), in each cluster (K=3). The numbers (1 to 19) correspond to the predefined populations. 53

Figure 2.7. Correspondence Factorial Analysis of individual *C. barbara* and *C. orni* from the Iberian Peninsula from sympatric and allopatric populations. The analysis was performed with both species and the scatterplot of the individual scores for the first two factors is shown in a). The individuals from *C. orni* were then excluded from the scatterplot for a better perception of the dispersion of the points (b). The opposite was done in (c). 55

Figure 3.1. Sampled populations of each species (● – allopatric populations of *Cicada barbara*; ○ – allopatric populations of *C. orni*; ▲ – sympatric populations of *C. barbara* and *C. orni*). 72

Figure 3.2. Oscillograms (amplitude-time waveform) and frequency-time spectrograms of the calling songs of *Cicada barbara* and *C. orni* from Iberian Peninsula. In the spectrogram, the intensity of each frequency is denoted by a greyscale from light grey (low intensity) to dark grey (high intensity). The peak frequency is the frequency with highest intensity. 77

Figure 3.3. Variation of the syllable rate of the calling song of *Cicada barbara* males with the ambient temperature for each population from the Iberian Peninsula and North Africa. For abbreviations of population names see Table 3.1. 82

Figure 3.4. Variation of the syllable rate, echeme rate, echeme duration and echeme period of the calling song of *C. orni* males with the ambient temperature for each population from the Iberian Peninsula and France. For abbreviations of population names see Table 3.1. 83

Figure 3.5. Boxplots of the acoustic variables analysed for each population of *C. barbara*. The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum, the circles are outliers, and the stars are extremes. For abbreviations of population names see Table 3.1. 84

Figure 3.6. Oscillogram (amplitude-time waveform) and frequency-time spectrogram of the amplitude modulated signal of one male of *Cicada barbara lusitanica* from Moura (a) and one male of *C. barbara barbara* from Fès (b). Sections I and II of a phrase are shown 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. 85

Page

Figure 3.7. Boxplots of the acoustic variables analysed for *C. orni* for each population in each region: Iberian Peninsula (1 – CoAlt; 2 – CoArr; 3 – CoCra; 4 – CoMon; 5 – CoMte; 6 – CoPie; 7 – CoSou; 8 – CoAlg; 9 – CoTol); France (10 – CoMol; 11 – CoNar; 12 – CoStH) and Greece (13 – CoAte; 14 – CoEvi; 15 – CoIte; 16 – CoKos; 17 – CoNea; 18 – CoPar; 19 – CoSka) (Abbreviations of population names as in Table 3.1). 87

Figure 3.8. Component scores of *C. orni* individuals from the Iberian Peninsula, France and Greece for components 1 and 3 obtained from Principal Component Analysis based on the data matrix of the acoustic variables (*cf.* Table 3.7) for each individual. 89

Figure 3.9. Syllable rate *versus* peak frequency of the calling songs of *C. barbara* (Cb) and *C. orni* (Co) in allopatry (Allop) and in sympatry (Symp). 91

Figure 3.10. Boxplots of the coefficients of variation (CV), in percentages, for each acoustic variable within individuals and the value of among-individuals CV in the population of *C. barbara* of Crato (1999). 92

Figure 3.11. Boxplots of the coefficients of variation (CV), in percentages, for each acoustic variable within individuals and the value of among-individuals CV in the population of *C. orni* of Crato (1999). 93

Figure 3.12. Scatterplot of syllable rate and temperature for each individual of *C. barbara* from Crato (1999) recorded two or more times during the day. 95

Figure 3.13. Scatterplot of each acoustic variable and temperature for each individual of *C. orni* from Crato (1999) recorded two or more times during the course of the day. 95

Chapter 1 – General introduction

I had two distinct objects in view, firstly, to show that species had not been separately created, and secondly, that natural selection had been the chief agent of change … Some of those who admit the principle of evolution, but reject natural selection, seem to forget, when criticising my book, that I had the above two objects in view; hence if I have erred in giving to natural selection great power, which I am far from admitting, or in having exaggerated its power, which is in itself probable, I have at least, as I hope, done good service in aiding to overthrow the dogma of separate creations.

Charles Darwin, *The Descent of Man* (1871)

1. General introduction

1.1. Species and speciation

A great amount of attention of evolutionary biologists has been dedicated to understanding the speciation process and to the definition of species (reviews in: Claridge *et al.*, 1997a; Wheeler & Meier, 2000; Barton, 2001 and articles in the same issue; Coyne & Orr, 2004; Fox & Wolf, 2006). Several mechanisms and processes leading to speciation (*i.e.*, the splitting of one species into two or more) have been suggested, the importance of each being highlighted or questioned as new theoretical models and new empirical evidence have accumulated. Darwin ([1859] 1964) viewed speciation mainly as the result of natural or sexual selection acting on the individuals due to competition for resources or mates and which produced character divergence. Workers after Darwin disagreed with this view of speciation in sympatry and emphasised the role of geographical isolation in speciation (see Coyne & Orr, 2004). Dobzhansky (1951) stressed the importance of the evolution of “reproductive isolating mechanisms” that prevent gene exchange between taxa. In Mayr’s synthesis of the genetics, natural history and biogeography of speciation, he defended the theory of geographic speciation: species arise when geographically isolated (allopatric) populations acquire during the period of isolation characters that promote or guarantee reproductive isolation after the external barriers break down (Mayr, 1963; 1970). Dobzhansky (1951), recognizing that reproductive isolation is stronger between sympatric than between allopatric species, formulated the hypothesis later called “reinforcement” of species isolating mechanisms. In his view, reproductive isolation is initiated in allopatry (by accumulation of genetic differences) but is only completed in sympatry, when these diverging incipient species meet and natural selection strengthens (“reinforces”) the premating isolating mechanisms (*i.e.*, mechanisms that prevent interspecific crosses, for example: seasonal, habitat or behavioural isolation), preventing the wastage in unfit hybrids. In the past few decades evidence has accumulated that speciation may occur despite gene flow (either in sympatry, parapatry or allopatry) and that natural selection may be more important than isolation (Via, 2001; Rieseberg *et al.*, 2003; Dieckmann *et al.*, 2004; Bürger *et al.*, 2006). In the genic view of the process of speciation (Wu, 2001), diverging populations have a number of divergent loci that contribute to differential adaptation, while gene exchange continues in other genomic regions. As the populations diverge, there will be more loci involved in differentiation, and a point may be reached where the populations are effectively reproductively isolated.

Studies on the evolution of postmating isolation (*e.g.*, zygote mortality, hybrid sterility or hybrid inviability), especially using the *Drosophila* model, have revealed patterns, for example that of a strong positive correlation of postmating isolation with time (measured by genetic distance) (Johnson, 2006). In contrast, the evolution of premating barriers is shown to be in many cases independent of genetic divergence between species (Tregenza & Bridle, 1997). A pattern of greater degree of premating isolation for sympatric pairs of species than for allopatric pairs with the same time level of divergence (estimated with allozymes) has been demonstrated in *Drosophila* (Coyne & Orr, 1989). This is the expected pattern resulting from reinforcing selection (described above). However, this interpretation should be taken with caution since the estimation of genetic distance based on neutral loci may be biased downwards in sympatric taxa due to gene exchange at neutral loci (Via, 2001).

Inferring which isolating barriers and which mechanism or mechanisms (natural selection, sexual selection or random genetic drift; or a combination of these) were involved in the speciation of a certain taxa, and if it occurred in allopatry or in sympatry, is a difficult task and requires the study of many sources of information, including historical, biogeographical, genetic, ecological and ethological data. Several approaches are possible:

- Comparative analysis of the types of isolating barriers present in taxa in different stages of evolutionary divergence, and in a variety of ecological and geographical situations may be useful in finding speciation patterns (Via, 2001; Turelli *et al.*, 2001).
- Areas of sympatry (*i.e.*, areas where taxa occur together) constitute an essential test for reproductive isolation. If two divergent populations, that are not completely reproductively isolated, evolve in allopatry and secondarily come into contact, they may interbreed and several outcomes are possible: extinction of one population (which one will be influenced by population growth parameters); permanent mixing of the gene pools, with the possible origin of a new species (Arnold, 1997); stable coexistence (hybrid zone) (Hewitt, 2001); or reinforcement of premating barriers (Butlin, 1995).
- The study of closely related species that differ markedly in sexually selected mating signals and associated preferences, whereas differing little in other traits, may reveal a pattern of speciation by sexual selection. However, other modes of divergence must be excluded (Panhuis *et al.*, 2001).
- Mate choice and hybridization studies can be conducted in captivity to test premating and postmating barriers (*e.g.*, Tregenza *et al.*, 2000). However, these are not possible to carry out in many groups of organisms and, even when possible, may not be indicative of the response of individuals in the wild.
- The molecular tools currently available allow assessment of the levels of genetic divergence (and hence of gene flow) between populations, evaluation of the extent of

hybridization and introgression and also identification of regions of the genome under selection (Schlötterer, 2004). Identifying genes responsible for reproductive isolation and species differences constitutes a recent and substantial advance in the field but is currently almost exclusively restricted to a few model organisms (Michalak & Noor, 2006).

Asking *how* species originated implies that we ask *what* a species is. Reproductive isolation is the critical criterion in the classical Biological Species Concept, which describes species as "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (Mayr, 1970). Although widely adopted in the past, its application is problematic in most cases, namely in allopatric populations for which the interbreeding criterion cannot be tested. Paterson (1985) suggested that, instead of describing a species in relation to another (through its isolation), a species should be characterized by its unique Specific Mate Recognition System (SMRS). Claridge *et al.* (1997b) considered that the two concepts (isolation and recognition) are similar and species would thus be characterized by distinct SMRSs resulting in reproductive isolation between species. Another concept, which is widely referred to, is the Phylogenetic Species Concept, based on the recognition of diagnosably distinct clades (Cracraft, 1997) and which can be applied to allopatric forms but which, according to Claridge *et al.* (1997b), does not allow the recognition of complexes of sibling species in many groups of organisms.

1.1.1. Acoustic mating signals

Mating signals may diverge between populations via several processes, including: direct adaptation to the signalling environment in order to increase conspicuousness to the receiver; differential selection by predators or parasitoids that may be attracted by the signal; pleiotropic effects from evolutionary changes in other characters of the organism; selection to avoid heterospecific matings (reinforcement); sexual selection; epigenetic influences (*e.g.* altitude and latitude), especially on body size; and also random genetic drift (Villet, 1995; Wilczynski & Ryan 1999; Turelli *et al.*, 2001; Zuk *et al.*, 2001; Simmons, 2004; Seddon, 2005). The associated preferences (or the sensory biases) of the receivers for signal characteristics may also be subject to the effects of natural or sexual selection (Panhuis *et al.*, 2001), as well as to the morphological and physiological constraints imposed on the signal receiving structures. These diversifying effects may produce significant differences in mating signals and preferences among populations, which can lead to reproductive isolation (premating barriers) between populations and subsequent speciation. However, if there is substantial divergence of signal traits without a corresponding divergence in preferences, that divergence may not result in isolation (Panhuis *et al.*, 2001).

Acoustic mating signals have been widely studied in several groups of animals (Gerhardt & Huber, 2002; Simmons *et al.*, 2003), since they are in many cases conspicuous to the human observer and are relatively easy to analyse. Cicadas (Insecta, Hemiptera, Cicadidae) are one such group, in which the calling songs of males have been shown to act as a mating signal, and which have been subjected to many evolutionary studies on speciation, isolating mechanisms, mate recognition systems and sexual selection (*e.g.*, Doolan & MacNally, 1981; Williams & Simon, 1995; Marshall & Cooley, 2000; Buckley *et al.*, 2001; Cooley & Marshall, 2001; 2004; Sueur & Aubin, 2003a; Gogala & Trilar, 2004; Villet *et al.*, 2004). In several cicada species it has been demonstrated that females are attracted to the conspecific against the heterospecific male song (*e.g.*, Villet, 1992; Daws *et al.*, 1997; Cooley & Marshall, 2001).

When several sound-producing species live in sympatry, acoustic interference between species is a potential problem, in which the receiver's ability to perceive signals may be affected by the background noise resulting from the calling activity of the other species (McGregor, 1991; Villet, 1995). In particular, closely related species are a special problem due to the similarity in sound producing systems. Differences in calling song traits, such as frequency content or temporal patterns, between closely related species allow them to avoid this acoustic interference. For example, the cicada species *Okanaga rimosa* and *O. canadensis* differ in the temporal patterns of the song but not in the frequency spectra (Stölting *et al.*, 2004). By contrast, *Magicicada tredecim* and *M. neotredecim* differ only in the carrier frequency (Marshall & Cooley, 2000). Other mechanisms that allow cicadas to avoid acoustic interference are spatial segregation (different habitats or different singing sites) (*e.g.*, Dybas & Lloyd, 1962; Claridge *et al.*, 1979) and temporal segregation (calling activity at different seasons or times of day) (*e.g.*, Wolda, 1993; Sueur & Puissant, 2002).

Patterns of signal divergence between closely related species in sympatric areas but not in allopatry have been described in several anurans and, more rarely, in acoustic insects (Gerhardt, 1994; Marshall & Cooley, 2000). These patterns are referred to as "reproductive character displacement" (see Loftus-Hills & Littlejohn, 1992; Noor, 1999) and may be driven by selection against wasteful heterospecific matings (reinforcement). However, they may also be led by other processes, such as adaptation to the acoustic environment, as seen above. Alternatively, such patterns may be the result of clinal variation. Marshall & Cooley (2000) reported a case of reproductive character displacement in one pair of north-American periodical cicadas, *Magicicada tredecim* and *M. neotredecim*. These authors found that *M. neotredecim* produced higher dominant frequency calls in sympatric areas than in allopatry

and *M. tredecim* maintained the same frequency over its distribution area. This example is an interesting and a good case study to test the hypotheses of reinforcement *versus* acoustic interference, since data on hybridization, female choice and evolutionary history of the species is available. Mitochondrial DNA and abdominal colour data suggest that there is strong reproductive isolation between the species (Marshall & Cooley, 2000; Simon *et al.*, 2000). Female *M. neotredecim* do indeed discriminate against male *M. tredecim* and there is a shift in preference associated with shift in the signal (Cooley *et al.*, 2006). Cooley *et al.* (2006) recognise that the acoustic interference hypothesis is supported by asymmetric character displacement and by a shift in preference associated with shift in the signal, but they also believe that the asymmetries may be explained by unequal abundance of species (since *M. neotredecim* is much less abundant, hybridization and displacement could be asymmetrical). The same authors suggest that initially, at the contact zone, *M. tredecim* calls would probably mask the calls of *M. neotredecim* and thus, hybridization may have occurred and associated costs of hybridization (supposing the hybrids were unfit) would be a selective force causing reproductive character displacement. This process would not leave any evidence of mitochondrial DNA introgression due to the asymmetrical crosses.

1.2. General overview of cicada biology, ecology and behaviour

Cicadas are insects with incomplete metamorphosis (hemimetabolous, exopterygota), and nymphs and adults have sucking mouth parts, feeding from the root (nymphs) or stem (adults) fluids. They are xylem feeders and adults are associated with plant communities rather than with individual plant species (Claridge *et al.*, 1979).

Cicadas are mainly distinguished by the ability of adult males to produce loud airborne acoustic signals during pair formation by means of a tymbal mechanism (*e.g.*, Pringle, 1954; Popov, 1975; Young & Bennet-Clark, 1995; Fonseca, 1996; Bennet-Clark, 1997, 1998a, 1999). Tymbals are membranes located dorsolaterally in the first abdominal segment (Figure 1.1), one on each side, that are distorted by the action of powerful muscles (tymbal muscles), which are driven by the nervous system (Pringle, 1954). Tymbal muscles may contract alternately or simultaneously depending on the species (Young, 1972). The tymbal is made of an unsclerotised membrane containing resilin, as well as of several sclerotised elements, with various degrees of flexibility, including the tymbal plate (on which the tymbal muscle is inserted) and the dorsal-ventral ribs (which are buckled inwards, producing sound, when the tymbal plate is distorted by the action of the muscle) (Young & Bennet-Clark, 1995). Depending on the species, the ribs may be present in different numbers and morphologies and may buckle in synchrony or sequentially (Fonseca & Bennet-Clark, 1998). The inward

distortion of the tymbal produces one or several pulses of sound (IN click) and is followed by a return to the original position with the production of another pulse of sound (OUT click) (Pringle, 1954; Fonseca, 1996). Associated accessory muscles, particularly the tensor muscles, are involved in modifying the curvature of the tymbals and, hence, the sound output (Pringle, 1954; Fonseca & Hennig, 1996). The sound produced is magnified (in amplitude) in the resonating chamber located in the abdomen (large air sac). It is then radiated by the paired tympana, which are situated ventrally on the second abdominal segment but also by the tymbals or by the abdomen walls, depending on the species (Bennet-Clark & Young, 1992).

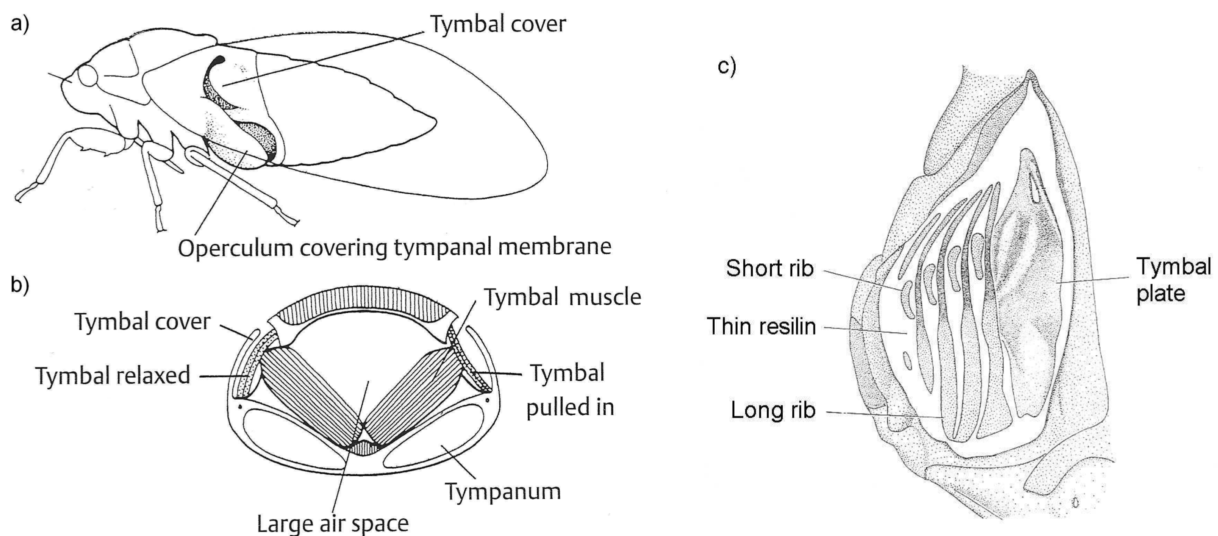


Figure 1.1. a) Lateral view of a male cicada with location of the tymbal and tympanal organs; b) Schematic cross section at base of abdomen (Adapted from McGavin, 2001); c) Lateral view of the left tymbal showing the sclerotized elements (grey) and the unsclerotized membrane containing resilin (white) (Adapted from Young & Bennet-Clark, 1995).

Different types of sound are produced by cicadas, with the long-range calling song being the most common. Calling songs are typically species-specific and have been commonly used by researchers to recognize cicada species, especially when no other such distinctive characters are available. These were suggested as taxonomic characters in cicadas (*e.g.*, Lei *et al.*, 1994), as well as in other groups of acoustic insects (*e.g.*, Reynolds, 1988). Most species can also produce other signals, *e.g.*, intra-specific courtship sounds, aggression sounds, inter-specific protest, distress or alarm sounds (Claridge, 1985; Sueur, 2003; Stölting *et al.*, 2004; Boulard, 2006). Despite these designations, the functions of some of these sounds are often unknown.

A key aspect of cicada biology is the strikingly long life cycle of some species. The nymphal stages may take several years to develop underground, feeding on root xylem fluids, until

they emerge and molt into winged adults, which live for just a few weeks for reproduction. Although there are cicadas with annual life cycles (Moulds, 1990), many other species have long life cycles, which can be explained by a diet low in nutrients (Heliövaara *et al.*, 1994), since cicadas feed on xylem fluid which is poor in protein and sugars. North-American periodical cicada species have 13- or 17-year life cycles, developing underground for precisely this number of years and emerging in synchrony in impressive large numbers, a strategy favouring predator satiation and increased mating success (Williams & Simon, 1995). A brood emerges in specific years (with intervals of 13 or 17 years) and is temporally reproductively isolated from the other broods that emerge in different years, although imperfect isolation leads to the possibility of gene flow through time (Simon, 1979; Martin & Simon, 1990; Williams & Simon, 1995). Allochronic speciation through shifts in reproductive timing by changes in life-cycle length (from 13- to 17-year and from 17- to 13-year life cycle) has been suggested in these species (Simon *et al.*, 2000; Cooley *et al.*, 2001).

1.2.1. Genus *Cicada* L. 1758

The evolutionary divergence of the genus *Cicada*, present in the Mediterranean region, has been studied at the morphologic, acoustic and genetic levels (*e.g.*, Quartau, 1988; Quartau *et al.*, 1997; 2001; Simões *et al.*, 2000; Quartau & Simões, 2006). It includes several closely related species, with different degrees of divergence at all levels. *Cicada barbara* (Stål) and *C. orni* Linnaeus are two of those species, very similar in morphology (sibling species), despite presenting some slight differences but not totally diagnostic in external characteristics, namely in wing spot intensity (Quartau, 1988; Ribeiro, 1998) and in thorax coloration, and some differences in the male genitalia, namely in the length of the pygofer (and its dorsal spine), of 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). Adults of both species appear in the beginning of the hot and dry season, after the development of the nymphs underground, and males sing during sunny days when ambient temperature is above 23ºC or even at night if temperature is well above 30ºC.

In both *C. barbara* and *C. orni*, males aggregate and sing simultaneously producing large “choruses”. The calling songs produced by males are distinct between these species, with *C. barbara* producing a continuous series of pulses without pauses (Boulard, 1982; 1995; Fonseca, 1991; Quartau & Rebelo, 1994), and *C. orni* producing a repetitive series of separate acoustic elements (echemes) which are composed of pulses, alternating with intervals of silence (Popov, 1975; Boulard, 1982; Fonseca, 1991; Quartau *et al.* 1999). Despite the clear differences in the calling songs between males of *C. barbara* and *C. orni*,

the structure of the tymbals and the mechanism of sound production by these species is similar, as expected in closely related species in the same genus, reflecting their common evolutionary history (Fonseca, 1991; 1994). In both species the tymbals alternate in the production of sound. The inward distortion of one of the tymbals produces usually three pulses of sound and the outward distortion another pulse, which is usually masked by the inward pulses of the other tymbal (Fonseca, 1991). Their calling position is similar to the resting position, with the abdomen slightly apart from the trunk (Boulard, 1995), the wings resting on the abdomen and no obvious abdominal movements (as described for *C. orni* by Claridge *et al.*, 1979; Boulard, 1995; Puissant & Sueur, 2001). Males of *C. barbara* and *C. orni* sing continuously for hours, almost without moving from their positions. This pattern of behaviour is found 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 & MacNally, 1981). In contrast, there are other species in which males are extremely mobile, alternating bouts of calling with short flights, such as the North-American *Magicicada* spp. (Cooley & Marshalll, 2001) and the Palaeartic *Euryphara contentei* (Quartau & Simões, 2004).

Behavioural observations of *C. orni* (personal observation) have shown that when a female lands on the trunk or branch where the male is singing, the male approaches the female, usually producing sound, and then ceasing at the time of copulation, which may last for more than 30 minutes. Males sometimes approach other silent males which then protest by emitting sound. Aggressive encounters between males involve production of sound and even physical contact between males, grabbing each other with their legs (personal observation).

Besides the calling song, other sounds are produced by these cicadas. These sounds have already been described but the functions and the designations of each type of sound are sometimes controversial. For *C. barbara* a “courtship” signal (or approaching song), modulated in amplitude, was described (Fonseca 1991; Boulard, 1995), even though this same signal was described as being an “alarm” signal by Quartau & Rebelo (1994), produced when the male is disturbed, for example by the presence of an human observer. A “male-to-male interaction” signal was described for *C. barbara* (Fonseca, 1991) and an “opposition” sound to another approaching calling male was described for *C. orni* (Boulard, 1995). Both species, when captured, produce a very irregular signal, with no time or amplitude pattern which has been called an “alarm”, “protest” or “stress” sound (Fonseca, 1991; 1996; Boulard, 1995). *C. orni* males produce a slower calling song when a shadow (of a bird, for example) falls on top of the male, and an “escape” sound when flying away (Boulard, 1995).

From the observations made by M. Boulard (personal communication) on the oviposition and nymph development of *C. orni* in captivity, nymphs can develop between two to five years, from the same batch of eggs. However, nothing is known about the development of nymphs of this species in the wild. The average duration of the larval development of *C. barbara* was reported to be two years by Giralda *et al.* (1998).

Both *C. barbara* and *C. orni* are found mainly in xerothermic habitats, in Mediterranean woodland or shrubland, mainly in olive, pine or oak trees, but even in eucalyptus, vineyards, and in gardens in cities (Quartau, 1995; Patterson *et al.*, 1997; Puissant & Sueur, 2001; Sueur *et al.*, 2004).

C. barbara is found in North Africa, in some western Mediterranean islands (Nast, 1972) and in the Iberian Peninsula. It is found in scattered high temperature environments. The Iberian populations have been described as a new subspecies, *Cicada barbara lusitanica* Boulard, in comparison with the nominal subspecies from North Africa (Tunisia), based on small differences of the male genitalia and the female ovipositor (Boulard, 1982). Mitochondrial DNA (cytochrome *b* gene) variation patterns also support the splitting of this species into *C. barbara lusitanica* and *C. barbara barbara* Stål (G. Pinto-Juma, personal communication). Boulard (1995) 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. Besides the calling song, the courtship song of both subspecies has been described and 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 been done.

C. orni is one of the most abundant and common cicadas found in southern and central Europe, in eastern Mediterranean Europe, as well as in western Asia and Middle East (Popov, 1975; Quartau & Fonseca, 1988; Schedl, 1973; 1999). Several authors have described the calling song of *C. orni* for a few local populations (*e.g.*, Popov, 1975; Joermann & Schneider, 1987; Fonseca, 1991; Boulard, 1995). Comparisons of the calling songs in populations from southern France with those from the former USSR did not show obvious geographic variation (Claridge *et al.*, 1979; Claridge, 1985). In contrast, differences between Portuguese and Greek populations were found at the morphologic, allozyme, mitochondrial and acoustic levels (Ribeiro, 1998; Quartau *et al.*, 1999; 2000a; 2001; G. Pinto-Juma, personal communication).

C. cretensis Quartau & Simões and *C. mordoganensis* Boulard, are two other closely related species to *C. orni* and are present in the eastern Mediterranean area only. They are very similar in morphology and have similar calling songs to *C. orni*, diverging slightly in the temporal pattern of the songs (Simões *et al.*, 2000; Quartau & Simões, 2005; 2006). In contrast, *C. lodosi* Boulard, another species found in the eastern Mediterranean area, is easily distinguishable in morphology and calling song from the others, since it is considerably bigger and produces a continuous call. Other species of the genus, *C. permagna* (from Turkey) and *C. cerisyi* (from Egypt) are only known from dried specimens in collections and their calling songs are unknown.

1.2.1.1. Reproductive isolation between Cicada barbara and C. orni

Experimental behavioural work with cicadas in general, and with *C. barbara* and *C. orni* in particular, is usually difficult in captivity (Fonseca & Revez, 2002a; Simões & Quartau, 2006). It has not been possible yet to test the responses by females (through phonotactic behaviour) to calling songs by males. There is evidence of species discrimination only by males in both *C. barbara* and *C. orni* (Fonseca & Revez, 2002a; Simões & Quartau, 2006). A playback study carried out with males of *C. barbara* (Fonseca & Revez, 2002a) showed that they discriminate the conspecific song from the *C. orni* song, with the latency of response (by singing) being lower with the conspecific song than with the heterospecific song. Also, altering the temporal pattern of the calling song of *C. barbara* significantly reduced the response by the males. The males did not respond to songs with pauses longer than 30 ms, approaching the characteristic *C. orni* song, and responded to a modified *C. orni* song without pauses. Males preferred frequencies between 6 and 9 kHz but the modified *C. orni* song (without pauses) proved as attractive as the *C. barbara* song. Conversely, a playback study with males of *C. orni* (Simões & Quartau, 2006) showed that the intensity of response by singing was very low to *C. barbara* song and very high both to *C. orni* and *C. mordoganensis* songs. Altering the temporal or the frequency patterns of songs, it was shown that the intensity of response was very low to echeme durations equal or below 20 ms (maintaining interecheme interval constant). No male *C. orni* responded to interval duration values equal or below 40 ms. Responses to peak frequency modifications were generally high below 6 kHz.

If the discrimination found in males of *C. barbara* and *C. orni* is present in females, the species may be reproductively isolated by this premating barrier. However, when two different species have no postmating isolating barriers and their isolation is based on premating barriers, particularly through different signalling systems, hybridization may be

favoured by certain factors that disrupt normal signal transmission or reception (Gee, 2005). Moreover, when one of the species is much more abundant than the other, in the absence of the appropriate stimuli from conspecific individuals, they may respond to inappropriate stimuli from a different species (Mayr, 1970). The genetic divergence between species is not always an indication of the existence of, or of the efficiency of postmating isolating barriers, since some species for which molecular data shows millions of years of divergence can still hybridize (Hewitt, 2000).

An allozyme study revealed only three of 19 loci to be diagnostic for the separation of *C. orni* and *C. barbara*, with very low Nei's genetic distances reported (Quartau *et al.*, 2000a; 2001). In contrast, mitochondrial DNA analysis revealed a high divergence between these species (G. Pinto-Juma, personal communication). Additionally, no evidence of hybridization between *C. orni* and *C. barbara* was found since they do not share any mitochondrial haplotype (G. Pinto-Juma, personal communication). However, since mitochondrial DNA is maternally inherited, it may not detect hybridization if this is asymmetrical (directionality in the hybridization crosses). Microsatellites are more efficient markers for the detection of hybridization due to their high polymorphism and high evolutionary rate (Bruford *et al.,* 1996; Chambers & MacAvoy, 1999). Hybrids may also be identified by the calling songs. The existence of hybrids with intermediate acoustic recognition signals between the parental species is commonly found in laboratory crosses of some species of insects (Claridge, 1985; Reynolds, 1988).

The Iberian Peninsula is an overlap area of the distribution ranges of *C. barbara* and *C. orni* with a few known sympatric localities, where males of both species are seen calling simultaneously from trunks and branches of the same trees. Some seasonal displacement has been found between adults of *C. orni* and *C. barbara* in allopatric and in sympatric areas. *C. orni* emerges earlier in the summer (June) than *C. barbara* (July/August) in both types of areas. In allopatric areas, the adults of *C. orni* disappear in September/October, but when in sympatry with *C. barbara*, they usually disappear much earlier (August). *C. barbara* shows only slight variation: it appears somewhat earlier in allopatry than in sympatry and disappears in September/October in both type of areas (Ribeiro, 1998; and personal observation). Also, at the beginning and at the end of each reproductive season, it is common that, in sympatric areas, one species is usually much more abundant than the other. As seen before, this situation may favour interspecific crosses if the isolating barriers are not totally efficient.

The sympatric areas between *C. barbara* and *C. orni* are also interesting to test the existence of patterns of reproductive character displacement. The only pairs of species of this genus

known to occur in sympatry have distinct calling songs (*C. barbara* and *C. orni* in Portugal and *C. lodosi* Boulard and *C. mordoganensis* in Turkey) (Quartau & Simões, 2006). In fact, *C. barbara* and *C. lodosi* produce a continuous sound and *C. orni* and *C. mordoganensis* produce songs with pauses. Even with these differences in the temporal pattern of the songs, if the frequency content of the songs is similar, songs of one species may be "masked" by songs of the other species and thus there may be selection for divergence of the songs.

1.3. Objectives and structure of the thesis

Cicada barbara and *C. orni* (Hemiptera, Cicadidae) are potentially good subjects to test hypotheses about population divergence and speciation since they possess an apparently well differentiated premating isolating barrier (different acoustic signals produced by the males), without having significant morphological differences, and, additionally, they exist in sympatry in some areas of the Iberian Peninsula where the efficiency of that isolating barrier can be tested.

Highly variable genetic markers, such as microsatellites, are expected to allow the detection of hybridization or introgression in this pair of species and to be highly informative about the genetic structure of populations. Acoustic analysis of the calling songs produced by the male cicadas may also allow the detection of hybrids and the study of the patterns of temporal and geographic variation of this important behavioural characteristic constitutes an initial step to understand which characteristics of the signal may be involved in sexual recognition.

The main objectives of the present study are:

- To evaluate evidence of hybridization between *C. barbara* and *C. orni*. Specific diagnostic microsatellite alleles and also acoustic signals with intermediate characteristics will be investigated in areas of sympatry as possible evidence of hybridization or introgression (*i.e.*, gene flow between species whose individuals hybridize).

- To compare sympatric and allopatric populations of each species in order to evaluate the possible influence of the interaction between species. Microsatellite allele frequencies and calling song characters will be compared between sympatric and allopatric situations to test the efficiency of the mechanisms of isolation and the existence of character displacement in the calling song.

- To study geographic variation in both *C. barbara* and *C. orni* populations. The variability and differentiation in microsatellite allele frequencies and in the calling song characters will be analysed for each species in order to describe the patterns of gene flow and isolation between populations.

- To test for the existence of reproductively isolated broods. The patterns of differentiation in microsatellite allele frequencies and in the calling song characters between cicadas from different years of emergence will be analysed with the aim of evaluating the gene flow and isolation between different years at the same locality.

The genetic analysis and the acoustic analysis are presented in two separate chapters (Chapter 2 and Chapter 3, respectively). Chapter 2 includes an introduction to microsatellite characteristics, its applications and limitations, as well as a methodological introduction to techniques and data analysis (2.1). It is then subdivided in two sections, the first one dedicated to microsatellite isolation (2.2), the second to microsatellite analysis (2.3), this one with material and methods, results and discussion. Since the microsatellite data revealed the existence of non-amplifying alleles ("null alleles"), a considerable part of the analysis was dedicated to discussing and evaluating the effects of this problem (*Single-locus analysis*). Chapter 3 is composed of an introduction to the acoustic techniques, material and methods, results of the analyses and discussion of the results. Chapter 4 presents a general discussion, integrating results from both types of analyses, as well as some final remarks about the work. Finally, Chapter 5 lists all the bibliographic references of the manuscript.

Chapter 2 – Genetic analysis

Atravessar o Alentejo por este braseiro, debaixo de um céu mais branco do que azul, entre um restolho que refulge, com raras azinheiras na terra nua e fardos de palha por recolher, sob o incessante reco das cigarras, seria logo uma história completa.

José Saramago, *A Jangada de Pedra*

2. Genetic analysis

2.1. Introduction to microsatellites

2.1.1. Characteristics, applications and limitations of microsatellites

Microsatellites have been a marker of choice in many population genetic studies over the past decade (Bruford *et al.*, 1996; Jarne & Lagoda, 1996; Goldstein & Schlötterer, 1999; Chambers & MacAvoy, 2000; Schlötterer, 2004). These markers are short tandem simple nuclear sequences of 2–6 base pairs (bp), biparentally inherited and highly abundant in eukaryotic genomes and relatively evenly spaced throughout the genome, although its abundance varies greatly across taxonomic groups (Primmer *et al.*, 1997; Schlötterer, 1998; Hancock, 1999; Nève & Meglécz, 2000). Microsatellites are believed to evolve by a mutation mechanism caused by polymerase slippage during replication, which causes gain or loss of repeat units (Tautz & Schlötterer, 1994; Ellegren, 2000), although there is also evidence for the effect of unequal crossing-over or gene conversion on the length change of tandem repeats (Li *et al.*, 2002). The different microsatellite alleles are, thus, recognized by their different sizes. The mutation rates of microsatellites are very high compared with point mutation rates in coding gene loci, varying between 10^{-6} and 10^{-2} mutations per locus per generation *(*Hancock, 1999; Ellegren, 2000; Li *et al.*, 2002), which makes microsatellite loci potentially highly polymorphic.

The total length of a microsatellite is usually small (less than 100 bp) and microsatellites are usually flanked by conserved DNA sequences. Therefore, primers for Polymerase Chain Reaction (PCR) may be designed from the flanking DNA to amplify the microsatellite locus, and PCR products may be screened for size variation to detect alleles, which is relatively cheap and easy compared to sequencing. The alleles are expressed codominantly, which means that information is available about the size of both alleles. Separation on polyacrylamide gels allows the high resolution of alleles (to one base pair), and several loci can be analysed together on the same gel. They are nowadays easily scored by automated sequencing machines with nonradioactive labelling (Chambers & MacAvoy, 2000). PCR analysis of small fragments also allows the use of noninvasive sampling in conservation studies or of degraded samples from museum specimens (Morin & Woodruff, 1996). They are generally selectively neutral markers, being compatible with the assumptions of most population genetic theory. Microsatellites have the advantage over mitochondrial DNA of being nuclear with codominant expression. In fact, mitochondrial DNA, being maternally

inherited without recombination, is a biased description of gene flow and is restricted to female-mediated processes (Pope *et al.*, 1996).

These characteristics make microsatellites valuable markers in linkage mapping, in forensic or museum samples analysis, in paternity and kinship analysis and also in studies of genetic variation in natural populations (Bruford & Wayne, 1993; Bruford *et al.*, 1996; Beaumont & Bruford, 1999; Chambers & MacAvoy, 2000; Beaumont, 2003), namely to understand phylogeographic and migration patterns (*e.g.*, Bowcock *et al.*, 1994; Estoup *et al.*, 1996; Massonnet & Weisser, 2004), to study fine-scale genetic structuring (*e.g.*, Surridge *et al.*, 1999), to estimate effective population size (*e.g.*, Xu & Fu, 2004) or to detect population bottlenecks (*e.g.*, Luikart *et al.*, 1998; Keller *et al.*, 2001). They are fundamentally used for intra-specific studies but can also be used to assess processes such as hybridization (*e.g.*, Roy *et al.*, 1994; Gottelli *et al.*, 1994; Goodman *et al.*, 1999; Beaumont *et al.*, 2001; Sætre *et al.*, 2001; Randi & Lucchini, 2002).

There are however several limitations to the use of microsatellites. Usually there must be an initial investment in the isolation of loci by construction of a genomic library enriched for microsatellites, although there are examples of successful application of primers from other species (Bruford *et al.*, 1996). This process of isolation of microsatellites and design of primers in a focal species and then using those primers in other related species may lead to biased estimates of genetic diversity due to ascertainment bias, *i.e.*, loci are generally chosen for their high polymorphism, and thus, the focal species may exhibit longer alleles and higher polymorphism than a related species (Ellegren *et al.*, 1997).

Even though microsatellite allele scoring is usually simple, it may present some difficulties, especially in dinucleotide repeat unit microsatellites, due to replication slippage during PCR, which causes “stutter” bands on the gel (Haberl & Tautz, 1999). Nevertheless, these “stutter” bands may also be very helpful in identifying specific amplification products (Schlötterer, 1998) and dinucleotide loci are often used due to their higher frequency in the genome than tri- or tetranucleotide loci (Chamber & MacAvoy, 2000).

The analysis of microsatellite data suffers limitations due to the uncertainty about the most appropriate mutational model to apply to microsatellite evolution. The two main mutation models proposed to explain the allele distributions in microsatellites are the infinite-allele model, in which each mutation gives rise to a new allele, and the stepwise mutation model, in which a mutation either adds or deletes a single unit from the current allele (Jarne & Lagoda, 1996). The infinite-allele model is in many cases found to be inadequate, since the range of

variation in number of repeats of microsatellites is found to be restricted (Goldstein *et al.*, 1995). The stepwise mutation model is apparently more in accordance with the mutation process of slippage during replication and fits the allele frequency distributions of some microsatellite loci (Valdes *et al.*, 1993; Shriver *et al.*, 1993), but a strict single step stepwise mutation model does not explain the allele frequency distribution in many other microsatellite loci. Modifications of the original stepwise model have been proposed allowing stepwise mutations by more than one unit (Di Rienzo *et al.*, 1994), allowing biased mutation dependent on allele size (Garza *et al.*, 1995), or allowing upper and lower range constraints (Feldman *et al.*, 1997). Additionally, base substitutions and large insertion/deletions may also occur in microsatellite loci (*e.g.*, Angers & Bernatchez, 1997; Macaubas *et al.*, 1997). Microsatellites thus have complex mutation patterns and the rate of mutation in microsatellite loci is found to be dependent on several factors, such as species, repeat types and repeat length (Goldstein & Schlötterer, 1999; Ellegren, 2000; 2004; Webster *et al.*, 2002).

The high mutation rate and the mutational processes of microsatellites makes them prone to size homoplasy, that is, the alleles may be identical-in-state (*i.e.*, they may have the same size) but not identical-by-descent (*i.e.*, they do not have the same historical affinities) due to convergent mutations (Jarne & Lagoda, 1996; Estoup *et al.*, 2002). This size homoplasy causes overestimation of relatedness, which is a cause of concern in the analysis of microsatellite data, limiting their utility especially above the species level but also on studies of ancient population patterns (Bruford *et al.*, 1996; Chambers & MacAvoy, 2000).

Microsatellites are usually considered as evolutionary neutral markers, but there is evidence of functional significance of some microsatellite loci, for example in chromatin organization or in regulation of several DNA metabolic processes (recombination, replication, cell cycle, gene activity) (Li *et al.*, 2002). They may also be involved in selective sweeps (genetic hitchhiking of advantageous mutations) and in background selection (continual removal of deleterious mutations) (Schug *et al.*, 1998).

2.1.2. Microsatellite isolation and genotyping

One of the initial difficulties in using microsatellites is the laborious and time-consuming process of cloning of new microsatellites for each new species under study. The applicability of specific primers for amplification of microsatellites in other taxa is variable, depending on the rate of evolution of the sequences flanking the microsatellites and on the times of divergence of the taxa. Often amplification is possible in closely related species but, even in

that case, the loci may be monomorphic (not variable) (Schlötterer, 1998; Chambers & MacAvoy, 2000).

In the specific case of cicadas, there were no primers published for any cicada species before the start of the present study. To our knowledge, only Wilson, Simon & Sunnucks (unpublished results) had developed primers for microsatellites for a cicada, *Magicicada cassini*, a North-American species belonging to the same Family (Cicadidae) as genus *Cicada*, but belonging to a different Subfamily (Tibicininae). These primers were tried in *Cicada barbara* and *C. orni* in the present study, with permission from the authors, but some of the loci did not amplify and the ones that did amplify were monomorphic.

Microsatellite cloning and isolation involve the construction of a genomic library enriched for the presence of microsatellite DNA (Bruford *et al.*, 1996; Schlötterer, 1998). Several methods were developed and are reviewed in Zane *et al.* (2002). The protocol which was followed in the present work briefly consists in digesting genomic DNA with an enzyme, size selecting the DNA and then hybridising the DNA to a microsatellite probe and capturing the hybrid molecules on an avidin matrix. The enriched DNA is ligated in to a cloning vector which is taken up by the bacteria following a heat-shock procedure. The colonies that grow are then screened for recombinants containing a microsatellite repeat and are sequenced. The sequences that do indeed contain microsatellite repeats, and which have sufficient DNA flanking them, may then be used to design primers. Several criteria should be followed when designing primers, general criteria related to stringency (sufficient size and complexity of the primer) (Hoelzel & Green, 1998) and criteria specific to microsatellites: primers should not be placed very close to the microsatellite motif; they should be designed to amplify regions from about 100 to 250 bp; they should have the same amplification protocol, allowing the amplification of several loci at the same time in the same amplification block or even in the same tube (multiplexing); they should have non-overlapping fragment size or they should be labelled with different fluorescence labels, enabling simultaneous electrophoresis (Bruford *et al.*, 1996; Scribner & Pearce, 2000).

Primers are then tested on a set of individuals and polymorphic loci are selected. Scoring of microsatellite gels is usually a simple and reliable process. Alleles can be accurately sized either manually or aided by automated systems with internal size markers. Once scored, the data are available for analysis.

2.1.3. Microsatellite data analysis

2.1.3.1. Hardy-Weinberg equilibrium

A basic principle of population genetics is the Hardy-Weinberg principle, which states that in a large population with random mating and with no selection, no mutation and no migration, the genotype frequencies and the allele frequencies at a locus are constant from generation to generation at a particular equilibrium value (corresponding to the Hardy-Weinberg proportions) and that there is a simple mathematical relationship between the genotype frequencies and the allele frequencies (p and q, in the case of a locus with two alleles) at that locus: frequency of genotype AA is p^2, of genotype Aa is $2pq$ and of genotype aa is q^2 (see Hartl & Clark, 1997; Hedrick, 2000).

Deviation from Hardy-Weinberg proportions in a population is measured by one of the F-statistics, F_{IS}, also called the inbreeding coefficient, which measures the reduction (or increase) in heterozygosity (proportion of heterozygous genotypes) of that population when compared to Hardy-Weinberg expectation (Hartl & Clark, 1997; Hedrick, 2000). It is higher than zero if there is a deficiency of heterozygotes and lower than zero if there is an excess of heterozygotes in the population (Hedrick, 2000). Considering a hierarchical population structure, with a population divided in several subpopulations, F-statistics express the correlation coefficients between alleles within a certain subdivision level relative to the alleles within a higher level: alleles within individuals relative to the alleles within a subpopulation (F_{IS}), within individuals relative to the total population (F_{IT}), and within a subpopulation relative to the total population (F_{ST}) (Excoffier, 2003). These coefficients are higher than zero if the gametes of common ancestry combine more frequently than expected at each subdivision level relative to a higher level (Neigel, 1996). F_{ST} is a measure of the genetic differentiation over subpopulations and F_{IS} and F_{IT} are measures of the deviation from Hardy-Weinberg proportions within subpopulations (F_{IS}) and in the total population (F_{IT}) (Hedrick, 2000).

Deviations from Hardy-Weinberg equilibrium can be due to inbreeding, subpopulation structure (Wahlund effect) or natural selection (Rousset & Raymond, 1995; Hedrick, 2000). Heterozygote deficits can, however, also be due to lack of PCR amplification or detection of one or several alleles (“null alleles”), creating an homozygote excess. Since in the present study the data from several loci had evidence of null alleles this issue is described with greater detail in the next subsection.

2.1.3.2. Null alleles

There was early awareness of the possible occurrence of "null alleles" in microsatellite data which could cause serious problems in the interpretation of population genetic results (*e.g.*, Chakraborty *et al.* 1992; Bruford & Wayne, 1993; Callen *et al.* 1993; Paetkau & Strobeck 1995; Pemberton *et al.*, 1995; Goodman *et al.*, 1999; Dakin & Avise, 2004). Null alleles are alleles that do not amplify during PCR due to mutations in the annealing site of the primers (Schlötterer, 1998; Chambers & MacAvoy, 2000). Mutations can be point substitutions or insertion/deletions (*e.g.*, Callen *et al.*, 1993 ; Paetkau & Strobeck, 1995; Ishibashi *et al.*, 1996; Lehmann *et al.*, 1996; Jones *et al.*, 1998). When electrophoresing PCR products, a heterozygote for an amplifying allele and for a null allele will appear as a homozygote and a homozygote for the null allele will produce no band, resulting in a heterozygote deficiency in the analysed population.

Null alleles can be detected by analysis of the inheritance of parental alleles or by testing if genotypic proportions are in accordance with Hardy-Weinberg equilibrium (*e.g.*, Pemberton *et al.*, 1995; McGoldrick *et al.*, 2000; Dawson *et al.*, 2005). Null alleles are expected to cause an excess of homozygotes for all homozygote size classes (Van Oosterhout *et al.*, 2004). Other explanations for heterozygote deficiency such as inbreeding or hidden substructure (Wahlund effect) are expected to exhibit different characteristics (Overall & Nichols, 2001). For example, inbreeding can be most likely ruled out if the deviation is not common to the majority of loci analysed (Schlötterer, 1998). For the Wahlund effect some heterozygote frequencies may be reduced and others remain unaffected or may increase (Hedrick, 2000). These types of patterns should be investigated and distinguished from random deviation to evaluate the most probable explanation. In this context, knowledge about population demography is also valuable but is very often absent.

Null alleles have been detected in microsatellite studies of numerous groups of organisms, in some cases, occurring at high frequencies in some loci (*e.g.*, 53% in Onyabe & Conn, 2001, 64% in Newman & Squire, 2001, and 70% in Hoare *et al.*, 1998). When null alleles are detected, several analytical approaches are possible, depending on the number of loci available and other constraints. First, if there are additional loci available or potentially available for screening, the loci with null alleles can simply be discarded. Second, if loci are scarce but time and money are not limiting, new primer variants can be designed. However, this process may be problematic, since the new priming site may also harbour the mutations that cause non-amplifying alleles (Pemberton *et al.*, 1995; Ishibashi *et al.*, 1996). Even so, some authors have redesigned primers with apparent success (*e.g.,* Callen *et al.*, 1993;

Paetkau & Strobeck, 1995; Ishibashi *et al.*, 1996; Lehmann *et al.*, 1996; Jones *et al.*, 1998). When the above approaches are not possible, the only option remaining is to take a statistical approach, estimating the null allele frequencies and adjusting the frequencies for all other alleles accordingly. Methods have been proposed to estimate the null allele frequencies from apparent deficiency in heterozygotes (Chakraborty *et al.*, 1992; Brookfield, 1996; Goodman *et al.*, 1999; Van Oosterhout *et al.* 2003). Several authors estimate null allele frequencies for their microsatellite data, but rarely are these values used to adjust allele and genotype frequencies (but see Goodman *et al.,* 1999; Keyghobadi *et al.*, 1999, 2005; Astanei *et al.*, 2005; Barker, 2005; Michel *et al.*, 2005; Karhu *et al.*, 2006).

2.1.3.3. Population genetic analysis

Population genetic analysis aims at assessing genetic variation and its structure within and among populations. Several methodologies and a number of software packages have been developed for analysis of microsatellite data, and the constant improvement in computer processing enables the use of ever more complex methodologies.

As seen before (2.1.3.1), departures from Hardy-Weinberg equilibrium can give indication about evolutionary processes affecting the loci under study. However, only a few generations (and in some cases only one generation) of random mating are enough to restore Hardy-Weinberg proportions and, thus, even recent demographic events may not be detected (Hartl & Clark, 1997; Chikhi & Bruford, 2005). The association between alleles from different loci (linkage) is also an important source of information about evolutionary processes affecting the loci. Linkage disequilibrium (non-random association) can be generated by genetic drift, mutation, admixture, selection and inbreeding (Hudson, 2003). Linkage equilibrium (random association) is eventually attained in a large population with random mating and no mutation, migration or selection, but this may require a large number of generations, contrary to the attainment of Hardy-Weinberg equilibrium (Hartl & Clark, 1997) and thus may carry information for a longer period of time.

Several measures of population differentiation have been developed, based on different mutational and demographic models. A popular measure of population differentiation is F_{ST} (see 2.1.3.1), which is easily estimated from allelic frequencies and is not too dependent on the theoretical demographic model (Chikhi & Bruford, 2005). However, values of F_{ST} are not comparable from study to study, and in hypervariable loci F_{ST} values may be artificially low due to the high variation within subpopulations (Chikhi & Bruford, 2005; Allendorf & Luikart,

2007). It also assumes an infinite-allele model of mutation. Other measures of differentiation related to F_{ST} but that use allele-size information (R_{ST}; Slatkin, 1995), assume a stepwise mutation model, which is theoretically more in accordance with the mutation patterns of microsatellites. However, F_{ST}-based estimates are found to be more reliable than R_{ST}-based estimates when sample sizes are small and/or the number of loci scored is low (Gaggiotti *et al*., 1999). Also, genetic distances developed to conform to the stepwise mutation model have very large variances (Paetkau *et al*., 1997). F_{ST}-like-based approaches are commonly used and are usually powerful in detecting and analysing population structure (Chikhi & Bruford, 2005). Analysis of molecular variance (AMOVA) is used for hierarchical population analysis as it allows the partition of the total variance of allele frequencies into the components of the hierarchical subdivision (Excoffier *et al*., 2005). Patterns of dispersal (gene flow) between populations can be investigated using pairwise F_{ST} values to determine whether genetic differentiation correlates with geographic distance (Slatkin, 1993; Rousset, 1997). Models of dispersal include the “island” model, in which dispersal is equally probable from and to any of the populations, and the “isolation by distance” model, in which the probability of dispersal from one site to another declines with increasing geographical distance, and thus, genetic similarity is expected to be higher between individuals from closer populations (Wright, 1943; Rousset, 2004).

Several genetic distances have been developed and are widely used together with clustering or ordination methods to find relationships among populations. These methods may be useful as an exploratory tool and are simple to use. However, they are not testable statistically and the results in terms of genetic relationships among populations often depend on the distance measure and on the graphical representation used, making them questionable and unreliable if used singly (*e.g.*, Queney *et al*., 2001). Furthermore, population structure in these methods is defined *a priori* and any hidden patterns within populations are not detected. Individual-based methods have been developed to infer hidden structure without using any *a priori* information regarding origin of the individuals. Again exploratory tools, such as clustering or multivariate ordination procedures based on genetic distances among individuals (*e.g.,* Bowcock *et al*., 1994), have been used but have the same drawbacks as described above.

More powerful methods for inference of hidden structure include likelihood and probabilistic methods (Pritchard *et al*., 2000; Dawson & Belkhir, 2001; Corander *et al*., 2003; Wu *et al*., 2006). In these methods, the parameters of a given model are inferred directly and jointly and it is possible to visualize the distribution of the most probable parameters given the data (Beaumont, 2004; Allendorf & Luikart, 2007). Inference is done by maximum-likelihood

methods, which find the parameters that maximize the probability of obtaining the observed data under the model, or by Bayesian methods, which give the probability distribution ("posterior distribution") for the parameter of interest by using the data and incorporating prior subjective or incomplete knowledge about the probability distribution ("prior distribution") of one or more parameters of the model (Luikart & England, 1999; Beaumont, 2004; Beaumont & Rannala, 2004). Bayesian methods require intensive computer processing capabilities and only recently have begun to be widely used in a number of population genetics problems (Beaumont & Rannala, 2004), including the detection of population structure and the assignment of individuals to populations (Rannala & Mountain, 1997; Pritchard *et al.*, 2000; Dawson & Belkhir, 2001; Corander *et al.*, 2003).

2.2. Isolation of microsatellites

Two separate partial genomic libraries were constructed, one for each species, using the protocol by Hammond *et al.* (1998). The detailed process is described below.

2.2.1. Preparing DNA for enrichment

DNA digestion and size selection

Total DNA was extracted from frozen muscle tissue of the thorax of five specimens of each species (from Monforte, Alcalar, Arrabida, Foz Côa and Crato for *Cicada barbara*; and from Sesimbra, Arrábida, Monte da Caparica, Monforte and Athens for *Cicada orni*) using the QIAamp DNA Mini Kit (QIAGEN). The DNA was digested with *Mbo I* restriction enzyme (Promega). Selected fragments (300–900 bp) were isolated from a 1.5% agarose gel by cutting the gel and electroeluting into dialysis tubing and purifying using Centricon Microconcentrators (Amicon).

Attachment of linker sequences and first PCR

The linker sequence SAULA/SAULB (a double stranded molecule made by annealing SAULA to SAULB and which has a GATC overhang, compatible with the ends produced by digestion of genomic DNA with *Mbo I*) was ligated to the size selected DNA with *T4 ligase* enzyme (Promega). These linkers act as a priming site in the PCR (5 minutes at 72ºC, then 32 cycles of 1 minute at 95ºC, 1 minute at 67ºC and 2 minutes at 72ºC, followed by an extension period of 5 minutes at 72ºC) to amplify the size selected DNA.

2.2.2. Enrichment for microsatellite repeats

Probes made of CA or GA repeats coupled to biotin molecules were hybridized to the fragments amplified in the PCR. Only the hybrids bind to the Vectrex Avidin D matrix (Vector Laboratories), which allows the remaining DNA to be washed away. In a final step of elution the fragments that contain microsatellites were obtained. These fragments were amplified in a PCR under the following conditions: 2 minutes at 95ºC, then 32 cycles of 1 minute at 95ºC, 1 minute at 66ºC and 2 minutes at 72ºC, followed by an extension period of 5 minutes at 72ºC.

2.2.3. Cloning

The fragments were digested with *Mbo I* enzyme and the linkers SAULA/SAULB were removed using Centricon Microconcentrators. The DNA was inserted into the plasmid vector pUC18, using a kit (Amersham Pharmacia Biotech) that contains the plasmid cleaved with the restriction enzyme *Bam H1*, which produces a GGATCC overhang (compatible with the ends produced by *Mbo I*). The vector:insert ratio used was approximately 1:1. Ligation products were cloned into One Shot® competent cells TOP 10 (Invitrogen) and the corresponding protocol was used.

2.2.4. Screening for colonies with microsatellites

The protocol followed for screening the colonies was that by Lunt *et al.* (1999). Single colonies were transferred to culture broth and grown for one hour. PCR amplification was performed with the vector primers (forward and reverse) and a microsatellite-specific primer (CA or GA repeat composed of 10 repeats with a T at the 3' end). The colonies considered positive were those that showed an additional smaller band when the PCR product was run in an agarose gel.

Of the 432 colonies screened, 71 were positive and were sequenced directly from the PCR products using a Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems), following the cycle sequencing conditions: 25 cycles of 10 s at 96ºC, 5 s at 50ºC and 4 minutes at 60ºC. The fragments were separated on an ABI 377 Sequencer. Fifty nine colonies had repeat sequences (14% enrichment).

2.2.5. Primer design and testing

Ten primer pairs were designed using the software programme OMIGA™ 2.0 (Oxford Molecular Ltd.). Amplification of microsatellite loci was performed on a Gene Amp 9700

Perkin Elmer thermal cycler under the following conditions: 2 minutes at 94ºC, then 30 cycles of 30 s at 94ºC, 40 s at 58–62ºC (Table 2.1) and 1 minute at 72ºC, followed by an extension period of 72ºC for 10–30 minutes. Optimised PCR reaction mixes contained approximately 100 ng of template DNA, 1.5 or 2 mM $MgCl_2$ (Table 2.1), 0.2 mM dNTPs, 0.2 μM of each primer (forward primer 5' end-labelled with a fluorescent dye), 0.25 U of Taq polymerase (GIBCO Life Technologies) and the manufacturer's buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl) in a final reaction volume of 10 μl. Amplified microsatellites were run on polyacrylamide gels (Gene-PAGE, Amresco) on an ABI 377 Automated Sequencer using an internal lane standard (GENESCAN-350 TAMRA from PE Applied Biosystems).

A total number of 63 specimens of *C. barbara* and *C. orni* (Table 2.1) were analysed to find the degree of polymorphism for each microsatellite locus. Six out of the ten loci tested were polymorphic in *C. barbara* and four of these were also polymorphic in *C. orni* (Table 2.1). All loci were obtained from *C. barbara* library except Cio08. The average number of alleles ($\pm$ standard deviation) was 7.2 $\pm$ 1.94 for *C. barbara* and 8.5 $\pm$ 2.38 for *C. orni*. One of the loci (Cib03) had non-overlapping allele size ranges between the species. In other loci (Cib01, Cib07 and Cio08) there were some exclusive alleles for each species. Observed heterozygosity was significantly lower than expected for all except one species-locus combination (that of *C. barbara* for Cib06).

Table 2.1. Characterization of *Cicada* L. microsatellite loci. The motif, primer sequences, annealing temperatures (T_a) and $MgCl_2$ concentrations are given for each locus. Allele size ranges, number of alleles and heterozygosity (observed – H_o and expected – H_e) were determined from genotyping 31 specimens of *C. barbara* (Cb) from seven sites in the Iberian Peninsula (Foz Côa, Crato, Monforte, Arrábida, Alcalar, Toledo and Cordoba) and one site in Northwest Africa (Ceuta), and 32 specimens from *C. orni* (Co) from six sites in the Iberian Peninsula (Lisboa, Monte da Caparica, Sesimbra, Arrábida, Crato and Monforte) and one site in Greece (Athens).

Locus	Repeat in sequenced allele	Primer sequence (5' to 3')	GenBank acession no.	T_a (ºC)	$MgCl_2$ (mM)	Allele size range (bp)		Number of alleles		H_o (H_e)	
						Cb	Co	Cb	Co	Cb	Co
Cib01	$(GA)_{18}$	F: GATAAAATCAGTGGAGTGC R: AGTCGATACAATCGAACC	AF437631	58	2	164–194	162–208	10	6	0.759 (0.828)	0.120 (0.607)
Cib03	$(GT)_2T(GT)_{13}$	F: ATATCTGATGGACCCTCG R: AGGTATCATGCCTTATTGC	AF437632	62	2	257–267	235–253	5	7	0.333 (0.625)	0.607 (0.759)
Cib06	$(GT)_2TT(GT)_{11}TGT$	F: CGCAGCAGACGATTTTATCC R: GCTCAATAACTGCCCATAACG	AF437633	60	1.5	256–294	-	8	-	0.793 (0.849)	-
Cib07	$(GT)_{15}$	F: TTGGAATTCAGAGTGTCG R: GTGTCCTGTGTATCCTACG	AF437634	60	1.5	133–165	129–197	7	11	0.160 (0.706)	0.484 (0.740)
Cio08	$(CA)_7TA(CA)_5$	F: CGATGGTTGTAATTTAGTGG R: TCAATAGCATATCTTGCTCC	AF437636	60	1.5	201–217	181–213	8	10	0.667 (0.811)	0.241 (0.735)
Cib10	$(CA)_{16}$	F: AGAAGAAGGTGGACAACC R: TGACGACTTGAAAACAGC	AF437635	60	1.5	141–165	-	5	-	0.160 (0.675)	-

2.3. Microsatellite analysis

The six microsatellites characterized previously were used to describe the genetic structure of the two species *Cicada barbara* and *C. orni*. For comparative purposes, two populations of two other species of the same genus found in Greek islands were also included in the analysis: *C. cretensis* Quartau & Simões, and *C. mordoganensis* Boulard.

Large deviations from Hardy-Weinberg equilibrium were found at almost every microsatellite locus in these species and the most probable explanation is the presence of null alleles (see 2.3.2.a). Since time and financial constraints did not allow new primer design, a statistical approach was taken, in a single-locus analysis (see 2.3.1.3.a.), adjusting the data to take into account the presence of null alleles and comparing the results in terms of variability and differentiation obtained with the adjusted dataset with that from the original data.

A multi-locus analysis was also carried out (see 2.3.1.3.b) in order to describe the spatial and temporal genetic variation in these species, and to find if there was any evidence of hybridization between the species.

2.3.1. Material and methods

2.3.1.1. Sampling sites

A total of 589 specimens of *C. barbara* (nine populations from the Iberian Peninsula and three from Northwest Africa), 475 of *C. orni* (seven populations from the Iberian Peninsula, two from France and four from Greece), 38 of *C. cretensis* and 38 of *C. mordoganensis* (both populations from Greece) were analysed (Figure 2.1, Table 2.2). To enable temporal analysis, sampling was carried out during more than one emergence season in some localities: Crato, Portel and Sousel for *C. barbara* and Portel, Algeciras, Narbonne and St. Hippolyte for *C. orni* (Table 2.1).

2.3.1.2. Sample processing

Cicadas were preserved in 96% ethanol (53% of the samples), frozen (43%) or dry (4%) (Table 2.2). Total DNA was extracted from thorax muscle tissue (76% of the samples) or from one of the legs (24%) using an extraction kit [QIAamp® DNA Mini Kit (QIAGEN) or NucleoSpin® Tissue kit (MACHEREY-NAGEL) (68% of the samples)], or a method adapted from Livak (1984) (32% of the samples): each sample was homogenized on 150 μl of lysis

buffer (10% SDS/5M NaCl/0.5M EDTA/0.5M Tris-HCl) and 9 μl Proteinase K (10mg/ml), incubated at 55ºC overnight; 100 μl of 3M potassium acetate was added and the homogenate was incubated on ice for one hour, and centrifuged for 10 minutes at 13 000 rpm; 220 μl supernatant was transferred to a new tube and the DNA was precipitated with 440 μl of 100% ethanol, incubated at -20ºC for two hours and centrifuged for 10 minutes at 13 000 rpm; the pellet was washed with 400 μl of cold 70% ethanol, dried and resuspended in 50 μl of water.

The six loci previously isolated were amplified as described in 2.2.5, except for Cib07, which was reoptimised as follows: cycles – 2 minutes at 94ºC, then 30 cycles of 30 s at 94ºC, 40 s at 58ºC and 1 minute at 72ºC, followed by an extension period of 72ºC for 20 minutes; mix – 2 mM $MgCl_2$, 0.3 mM dNTPs, 0.25 μM of each primer, 0.25 U of Taq polymerase (GIBCO Life Technologies) and the manufacturer's buffer at 1.5x (20 mM Tris-HCl pH 8.4, 50 mM KCl) in a final reaction volume of 10 μl. For cases where other loci/individuals did not amplify, similar PCR adjustments were made, allowing the amplification of loci for some individuals but only in a minority of cases. Amplified microsatellites were run on polyacrylamide gels (Gene-PAGE, Amresco) on an ABI 377 Automated Sequencer using an internal lane standard (GENESCAN-350 TAMRA from PE Applied Biosystems). Allele sizes were obtained using GENESCAN® ANALYSIS version 3.1.2 software and the electropherograms were analysed on GENOTYPER® version 2.5 software.

Figure 2.1. Sampled populations of each species (● – allopatric populations of *Cicada barbara*; ○ – allopatric populations of *C. orni*; ▲ – sympatric populations of *C. barbara* and *C. orni*; ■ – population of *C. cretensis*; □ – population of *C. mordoganensis*).

Table 2.2. Number of cicadas in each sampled population of *Cicada barbara*, *C. orni*, *C. cretensis* and *C. mordoganensis*. Percentage of samples of each population stored with each type of method, of each type of tissue, and extracted with each type of extraction method.

Population	Locality	Year	N	Storage			Tissue		Extraction	
C. barbara				Dry	Eth	Froz	Leg	Muscle	Livak	Kit
	Iberian Peninsula									
CbAlc95	Alcalar (Algarve, Portugal)	1995	22	0	0	100	0	100	68	32
CbCas96	Casalinho (Arrábida, Portugal)	1996	23	0	0	100	0	100	35	65
CbCra95	Crato (Alto Alentejo, Portugal)	1995	10	0	0	100	0	100	0	100
CbCra96	" " "	1996	23	0	0	100	0	100	13	87
CbCra99	" " "	1999	15	0	40	60	0	100	0	100
CbCra00	" " "	2000	16	0	0	100	0	100	0	100
CbCra01	" " "	2001	35	0	6	94	6	94	94	6
CbCra02	" " "	2002	44	0	100	0	0	100	0	100
CbFoz99	Foz Côa (Trás os Montes, Portugal)	1999	28	0	11	89	0	100	50	50
CbMon95	Monforte (Alto Alentejo, Portugal)	1995	29	0	0	100	0	100	0	100
CbMou01	Moura (Baixo Alentejo, Portugal)	2001	25	0	0	100	0	100	100	0
CbPor01	Portel (Baixo Alentejo, Portugal)	2001	42	0	0	100	0	100	86	14
CbPor02	" " "	2002	35	0	100	0	0	100	0	100
CbSou01	Sousel (Alto Alentejo, Portugal)	2001	41	0	15	85	15	85	85	15
CbSou02	" " "	2002	54	0	100	0	80	20	0	100
CbSev01	Sevilha (Andaluzia, Spain)	2001	32	0	50	50	3	97	81	19
	Northwest Africa									
CbCeu99	Ceuta (north coast of Morocco)	1999	35	0	29	71	0	100	69	31
CbFes01	Fes (Morocco)	2001	53	0	100	0	32	68	40	60
CbMek01	Meknes (Morocco)	2001	27	0	100	0	0	100	67	33
C. orni										
	Iberian Peninsula									
CoAlt98	Alter-do-Chão (Alto Alentejo, Portugal)	1998	21	0	0	100	0	100	43	57
CoCra01	Crato (Alto Alentejo, Portugal)	2001	22	0	9	91	0	100	91	9
CoMte95	Monte-da-Caparica (Área Grande Lisboa, Portugal)	1995	12	0	0	100	0	100	50	50
CoPie96	Piedade (Arrábida, Portugal)	1996	10	0	0	100	0	100	50	50
CoPor01	Portel (Baixo Alentejo, Portugal)	2001	18	0	6	94	0	100	56	44
CoPor02	" " "	2002	33	0	100	0	0	100	0	100
CoSou02	Sousel (Alto Alentejo, Portugal)	2002	41	0	100	0	2	98	0	100
CoAlg01	Algeciras (Andaluzia, Spain)	2001	30	0	100	0	0	100	70	30
CoAlg02	" " "	2002	32	0	100	0	78	22	0	100
	South of France									
CoNar01	Narbonne (Languedoc-Roussillon, France)	2001	39	0	100	0	18	82	62	38
CoNar02	" " "	2002	39	0	100	0	0	100	0	100
CoStH01	St. Hippolyte (Languedoc-Roussillon, France)	2001	26	0	100	0	50	50	50	50
CoStH02	" " "	2002	42	0	100	0	0	100	0	100
	Greece									
CoAte97	Athens (Greece)	1997	36	0	0	100	44	56	0	100
CoKit02	Kithira Island (Greece)	2002	28	86	14	0	100	0	0	100
CoNax99	Naxos Island (Greece)	1999	21	48	52	0	100	0	0	100
CoSky02	Skyros Island (Greece)	2002	25	72	28	0	100	0	0	100
C. cretensis										
CcCre00	Crete Island (Greece)	2000	38	0	100	0	100	0	0	100
C. mordoganensis										
CmSam97	Samos Island (Greece)	1997	38	0	0	100	74	26	0	100

2.3.1.3. Data analysis

2.3.1.3.a) Single-locus analysis

Deviations from Hardy-Weinberg equilibrium for each locus and population were assessed using the exact probability tests available in GENEPOP Version 3.1d (Raymond & Rousset, 1995). Since there were significant heterozygote deficiencies detected for a number of loci and populations, the data were checked for the presence of genotyping errors using MICRO-CHECKER Version 2.2.1 (Van Oosterhout *et al.*, 2004). This program detects the presence of null alleles and scoring errors due to allelic stuttering or large-allele dropout from the comparison of observed and expected homozygote frequencies for each allele class, from the distribution of excess homozygotes over the allele classes and from allele size differences in heterozygotes (Van Oosterhout *et al.*, 2004). For the cases where a null allele was detected, frequencies were calculated as maximum likelihood estimates using the same program.

Due to the high number of non-amplified samples in some locus/population combinations and since these results were repeated after several different PCR trials and conditions it was expected that at least some genotypes would be null allele homozygotes. Therefore, the estimator chosen was Brookfield's (1996) estimator 2. However the question of how many of the non-amplified samples are null homozygotes remains because additional amplification problems may have occurred. Therefore, adjusted allele and genotype frequencies were calculated (using MICRO-CHECKER) considering three alternatives:

1 – all the non-amplified samples in each population (for which there was null allele evidence) were null homozygotes – dataset MAX;

2 – half of the non-amplified samples in each population are null homozygotes and the remaining are the result of other amplification problems – dataset MED;

3 – none of the non-amplified samples are null homozygotes (in this last case, Brookfield estimator 2 corresponds to Brookfield's (1996) estimator 1) – dataset MIN.

For each of these alternatives, a new matrix with the adjusted genotypes in each population/locus was made and the adjusted matrices were used in the data analysis that followed, along with the original matrix (dataset ORIG). For construction of the input files for the analysis programs, the null allele was scored as a value not present in the original data, so these datasets could not be used for allele size-based analysis. Also, adjusted datsets could not be used in multi-locus analysis, since adjusted genotypes including the null allele cannot be assigned to a particular individual and, therefore, it is not possible to generate a multi-locus genotype.

To rule out possible influences of sample storage method, of the tissue extracted and of the extraction method on the amplification of the loci, logistic regressions were applied, creating a model for each locus where the dependent binary variable was the amplification success/failure of the sample for that locus and the independent variables were the categorical variables storage, tissue, type of extraction and species, recoded as dummy variables. The significance of each variable in the model was assessed at the 0.05 level and the effect was evaluated by the positive or negative value of the corresponding parameter estimate. This analysis was performed using R (R Development Core Team, 2003).

To assess if the high proportions of non-amplified samples (NAS) were related to the deviations from Hardy-Weinberg equilibrium, Spearman correlations (r_S) were calculated between the proportion of NAS and F_{IS} using SPSS for Windows 10.0.1 (SPSS Inc, 1999). Exact tests for genotypic linkage disequilibrium were performed in GENEPOP using the original dataset. Using the four datasets (ORIG, MIN, MED and MAX), measures of population genetic variability for each microsatellite locus (number of alleles per locus and expected heterozygosity) were calculated using GENETIX Version 4.02 (Belkhir, Laboratoire Génome, Populations, Interactions, CNRS Montpellier). Friedman nonparametric tests for related samples were performed in SPSS to compare the measures of variability among datasets for each locus. Wilcoxon two-sample tests were applied where Friedman tests were significant.

Measures of variability were compared among *C. barbara* and *C. orni* and among regions for each species using nonparametric tests for independent samples (Mann-Whitney *U* tests) for each locus and for each dataset using SPSS. Single-locus F_{ST} estimates among populations of *C. barbara* and among populations of *C. orni* were calculated as θ (Weir & Cockerham, 1984) for all datasets using GENETIX. Using the same software, the correlations between pairs of F_{ST} matrices for the datasets for each locus were computed via Mantel tests and the significance was obtained by 1000 permutations of populations. Population differentiation was estimated within and among regions for *C. barbara* and for *C. orni* with a locus-by-locus AMOVA (Analysis of Molecular Variance) using ARLEQUIN 3.01 (Excoffier *et al.*, 2005) for each dataset. The significance of covariance components and of the associated fixation indices were obtained after 1000 permutations.

2.3.1.3.b) Multi-locus analysis

The previous analysis only allowed a single-locus approach. The multilocus data were also analysed, excluding the loci with null alleles whenever the assumptions of the analysis required loci in Hardy-Weinberg equilibrium.

The average number of alleles per locus for each population was estimated by bootstrapping 1000 times, using AGARst (Harley, University of Cape Town, South Africa, 2001), to account for sample size. Some populations were excluded due to small sample size (CbCra95, N=10; CoMte95, N=12; CoPie96, N=10). CoKit02 was excluded due to very low number of amplified individuals for all loci except Cib01. The average expected heterozygosity (across loci) was calculated for each population for two datasets: one with data for all loci and another with data for only those loci in Hardy-Weinberg equilibrium (Cib01, Cib06 and Cio08 for *C. barbara* and Cib03 and Cib07 for *C. orni* – see 2.3.2a). Spearman correlations between results from both datasets were calculated for *C. barbara* and *C. orni* separately. Differences in expected heterozygosity and in the average number of alleles per locus (bootstrapped) between regions were assessed by Mann-Whitney *U* tests using SPSS. Dunn Sidak's method for the correction of significance values in multiple tests (Dytham, 2003) was used by reducing the critical P value from 0.05 to 1-(0.951/k), where k is the number of tests performed. Allele frequencies for each population and species were calculated using GENETIX 4.05 (Belkhir *et al.*, 1996-2004).

Genetic differentiation among populations was estimated with θ (Weir & Cockerham, 1984), a relatively unbiased estimator of Wright's F_{ST}, using GENETIX. The significance of F_{ST} values was tested using 1000 permutations of multilocus individuals between each pair of populations.

In order to test if the model of "isolation by distance" applies to these populations, the correlation between the matrix of population-pairwise $F_{ST}/(1-F_{ST})$ values and the natural logarithm (ln) of geographical distances (Rousset, 1997) was analysed using a Mantel test with 1000 permutations of populations. This analysis was done in GENETIX for the dataset of *C. barbara* populations from the Iberian Peninsula and Northwest Africa and for the dataset of *C. orni* from the Iberian Peninsula and Southwest of France. Separate analyses were done for the Iberian Peninsula populations alone.

Factorial Correspondence Analysis (Benzécri, 1973) of the individual multilocus scores, as implemented in the program GENETIX, was used as an exploratory tool to assess the

similarity/dissimilarity between individuals. In this analysis (Belkhir *et al.*, 1996-2004), the genotypic data matrix is converted into a new matrix where a score is attributed to each individual in each allele of each locus (0 for absence, 1 for presence in heterozygote state and 2 for presence in homozygote state). The objects are then the individuals and the variables are the alleles at the different loci. The algorithm finds independent directions which are defined by the eigenvectors of the matrix and which determine the factorial axes, with the first axis being the one with stronger contribution for the total inertia. The coordinates of the individuals in each factor can then be plotted to visualize the similarity/dissimilarity between them. Two datasets were used: one with all *C. barbara* individuals (six loci) and another with *C. orni*, *C. mordoganensis* and *C. cretensis* (four loci).

Population structure and assignment tests were carried out using STRUCTURE 2.1 (Pritchard *et al.*, 2000; Falush *et al.*, 2003) and PARTITION 2.0 (Dawson & Belkhir, 2001). These methods use a model-based approach to define the probability of generating the data assuming K hidden partitions, without using any *a priori* information regarding the origin of the individuals. The multilocus genotype data of the individuals are used to estimate the number of K source populations (partitions) and to assign the individuals to the populations by applying a Bayesian approach. The modelling assumptions are that populations are in Hardy-Weinberg equilibrium and loci are in linkage equilibrium.

In the STRUCTURE analysis, one dataset consisted of the genotypes of all the individuals of *C. barbara* (N=589) for the three microsatellite loci which were in Hardy-Weinberg equilibrium (Cib01, Cib06 and Cio08). The values of K tested ranged from 1 to 19 (the maximum number of populations present, if we consider the different years of capture as independent populations). Another dataset consisted of all West-European individuals of *C. orni* (Iberian Peninsula and France; N=365) for the two microsatellite loci in Hardy-Weinberg equilibrium (Cib03 and Cib07), and the values of K tested ranged from 1 to 13. A model with admixture was used. The runs used 400 000 iterations after a burn-in period of 100 000. Each set was run three times to evaluate the consistency of the runs. The program gives the estimated logarithmic probability of data, ln Pr(X|K), for each K, which allows the estimation of the posterior probabilities of K (p.12 in Pritchard & Wen, 2003). However, since for the *C. barbara* dataset the values of ln Pr(X|K) did not show a clear maximum for one specific K, we calculated the statistic ΔK (a quantity based on the rate of change of the log probability of data with respect to the number of clusters), described by Evanno *et al.* (2005). These authors found this to be a good predictor of the real number of clusters in simulated data. Since this methodology did also not allow a clear definition of K= 3 or 4, runs of of 1.3×10^6 iterations after a burn-in of 200 000 (repeated 10 times) were done for K from 1 to 5.

In PARTITION the posterior probability distribution of the number of source populations K was estimated, from K=1 to a maximum value chosen (Kmax), as well as the posterior co-assignment probabilities (probability that a set of individuals all belong to the same source population). The same microsatellite dataset of *C. barbara* (but excluding the individuals with missing values, N=543) was analysed. The maximum K was set to 19. The number of observations of the Markov chain was 10 000, with 10 iterations between observations. Burn-in was set to 1000. The dataset of *C. orni* could not be used in this analysis due to the great amount of missing values.

Temporal variation within populations was tested with an Analysis of Molecular Variance (AMOVA) using ARLEQUIN 3.01 (Excoffier *et al*., 2005) on the *C. barbara* populations that were sampled in more than one year (Group 1: Crato – CbCra95, CbCra96, CbCra99, CbCra00, CbCra01 and CbCra02; Group 2: Portel – CbPor01 and CbPor02; and Group 3: Sousel – CbSou01 and CbSou02). The significance of covariance components and of the associated fixation indices were obtained after 1000 permutations. A locus-by-locus AMOVA was also carried out. The same analysis was done for *C. orni* populations from the Iberian Peninsula with more than one year of capture (Group 1: Portel – CoPor01 and CoPor02; and Group 2: Algeciras – CoAlg01 and CoAlg02), as well as for *C. orni* populations from France (Group 1: Narbonne – CoNar01 and CoNar02; and Group2: St Hippolyte – CoStH01 and CoStH02).

STRUCTURE was used to test for hidden substructure between years of capture. Crato, Portel and Sousel populations of *C. barbara* (N=315) were analysed for three microsatellite loci (Cib01, Cib06 and Cio08) and the values of K ranged from 1 to 10. The Portel, Algeciras, Narbonne and St Hippolyte populations of *C. orni* (N=259) were analysed for loci Cib03 and Cib07, with K values ranging from 1 to 8.

In order to assess the hypothesis of hybridization between *C. barbara* and *C. orni*, the percentage of private alleles in each species in each of the four common loci was compared between sympatric and allopatric populations of Iberian Peninsula using Wilcoxon signed rank tests in SPSS. Additionally, a Factorial Correspondence Analysis was applied to the data matrix of *C. barbara* and *C. orni* of both sympatric and allopatric populations of the Iberian Peninsula using GENETIX. Particularly useful in detecting hybridization are loci at which parental taxa have very different allele frequencies (Allendorf & Luikart, 2007). In the pair of species *C. barbara*/*C. orni* locus Cib03 may constitute a useful diagnostic locus since it shows almost complete non-overlapping alleles between species.

2.3.2. Results

a) Single-locus analysis

(i) *Hardy-Weinberg analysis*

A large proportion of population/locus combinations (105 of 190) showed a significant heterozygote deficit ($F_{IS}>0$; $p<0.05$) relative to that expected under Hardy-Weinberg equilibrium (Table 2.3). A pattern of species-dependent deviation from equilibrium was apparent for loci Cib01 and Cio08 (for which most *C. orni* populations showed deviations and almost no *C. barbara* did) and in locus Cib03 (for which the opposite occurred). Species-dependent deviation may indicate that the amplification of certain alleles was prevented in one of the species due to one or more substitutions or indels in the annealing site of the primers. Interestingly, in both loci Cib03 and Ci08 deviations occurred in the species in which the locus was originally cloned. For locus Cib07 almost every population of *C. barbara* and half the populations of *C. orni* showed deviation. Cib10 was highly deviated from the equilibrium in all populations of *C. barbara*. Additionally, in Cib07 for *C. barbara* and in Cib01 and Cio08 for *C. orni*, significant positive correlations were found between the proportion of samples that did not amplify in a population and its F_{IS} value ($r_S=0.699$, $p=0.0009$; $r_S=0.551$, $p=0.022$; and $r_S=0.663$, $p=0.003$, respectively).

Logistic regression models for each locus, except Cib06, indicated that some of the variables considered (storage, tissue, extraction and species) had a significant effect on the probability of amplification, but results were generally not consistent across loci (Table 2.4a). For locus Cib01, the probability of amplification was lower in ethanol than in frozen preserved samples, higher in leg than in thorax samples, lower in *C. cretensis* and higher in *C. barbara*. For locus Cib03, amplification was lower in dry than in frozen samples and higher in *C. barbara*. For locus Cib07, the probability of amplification was lower in dry than in frozen samples, lower in leg than muscle and lower with Livak than with the kit extraction method. In locus Cio08 it was lower in dry than in frozen samples. In locus Cib10 amplification was on this occasion higher in ethanol preserved than in frozen samples.

Table 2.3. Values of F_{IS} calculated for each population/locus combination. Shaded values are statistically significant: *p<0.05; **p<0.01; ***p<0.001. N – evidence of null alleles detected by the program MICRO-CHECKER; S – evidence of stuttering detected by the same program.

	Loci											
Population	**Cib01**		**Cib03**		**Cib06**		**Cib07**		**Cio08**		**Cib10**	
CbAlc95	+0.141		+0.661***	N	+0.134		+0.671***	N	+0.127		+0.526***	N
CbCas96	-0.109		+0.684***	NS	+0.282	N	+0.357		+0.300		+0.781***	N
CbCra95	-0.029		+0.660*	N	+0.138		+0.738***	N	+0.027		+0.855***	N
CbCra96	-0.027		+0.597**	NS	+0.047		+0.499***	N	-0.015		+0.772***	N
CbCra99	+0.067		+0.461		+0.198		+0.895***	N	+0.186		1***	N
CbCra00	+0.098		+0.326		+0.309*	N	+0.860***	N	-0.083		1***	N
CbCra01	-0.055		+0.427**	NS	-0.036		+0.791***	N	+0.348***	NS	+0.671***	NS
CbCra02	+0.020		+0.408***	N	-0.076		+0.606***	N	+0.138		+0.429***	N
CbFoz99	-0.043		+0.515**	N	+0.052		+0.770***	N	+0.220	N	+0.593***	N
CbMon95	-0.075		+0.348		+0.064*		+0.855***	N	+0.043		+0.603***	N
CbMou01	+0.078		+0.494**	N	-0.062		1***	N	+0.153*		+0.523***	N
CbPor01	-0.121		+0.482***	NS	+0.155	N	+0.857***	N	-0.062		+0.616***	N
CbPor02	+0.066		+0.679***	NS	+0.098		+0.676***	N	+0.044		+0.733***	N
CbSou01	-0.012		+0.357**	N	+0.035		+0.951***	N	+0.128		+0.670***	NS
CbSou02	-0.025		+0.446***	N	-0.010		+0.594***	N	+0.016*		+0.798***	N
CbSev01	+0.050		+0.645***	NS	+0.108		1***	N	-0.025		+0.589***	N
CbCeu99	-0.050		+0.552***	N	+0.263**	N	+0.742***	N	+0.103*		+0.443***	N
CbFes01	+0.124	N	+0.594***	N	+0.181*	N	+0.939***	NS	+0.061*		+0.611***	NS
CbMek01	-0.070		+0.643***	N	+0.057		+0.836***	N	+0.010		+0.754***	N
CoAlt98	1*		+0.329	N			+0.188		+0.415**	N		
CoCra01	+845***	N	-0.043				+0.488***	N	+0.589***	N		
CoMte95	-0.029		+0.054				+0.273		1**	NS		
CoPie96	-		+0.031				+0.252		+0.308*			
CoPor01	+0.782**	N	+0.409**	N			+0.256		+0.426*	N		
CoPor02	+0.280*	N	+0.173				+0.399**	N	+0.681***	NS		
CoSou02	+0.331***	N	-0.005				+0.200***	N	+0.605***	N		
CoAlg01	+0.382**	N	-0.141				+0.372**	N	+0.567***	N		
CoAlg02	+0.465**	N	-0.160				+0.464***	N	+0.102			
CoNar01	+0.796***	N	-0.075				+0.060		+0.544***	N		
CoNar02	+0.769***	N	+0.082*				+0.135		+0.659***	NS		
CoStH01	+0.736***	NS	-0.158				+0.091		+0.797***	N		
CoStH02	+0.729***	N	+0.058				+0.167*		+0.510***	NS		
CoAte97	+0.277*	N	+0.032				+0.167*	NS	+0.729***	N		
CoKit02	+0.085		+0.020				-		+0.833***	N		
CoNax99	+0.640**	N	+0.439*	N			+0.152		+0.457***	N		
CoSky02	+0.254		+0.721***	NS			+0.144*		+0.788***	NS		
CcCre00	+0.600		-0.007				+0.218*	N	+0.038			
CmSam97	+0.188		+0.069				+0.391***	N	+0.007			

Dried samples thus presented lower amplification than other storage methods for three out of four loci (the remaining two loci were only amplified in *C. barbara*, which did not have any dried samples). Kithira, Naxos and Skyros island samples (*C. orni*) were the only populations with dried specimens, which could explain the high proportions of non-amplified samples for three out of four loci in Kithira population (79% of non-amplified samples in both loci Cib03 and Cio08 and 86% in locus Cib07). Separate logistic regressions for these populations were carried out and loci Cib03 and Cio08 showed a lower probability of amplification with dried samples than with ethanol samples (Table 2.4b), but no effect of storage was detected for loci Cib01 and Cib07. The probability of amplification was different among islands and was higher in Skyros and Naxos than in Kithira for loci Cib03 and Cib07 and higher in Skyros than in Kithira for locus Cio08 (Table 2.4b). Another case of locus-dependent amplification was for the *C. cretensis* population, which had a high proportion of non-amplified samples (92%) at one of the loci (Cib01), whereas showing a proportion of non-amplified samples ranging from 5 to 8% in the other three loci.

Table 2.4. Estimated coefficients β and significance values of the logistic regression models of the binary variable amplification success/failure of the locus onto the variables type of storage, type of tissue, type of extraction and species (coded as dummy variables). Significance levels: *p<0.05; **p<0.01; ***p<0.001.

a) Analysis with all cicadas

	locus Cib01		**locus Cib03**		**locus Cib06**	
Variables	β	p	β	p	β	p
Intercept	1.89	0.0029 **	1.69	0.0018**	3.75	8.13e-16***
Storage Eth	-1.08	0.0002 ***	-0.40	0.1810	-0.28	0.5523
Storage Dry	-0.19	0.7693	-2.91	3.68e-08***		
Tissue Leg	0.86	0.0067 **	0.28	0.4314	-0.38	0.6062
Extraction Livak	-0.07	0.7924	0.21	0.4933	-0.91	0.0627
Species *C. cretensis*	-4.13	5.35e-06 ***	1.32	0.1513		
Species *C. orni*	-0.07	0.9149	1.02	0.0965		
Species *C. barbara*	2.69	0.0003 ***	1.21	0.0410*		

	locus Cib07		**locus Cio08**		**locus Cib10**	
Variables	β	p	β	p	β	p
Intercept	2.47	1.27e-07***	17.20	0.9785	1.57	2.93e-13***
Storage Eth	-0.08	0.6548	0.039	0.8720	0.64	0.0371*
Storage Dry	-1.66	2.28e-05***	-2.18	1.62e-06***		
Tissue Leg	-1.22	7.64e-08***	0.52	0.0988	-0.43	0.3139
Extraction Livak	-1.01	6.64e-08***	0.44	0.0965	0.34	0.2225
Species *C. cretensis*	1.29	0.0895	-14.87	0.9814		
Species *C. orni*	-0.06	0.8986	-16.09	0.9799		
Species *C. barbara*	-0.74	0.1236	-13.99	0.9825		

b) Analysis with the cicadas *C. orni* from the Greek islands Kithira, Naxos and Skyros

	locus Cib01		locus Cib03	
Variables	β	P value	β	P value
Intercept	2.19	0.0308 *	0.14	0.8621
Storage Dry	0.45	0.6023	-1.79	0.0280 *
Island Naxos	-0.59	0.5632	2.09	0.0044 **
Island Skyros	-0.05	0.9590	4.55	6.79e-05 ***

	locus Cib07		locus Cio08	
Variables	β	P value	β	P value
Intercept	-1.81	0.0154 *	0.46	0.5289
Storage Dry	0.02	0.9782	-2.22	0.0018 **
Island Naxos	1.51	0.0395 *	0.96	0.1862
Island Skyros	2.55	0.0002***	2.53	0.0004 ***

When combined, these results strongly suggested the presence of null alleles in the dataset and MICRO-CHECKER detected evidence for null allele presence for most of the populations/loci for which they were previously suspected to occur by F_{IS} value (Table 2.3). Allelic stuttering was implicated as a possible cause of scoring errors in some cases, but was not a general problem (Table 2.3) and no evidence of large-allele dropout was found. The estimated null allele frequencies, for those populations that showed evidence for null alleles, ranged from 0.066 to 0.950 with the Brookfield 2 estimator. Comparing with the frequencies of the most common allele in the original data (Table 2.5), these values are sometimes extremely high, suggesting that either the null allele is the most frequent or that several classes of null alleles could be present. Generally, the adjusted datasets presented substantially lower deviations from Hardy-Weinberg equilibrium (*i.e.*, lower F_{IS} values) than the original dataset, although some significant values remained (Appendix I). In some cases the deviations were due to heterozygote excess instead of the initial heterozygote deficits.

(ii) *Molecular characteristics*

Inspecting the sequences of the primers, some contained repetitive elements: AAAA in Cib01*F*, ATAT in Cib03*F*, GCAGCA and TTTT in Cib06*F*, GTGT in Cib07*F* and in Cib07*R* (twice), GAAGAA in Cib10*F* and AAAA in Cib10*R*. It is possible that some of these motifs are more affected by insertion/deletions (Jones *et al.*, 1998) but no correlation could be seen with the presence of null alleles since only Cio08 did not include repetitive elements. It has been suggested that the sequences closer to a microsatellite are more prone to mutations due to slippage events in the microsatellite (Pemberton *et al.*, 1995). Examination of the number of bases that separate both priming sites from the microsatellites (Cib01: 38+63=101; Cib03: 188+4=192; Cib06: 36+153=189; Cib07: 34+43=77; Cio08: 59+77=136; Cib10: 51+37=88), revealed that Cib07 and Cib10 possessed the lowest global values and are the loci with highest F_{IS} values (Table 2.3).

Table 2.5. Estimated null allele frequencies using the Brookfield 2 algorithm in MICRO-CHECKER (Null) and frequencies of the most common allele in the original data (MCA).

	Loci											
	Cib01		Cib03		Cib06		Cib07		Cio08		Cib10	
Population	**Null**	MCA	**Null**	MCA	**Null**	MCA	**Null**	MCA	**Null**	MCA	**Null**	MCA
CbAlc95			**0.380**	0.600			**0.459**	0.342			**0.296**	0.333
CbCas96			**0.384**	0.595	**0.199**	0.296					**0.620**	0.625
CbCra95			**0.389**	0.444			**0.438**	0.333			**0.323**	0.450
CbCra96			**0.200**	0.500			**0.345**	0.262			**0.516**	0.526
CbCra99							**0.586**	0.333			**0.663**	0.364
CbCra00					**0.458**	0.292	**0.752**	0.389			**0.541**	0.500
CbCra01			**0.260**	0.742			**0.576**	0.407	**0.248**	0.348	**0.426**	0.452
CbCra02			**0.279**	0.646			**0.477**	0.583			**0.291**	0.500
CbFoz99			**0.311**	0.558			**0.530**	0.435	**0.219**	0.250	**0.436**	0.479
CbMon95			**0.113**	0.500			**0.592**	0.591			**0.570**	0.381
CbMou01			**0.167**	0.500			**0.836**	0.333			**0.344**	0.326
CbPor01			**0.224**	0.500	**0.066**	0.250	**0.682**	0.393			**0.363**	0.423
CbPor02			**0.339**	0.576			**0.379**	0.288			**0.354**	0.412
CbSou01			**0.274**	0.526			**0.749**	0.360			**0.612**	0.241
CbSou02			**0.230**	0.683			**0.504**	0.488			**0.470**	0.500
CbSev01			**0.283**	0.629			**0.836**	0.467			**0.301**	0.387
CbCeu99			**0.349**	0.547	**0.292**	0.2742	**0.473**	0.583			**0.361**	0.210
CbFes01	**0.130**	0.206	**0.224**	0.740	**0.214**	0.296	**0.665**	0.295			**0.375**	0.615
CbMek01			**0.401**	0.604			**0.699**	0.529			**0.529**	0.432
CoAlt98			**0.132**	0.357					**0.380**	0.417		
CoCra01	**0.490**	0.806					**0.277**	0.405	**0.429**	0.421		
CoMte95									**0.753**	0.429		
CoPie96												
CoPor01	**0.570**	0.808	**0.464**	0.464					**0.469**	0.429		
CoPor02	**0.323**	0.672					**0.222**	0.453	**0.522**	0.596		
CoSou02	**0.525**	0.690					**0.181**	0.320	**0.452**	0.343		
CoAlg01	**0.512**	0.841					**0.395**	0.260	**0.417**	0.481		
CoAlg02	**0.441**	0.750					**0.417**	0.222				
CoNar01	**0.552**	0.683							**0.435**	0.379		
CoNar02	**0.782**	0.722							**0.698**	0.326		
CoStH01	**0.620**	0.611							**0.502**	0.523		
CoStH02	**0.719**	0.717							**0.579**	0.466		
CoAte97	**0.173**	0.429					**0.128**	0.214	**0.576**	0.179		
CoKit02									**0.950**	0.167		
CoNax99	**0.409**	0.833	**0.489**	0.500					**0.669**	0.208		
CoSky02			**0.266**	0.833					**0.588**	0.447		
CoCre00							**0.222**	0.143				
CmSam97							**0.424**	0.564				

No significant associations between alleles in different loci for each population were found consistently among locus pairs. Ten significant values from 285 comparisons (six loci in 19 populations) in *C. barbara* and two from 102 (four loci in 17 populations) in *C. orni* were significantly lower than the 5% expected by chance. No significant associations were found in *C. cretensis* and *C. mordoganensis*.

(iii) *Microsatellite variability*

The total number of alleles per locus in the original dataset ranged from 16 to 35. Loci Cib06 and Cib10 were monomorphic in *C. orni*, *C. cretensis* and *C. mordoganensis* with the same allele sizes in the three species (252 bp for Cib06 and 185 bp for Cib10).

The population expected heterozygosity and number of alleles (Appendix II) showed significant differences among ORIG, MIN, MED and MAX datasets for loci Cib03, Cib06, Cib07 and Cib10 for *C. barbara* (Friedman tests, $p<0.05$). For *C. orni*, the only non-significant result was found for expected heterozygosity for locus Cib07. The differences in heterozygosity were not due to any dataset in particular across loci (Wilcoxon tests). The difference in number of alleles was expected since the adjusted datasets included one new allele wherever the presence of a null allele was suspected.

When comparing measures of variability among species and regions within species for each locus, the results were not always consistent among datasets (Table 2.6). Furthermore, the results were not consistent among loci – Cib01 and Cio08 showed higher variability in *C. barbara* than in *C. orni* and the opposite was found for Cib03 and Cib07. Comparing the variability between Iberian and African populations of *C. barbara*, the results were consistent among datasets – higher variability in Africa than in Iberian Peninsula for locus Cib01 (both number of alleles and heterozygosity) and for locus Cio08 (in number of alleles). All the remaining tests were not significant (Table 2.6). In the case of *C. orni*, the results on the comparison of variability among regions were again not consistent among datasets or loci (Table 2.6).

(iv) *Population differentiation*

The single-locus F_{ST} estimates were highly correlated among all datasets ORIG, MIN, MED and MAX [correlation values ranged from 0.620 to 1 in *C. barbara* (19 populations) and from 0.613 to 0.998 in *C. orni* (17 populations); Mantel tests, $p<0.01$]. Average F_{ST} values for each dataset and each locus are shown in Appendix III.

From the AMOVA (Table 2.7), most of the genetic variation in *C. barbara* was found within populations for all loci (>80%). For loci Cib01 and Cib03 the differentiation among regions was significant for all datasets ($p<0.05$) and the percentage of variation varied from 5.7 to 12.1%, higher than the within regions differentiation. For the remaining loci the among regions differentiation was very low (<2%) and lower than the within regions differentiation.

Table 2.6. P-values of nonparametric tests comparing the measures of variability (number of alleles – Na, and expected heterozygosity – H_e) among species (Cb – *C. barbara* and Co – *C. orni*) and among regions for each locus for each dataset (Mann-Whitney tests for two-sample comparisons and Kruskal-Wallis tests for multiple-sample comparisons). Significance levels: *p<0.05; **p<0.01; ***p<0.001. Relations between average ranks indicated between parenthesis. For *C. orni* regions, Mann-Whitney tests were applied between pairs of regions when Kruskal-Wallis tests were significant.

Locus	**Measure of variability**	**Dataset**	***C. barbara* (N=19) *vs. C. orni* (N=17)**	***Cicada barbara***	***Cicada orni***
				Iberian Peninsula (N=16) *vs.* Africa (N=3)	Iberian Peninsula (N=9) *vs.* France (N=4) *vs.* Greece (N=4)
Cib01	Na	ORIG	0.0054** (Cb>Co)	0.0021** (IbPen<Africa)	0.315
		MIN	0.0933	0.0021** (IbPen<Africa)	0.297
		MED	0.0933	0.0021** (IbPen<Africa)	0.297
		MAX	0.0933	0.0021** (IbPen<Africa)	0.297
	H_e	ORIG	0.0000** (Cb>Co)	0.0021** (IbPen<Africa)	0.223
		MIN	0.0000** (Cb>Co)	0.0021** (IbPen<Africa)	0.043 (IbPen<France)
		MED	0.0000** (Cb>Co)	0.0021** (IbPen<Africa)	0.607
		MAX	0.0000** (Cb>Co)	0.0021** (IbPen<Africa)	0.299
Cib03	Na	ORIG	0.0161* (Cb<Co)	0.1094	0.306
		MIN	0.3795	0.1094	0.185
		MED	0.3795	0.1094	0.185
		MAX	0.3795	0.1094	0.185
	H_e	ORIG	0.0000** (Cb<Co)	0.4871	0.059
		MIN	0.3462	0.3591	0.058
		MED	0.5308	0.3591	0.043 (IbPen>France)
		MAX	0.7072	0.8751	0.043 (IbPen>France)
Cib06	Na	ORIG		0.1713	
		MIN		0.1713	
		MED		0.2105	
		MAX		0.2105	
	H_e	ORIG		0.6336	
		MIN		0.4211	
		MED		0.3034	
		MAX		0.4871	
Cib07	Na	ORIG	0.8511	0.7926	0.445
		MIN	0.3969	0.7926	0.444
		MED	0.3969	0.7926	0.444
		MAX	0.3969	0.7926	0.444
	H_e	ORIG	0.8757	0.7121	0.524
		MIN	0.1964	0.4211	0.459
		MED	0.1215	0.7926	0.438
		MAX	0.0090** (Cb<Co)	0.8751	0.438
Cio08	Na	ORIG	0.1139	0.0041** (IbPen<Africa)	0.048 (France<Greece)
		MIN	0.7306	0.0144** (IbPen<Africa)	0.085
		MED	0.7543	0.0144** (IbPen<Africa)	0.085
		MAX	0.7543	0.0144** (IbPen<Africa)	0.085
	H_e	ORIG	0.0252* (Cb>Co)	0.1094	0.068
		MIN	0.8023	0.1383	0.033 [Greece>(IbPen+France)]
		MED	0.0194* (Cb>Co)	0.2105	0.258
		MAX	0.0002** (Cb>Co)	0.2105	0.739
Cib10	Na	ORIG		0.2539	
		MIN		0.2539	
		MED		0.2539	
		MAX		0.2539	
	H_e	ORIG		0.7926	
		MIN		0.7926	
		MED		1.0000	
		MAX		0.7926	

In *C. orni* the differentiation among Iberian Peninsula and France for loci Cib01 and Cib03 was low (percentage of variation ranged from 2.3 to 6.1%) but significant (p<0.05) in all datasets and it was very low (<3%) and non-significant for loci Cib07 and Cio08 in all datasets. The differentiation was higher within regions than among regions for all loci and all datasets except Cib01 and Cib03 in datasest ORIG and MIN and for Cib01 in dataset MED. The differentiation among Western populations (Iberian Peninsula+France) and Greece was generally high (4.6 to 28.2%) and it was significant for all loci and all datasets. The variation among regions was higher than within regions in all loci in all datasets except for loci Cio08 in datasets MED and MAX.

Table 2.7. Percentage of variation among regions (Groups), among populations within regions (A-pop) and within populations (W-pop) obtained from the AMOVA for each dataset (ORIG, MIN. MED and MAX). For *C. barbara* two groups were defined (Iberian Peninsula and Northwest Africa) and for *C. orni* two separate AMOVA were carried out, one comparing Iberian Peninsula and France and another comparing Western Europe (Iberian Peninsula and France populations) and Greece.

		C. barbara				***C. orni***							
		Iberian Peninsula/Africa				Iberian Peninsula/France				Iberian Pen.+France/Greece			
Locus		ORIG	MIN	MED	MAX	ORIG	MIN	MED	MAX	ORIG	MIN	MED	MAX
Cib01	Groups	8.00	8.00	7.96	7.92	4.97	6.07	5.70	5.81	28.15	22.34	17.28	20.91
	A-pop	1.84	1.82	1.85	1.91	3.18	2.99	4.81	8.92	6.86	7.05	8.59	11.10
	W-pop	90.16	90.17	90.20	90.17	91.85	90.94	89.49	85.27	64.99	70.60	74.13	67.99
Cib03	Groups	12.15	6.95	6.23	5.73	2.30	2.32	2.41	2.40	12.54	12.18	11.88	11.64
	A-pop	6.27	4.81	4.91	5.50	1.88	2.11	2.61	3.38	9.08	8.28	8.83	9.83
	W-pop	81.58	88.25	88.86	88.77	95.82	95.57	94.99	94.22	78.38	79.55	79.29	78.53
Cib06	Groups	1.61	1.59	1.65	2.05								
	A-pop	4.87	4.69	4.90	5.30								
	W-pop	93.52	93.72	93.45	92.66								
Cib07	Groups	0.19	-0.46	-0.31	-0.87	-0.20	0.47	2.02	2.86	11.31	10.59	10.24	10.61
	A-pop	8.57	6.01	4.77	7.52	5.72	5.35	6.09	6.47	8.08	8.28	9.38	9.95
	W-pop	91.24	94.45	95.54	93.35	94.48	94.18	91.89	90.67	80.61	81.13	80.38	79.45
Cio08	Groups	0.05	0.03	0.02	-0.03	1.68	1.18	0.86	1.53	10.14	6.68	5.60	4.61
	A-pop	4.33	4.33	4.51	4.86	3.67	3.64	4.85	7.06	5.81	5.50	7.50	11.93
	W-pop	95.62	95.65	95.47	95.17	94.66	95.18	94.29	91.41	84.06	87.83	86.89	83.46
Cib10	Groups	0.51	0.34	0.21	0.03								
	A-pop	4.93	2.56	1.78	2.38								
	W-pop	94.56	97.10	98.01	97.59								

b) Multi-locus analysis

i) Microsatellite variability

Average number of alleles across loci (bootstrapped) ranged from 4.22 to 6.83 in populations of *C. barbara* (six loci), from 3.66 to 6.68 in populations of *C. orni* (four loci), and were 6.10 for *C. cretensis* (four loci) and 5.04 for *C. mordoganensis* (four loci) (Table 2.8). The average expected heterozygosity ranged from 0.603 to 0.769 in *C. barbara*, from 0.545 to 0.805 in *C. orni* and was 0.620 for *C. cretensis* and 0.659 for *C. mordoganensis*. The number of alleles and heterozygosity values for the dataset with all loci and for that containing only the loci in Hardy-Weinberg equilibrium were significantly correlated for *C. barbara* (r_S= 0.857, p<0.001 and r_S=0.649, p=0.004, respectively). For the *C. orni*+*C. cretensis*+*C. mordoganensis* dataset the number of alleles was also highly correlated between the dataset with all loci and the dataset with only the loci under Hardy-Weinberg equilibrium (r_S=0.944, p<0.001). Expected heterozygosities were not significantly correlated (r_S=0.494, p=0.052).

Northwest African populations of *C. barbara* had higher values of average number of alleles and higher heterozygosity than Iberian ones in both datasets. The difference was significant for the number of alleles in both datasets (Mann-Whitney tests, p=0.002 for both) and for the expected heterozygosity for the dataset with only the loci in Hardy-Weinberg equilibrium (Mann-Whitney test, p=0.010) but non-significant for the expected heterozygosity for the dataset with all loci (Mann-Whitney test, p=0.164). Iberian, French and Greek populations of *C. orni* did not show significant differerences in average number of alleles and heterozygosity, although it was noticeable that the mainland Greek population of Athens had the highest values of number of alleles and heterozygosity in both datasets (Table 2.8). The average number of alleles for the Kithira population (7.75) was not bootstrapped, due to the high amount of missing data (non-amplifying samples), and this high value is attributed to one of the loci (Cib01), which had 14 alleles.

Allele frequencies for each locus in each species are shown in Figure 2.2. Taking a threshold frequency of 0.05, for the four common loci, 13 private alleles were found in *C. barbara*, nine in *C. orni*, eight in *C. cretensis* and five in *C. mordoganensis*. Considering only the first two species, 17 private alleles were found in *C. barbara* and 10 in *C. orni*. In locus Cib03 the allele size ranges were almost non-overlapping between these species.

Table 2.8. Average number of alleles (bootstrapped) and average expected heterozygosity (H_e) across loci for each population of *Cicada barbara*, *C. orni, C. cretensis* and *C. mordoganensis* using all loci and using only the loci in Hardy-Weinberg equilibrium (for *C. barbara*: loci Cib01, Cib06 and Cio08; and for the other species: loci Cib03 and Cib07). *populations not included in the bootstrapping; presented values for these populations are not bootstrapped.

		All loci		Loci in HW equilibrium	
Population	**N**	**No. alleles**	**H_e**	**No. alleles**	**H_e**
C. barbara					
Iberian Peninsula					
CbAlc95	22	5.75	0.746	6.59	0.790
CbCas96	23	4.60	0.603	5.75	0.741
CbCra95*	*10*	*5.33*	0.730	*6.67*	0.798
CbCra96	23	5.65	0.744	6.79	0.819
CbCra99	15	4.77	0.673	5.92	0.781
CbCra00	16	4.78	0.706	5.77	0.770
CbCra01	35	5.31	0.695	6.69	0.802
CbCra02	44	5.87	0.717	7.11	0.828
CbFoz99	28	5.69	0.730	6.92	0.807
CbMon95	29	4.22	0.662	4.91	0.721
CbMou01	25	5.82	0.755	6.68	0.802
CbPor01	42	5.77	0.728	7.25	0.790
CbPor02	35	5.75	0.742	6.36	0.779
CbSou01	41	5.15	0.715	5.48	0.726
CbSou02	54	5.18	0.681	5.11	0.735
CbSev01	32	5.76	0.737	6.34	0.796
Northwest Africa					
CbCeu99	35	6.83	0.769	8.25	0.846
CbFes01	53	6.43	0.711	7.9	0.816
CbMek01	27	6.37	0.748	8.08	0.846
C. orni					
Iberian Peninsula					
CoAlt98	21	4.19	0.564	4.84	0.707
CoCra01	22	4.15	0.638	4.51	0.738
CoMte95*	*12*	*3.25*	0.590	*3.00*	0.578
CoPie96*	*10*	*3.50*	0.445	*5.00*	0.688
CoPor01	18	3.96	0.617	4.72	0.716
CoPor02	33	3.89	0.619	4.5	0.690
CoSou02	41	4.67	0.691	5.18	0.742
CoAlg01	30	4.05	0.603	4.67	0.719
CoAlg02	32	4.38	0.578	5.84	0.745
Southwest of France					
CoNar01	39	3.66	0.606	4.08	0.617
CoNar02	39	3.90	0.630	4.24	0.663
CoStH01	26	4.17	0.618	4.65	0.612
CoStH02	42	3.74	0.614	4.18	0.654
Greece					
CoAte97	36	6.68	0.805	7.08	0.834
CoKit02*	*28*	*7.75*	0.656	*5.00*	0.498
CoNax99	21	5.62	0.664	6.14	0.748
CoSky02	25	4.01	0.545	4.34	0.509
C. cretensis					
CcCre00	38	6.10	0.620	6.41	0.494
C. mordoganensis					
CmSam97	38	5.04	0.659	6.33	0.727

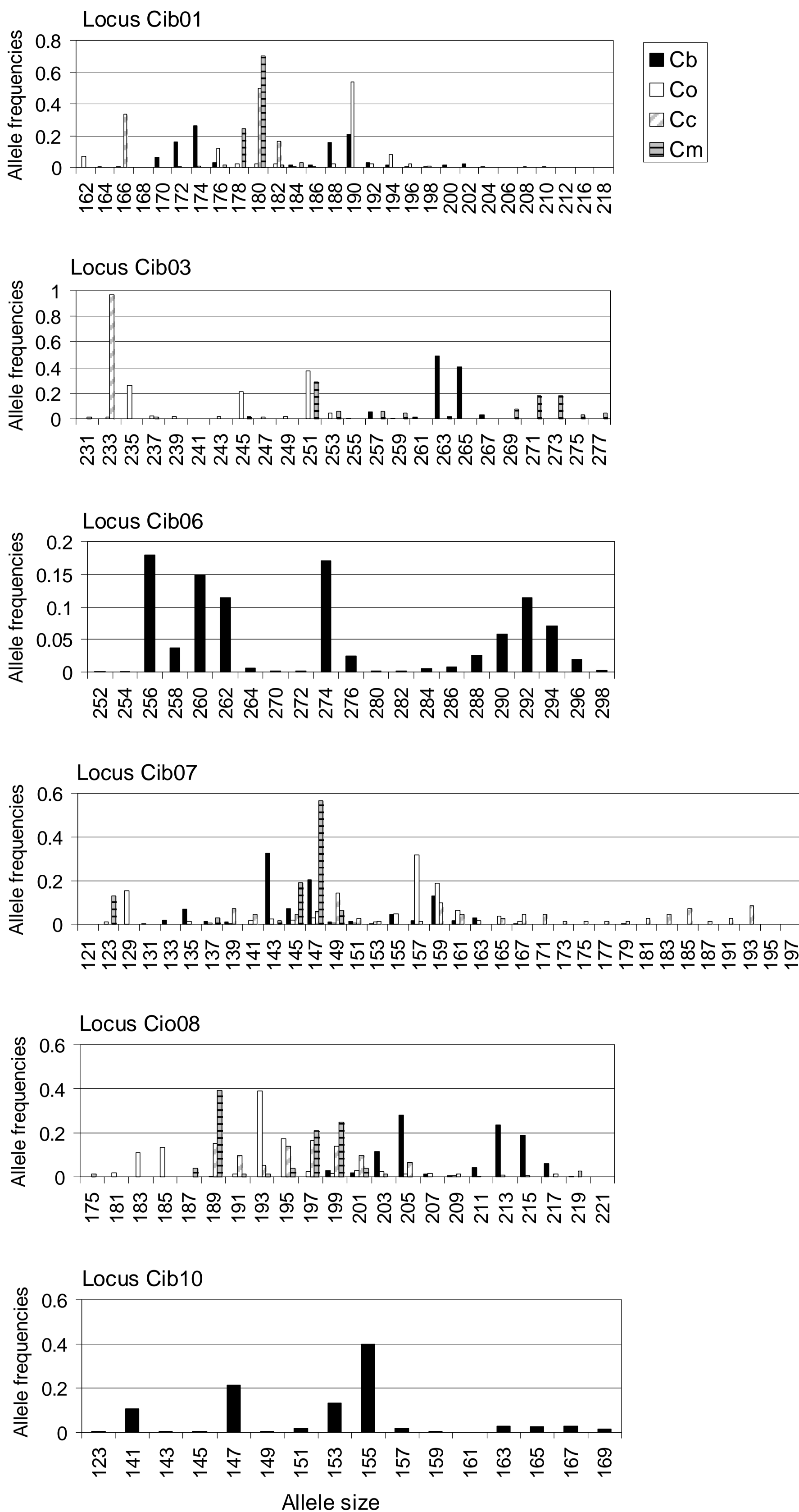


Figure 2.2. Allele frequencies observed in *C. barbara* (Cb), *C. orni* (Co), *C. cretensis* (Cc) and *C. mordoganensis* (Cm) for each microsatellite locus.

ii) Population structure

Multilocus F_{ST} values (calculated using the loci in Hardy-Weinberg equilibrium) between pairs of populations ranged from negative (effectively zero) among Iberian populations of *C. barbara* and among West-European populations of *C. orni* (Iberian Peninsula and Southwest France) to ~0.4 between some Greek and West-European populations (Appendix IV). Among Iberian populations of *C. barbara* the average (± standard deviation) F_{ST} was 0.036 (± 0.0250), among Northwest African populations it was 0.026 (± 0.0109) and between Iberian and Northwest African it was 0.063 (± 0.0219) (Appendix IVa). Ceuta, which showed higher mitochondrial differentiation from Moroccan populations than from Iberian populations (G. Pinto-Juma, personal communication), had an average F_{ST} (± standard deviation) of 0.052 (± 0.021) from Iberian populations and of 0.027 (± 0.015) from Moroccan populations. Among Iberian populations of *C. orni* the average F_{ST} was 0.033 (± 0.0296), among French populations the average was -0.003 (± 0.0081) and between Iberian and French populations it was 0.035 (± 0.0275). Many of these values were non-significant (Appendix IVb). In contrast, every F_{ST} value between West-European and Greek populations of *C. orni* was significant, with an average F_{ST} of 0.216 ± 0.0745 (Greek *versus* Iberian Peninsula populations) and 0.273 ± 0.0649 (Greek *versus* French populations). F_{ST} values were also high (0.236 ± 0.0909) and significant among Greek populations.

The correlation between mutilocus F_{ST} values calculated with all loci and with only the loci under Hardy-Weinberg equilibrium was high and significant for both *C. barbara* (Mantel test, $r=0.674$, $p<0.001$) and *C. orni* (Mantel test, $r=0.912$, $p<0.001$).

For *C. barbara*, and considering a threshold frequency of 0.05, Iberian populations had 21 private alleles (in six loci) and African populations had ten, which corresponds, respectively to 36.2% and 21.3% of private alleles from the total of alleles present in each region. All of them were alleles with frequencies equal to or lower than 0.2, except one allele of Cib01 that reached a frequency of 0.4 in one population in Iberia and two alleles of Cib06, one of which reached a frequency of 0.3 in one population of Northwest Africa and the other also a frequency of 0.3 in an Iberian population.

For *C. orni*, Greek cicadas had 26 (52%) private alleles (in four loci), six of which with frequencies higher than 0.2, whereas Iberian ones had five (16%) private alleles with frequencies lower than 0.2 and French ones had no private alleles. Within Greece, Athens had eight (32%) private alleles with frequencies lower than 0.2, Kithira had twelve (57%) (three of which with frequencies of 0.29, 0.35 and 0.88), Naxos had nine (38%) with

frequencies lower than 0.2 and Skyros had three (20%) (one of which with a frequency of 0.47).

The scatterplots of the population pairwise $F_{ST}/(1-F_{ST})$ and the natural logarithm of geographic distances for each species *C. orni* and *C. barbara* are shown in Figure 2.3. No correlation between genetic and geographic distances was detected for all *C. barbara* populations (Mantel test, r=0.221, p=0.109). When considering only the Iberian populations of *C. barbara*, there was a negative and nonsignificant correlation (Mantel test, r=-0.489, p=0.941).

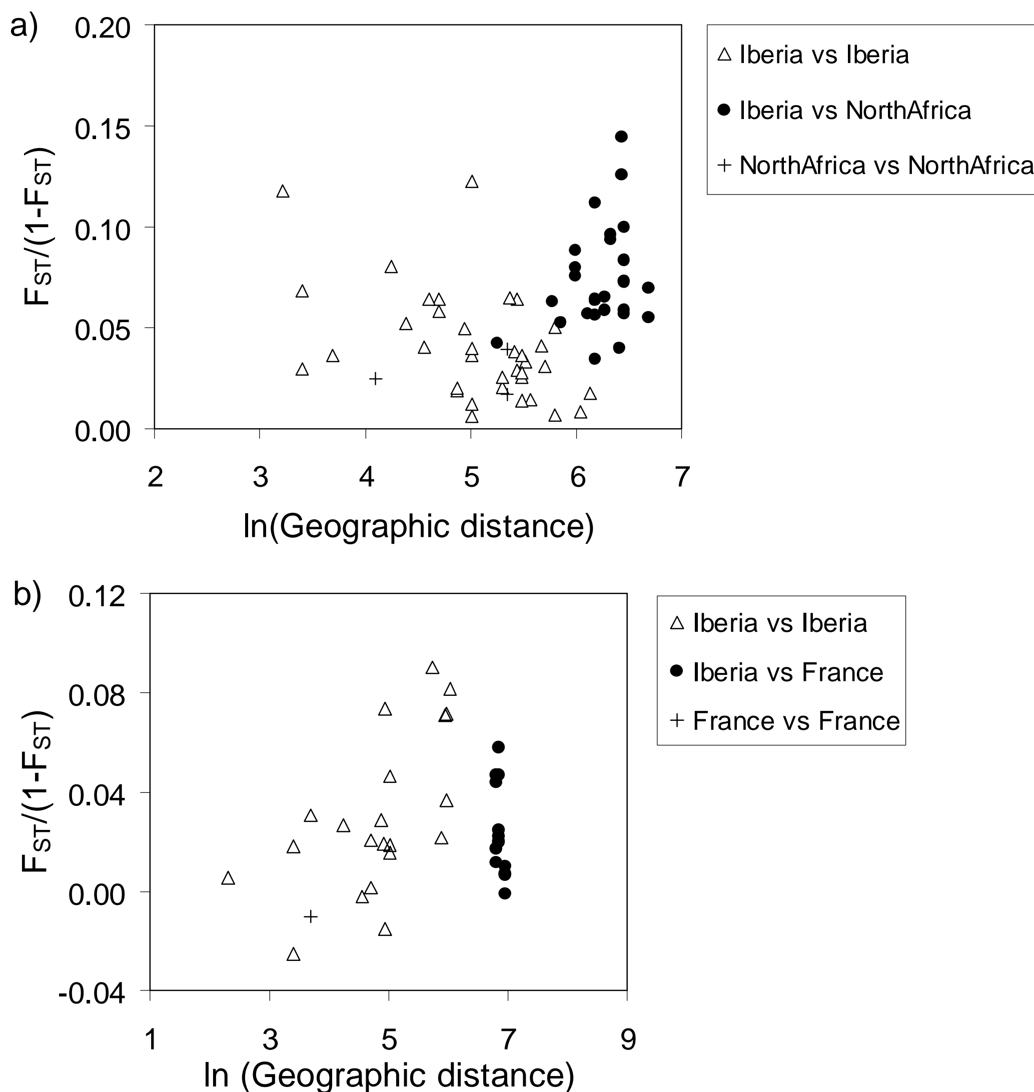


Figure 2.3. Scatterplots of $F_{ST}/(1-F_{ST})$ and natural logarithm of the geographic distance among *C. barbara* populations from Iberian Peninsula and Northwest Africa (a) and among *C. orni* populations from Iberian Peninsula and Southwest France (b). F_{ST} was calculated using only loci in Hardy-Weinberg equilibrium.

For *C. orni* populations in Iberian Peninsula and France the correlation was also nonsignificant (Mantel test, r=0.267, p=0.138). However, when considering only Iberian populations of *C. orni* there was a positive and significant correlation (Mantel test, r=0.602, p=0.033). The same patterns of differentiation were found when calculating F_{ST} using all loci.

Factorial Correspondence Analysis of *C. barbara* individuals did not clearly separate Iberian and African samples, although Iberian individuals tended to cluster apart from the African ones on the axis corresponding to factor 1 (Figure 2.4). Ceuta individuals were somewhat intermediate in their location along this same axis. The first two axes explained 4.99% of the genetic variance.

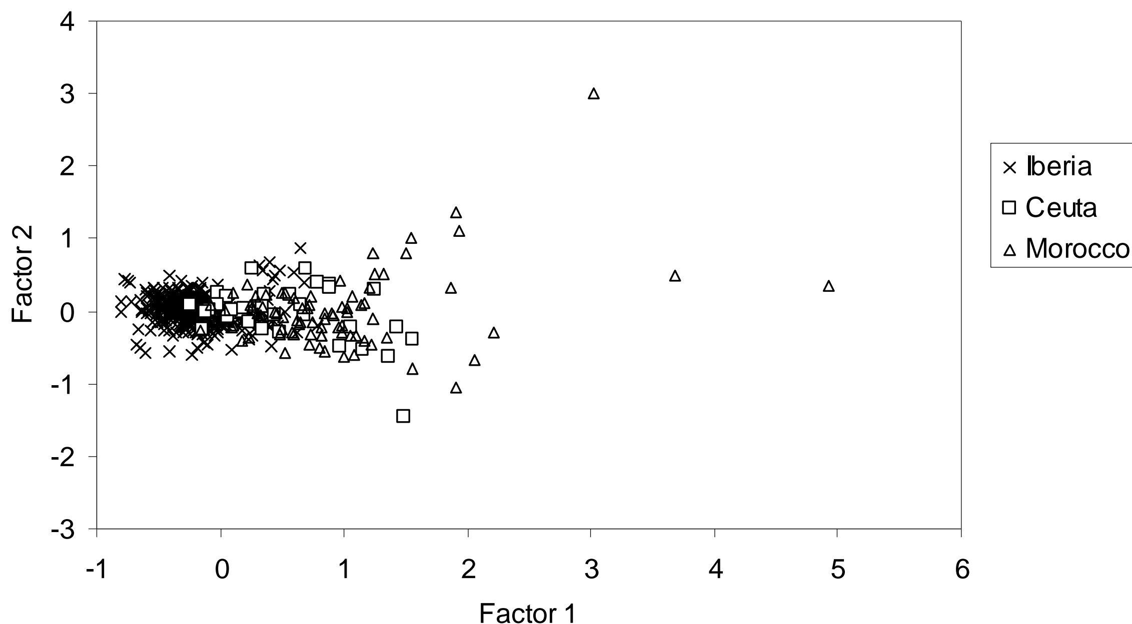
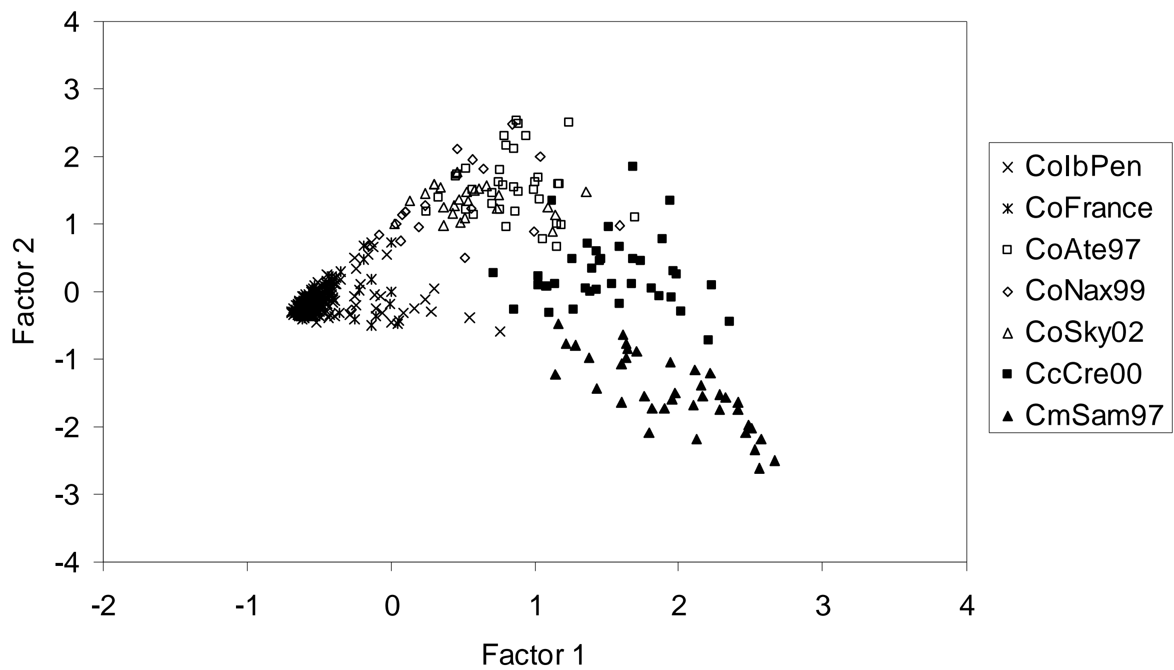


Figure 2.4. Correspondence Factorial Analysis of individual cicadas *C. barbara* from the Iberian Peninsula and Northwest Africa (Ceuta and Morocco).

The Factorial Correspondence Analysis of the dataset *C. orni* + *C. cretensis* + *C. mordoganensis* clearly separated *C. cretensis* and *C. mordoganensis* from Western European samples of *C. orni* (first factor) and *C. mordoganensis* from *C. cretensis* and from Greek *C. orni* (second factor) (Figure 2.5). Western European individuals segregated from Greek individuals of *C. orni* in both axes. These two axes explained 7.62% of the total genetic variance.

a)



b)

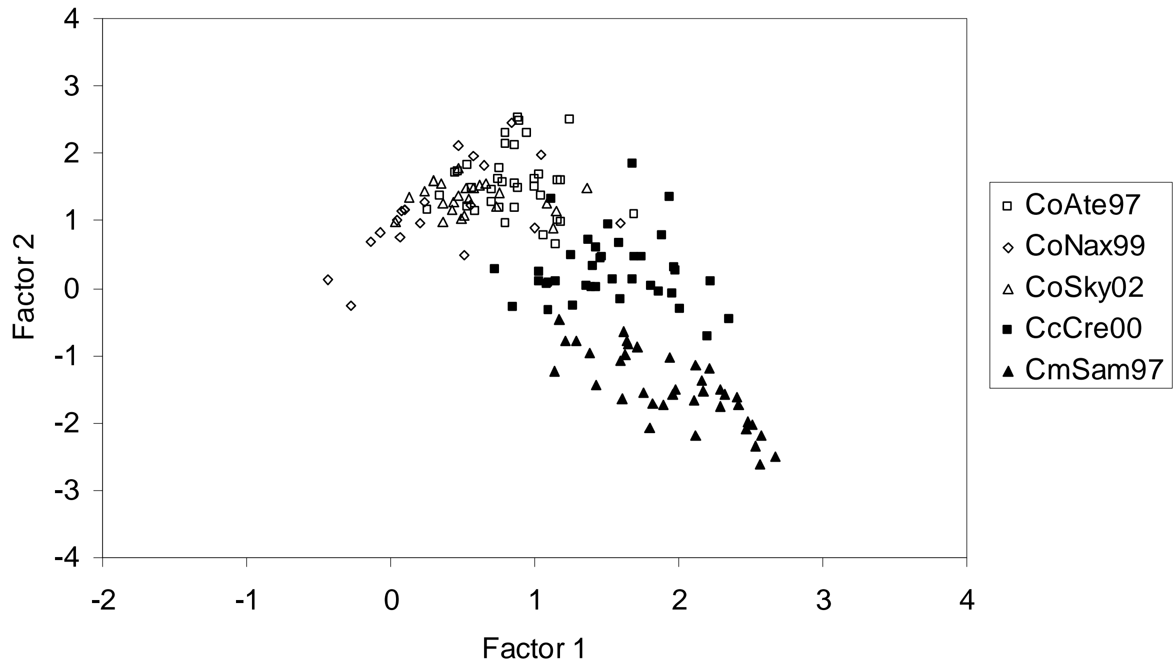


Figure 2.5. Correspondence Factorial Analysis of individual cicadas *C. orni* (from the Iberian Peninsula, France, and from the Greek populations of Athens, Naxos and Skyros) , *C. cretensis* and *C. mordoganensis*. In a) all the individuals are shown and in b) the Iberian and French individuals were excluded from the scatterplot.

STRUCTURE analysis of *C. barbara* revealed a most probable K of 3 using the posterior probabilities of K (p~1) and the Evanno *et al.* (2005) methodology. Most cicadas from the Northwest African populations (Ceuta – Pop 17, Fès – Pop 18, and Meknès – Pop 19) were assigned to a cluster distinct from the Iberian Peninsula populations (Figure 2.6). There was also some distinction within Iberia: Foz Côa (Pop 9), Monforte (Pop 10), Moura (Pop 11), Portel (Pop 12 and 13) and Sevilla (Pop 16) were mostly assigned to a second cluster and Sousel (Pop 14 and 15) mostly to a third cluster. The remaining populations had contributions from the three groups. PARTITION did not detect any structure in the *C. barbara* dataset (posterior probability of 0.98 for K=1, 0.02 for K=2, 0.0003 for K=3, and 0 for K>3). Bayes factor was 839.24, favouring the hypothesis K=1.

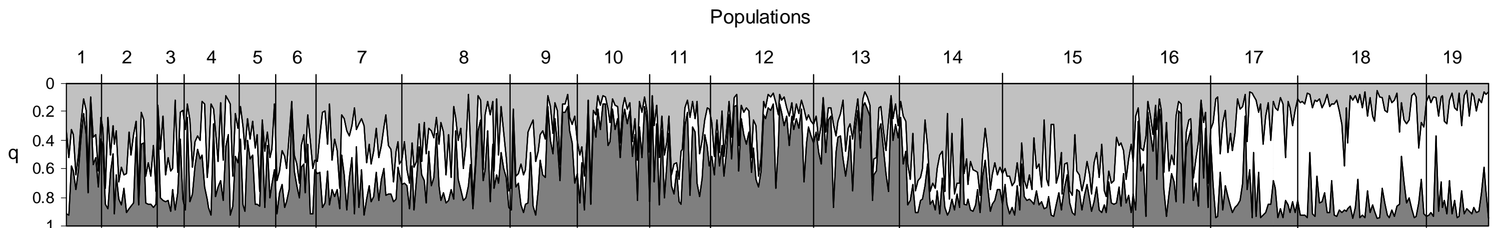


Figure 2.6. Estimated membership coefficients (q) for each individual *C. barbara* (each single vertical line), in each cluster (K=3). The numbers (1 to 19) correspond to the predefined populations. 1 – CbAlc95; 2 – CbCas96; 3 – CbCra95; 4 – CbCra96; 5 – CbCra99; 6 – CbCra00; 7 – CbCra01; 8 – CbCra02; 9 – CbFoz99; 10 – CbMon95; 11 – CbMou01; 12 – CbPor01; 13 – CbPor02; 14 – CbSou01; 15 – CbSou02; 16 – CbSev01; 17 – CbCeu99; 18 – CbFes01; 19 – CbMek01.

From STRUCTURE the most probable K for *C. orni* (Iberian Peninsula and Southwest France), was 1, and thus, no structure was detected in Western European *C. orni* populations.

AMOVA carried out on different years of capture for *C. barbara* detected most of the variation within years (93.30%) and the variation among populations (4.28%) was higher than among years within a population (2.42%) (Table 2.9). All three fixation indices were significant. A locus-by-locus AMOVA gave similar patterns for every locus, with very high percentage of variation within years in all loci, being the percentage of variation among populations higher than among years, with the exception of Cib10 (Table 2.9). No structure was found in this dataset using STRUCTURE (p~1 for K=1).

For Iberian *C. orni*, AMOVA gave a within-year variation of 96.05% and showed a higher among-year within-population variation (2.78%) than among populations (1.17%). The

French samples showed the same pattern, with 96.92% of within-years variation, 5.04% among years and -1.96% among populations. This same pattern was found in every locus in the locus-by-locus AMOVA except in Cib03, where variation among years was lower than variation among populations (Table 2.9). No population structure was detected using STRUCTURE when analyzing the populations *C. orni* from Iberian Peninsula and Southwest France with more than one year of sampling.

Table 2.9. Percentage of variation in AMOVA found among populations (Pop.), among years within populations (Years) and within years in *C. barbara* (Crato, Portel and Sousel), in *C. orni* from Iberian Peninsula (Portel and Algeciras) and in *C. orni* from France (Narbonne and St. Hippolyte) for single-locus and multilocus analysis.

	C. barbara			***C. orni* – Iberian Peninsula**			***C. orni* – France**		
Locus	Pop.	Years	Within years	Pop.	Years	Within years	Pop.	Years	Within years
Cib01	1.93	0.73	97.34	-0.33	0.09	100.23	-2.75	5.66	97.08
Cib03	7.19	3.38	89.43	3.34	3.32	93.34	0.28	-0.42	100.14
Cib06	4.56	1.06	94.38						
Cib07	5.24	4.10	90.66	1.73	4.57	93.70	-5.85	10.75	95.10
Cio08	6.90	0.92	92.18	-0.56	2.21	98.34	0.55	3.68	95.77
Cib10	0.28	4.62	95.10						
All loci	4.28	2.42	93.30	1.17	2.78	96.05	-1.96	5.04	96.92

The percentage of private alleles in each locus present in each species *C. barbara* and *C. orni* did not differ between sympatric and allopatric populations of Iberian Peninsula (Wilcoxon signed rank test, p=0.285 for *C. barbara* and p=0.593 for *C. orni*). The Factorial Correspondence Analysis showed a clear separation of both species in the first factor (Figure 2.7), except for a few individuals of *C. orni* which showed atypical genotypes for some loci, namely homozygous alleles which are more common in *C. barbara*. However, these individuals had also missing data for some loci and therefore it is not possible to conclude whether they are the result of introgression.

In locus Cib03, which shows almost complete non-overlapping alleles between *C. barbara* and *C. orni*, there is one allele typical of *C. orni* occurring in *C. barbara* in Monforte and one allele typical of *C. barbara* occurring in *C. orni* in Sousel, which are both sympatric areas. These alleles occurred in both populations in only one individual at heterozygous state.

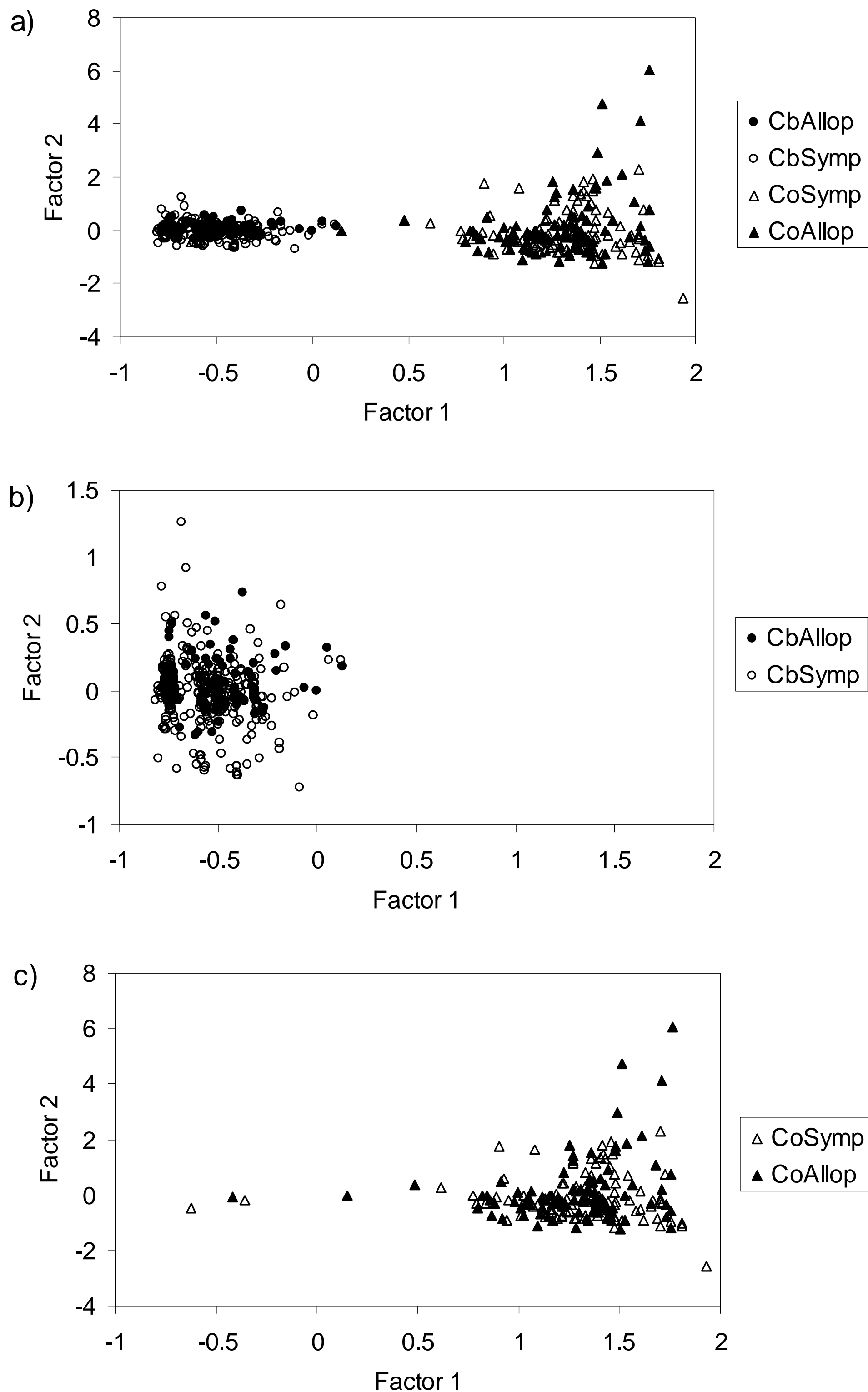


Figure 2.7. Correspondence Factorial Analysis of individual *C. barbara* and *C. orni* from the Iberian Peninsula from sympatric and allopatric populations. The analysis was performed with both species and the scatterplot of the individual scores for the first two factors is shown in a). The individuals from *C. orni* were then excluded from the scatterplot for a better perception of the dispersion of the points (b). The opposite was done in (c).

2.3.3. Discussion

a) Single-locus analysis

We found a high incidence of heterozygote deficiency (F_{IS} close to or even equal to 1) at some microsatellite loci for certain populations that was attributed to the presence of null alleles. High F_{IS} values have been reported in other insects, such as in the house fly (Endsley *et al.*, 2002; F_{IS}=0.675), in *Triatoma infestans* (Marcet *et al.*, 2006; F_{IS}=0.91) and in *Scaptodrosophila hibisci* (Drosophilidae) (Barker, 2005; F_{IS}=1).

One line of evidence in support of null alleles here, as opposed to inbreeding or substructure, is the lack of evidence for these demographic processes in allozymes screened for some of the same populations (Quartau *et al.*, 2001). However, it should be noted that in loci with lower variability (as for protein polymorphisms), deviations from Hardy-Weinberg equilibrium expectations are sometimes difficult to detect statistically (Chakraborty *et al.*, 1992). Further, not every locus showed heterozygote deficit and there was a clear species-dependent deviation. It could be argued that scoring errors did not detect heterozygotes, and in large datasets it is probable that genotyping errors occur (Hoffman & Amos, 2005). But such a strong systematic bias is improbable in populations of one species and not another. Furthermore, the number of non-amplified samples (from which no PCR product was obtained) in some loci was higher in those populations deviating most from Hardy-Weinberg equilibrium.

The non-amplified samples did not appear to be due to poor quality DNA, since all the extractions showed standard patterns when run on agarose gels, and the same samples amplified well for other loci. The effects of the storage method, tissue and extraction method on amplification rates, despite being significant for some loci, were not consistent, as shown by logistic regression. Nevertheless, the lower amplification success in dried samples detected for some loci cannot be ignored and reinforces the potential PCR amplification problems when using samples which may have lower quantities of DNA due to degradation. Since dried samples were only present in three populations (Kithira, Naxos and Skyros, all belonging to *C. orni* species), this cannot explain the deviations from Hardy-Weinberg equilibrium in the other populations. Furthermore, the occurrence of PCR amplification problems at a high level for some loci in certain populations (namely Kithira and in Crete) may indicate the existence of mutations in the binding sites of the primers specific to those populations. In Kithira (*C. orni*) Cib03, Cib07 and Cio08 had very high proportions of non-amplified samples (>75%), whereas in Crete (*C. cretensis*) Cib01 was the only one with a very high proportion of non-amplified samples (92%).

There are some issues that should be taken into account when designing primers in order to decrease the probability of null alleles: the presence of repetitive sequences in flanking regions of microsatellites may increase the probability of mutation (Jones *et al.*, 1998; Wang *et al.*, 2000); the distance from the microsatellite to the priming site may also be an important question (Pemberton *et al.*, 1995) and here we found that the loci with the priming sites closest to the microsatellite had the highest proportion of null alleles. However, Callen *et al.* (1993) did not find a relationship between the sequence or the position of the priming site and the occurrence of null alleles. Some organisms may have high rate of mutation in the microsatellite-flanking regions, as suggested by Keyghobadi *et al.* (1999) for butterfly species and by Hedgecock *et al.* (2004) for oysters, and this could be a possibility in this study. Importantly, there was no evidence for a higher incidence of null alleles in the non-focal species. All loci were obtained from a *C. barbara* library except *Cio08* which was obtained from a *C. orni* library. Further, no clear tendency has been detected in other organisms for the occurrence of lower frequency of null alleles in species for which the primers were cloned (Pemberton *et al.*, 1995).

The approach adopted here may be the only option currently available to analyse data that deviate very strongly from Hardy-Weinberg equilibrium. However, there are several considerations when taking this approach. First, when designating null alleles, several alleles may be included in this general designation and here they are treated as only one extra allelic class. Where authors have redesigned primers, it was found that some loci had several different sized null alleles (*e.g.*, Callen *et al.*, 1993; Ishibashi *et al.*, 1996; Lehmann *et al.*, 1996; Jones *et al.*, 1998). It was not possible to establish how many null allele classes occurred for each locus in this study. Thus, the variability in terms of number of alleles may be underestimated and comparisons among species must be taken with caution (adding to the caution we need in analysing microsatellites in general). The option of performing random permutations to obtain multi-locus genotypes and then analyse the multiple sets would be feasible in cases where the null allele frequencies are low and where the new allele class in the adjusted dataset comprises only one allele. In this case, where the new allele class may include several different null alleles, this approach is not possible. The persistence of significant deviations from Hardy-Weinberg equilibrium in some adjusted datasets may indicate other causes for deviation, as suggested by VanTreuren (1998) for oystercatchers. Also, it is possible that this adjustment approach does not cope with severe deviations.

Notwithstanding these considerations, it is still possible to report that the microsatellite variability is generally high in *C. barbara* and *C. orni*, especially if we consider that these

values may be underestimates because of null alleles. The number of alleles in *C. barbara* ranged from 10 to 23 and in *C. orni* from 14 to 26 and the expected heterozygosities from 0.594 to 0.833 in *C. barbara* and from 0.686 to 0.829 in *C. orni* (in the original dataset). Therefore, both species present similar values of microsatellite variability and diversity. These values are well within the range found in other species of insects. For example, Estoup *et al.* (1995) found seven to 30 alleles and average heterozygosities from 0.291 to 0.872 in honeybee (*Apis mellifera*) populations (seven loci; nine populations; N=20 to 60). Schlötterer *et al.* (1997) found five to 13 alleles and average expected heterozygosities from 0.50 to 0.87 in natural *Drosophila melanogaster* populations (ten loci; seven populations; N= 10 to 20).

When comparing single-locus variability among datasets, significant differences were found for most loci, indicating the need for care with the interpretation of results when using datasets with deviations from Hardy-Weinberg equilibrium. The comparisons between regions in *C. barbara* were consistent among datasets for each locus, but only two loci gave significant differences. The results were not always consistent among datasets when comparing species and regions in *C. orni*.

Microsatellite variability is generally higher in focal than in nonfocal species (Ellegren *et al.*, 1997; Hutter *et al.* 1998). Here, the two loci that were monomorphic in *C. orni*, *C. cretensis* and *C. mordoganensis* were originally designed for *C. barbara*. This may be explained by mutation(s) in the microsatellites that prevented replication slippage and thus the generation of novel alleles. The fact that these species have the same fixed allele may indicate that the mutation(s) occurred before the species diverged. For the remaining loci, no consistent difference in variability was found between species. Quartau *et al.* (2001) had found higher allozyme variability in *C. orni* than in *C. barbara*.

Single-locus population differentiation estimates were highly correlated among datasets and differences in population differentiation within and among regions for each species were generally consistent among datasets for each locus. F_{ST} estimates are thus robust to deviations from Hardy-Weinberg equilibrium.

b) Multi-locus analysis

Since there were very high frequencies of null alleles in some of the loci employed in this study, datasets with and without these problematic loci were used and results compared. Results were similar among datasets for both variability and differentiation analyses.

In the Iberian Peninsula both *C. barbara* and *C. orni* showed relatively low values of genetic differentiation (average F_{ST} values of 0.036 in *C. barbara* and of 0.033 in *C. orni*). Low values of population differentiation, assessed using microsatellites, were found in other insects in which dispersal and migration is presumed to be high. For example, Llewellyn *et al.* (2003), studying the genetic structure of the grain aphid (*Sitobion avenae*) in a 650 km north-south transect in Britain, found that most pairwise F_{ST} values were low (lower than 0.05) suggesting that it is a highly migratory species. Augustinos *et al.* (2005) found an average F_{ST} of 0.039 among populations of the olive fly across the Mediterranean area.

As seen before, low F_{ST} values calculated from microsatellite data may be artificially low due to the high variability within populations. Cicadas have in general poor dispersal abilities (Karban, 1981; Williams & Simon, 1995; de Boer & Duffels, 1996), even though long distance dispersal across hundreds of kilometers of ocean, probably through wind currents, has been suggested for the origin of some of the New Zealand cicadas (Arensburger *et al.*, 2004). The dispersal capabilities of the cicadas analysed in the present study are not yet known but, despite being powerful fliers, they are not seen flying more than a few tens of meters at a time. It has been recognized that the physical ability of animals to disperse is often a poor predictor of dispersal distributions (Rousset, 2004). A factor that is expected to contribute to low effective dispersal in these cicadas is the tendency of cicada males of these species to aggregate, forming singing choruses.

The evidence of isolation by distance (IBD) in *C. orni* but not in *C. barbara* may be explained by historical, demographic, biological and ecological factors affecting each species independently. *C. barbara*, being a thermophilic species with a more restricted distribution, may be more limited in its dispersal capabilities in the Iberian Peninsula in comparison to *C. orni*. *C. orni* is a widely distributed species in the Mediterranean area, existing even in urban areas. Peterson & Denno (1998) found, in a review of allozyme studies of phytophagous insects, that isolation by distance was weak in sedentary and in highly mobile species and pronounced in moderately mobile species. Another possibility for the lack of a pattern of IBD, as suggested by Slatkin (1993), is that the species is not in demographic equilibrium and may have only recently invaded the area it occupies. Nonsignificant IBD is expected in

populations that are newly established and/or have highly variable patterns of migration and gene flow (Moyle, 2006). Drift due to small population sizes and founder effects have been evoked to explain the lack of IBD in species like the aphid *Macrosiphoniella tanacetaria* (Massonet & Weisser, 2004). The ability of detecting IBD is also dependent on the scale and design of the sampling strategy (Rousset, 1997; Keyghobadi *et al.*, 2005). The hypothesis of *C. barbara* being a recent immigrant in the Iberian Peninsula has already been suggested by Quartau *et al.* (2001) but the present microsatellite data do not allow testing this hypothesis.

Higher microsatellite variability in Northwest African than in Iberian populations of *C. barbara* was detected, which could be due to higher population sizes in Africa, a reduction in population sizes in Iberian Peninsula (probably during the Pleistocene glaciations), or even to a more recent colonization of Iberian habitats. The differentiation estimates between Iberian *C. barbara lusitanica* and African *C. barbara barbara* (average estimate of 0.06) were higher than within each of these regions (average estimate <0.05). The percentage of private alleles in each region was also high, even if most of them were present at very low frequencies. The Strait of Gibraltar (14 km wide at the minimum distance) is an effective barrier to the dispersal of several flying species (Castella *et al.*, 2000; Broderick *et al.*, 2003) but does not completely prevent the occasional dispersal in some species. In fact, it was the route of dispersion for many species from Africa to Europe or vice-versa (*e.g.*, Franck *et al.*, 1998; Guillaumet *et al.*, 2006).

The fact that very few loci were used in the STRUCTURE analysis means that robust conclusions are not possible. The interpretation of K may not have always a clear biological meaning (Pritchard *et al.*, 2000). The fact that PARTITION did not detect any structure on *C. barbara* populations indicates that the clusters found by STRUCTURE are not likely to be very distinct. However, the assignment of a majority of Northwest African genotypes (from Ceuta, Fès and Meknès) to a cluster different from those of Iberian populations was supported by the STRUCTURE analysis. Ceuta, located on the Northwestern coast of Africa, is separated from the Iberian Peninsula by the Strait of Gibraltar and has a mountainous area (Rif) separating it from the inland Moroccan populations. It showed higher microsatellite differentiation (F_{ST}) from Iberian populations (*C. barbara lusitanica*) than from the Moroccan populations (C*. barbara barbara*). This is contradictory with mitochondrial DNA findings (G. Pinto-Juma, personal communication), which revealed that Ceuta samples had only Iberian mitochondrial haplotypes. This incongruence may be due to the different patterns of variation and evolutionary rates of microsatellites and mitochondrial DNA. Also, the maternal inheritance of mitochondrial DNA only allows the detection of female dispersal, while nuclear loci reflect both female and male dispersal. In insects, asymmetrical introgression of

mitochondrial and nuclear genomes has been reported, particularly in social species such as the honeybees, where nuclear genes disperse to other colonies through the males that inseminate queens (Franck *et al.*, 2000), but it has also been suggested for other species, such as the montane mayfly *Baetis bicaudatus* (Hughes *et al.*, 2003). Alternatively, mitochondrial DNA may be subjected to selective forces (William *et al.*, 1995) that make the Iberian haplotype selectively superior to the local haplotype in the coastal Northwestern Africa. This hypothesis of selection acting on mitochondrial DNA was suggested, for example, for honeybees by Franck *et al* (1998), to explain the spread of the 'african' mitochondrial haplotype in Iberian Peninsula, while with microsatellites there was a clear disruption between Africa and Western Europe (including the Iberian Peninsula). The origin of the Ceuta population of *C. barbara* and subsequent patterns of gene flow remains to be elucidated.

No significant difference was found in variability between Iberian and French populations of *C. orni* and no genetic structure was found in Western European populations of this species. The differentiation between Iberian Peninsula and France was low and similar to that within Iberian Peninsula, which suggests substantial gene flow between these regions or a recent divergence. Although the Pyrenees are a significant barrier to several species (Hewitt, 2000), this does not seem to be the case for *C. orni*. It is most likely that the dispersal occurs along the coast (most likely in the Mediterranean coast) and not across the Pyrenees itself, as suggested for noctuids by Bues *et al.* (1996).

The high microsatellite variability and diversity in the mainland population of Greece (Athens) could be a reflection of its very large population size (J.A. Quartau and P. Simões, personal communication). The Balkans have been suggested to be the source of several species that have expanded westwards after the glaciations (Hewitt, 2000). However, the present data are not enough to test this hypothesis. Greek populations were highly differentiated from Western European populations. In fact, they were clearly separated in the Factorial Correspondence Analysis. Also the proportion of private alleles was very high in both regions, especially in Greece.

The differentiation among Aegean islands and the very high proportion of private alleles, present on each of them in high frequencies, also indicates low levels of gene flow between islands. Microsatellite variability and heterozygosity values in these islands were lower than in the mainland Greek population, Athens, which may indicate bottleneck effects during the process of colonization of the islands. However, the values were still high and similar to, or even higher than, in the other Western European populations of *C. orni*. Consequently, it

seems that the bottleneck effects were not very strong or that the colonization was ancient and the populations were allowed to recover high levels of variability. Vicariance scenarios due to sea level change in the Aegean islands have been proposed to explain the patterns of genetic variation in several animal species (Beerli *et al*., 1996; Kasapidis *et al*., 2005; Parmakelis *et al*., 2006) and they imply very ancient (millions of years) divergence times. When compared with the differentiation among *C. barbara* subspecies across the Strait of Gibraltar, Aegean populations of *C. orni* seem much more differentiated, probably reflecting older colonization and isolation, with higher drift and/or lower dispersal than in the *C. barbara* case. However, comparison of differentiation values among species is always problematic since, even in closely related species, differences in life-history characteristics, dispersal capabilities, mutation rates and evolutionary history may produce very different variation patterns among populations.

In *C. barbara*, variation among years within populations was lower than among populations in the AMOVA analysis. No structure was found among years in STRUCTURE. In contrast, in *C. orni* the variation among years was generally higher than among populations in the AMOVA, but no structure was detected in STRUCTURE. Gene flow between years is thus high, indicating that no independent broods exist in these cicadas, that is, they have no fixed period of nymphal development.

C. barbara and *C. orni* were highly divergent at microsatellite loci and both had private alleles in most loci, with one locus presenting almost non-overlapping allele size ranges. This contrasts with results from allozyme studies, where only three of 19 loci were found to be diagnostic for the separation of *C. orni* and *C. barbara* and the genetic distances (Nei's distances) were very low (Quartau *et al*., 2001). The lower divergence between *C. orni* and *C. mordoganensis* than between *C. orni* and *C. barbara* had already been seen with allozymes (Seabra *et al*., 2000) and is also observed in morphology and calling songs (Boulard, 1995; Quartau, 1988; Simões *et al*., 2000). Nevertheless, *C. mordoganensis* and also *C. cretensis* clearly separated from Greek *C. orni* in the Factorial Correspondence Analysis. It is interesting to note that these three species that occur in Eastern Mediterranean areas and that have very similar calling songs, have not yet been found in sympatry.

In the sympatric areas of *C. barbara* and *C. orni* no evidence of hybridization from microsatellites was found. However, the fact that locus Cib03, which shows almost complete non-overlapping alleles between *C. barbara* and *C. orni*, had one allele typical of *C. orni* occurring in *C. barbara* in Monforte and one allele typical of *C. barbara* present in *C. orni* in Sousel, both sympatric areas, may be an indication of occasional hybridization. Nonetheless,

from these results, and from acoustic behavioural studies that did not find any evidence of hybrids (see Chapter 3), we can say that the premating isolating mechanisms (or mate recognition systems) are efficient in bringing together only conspecific mates. The calling songs of males are most certainly an important mechanism of recognition/isolation. These species do not seem to have divergent habitat preferences in most sympatric areas, contrary to what is observed, for example, in sympatric *Magicicada* species, which prefer distinct habitats for oviposition (Dybas & Lloyd, 1962) or in *Tibicina* species, which prefer different habitat structures, particularly different height of vegetation (Sueur & Puissant, 2002). However, a more detailed ecological and behavioural study is needed to elucidate this question.

Chapter 3 – Acoustic analysis

I have often reflected with surprise on the diversity of the means for producing music with insects, and still more with birds. We thus get a high idea of the importance of song in the animal kingdom.

Charles Darwin, Letter to Muller (1868)
(http://darwin-online.org.uk)

3. Acoustic analysis

3.1. Introduction to the acoustic analysis

3.1.1. Acoustic communication – sound production, transmission and reception in cicadas

Long-distance acoustic communication through air implies the presence of a sound producing system capable of delivering low-frequency sounds (which are less attenuated by distance than high-frequency sounds) at high intensity. Despite their relatively small size, many insects produce such sounds (Bennet-Clark, 1998b; Gerhardt & Huber, 2002). These insects use resonating systems, which increase the amplitude of the sound and act as a narrow band-pass filter, concentrating the energy in a narrow frequency band (Bennet-Clark & Young, 1992; Fonseca, 1994; Bennet-Clark, 1999). Male cicadas use as a resonating system the coupling between the tymbals and the abdominal air sacs (Young, 1990; Bennet-Clark & Young, 1992), and some cicada species are able to produce the highest intensity sounds among insects (Bennet-Clark, 1998a).

A natural occurring biological acoustic signal is characterised by its periodicity (representing the repeated operation of a sound-producing mechanism like the vibration of a membrane), its intensity (dependent on the amplitude of vibration) and its modulation patterns (Beeman, 1998). Biological signals are usually complex, containing energy at multiple frequencies. To a large extent, the natural song frequency of cicadas is dependent on the dimension and structure of the tymbals, namely the morphology, number and position of the ribs, as well as on the morphology and dimensions of the abdominal cavity and of the tympana (Young, 1972; Bennet-Clark & Young, 1992). Other characterisitics of the songs of cicadas like the time and amplitude modulation patterns are mostly dependent on the motor pattern generated by the nervous system (Fonseca, 1994).

During sound transmission, sound intensity decays proportionally to the square of the distance (Henrique, 2002). In a non-uniform media, like the air, the signal will be further attenuated by absorption (Bennet-Clark, 1998b). Additionally, temperature and wind gradients cause heterogeneities in the medium that cause refraction of the sound waves. The obstacles present in the environment (such as topographic elements and vegetation) will cause absorption, reflection and diffraction of the sound waves, causing directionality changes (Henrique, 2002; Eliopoulos, 2006). The conjunction of these phenomena will cause the attenuation and change of the frequency content of the sound. Furthermore, the ambient

noise level is a factor that may reduce the conspicuousness of a signal, especially if the band frequencies of the noise include the band frequencies of the signal. Usually, conspecifics are the cause of most significant masking sounds (Hopp & Morton, 1996).

These interference factors constitute important selective forces in shaping the signal characteristics and also the behaviour of the animals during sound emission, to enhance the conspicuousness of the signal used in long-distance communication. For example, in an open habitat, animals will avoid refraction of their emitted sounds caused by temperature and wind gradients if calling from an elevated spot. In contrast, forests have relatively homogeneous air below the canopy but tree trunks and branches may reflect and scatter some frequencies, causing reverberation (decay of sharp signal onsets and offsets) (Hopp & Morton, 1996), and thus, longer time intervals between components of the song are considered adaptive because there will be less superimposition if the components are more separated (Kopuchian *et al.*, 2004).

Both sexes in cicadas have auditory organs located ventrally in the second abdominal segment; the auditory organs include two tympanic membranes, backed by the abdominal air sacs, which allow the vibration of the membrane. Tympana and air sacs are larger in males than in females, for in males they are also involved in sound production and radiation (Fonseca, 1994; Boulard, 2006). In contrast, females use much of the abdominal space to produce eggs and thus the air sacs are smaller. Nevertheless, the auditory abilities of females are not diminished (Fonseca, 1994). The tympanal ears of cicadas have thousands of auditory receptor cells, allowing a high frequency resolution (Fonseca *et al.*, 2000). It has been shown that some cicadas also have high frequency resolution in the central nervous system (*cf.* Fonseca *et al.* 2000 for *Tettigetta josei* and Fonseca & Revez, 2002a for *C. barbara*).

The properties of the acoustic signals that may have important communication roles in animals include temporal properties (duration, repetition and sequence of sound elements) and spectral properties (frequency, bandwidth, harmonic structure and noisiness) (Beeman, 1998). In cicadas, a few studies on the influence of song characteristics of the males on the species-recognition and choice by females have been done using the phonotactic behaviour of females. For example, Doolan & Young (1989) found that in *Cystosoma saundersii* it is the carrier frequency of the calling song that is most important in long-range communication, whereas the temporal patterns are more important in short-range communication. Stölting *et al.* (2004) found that in *Okanagana rimosa* the repetition rate of the calling song is an important characteristic for species discrimination by females. Whenever the female

response is not detectable experimentally, the male response by singing to other males' songs has also been tested. For instance, males of *Cicada barbara* (Fonseca & Revez, 2002a), *C. orni* (Simões & Quartau, 2006) and *Tibicina haematodes* (Sueur & Aubin, 2002) are capable of temporal and frequency discrimination of the calling songs.

3.1.2. Sound recording, processing and analysis

In the field, obtaining good quality recordings of insect sounds is sometimes a difficult task due to several influencing factors, namely the presence of background noise, the adverse environmental conditions that affect sound propagation (see 3.1.1) and the influence of the presence of the researcher on the behaviour of the animal under study. Patterns of sound propagation in natural habitats cause changes in the spectral structure (relative amplitudes of different frequency components) of animal signals, which may be considerable depending on the distance from the animal at which recordings are made (Gerhardt, 1998). In the field, when recording the sound produced by an animal, one must approach it carefully to minimise disturbance and the microphone should be placed as close as possible to the emitting animal to get as loud a recording in relation to background noise as possible (Ragge & Reynolds, 1998). However, there is also a minimum distance (at least one wavelength of the lowest frequency in the sound) below which near-field effects will compromise the recording (Gerhardt, 1998).

Bioacoustic recordings in the field require portable equipment (a microphone connected to an audio recorder) with the adequate frequency response to the signal one plans to record. Nowadays, digital recorders are the choice among bioacoustic researchers since they deliver good quality recordings with great accuracy and low noise in a reliable storage format (Eliopoulos, 2006). Digital audio tape (DAT) recorders store magnetically on tape an audio signal in digital format. Microphones convert the pressure variations composing the acoustic signal into an electrical signal, which is then transferred to the sound recorder. The electric signal, after being filtered to remove frequencies above 20 kHz, is passed through an A/D converter that measures the voltage of the audio waveform several thousands of times per second (sampling rate or frequency) and generates a binary number (number of bits used represent the sample size) for each of those voltage measurements, which are then stored on tape (Eliopoulos, *op.cit.*).

The sounds recorded in the field are then taken to the laboratory for transfer and analysis on a computer with the adequate hardware and software. Digital transfer from DATs to computers is allowed by specific boards with digital input/output (I/O) capabilities or digitising

may be done by an A/D converter hardware. In the latter case, again a digital sampling is performed, the precision of which depends on the sampling frequency and the number of bits used. For an accurate representation, the sampling rate must be more than twice the highest frequency contained in the signal (Eliopoulos, *op.cit.*).

Using the appropriate software, the physical properties of the sound can then be represented in the temporal and the spectral domains (Gerhardt, 1998). In the temporal analysis, the amplitude of the signal is displayed as a function of time *(amplitude-time waveform* or *oscillogram)* and several measurements may be taken from it, such as duration of the acoustic elements, repetition rates, patterns of amplitude modulation, or rise-fall characteristics of the waveform (Gerhardt, 1998). In the spectral analysis, the relative amplitudes of the frequency components of the signal are displayed as a function of frequency (*frequency spectrum*). This is obtained by Fourier analysis of the time waveform, an analysis that uses a mathematical integration technique (discrete Fourier transform) to decompose the signal into sinusoids. This technique is available in most digital signal processing systems as an efficient computational algorithm known as fast Fourier transform (FFT) (Smith, 1997). The periodicities of the signal are shown by the concentration of energy in the carrier (or dominant) frequency or frequencies. If the signal is modulated, sidebands occur around the carrier frequency or frequencies at intervals equal to the rate of modulation.

The energy distribution of the signal can also be displayed as a function of both frequency and time (*frequency-time spectrogram*, also called *sonogram* or *sonagram*), where the amplitudes of a signal component are expressed visually by the darkness of an area displayed on two axes representing time and frequency (Beeman, 1998). Digital spectrograms are obtained by calculating discrete Fourier transforms for each of successive small segments of the waveform (Beeman, *op. cit.*). Spectrograms are very useful in finding spectral-temporal patterns and general similarity among sounds. The digital format allows automated analysis of the acoustic properties, for example to obtain measurements of onset and offset characteristics of acoustic elements in both time and frequency domains by locating the time and frequency boundaries at which matrix values of the spectrogram exceed specified threshold levels (Beeman, *op. cit.*; Specht, 2002).

Sound comparisons have been used to relate signal characteristics to species identity, geographical location or individual variation. These comparisons can be made qualitatively (auditory comparisons or visual comparison of spectrograms) or by reducing sounds to significant parameters that may be compared statistically (univariate and multivariate statistical techniques). For more complex signals, similarity can be measured by comparing

the complete numerical matrices of the digital spectrogram through cross-correlation (Beeman, 1998).

The calling songs of the cicadas under study (*Cicada barbara* and *C. orni*) are pulse-repetition signals, with relatively simple frequency spectra, with a dominant (or carrier) frequency and a few lower amplitude sidebands. The variability of the signals can be described using a small set of acoustic variables (described below in 3.2.2) that are measured automatically from the frequency-time spectrogram. In both species, a group of sound pulses (called syllable) is produced each time a tymbal buckles and returns to its original position. These syllables can be counted manually in an oscillogram, displayed on the screen with the appropriate resolution and magnification. The terminology used for describing the acoustic elements in this study is adapted from Broughton (1976) and Fonseca (1991).

When analysing variability levels of the song properties it is important to be aware that they may be constrained by environmental characteristics, such as temperature (Gerhardt & Huber, 2002). Additionally, the acoustic signal may be distorted both in temporal and spectral properties due to environmental effects during transmission and due to background noises (Villet, 1995).

3.2. Material and methods

3.2.1. Field procedures and sound recording

Male cicadas were located by their calling songs during the hottest season (June–September, 1995–2003) and recordings were made for up to three minutes each using a Sony DAT recorder (TCD-D10 ProII and TCD-D8; frequency ranges of 20–22 000 Hz and 20–20 000 Hz, respectively; sampling frequency 44.1 kHz) connected to a uni-directional dynamic Sony F-780 microphone or a Telinga Pro4PiP microphone (frequency responses 50–18 000 Hz and 40–18 000 Hz, respectively). A few recordings were made using a UHER 4200 Report Monitor (sampling frequency 44.1 kHz, frequency range 20–25 000 Hz) with an AKG D202 dynamic microphone. In most cases, the microphone was placed at 30 cm to 1 m from the cicada. Recordings were performed usually between 9 am and 7 pm and ambient temperature was taken on the shade near the place where cicadas were singing at the time of each recording, with measured temperatures ranging from 23ºC to 41ºC.

Recordings of *C. barbara* were made on a total of 122 individuals from 15 localities, 11 from the Iberian Peninsula and four from Northwest Africa (Figure 3.1 and Table 3.1). Recordings of the calling songs of *C. orni* were obtained from 19 localities, nine from the Iberian Peninsula region, three from south of France and seven from continental Greece, in a total of 178 individuals (Figure 3.1; Table 3.1).

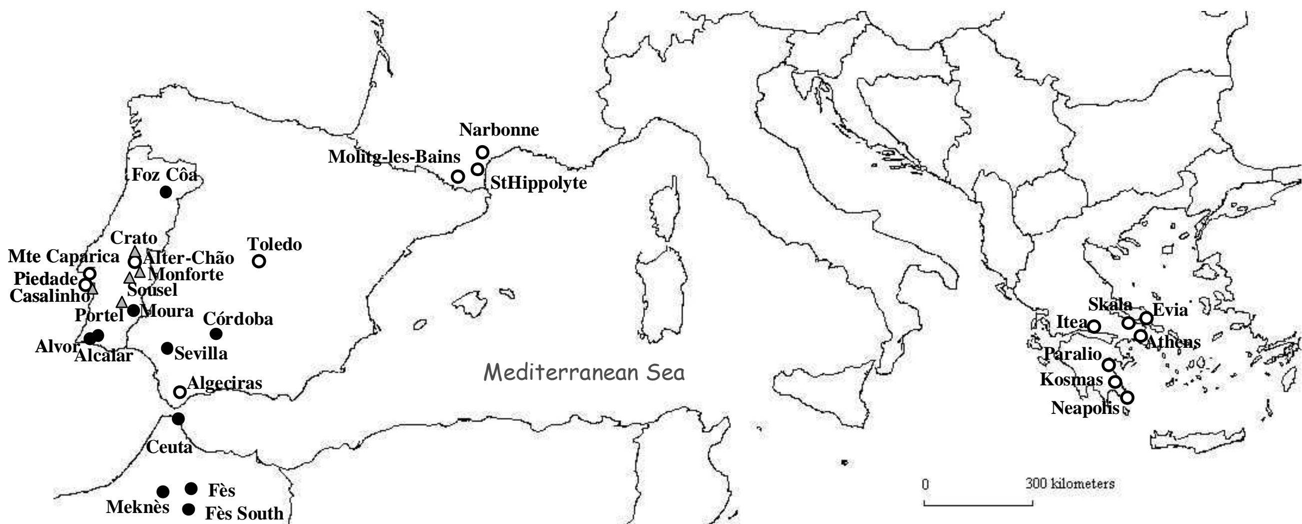


Figure 3.1. Sampled populations of each species (● – allopatric populations of *Cicada barbara*; ○ – allopatric populations of *C. orni*; ▲ – sympatric populations of *C. barbara* and *C. orni*).

To test for any differences in the calling song characters between allopatric and sympatric populations of both species, ten populations from *C. barbara* and eight from *C. orni* from the Iberian Peninsula were analysed in allopatric (five for *C. barbara* and four for *C. orni*) and in sympatric areas (five for *C. barbara* and four for *C. orni*), for a total of 153 individuals (Table 3.1).

To assess the temporal variation at the individual level in the calling song properties, males of both species from Crato in 1999 were recorded at least twice during the day (Table 3.2), with recordings separated by an hour or more, from 9 am to 7 pm. Ambient temperatures were taken at each recording time. Temporal variation at the population level was also assessed by analysing recordings made in different years at the same localities (three populations of *C. barbara* and five populations of *C. orni*) (Table 3.3).

Table 3.1. Number of males (N) of *Cicada barbara* (Cb) and *C. orni* (Co) recorded at each locality when emitting the calling songs, with range of ambient temperatures at recording (T) and indication of populations in allopatry (A) or sympatry (S) relative to the congeneric species.

Population	Locality	N	Date (day/month/year)	T (ºC)	Allopatry/ Sympatry
C. barbara					
	Iberian Peninsula				
CbAlc	Alcalar (Algarve, Portugal)	10	23/8/1995	-	A
CbAlv	Alvor (Algarve, Portugal)	3	28/8/1995	-	A
CbCas	Casalinho (Estremadura, Portugal)	5	27/7/1995	-	S
CbCra	Crato (Alto Alentejo, Portugal)	14	6–8/7; 15/7; 1–3/8/1999	26–41	S
CbFoz	Foz Côa (Beira Alta, Portugal)	3	11/7/1999	34	A
CbMon	Monforte (Alto Alentejo, Portugal)	6	22–24/7/1995	-	S
CbMou	Moura (Baixo Alentejo, Portugal)	11	28/8/2001	31–34	A
CbPor	Portel (Baixo Alentejo, Portugal)	10	24/7; 10/8/2001	31–35	S
CbSou	Sousel (Alto Alentejo, Portugal)	11	8/9/2001	33–35	S
CbCor	Córdoba (Andalucía, Spain)	5	6/9/2000	34	
CbSev	Sevilla (Andalucía, Spain)	7	6/8/2001	38–41	A
	Northwest Africa				
CbCeu	Ceuta (Spain)	12	21–22/7/1999	29–35	
CbFes	Fès (Morocco)	10	2/8/2001	31–35	
CbFesS	Fès South (Morocco)	6	3/8/2001	34	
CbMek	Mèknes (Morocco)	9	4/8/2001	31–37	
C. orni					
	Iberian Peninsula				
CoAlt	Alter-do-Chão (Alto Alentejo, Portugal)	7	6–9/8/1997	25–30	A
CoArr	Arrábida (Estremadura, Portugal)	4	18/8; 10/9/1997	30	S
CoCra	Crato (Alto Alentejo, Portugal)	8	27/6/2001	24–26	S
CoMon	Monforte (Alto Alentejo, Portugal)	16	25/7–7/8/1997	23–38	S
CoMte	Monte-da-Caparica (Estremadura, Portugal)	7	16–22/9/1997	25–30	A
CoPie	Piedade (Arrábida, Estremadura, Portugal)	10	19/7–12/8/1995	-	A
CoSou	Sousel (Alto Alentejo, Portugal)	11	27/6/2003	27–30	S
CoAlg	Algeciras (Andalucía, Spain)	10	5/8/2001	31–34	A
CoTol	Toledo (Castilla-La Mancha, Spain)	2	5/7/2000	-	

Table 3.1 (cont.)

Population	Locality	N	Date (day/month/year)	T (ºC)
	South of France			
CoMol	Molitg-les-Bains (Languedoc-Roussillon, France)	10	17/7/2001	27
CoNar	Narbonne (Languedoc-Roussillon, France)	8	16/7/2001	26
CoStH	St Hippolyte (Languedoc-Roussillon, France)	11	14; 17/7/2001	26–33
	Greece			
CoAte	Athens (Athika, Greece)	18	9 –10/7/1997; 15/7/1998	-
CoEvi	Evia (Athika, Greece)	12	29/6/2002	-
CoIte	Itea (Athika, Greece)	24	26; 29/6/2002	-
CoKos	Kosmas (Peloponnese, Greece)	2	24/6/2002	-
CoNea	Neapolis (Peloponnese, Greece)	7	25/6/2002	-
CoPar	Paralio (Peloponnese, Greece)	7	24/6/2002	-
CoSka	Skala (Athika, Greece)	4	29/6/2002	-

3.2.2. Sound analysis

Sound recordings were digitised through the software AVISOFT-SASLAB PRO (Specht, 2002) at a sampling rate of 44.1 kHz and a resolution of 16 bits. For each recording, a sound fragment of about 60 s was used to produce oscillograms, frequency-time spectrograms and mean amplitude spectra, using fast Fourier transformation with a frame lenght of 512 points and a Hamming Window (bandwidth: 112 Hz; and resolution: 86 Hz) and 50% overlap for temporal resolution, allowing temporal and spectral analyses.

AVISOFT software allowed the automatic measurement of the gross-temporal variables of *C. orni* song derived from the frequency-time spectrogram, namely the number and duration of the acoustic elements (the echemes), and the duration of the interval between them (inter-echeme interval) (Figure 3.2). Echeme rates, periods and ratios of echeme duration per interval duration were calculated from them. Peak frequency (the frequency of maximum amplitude on the spectrum – also called carrier frequency or dominant frequency) was also obtained automatically from the mean spectrum of each echeme in *C. orni* or from the mean spectrum of the entire sound wave in *C. barbara*. For *C. orni*, time and frequency measurements of the echemes were then averaged and the mean was taken as the value of the variable for that individual. A fine-temporal property of the signal, the syllable rate, was obtained for both species by manual screen measurements (using the computer cursor) on the oscillograms. A syllable is a group of sound pulses recognizable in the oscillogram

(Figure 3.2) that corresponds to the movement of one of the two tymbals. The tymbals alternate in sound production, although there is some overlap occurring between the out-movement of one tymbal and the in-movement of the opposite tymbal (Fonseca, 1991). The number of syllables was counted in 30 echemes in *C. orni* (first and last syllables in each echeme were discarded due to their different characteristics – see Figure 3.2) and in 30 fragments of about 0.1 s in *C. barbara*. The average number of syllables per unit of time was calculated for each male.

Table 3.2. Number of recordings (N) of the calling song of males of *C. barbara* (Cb) and *C. orni* (Co) obtained at different times of day and temperatures in the locality of Crato in 1999.

Species	Male ID	Date (day/month)	N	Hours of day	T (ºC)
C. barbara	Cb720	06/07	7	11–18	27.5–33.0
	Cb725	07/07	2	10–11	32.0–33.0
	Cb730	08/07	5	10–14	33.0–40.0
	Cb731	08/07	2	17–18	40.0–41.0
	Cb766	14/07	6	11–18	33.0–38.0
	Cb771	15/07	3	16–18	38.5–39.5
	Cb851	01/08	5	11–17	28.0–35.0
	Cb852	01/08	3	12–17	30.5–35.0
	Cb854	01/08	2	16–17	34.0–35.0
	Cb855	02/08	5	10–17	25.5–34.0
	Cb856	02/08	6	11–17	27.0–34.0
	Cb858	03/08	5	11–16	25.0–32.5
C. orni	Co699	01/07	6	10–17	35.0–39.0
	Co700	01/07	4	12–16	36.5–39.0
	Co707	30/06	2	10–11	34.0
	Co709	30/06	3	13–14	36.0–39.0
	Co712	30/06	2	15–16	38.0–40.0
	Co713	30/06	2	17–18	38.0–39.0
	Co715	02/07	8	11–19	28.0–35.0
	Co716	02/07	6	12–18	30.0–34.5
	Co719	06/07	7	11–18	26.5–32.5
	Co724	07/07	3	09–11	30.0–33.5
	Co726	07/07	4	15–18	38.0–39.0
	Co729	08/07	7	09–17	30.0–41.0
	Co765	14/07	7	11–18	33.0–38.0
	Co767	15/07	4	10–13	30.0–40.5
	Co768	15/07	3	10–12	30.5–40.0

Table 3.3. Number of males (N) of *C. barbara* (Cb) and *C. orni* (Co) recorded in each locality in each year and range of temperatures at recording (T).

Species	Locality	Year	Population	N	T (ºC)
C. barbara	Crato (Alto Alentejo)	1995	CbCra95	4	-
		1996	CbCra96	5	-
		1997	CbCra97	7	25.0–35.0
		1999	CbCra99	12*	25.0–41.0
		2001	CbCra01	3	35.0
	Portel (Baixo Alentejo)	1999	CbPor99	8	35.0–37.0
		2001	CbPor01	10	31.0–35.0
	Sousel (Alto Alentejo)	2001	CbSou01	11	33.0–35.0
		2003	CbSou03	8	29.0–33.0
C. orni	Alter-do-Chão (Alto Alentejo)	1997	CoAlt97	7	25.0–30.0
		1998	CoAlt98	8	30.0–34.0
	Crato (Alto Alentejo)	1996	CoCra96	4	-
		1997	CoCra97	3	25.0–30.0
		1998	CoCra98	3	26.0
		1999	CoCra99	15*	26.5–41.0
		2001	CoCra01	8	24.0–26.0
	Monforte (Alto Alentejo)	1995	CoMon95	4	-
		1996	CoMon96	2	-
		1997	CoMon97	16	23.0–35.0
	Monte-da-Caparica (Área Grande Lisboa)	1996	CoMte96	4	-
		1997	CoMte97	7	25.0–30.0
		1998	CoMte98	3	25.0–29.0
	Piedade (Arrábida, Estremadura)	1995	CoPie95	10	-
		1996	CoPie96	3	-

* Each of these males was recorded more than once during the course of the day (at different times of day/different temperatures; see Table 3.2): CbCra99 – total of 51 recordings; CoCra99 – total of 68 recordings.

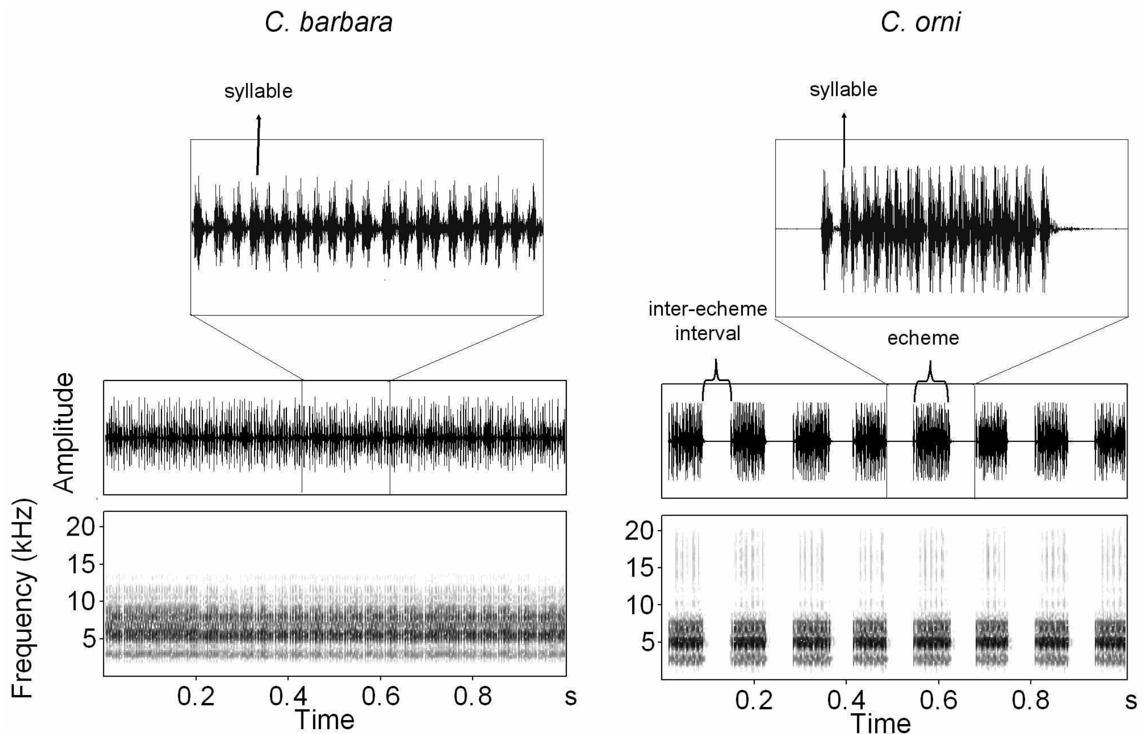


Figure 3.2. Oscillograms (amplitude-time waveform) and frequency-time spectrograms of the calling songs of *Cicada barbara* and *C. orni* from Iberian Peninsula. In the spectrograms, the intensity of each frequency is denoted by a greyscale from light grey (low intensity) to dark grey (high intensity). The peak frequency is the frequency with highest intensity.

Due to the strong background noise of the chorusing cicadas in the Greek samples, which prevented the distinction of syllables, syllable rate could not be calculated for these samples. Moreover, missing data relative to temperature in some sites did not allow taking this environmental factor into account.

Amplitude modulated signals, described by Fonseca (1991) and Boulard (1995) as being the courtship song of this species, and by Quartau & Rebelo (1994) as an alarm signal, were also 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 Northwest 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 digitising 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. The peak frequency and duration of each section in each phrase were obtained using AVISOFT. Also, 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 individual.

3.2.3. Statistical analysis

Assumptions of normality and homogeneity of variances for some variables were rejected (Kolmogorov-Smirnov and Levene tests, respectively) and, thus nonparametric tests were applied. All statistical tests were made using either STATISTICA 6.0 software (StatSoft 2001), MINITAB version 14 software (Minitab Inc, 2004) or SPSS Version 10.0 (SPSS Inc, 1999).

The significance of multiple tests was assessed 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. For *C. barbara*, since there are two variables tested, k=2 and the critical P is 0.025. For *C. orni*, with seven variables, k=7 and the critical P is 0.0073.

Descriptive statistics of each variable for each population were analysed, including the coefficient of variation (CV). Corrected for small samples (Sokal & Rohlf, 1981), CV is calculated as 100x(1+1/4N)xSD/Average; N – sample size; SD – standard deviation. This coefficient allows the comparison of variation between sets of data even when the averages differ greatly.

Spearman rank order correlation coefficients between all pairs of acoustic variables and between these variables and the ambient temperature were calculated.

3.2.3.1. Geographic variation in the calling song of C. barbara and C. orni

Kruskal-Wallis and Mann-Whitney *U* non-parametric tests were used to compare the acoustic variables among populations and among regions. For *C. barbara* two regions were considered, the Iberian Peninsula and Northwest Africa, theoretically corresponding to the subspecfic division of *C. barbara lusitanica* and *C. barbara barbara*, respectively (Boulard, 1982). However, since the specimens from Ceuta in Northwest Africa revealed the same mitochondrial DNA pattern as the Iberian ones (G. Pinto-Juma, personal communication), tests were carried out grouping Ceuta with the remaining Northwest African localities or with the Iberian ones. For *C. orni*, three regions were considered: Iberian Peninsula, South of France and Greece.

To control the effect of temperature, parametric ANCOVA with covariate temperature was also applied but some populations were excluded when no record of temperature was available.

Discriminant Function Analysis (DFA) was carried out to determine statistical significant discriminant functions that might separate the groups (in this case the regions) for each species. For *C. barbara* both peak frequency and syllable rate were used as predictors and two regions were defined *a priori* as the grouping variable. For *C. orni* the acoustic variables were first analysed using Principal Components Analysis (PCA) to reduce them to a smaller number of factors (or components). The percentages of variance explained by the components and the correlation coefficients between the variables and the components (component loadings) were obtained, as well as the component scores for the individuals, which were used to compare groups (in this case the regions) using nonparametric Kruskall-Wallis tests and Mann-Whitney tests. The component scores were then used in the DFA, with the three regions used as the grouping variable.

3.2.3.2. Amplitude modulated signal in C. barbara

The amplitude modulated signal of *C. barbara* was analysed using non-parametric tests 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 again calculated using Spearman correlation coefficient.

3.2.3.3. Comparisons between species in allopatry and sympatry

Mann-Whitney tests were used to compare each acoustic variable among species and also among sympatric and allopatric populations of each species. Discriminant Function Analysis (DFA) was applied to the data matrix of the two acoustic variables common to both species (peak frequency and syllable rate) measured for 153 individuals. In this case four groups were considered: *C. barbara* in allopatry, *C. barbara* in sympatry, *C. orni* in sympatry and *C. orni* in allopatry.

3.2.3.4. Temporal variability at the individual and population level

The within-individual and the among-individual variability of each acoustic variable in cicadas from Crato recorded in 1999 were calculated using the coefficients of variation (CV). Nonparametric Spearman correlations were calculated between each acoustic variable and

the ambient temperature for all observations of all individuals as well as for each individual separately (those with more than three observations).

Nonparametric Kruskal-Wallis or Mann-Whitney tests were used to compare the acoustic variables from different years on the same locality and from different populations. Parametric ANCOVA with covariates temperature and hour was also applied. However, this control of temperature was not complete because some localities in some years lacked a record of ambient temperature.

3.3. Results

3.3.1. Description of the calling songs of *Cicada barbara* and *C. orni*

The calling song of *C. barbara* consists of a continuous emission of pulses (Figure 3.2) produced by the tymbals. Each tymbal is responsible for the repetitive production of a group of pulses, called syllable. In the calling song, each syllable lasts about 5 ms (*ca.* 200 syllables per second; Table 3.4). In the males analysed (N=122), two were apparently either using only one tymbal (results similar to Fonseca (1991) when experimentally destroying one of the tymbals) or there was a synchronous action of both tymbals, as revealed by the oscillograms with half of the number of syllables produced per unit of time. In this case, the peak frequency was not significantly different from the cases in which both tymbals were working (Mann-Whitney test, $p>0.05$), but the variable 'syllable rate' had to be discarded. The peak frequency for all the individuals analysed varied from 5080 to 7660 Hz (Table 3.4).

Coefficients of variation (CV) among individuals within populations of *C. barbara* ranged from 0.9 to 13.6% for peak frequency (7.1% for all individuals) and from 3.3 to 12.8% for syllable rate (9.3% for all individuals) (Table 3.4). Iberian Peninsula and Moroccan populations had similar levels of CV.

The calling song of *C. orni* is made up of a regular repetition of echemes (Figure 3.2), which are composed of a variable number of syllables. This signal can be described in the time domain as having echemes with 0.08 ± 0.03 (average ± standard deviation) seconds of duration separated by intervals of 0.15 ± 0.07 seconds (Table 3.5). The spectral characteristics of the signal showed a peak frequency of 4820 ± 486 Hz.

Peak frequency of *C. orni* presented lower among-individual CV (10%) than the gross-temporal variables (CV ranging from 28% to 74%), whereas the fine-temporal characteristic, the syllable rate, had a CV of 13%.

Table 3.4. Descriptive statistics of the acoustic variables of the calling song of *Cicada barbara*: N – number of individuals analysed; SD – standard deviation; CV – coefficient of variation; at the bottom of the table between parenthesis is the range (minimum–maximum). For abbreviations of population names see Table 3.1.

	Peak frequency (Hz)			Syllables rate (s^{-1})		
Population	**N**	**Average ± SD**	**CV (%)**	**N**	**Average ± SD**	**CV (%)**
CbAlc	10	6051 ± 319.6	5.3	10	198 ± 14.4	7.3
CbAlv	3	5740 ± 52.0	0.9	3	172 ± 12.6	7.4
CbCas	5	5922 ± 166.6	2.8	5	210 ± 9.7	4.6
CbCra	14	6081 ± 461.0	7.6	14	204 ± 21.8	10.7
CbFoz	3	5797 ± 730.4	13.6	2	219 ± 6.5	3.3
CbMon	6	6238 ± 348.5	5.6	6	207 ± 6.8	3.3
CbMou	11	6676 ± 342.3	5.1	10	195 ± 8.5	4.4
CbPor	10	6146 ± 398.1	6.5	10	203 ± 22.2	11.0
CbSou	11	6518 ± 282.9	4.3	11	206 ± 21.5	10.4
CbCor	5	6816 ± 187.3	2.7	5	188 ± 22.1	11.8
CbSev	7	6590 ± 593.2	9.0	7	210 ± 16.8	8.0
CbCeu	12	5991 ± 272.0	4.5	12	193 ± 20.7	10.6
CbFes	10	6612 ± 248.8	3.8	9	201 ± 17.3	8.6
CbFesS	6	6515 ± 343.4	5.3	6	199 ± 25.5	12.8
CbMek	9	6606 ± 163.2	2.5	9	209 ± 11.3	5.4
Total	**122**	**6317 ± 449.7 (5080–7660)**	**7.1**	**118**	**201 ± 18.7 (137.6–252.3)**	**9.3**

In both *C. barbara* and *C. orni* no significant correlations were found between the time variable(s) and the frequency variable. In contrast, echeme duration and the interval between echemes in *C. orni* were weakly but significantly correlated (r_S=-0.362; p<0.001) and echeme rate and echeme period were strongly correlated with inter-echeme interval (r_S=-0.839 and r_S=0.866, p<0.001, respectively) but not with echeme duration (r_S=-0.056, p=0.458; r_S=0.051, p=0.500, respectively).

Table 3.5. Descriptive statistics of the acoustic variables of the calling song of *C. orni*. N – number of individuals analysed; SD – standard deviation; Range – minimum and maximum; CV – coefficient of variation.

Variables	N	Average ± SD	Range	CV (%)
Peak frequency (Hz)	178	4820 ± 486.1	3851–6584	10.1
Syllable rate (s^{-1})	71	220 ± 27.5	149–264	12.6
Echeme rate (s^{-1})	178	4.7 ± 1.30	2.2–7.5	27.7
Echeme duration (s)	178	0.08 ± 0.03	0.03–0.21	36.1
Inter-echeme interval (s)	178	0.15 ± 0.07	0.06–0.36	48.8
Echeme period (s)	178	0.23 ± 0.07	0.13–0.45	31.0
Ratio echeme/ inter-echeme interval	178	0.72 ± 0.53	0.17–3.31	74.0

In *C. barbara* temperature was not correlated with peak frequency (r_S=0.066, p=0.528) and it was significantly correlated with syllable rate, although weakly (r_S=0.340, p<0.001) (see Figure 3.3). In *C. orni* temperature was not correlated with peak frequency, inter-echeme interval or ratio echeme/interval (r_S=-0.027, p=0.835; r_S=-0.036, p=0.776 and r_S=-0.275. p=0.028, respectively) but it was significantly correlated with syllable rate, echeme rate, echeme duration and echeme period (r_S=0.761, 0.505, -0.615, and -0.505, respectively, with p<0.001) (see Figure 3.4).

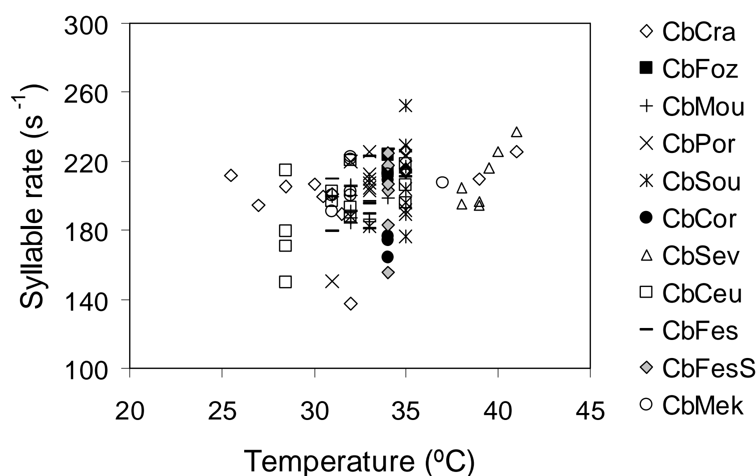


Figure 3.3. Variation of the syllable rate of the calling song of *Cicada barbara* males with the ambient temperature for each population from the Iberian Peninsula and Northwest Africa. For abbreviations of population names see Table 3.1.

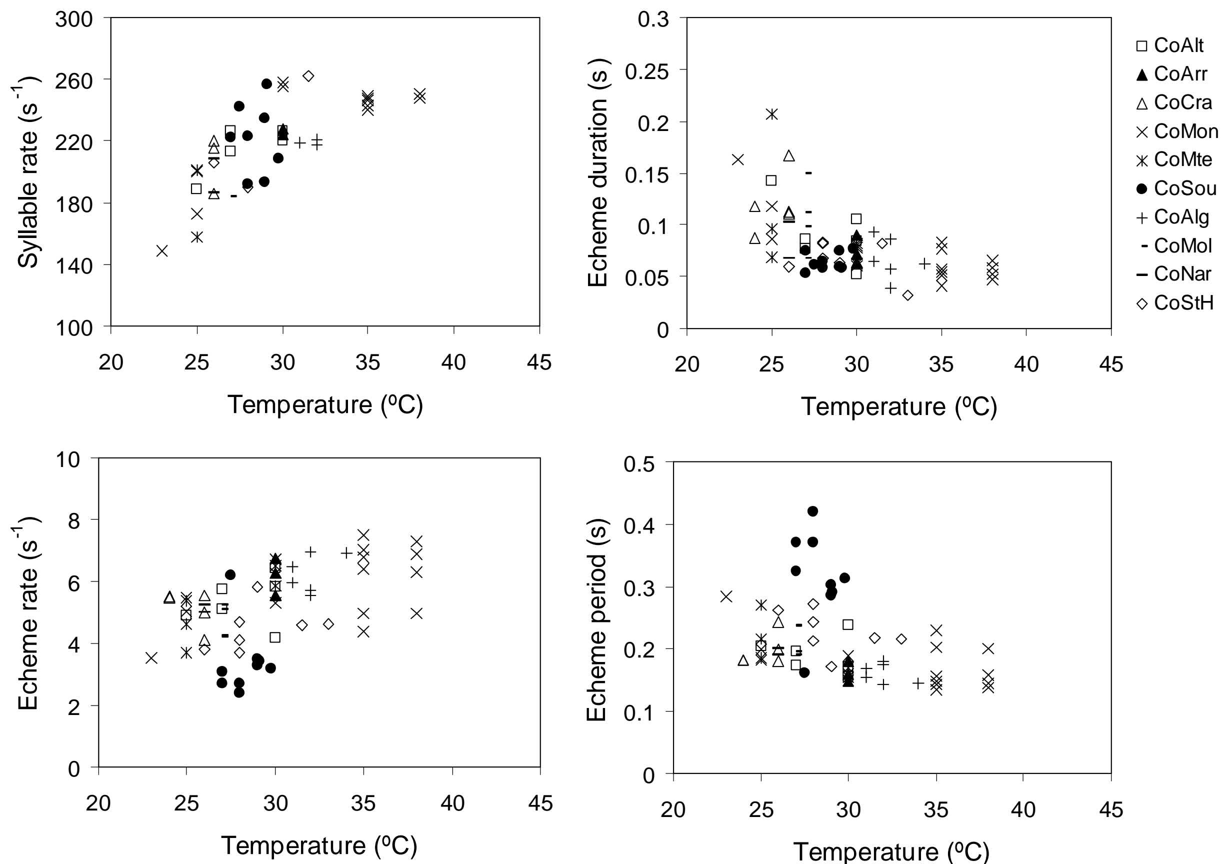


Figure 3.4. Variation of the syllable rate, echeme rate, echeme duration and echeme period of the calling song of *C. orni* males with the ambient temperature for each population from the Iberian Peninsula and France. For abbreviations of population names see Table 3.1.

3.3.2. Geographic variation of the calling song in *Cicada barbara*

Peak frequency of the calling songs differed significantly between populations of *C. barbara* (Kruskal-Wallis tests, p<0.001). The syllable rate did not differ significantly (Kruskal-Wallis tests, p=0.085), although cicadas from Alvor had clearly lower syllable rates than other populations (Figure 3.5). Within Northwest African populations, peak frequency was significantly lower in Ceuta than in the remaining three Northwest African populations (Mann-Whitney tests, p<0.001; Figure 3.5). These three populations did not differ between them in peak frequency (Kruskal-Wallis test, p=0.846). Ceuta had also significantly lower peak frequency than Iberian samples (Mann-Whitney test, p=0.022). Within the Iberian Peninsula, peak frequency differed significantly between populations (Kruskal-Wallis tests, p<0.001). When comparing Iberia and Morocco (excluding Ceuta individuals), peak frequency was significantly higher in Moroccan cicadas than in Iberian ones (Mann-Whitney test, p=0.003). The influence of temperature could not be tested to explain the low syllable rate values found in Alvor (Figure 3.5), as temperature data from this population were missing.

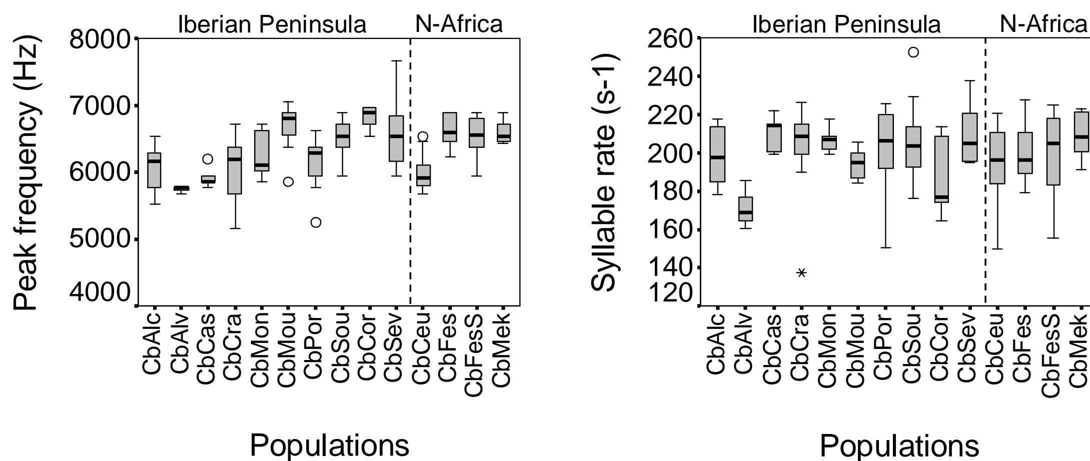


Figure 3.5. Boxplots of the acoustic variables analysed for each population of *C. barbara*. The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum, the circles are outliers, and the stars are extremes. For abbreviations of population names see Table 3.1.

In the Discriminant Function Analysis and when considering the two groups Northwest Africa (including Ceuta) and the Iberian Peninsula, the discriminant function obtained was not significant (Wilk's λ= 0.989, p=0.540). When considering the two groups Northwest Africa (excluding Ceuta) and Iberian Peninsula+Ceuta, DFA obtained a significant function (Wilk's λ= 0.899, p=0.002) with an overall correct classification rate of 68.6% (79.2% correct for Morocco samples and 66% for Iberian Peninsula+Ceuta samples). The structure matrix indicated that the peak frequency was the most important variable in determining the discriminant function (correlation=0.962). When excluding Ceuta and considering two groups, Morocco and Iberian Peninsula, the discriminant function was significant (Wilk's λ= 0.913, p=0.010). Peak frequency was again the variable more correlated with the discriminant function (correlation=0.949). The overall correct classification rate was in this case 67.3% (79.2% correct for Moroccan samples and 63.8% for Iberian samples).

3.3.3. Amplitude modulated signal in *Cicada barbara*

In the amplitude modulated signal, section I of the phrase had higher amplitude (Figure 3.6) and also significantly lower duration and higher syllable rate than Section II (Wilcoxon test, p=0.001 for both) (Table 3.6). In contrast, peak frequency was not significantly different between Section I and II (Wilcoxon test, p=0.701), although some slight frequency modulation is seen in the spectrograms (Figure 3.6). Temperature had no significant correlation with any of the acoustic variables measured in this modulated signal (Spearman correlations, p>0.1). There was no significant difference between African and Iberian

samples nor between Morocco samples and Iberian Peninsula+Ceuta samples in any variable (Mann-Whitney tests, p>0.1). The acoustic variable with the highest coefficients of variation was the temporal variable duration of sections (CV from 21.3% to 32.7%), whereas the peak frequency and the syllable rate were less variable (CV from 5.2% to 10.8%) (Table 3.6).

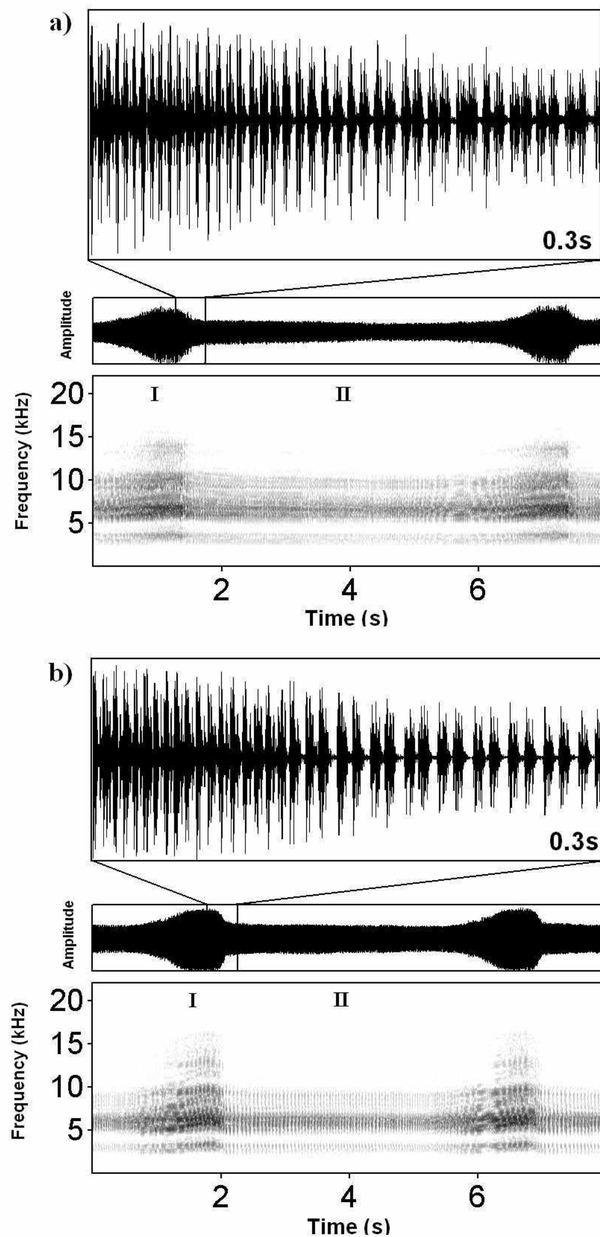


Figure 3.6. Oscillogram (amplitude-time waveform) and frequency-time spectrogram of the amplitude modulated signal of one male of *Cicada barbara lusitanica* from Moura (a) and one male of *C. barbara barbara* from Fès (b). Sections I and II of a phrase are shown and a fragment of 0.3 seconds of the transition between sections is shown enlarged on top, depicting the clear difference in syllable rates between Section I and II.

Table 3.6. Descriptive statistics for the acoustic variables measured in the phrases of the amplitude modulated signal of *Cicada barbara barbara* (Northwest Africa) and *C. barbara lusitanica* (Iberian Peninsula). SD – standard deviation; CV – coefficient of variation.

		C. barbara barbara			***C. barbara lusitanica***		
		Average ± SD	**Range**	**CV (%)**	**Average ± SD**	**Range**	**CV (%)**
Section I	**Peak frequency (Hz)**	6334.1 ± 322.6	5758–6854	5.2	5971.2 ± 627.9	5137.1–6771.7	10.8
	Duration (s)	0.9 ± 0.3	0.7–1.6	29.9	0.9 ± 0.2	0.7–1.1	21.3
	Syllable rate (s^{-1})	217.8 ± 17.8	175.1–229.1	8.4	225.1 ± 14.3	209.9–242.6	6.5
Section II	**Peak frequency (Hz)**	6415.6 ± 343.8	5791–6860	5.5	6137.1 ± 506.6	5287.1–6534	8.5
	Duration (s)	2.8 ± 0.9	1.9–4.4	32.7	2.7 ± 0.9	2.1–4.1	32.3
	Syllable rate (s^{-1})	123.1 ± 9.9	101.4–133.6	8.2	132.4 ± 7.3	120.5–140.6	5.7
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
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

3.3.4 Geographic variation of the calling song in *Cicada orni*

Peak frequency of the calling song in *C. orni* was significantly higher in Greek samples than in the other regions (Mann-Whitney tests, p<0.002; Figure 3.7), whereas samples from the Iberian Peninsula and France were similar to each other (Mann-Whitney test, p=0.682). Syllable rate was significantly higher in Iberian cicadas than in French cicadas (Mann-Whitney test, p= 0.002). No data on syllable rate was available from Greek samples. Echeme duration did not differ between the three regions (Kruskal-Wallis test, p=0.377), but echeme rate and echeme period were significantly different between the three regions (Mann-Whitney tests, p<0.006), with higher echeme rates (and shorter echeme periods) in Iberian cicadas, intermediate in French and lower echeme rates (and longer echeme periods) in Greek cicadas. Inter-echeme interval was longer in Greek cicadas than in the other regions (Mann-Whitney tests, p<0.002; Figure 3.7) but did not differ between the Iberian Peninsula and France (Mann-Whitney test, p=0.013). Ratio echeme/interval was significantly different between Iberian and Greek cicadas but not between Iberian and French or between French and Greek cicadas (Mann-Whitney tests). When controlling for temperature (ANCOVA), there were no significant differences between the Iberian Peninsula and France for any of the acoustic variables. No recorded temperature data were available for Greek cicadas, but the temperature was usually around 30ºC (J.A. Quartau & P.C. Simões, personal communication).

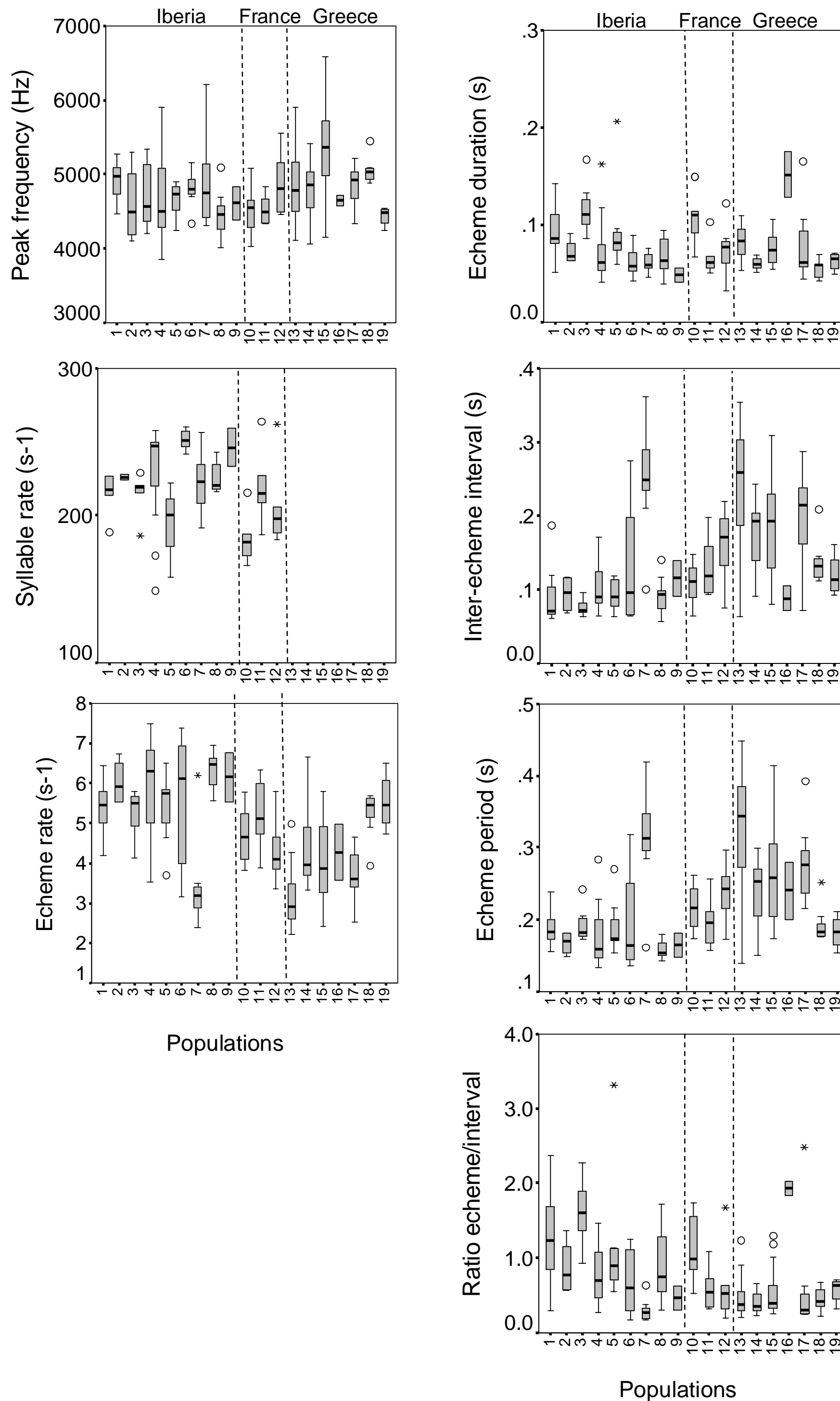


Figure 3.7. Boxplots of the acoustic variables analysed for *C. orni* for each population in each region: Iberian Peninsula (1 – CoAlt; 2 – CoArr; 3 – CoCra; 4 – CoMon; 5 – CoMte; 6 – CoPie; 7 – CoSou; 8 – CoAlg; 9 – CoTol); France (10 – CoMol; 11 – CoNar; 12 – CoStH) and Greece (13 – CoAte; 14 – CoEvi; 15 – CoIte; 16 – CoKos; 17 – CoNea; 18 – CoPar; 19 – CoSka; for abbreviations of population names see Table 3.1). The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum, the circles are outliers, and the stars are extremes.

In Greece, all the variables showed significant differences among populations (Kruskal-Wallis tests, $p<0.003$), except the ratio echeme/inter-echeme interval (Kruskal-Wallis tests, $p=0.131$), although Kosmas had substantially higher ratio than the other populations (Figure 3.7). In the Iberian Peninsula every time variable showed significant differences among populations (Kruskal Wallis tests, $p<0.006$), whereas peak frequency did not (Kruskal-Wallis test, $p=0.333$). French populations did not differ significantly among them for any variable, except for echeme duration ($p=0.006$). When controlling for temperature (ANCOVA) no differences between populations (both Iberian and French taken together) were found in peak frequency, syllable rate and echeme duration ($p=0.582$, $p=0.743$ and $p=0.069$, respectively), whereas significant differences were found for all the remaining variables ($p<0.002$).

Some populations deviated quite obviously from the others for some variables (Figure 3.7). Sousel (Portugal) had on average longer inter-echeme intervals than the remaining Portuguese populations, which was also reflected in shorter echeme rates, longer echeme periods and shorter ratios echeme/interval (similar to the values found in Greek populations). In Greece, it was Kosmas that showed longer echemes and shorter intervals than the other Greek populations investigated.

In the Principal Components Analysis the syllable rate could not be used due to the missing data in Greek samples. The first three components of the PCA explained 97.7% of the variation (54.5% explained by the first component, 82.2% by the first two). Component loadings indicated that the inter-echeme interval was the variable more correlated with the first component (0.986), echeme duration with the second (0.934) and peak frequency with the third (0.944) (Table 3.7). The component scores of individuals for the first and third components showed significant differences among regions (Kruskal-Wallis test, $p<0.03$). The second component did not show differences among regions (Kruskal-Wallis test, $p=0.073$). In the first component the scores were significantly different between all three regions, with lower mean scores in the Iberian Peninsula individuals, intermediate in French and higher in Greek (Mann-Whitney tests, $p<0.02$; Figure 3.8). In the third component the only significant difference was between Greece and France, with French samples presenting lower mean scores than Greek ones (Mann-Whitney test, $p=0.015$; Figure 3.8). These results reflect the univariate results shown above of Greek cicadas having longer inter-echeme intervals and higher peak frequencies than the remaining ones.

Table 3.7. Correlations (component loadings) between each variable and each component extracted in the Principal Components Analysis based on the matrix of the six acoustic variables (syllable rate excluded) for each *C. orni* individual.

	Components			
Variables	**1**	**2**	**3**	**4**
Peak frequency	0.288	0.160	0.944	0.0008
Echeme rate	-0.873	-0.420	0.01	0.217
Echeme duration	-0.323	0.934	-0.05	-0.001
Inter-echeme interval	0.986	0.002	-0.01	0.126
Echeme period	0.904	0.387	-0.126	0.127
Ratio echeme/interval	-0.730	0.659	-0.002	0.008

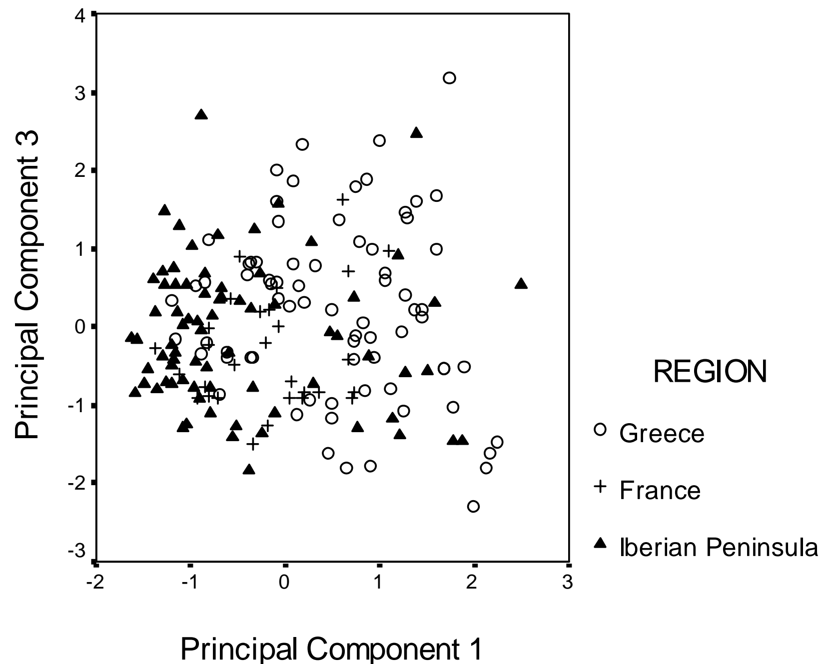


Figure 3.8. Component scores of *C. orni* individuals from the Iberian Peninsula, France and Greece for components 1 and 3 obtained from Principal Component Analysis based on the data matrix of the acoustic variables (*cf*. Table 3.7) for each individual.

Component 1 and 3 obtained in the PCA were used in the Discriminant Function Analysis. The first discriminant function obtained was significant (Wilk's λ= 0.751, $p<0.001$) and explained 96.6% of the variation, but allowed only an overall correct classification of 60.7% (65.3% correct for Iberian samples, 37.9% for French samples and 64.9% for Greek samples).

3.3.5. Comparisons between species

Additionally to the different gross-temporal patterns of the calling songs of *Cicada barbara* and *C. orni*, they also differed significantly in both peak frequency and syllable rate (Mann-Whitney, $p<0.001$; Table 3.8). Considering only the Iberian cicadas, the peak frequency was lower on average for *C. orni* [4710 Hz ± 452.44 (average ± standard deviation)] than for *C. barbara* (6250 ± 469.5), whereas the syllable rate was on average higher in *C. orni* (224 ± 25.0) than in *C. barbara* (202 ± 18.4). The among-individual coefficient of variation of the peak frequency was 7.5% in *C. barbara* and 9.6% in *C. orni* and of the syllable rate was 9.1% in *C. barbara* and 11.2% in *C. orni*.

The power spectrum is broad in both species. Most energy of the spectrum (bandwidth at -20 dB) is found between 2825 and 10 796 Hz in *C. barbara* and between 2126 and 9100 Hz in *C. orni* (see Figure 3.2).

Table 3.8. Descriptive statistics of the acoustic variables in *C. barbara* and *C. orni* from the Iberian Peninsula. N – number of individuals analysed; SD – standard deviation; Range – minimum and maximum; CV – coefficient of variation.

	C. barbara				***C. orni***			
Variables	N	Average ± SD	Range	CV(%)	N	Average ± SD	Range	CV(%)
Peak frequency (Hz)	80	6250 ± 469.5	5080–7660	7.5	73	4710 ± 452.4	3851–6199	9.6
Syllable rate (s^{-1})	77	202 ± 18.4	137.6–252.3	9.1	53	224 ± 25.0	148.7–260.3	11.2

No significant differences were found between sympatric and allopatric populations of *C. barbara* in peak frequency (Mann-Whitney test, $p<0.001$), but the syllable rate was significantly higher in sympatry than in allopatry (Mann-Whitney test, $p=0.021$; Figure 3.9). However, when controlling for temperature (ANCOVA), no difference was found between allopatric and sympatric situations ($p=0.354$).

In *C. orni*, no difference was found between allopatric and sympatric populations both in peak frequency and syllable rate (Mann-Whitney tests, $p=0.666$ and $p=0.886$, respectively). Also, controlling for temperature, no difference was found between these two groups of populations.

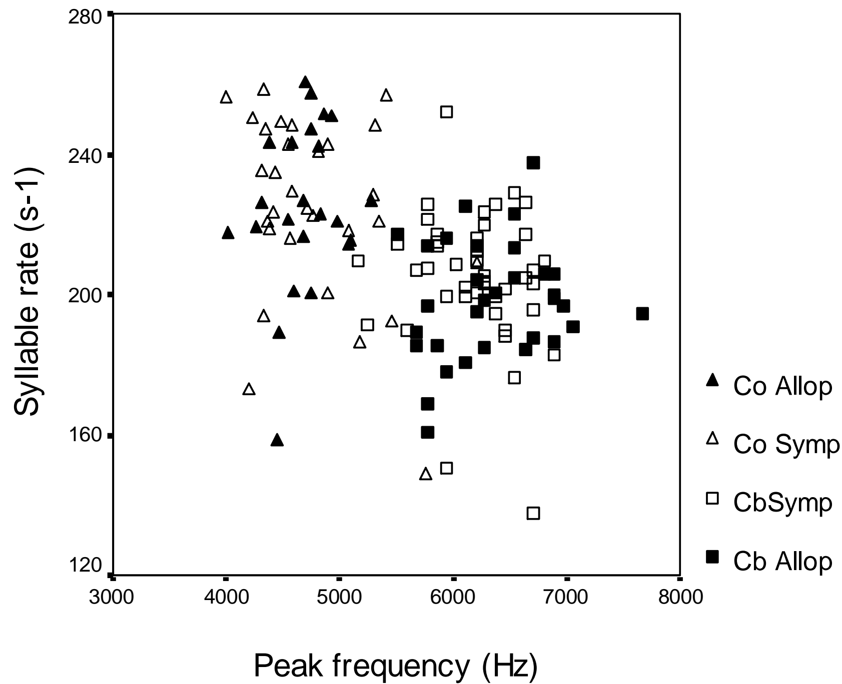


Figure 3.9. Syllable rate *versus* peak frequency of the calling songs of *C. barbara* (Cb) and *C. orni* (Co) in allopatry (Allop) and in sympatry (Symp).

Using both peak frequency and syllable rate as predictors, the Discriminant Function Analysis gave an overall correct classification rate of 54.6%. Two functions were computed but only the first one was significant (Wilk's λ= 0.246, p<0.001) with 99.8% of the variation explained by it. The structure matrix indicated that the peak frequency was the most important in determining Function 1 (correlation=0.981). The classification table showed that almost every *C. barbara* and *C. orni* were correctly classified in the respective species, with the exception of five samples (two sympatric *C. barbara* were classified as sympatric *C. orni*; one sympatric *C. orni* was classified as allopatric *C. barbara*; and two sympatric *C. orni* were classified as sympatric *C. barbara*). However, a substantial number of samples of either allopatric or sympatric groups were misclassified. Samples of *C. orni* in sympatry were mostly classified in the allopatric group.

3.3.6. Individual variation

In each species, there were significant differences among individuals from Crato 1999 on every acoustic variable (Kruskal-Wallis tests, p<0.003). In *C. barbara* the coefficients of variation were in general higher among individuals than within individuals (Figure 3.10). In *C. orni* the coefficients of variation were also generally higher among individuals than within individuals, except for echeme duration and interval, for which they were very similar (Figure 3.11).

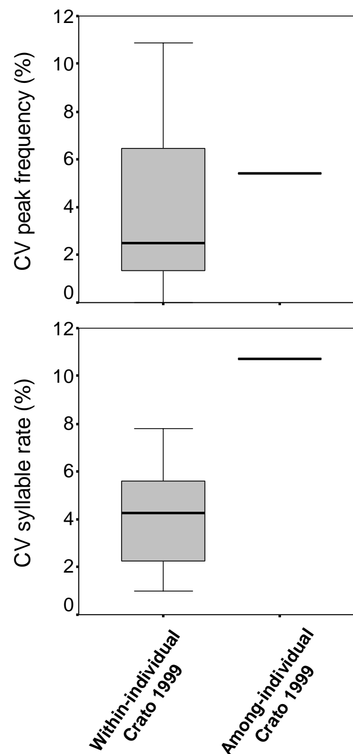


Figure 3.10. Boxplots of the coefficients of variation (CV), in percentages, for each acoustic variable within individuals and the value of among-individual CV in the population of *C. barbara* of Crato (1999). The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum.

The relationship of each acoustic variable with temperature and hour of the day was analysed for the recordings at Crato in 1999, where several individuals were recorded at least twice during the course of the day. Considering the total recordings of *C. barbara*, the temperature was positively and significantly correlated with the syllable rate (Table 3.9a). Controlling for the hour of day (partial correlations) the correlation was still positive but non-significant, and the non-significant correlation with peak frequency was maintained. However, when considering the correlations for each male individually (only considered with more than three recordings), the results were not consistent among individuals, except that, for the syllable rate, every individual had a positive correlation with temperature (Table 3.9a; Figure 3.12).

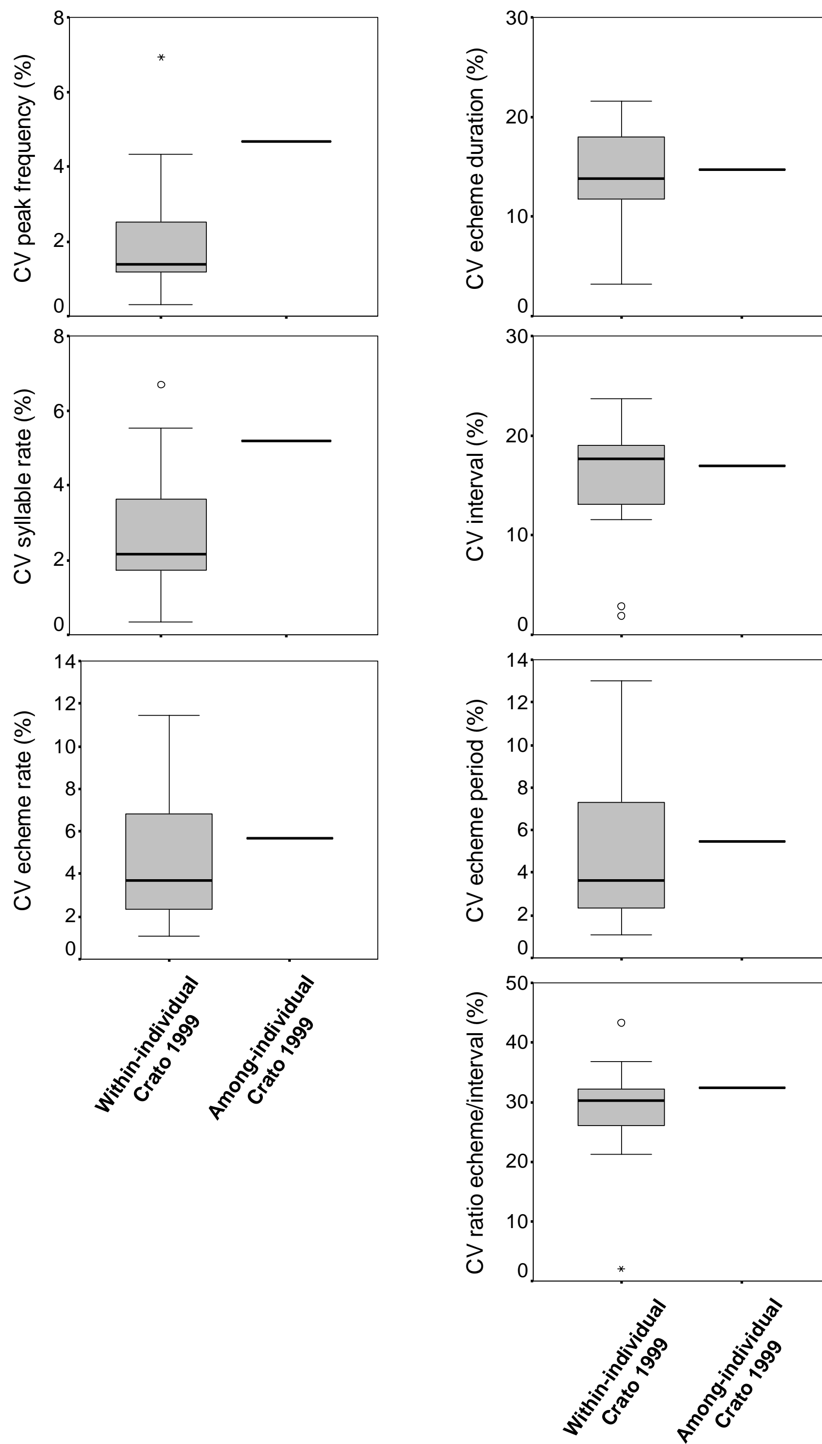


Figure 3.11. Boxplots of the coefficients of variation (CV), in percentages, for each acoustic variable within individuals and the value of among-individual CV in the population of *C. orni* of Crato (1999). The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum, the points are outliers, and the stars are extremes.

In *C. orni* the results were very discrepant among individuals (Table 3.9b). Using the total of observations, significant correlations with temperature were found for echeme duration (negative), inter-echeme interval (positive) and ratio echeme/interval (negative) (Table 3.9b; Figure 3.13). The correlation of syllable rate with temperature was very low and non-significant (Table 3.9b; Figure 3.13), but when excluding the male Co729 (see Figure 3.13) the correlation increased and became significant (r_S=0.357, p=0.005). When controlling for the time of day (partial correlation) the results were similar (Table 3.9b). Again, when excluding the male Co729, the partial correlation of syllable rate with temperature (controlling for time of day) was significant and positive (r_S= 0.503, p<0.001).

Table 3.9. Spearman correlation coefficients of each acoustic variable with the ambient temperature for each individual cicada and for the total of the observations of *Cicada barbara* (a) and of *C. orni* (b) from Crato in 1999. Partial correlations (parametric) of the acoustic variables and temperature controlled for hour of day are also given. *Significant after Dunn-Sidák correction.

3.9a)

	Individuals							All individuals	
	Cb720	Cb730	Cb766	Cb851	Cb855	Cb856	Cb858	Simple correlation	Partial correlation
Peak frequency	-0.612	-0.154	0.247	0.224	0	-0.926*	0.949*	-0.311	-0.336
Syllable rate	0.286	0.800	0.714	0.300	0.400	0.829	0.500	0.616*	0.263

3.9b)

	Individuals									All individuals	
	Co699	Co700	Co715	Co716	Co719	Co726	Co729	Co765	Co767	Simple correlation	Partial correlation
Peak frequency	0.714	0.800	0.59	0.771	0.883	0.316	0.847	-0.306	0.600	0.244	0.185
Syllable rate	0.429	-0.800	0.843	0.886	0.523	-0.632	-0.721	0.144	0.800	0.192	0.141
Echeme rate	0.714	-0.800	0.735	0.771	0.955*	0.632	0.342	0.811	0.200	-0.056	-0.174
Echeme duration	0.771	-0.800	-0.916*	0.086	-0.577	-0.632	-0.559	-0.559	0.400	-0.405*	-0.420*
Inter-echeme interval	-0.829	0.600	-0.048	-0.6	0.18	-0.316	0.577	0.324	-0.200	0.439*	0.522*
Echeme period	-0.714	0.800	-0.735	-0.771	-0.955*	-0.632	-0.342	-0.811	-0.200	0.055	0.142
Ratio echeme/interval	0.829	-0.800	-0.723	0.257	-0.541	0.316	-0.505	-0.342	0.200	-0.481*	-0.576*

The hour of day (controlled for temperature) was significantly correlated with echeme rate (r_S=0.415, p<0.001) and echeme period (r_S=-0.404, p<0.001) in *C. orni*.

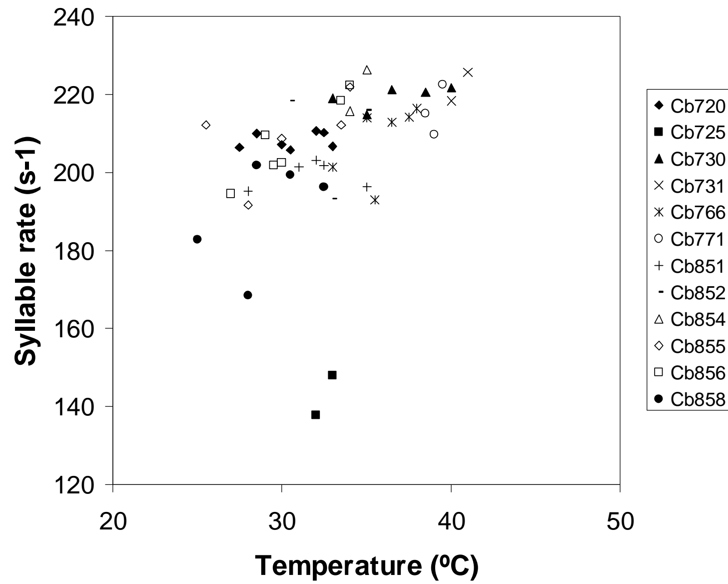


Figure 3.12. Scatterplot of syllable rate and temperature for each individual of *C. barbara* from Crato (1999) recorded two or more times during the day.

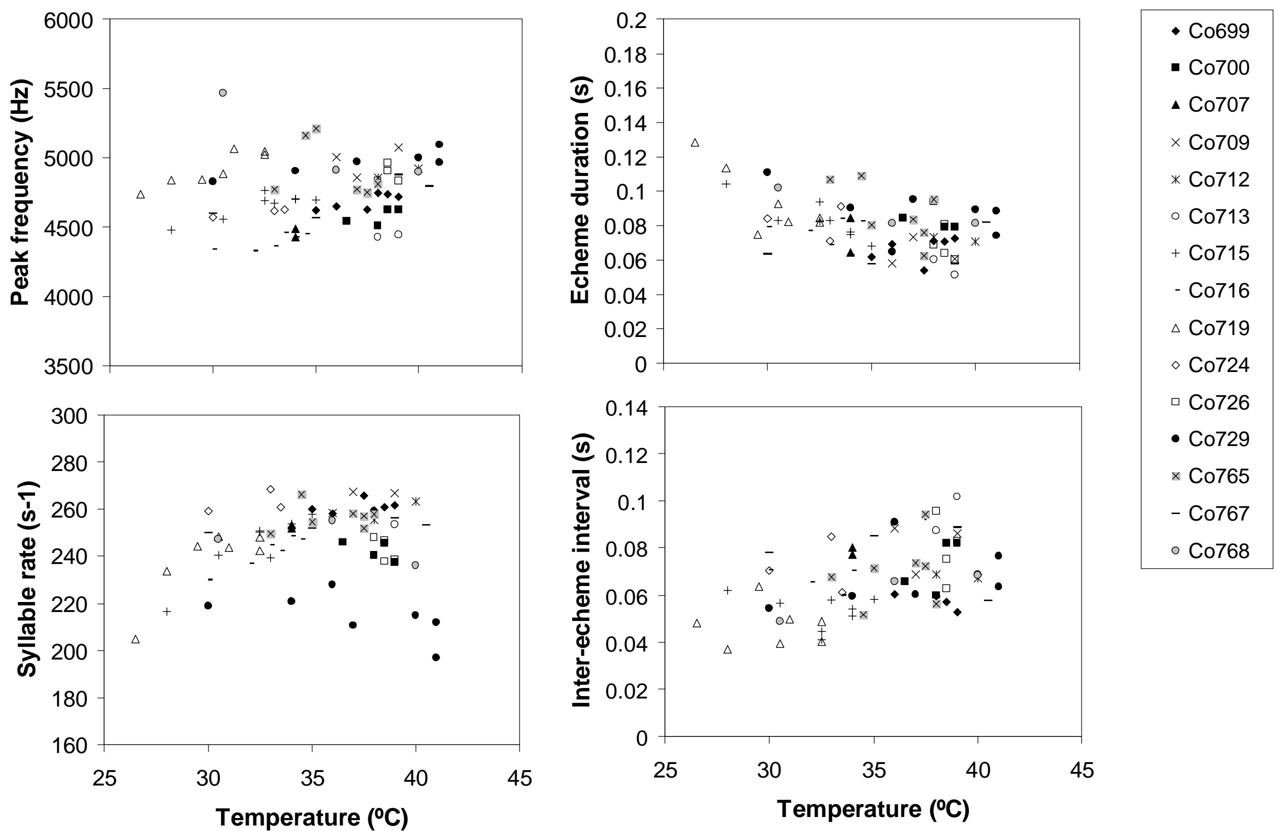


Figure 3.13. Scatterplot of each acoustic variable and temperature for each individual of *C. orni* from Crato (1999) recorded two or more times during the course of the day.

3.3.7. Temporal variation at the population level

No differences were found in any acoustic variable among years within each locality for *C. barbara* (Kruskal-Wallis test for Crato and Mann-Whitney tests for Portel and Sousel, Table 3.10a). When comparing localities (including all individuals from different years of the same locality in the same group), there were significant differences among localities in peak frequency, with peak frequency significantly higher in Sousel than in the other populations (Mann-Whitney tests), but there were no significant differences in syllable rate. When controlling for temperature and time of day (ANCOVA) results were similar (Table 3.10a).

For *C. orni* there were no significant differences in the acoustic variables among years in Mte Caparica and Monforte (Table 3.10b). In Alter-do-Chão, syllable rate showed significant difference among years (Mann-Whitney tests; Table 3.10b), which could be partly explained by temperature: some cicadas were recorded in 1997 at lower temperatures than the ones recorded in 1998. In Piedade it was peak frequency that showed significant differences among years. In Crato, peak frequency did not show significant differences among years but all the time variables did (Kruskal-Wallis tests; Table 3.10b) and results were similar when controlling for temperature and time of day (ANCOVA). There were significant differences among localities (different years taken together in each locality) for every temporal variable. Temperature data were missing for populations from Piedade and Mte Caparica and thus no ANCOVA could be done. When excluding Crato from the comparisons among localities, only echeme duration was significantly different among the four localities, with Monte-da-Caparica and Alter-do-Chão having longer echemes than Monforte and Piedade. It was not possible to test for the effect of temperature and hour due to missing data.

Table 3.10. P-values obtained from nonparametric tests (two-samples – Mann-Whitney test; more than two samples – Kruskal-Wallis test) comparing the years of sampling within each locality and comparing the localities for each species, *C. barbara* (a) and *C. orni* (b). ANCOVA results are also given, with covariates being temperature and hour of day. *Significant after Dunn-Sidák correction.

3.10a)

	Among years within locality			Among localities	Controlling for temperature and hour of day
	Crato (5 years)	Portel (2 years)	Sousel (2 years)	(3 localities)	
Peak frequency	0.447	0.965	0.545	0.004*	0.006*
Syllable rate	0.087	0.573	0.536	0.951	0.865

Table 3.10 (cont.)

3.10b)

	Among years within locality					Among localities	
	Alter-do-Chão (2 years)	Piedade (2 years)	Mte. Caparica (3 years)	Crato (5 years)	Monforte (3 years)	(5 localities)	(4 localities, Crato excluded)
Peak frequency	0.397	0.007*	0.165	0.624	0.394	0.749	0.700
Syllable rate	0.001*	0.018	0.951	0.000*	0.516	0.001*	0.024
Echeme rate	0.694	0.692	0.597	0.000*	0.159	0.000*	0.050
Echeme duration	0.336	0.692	0.223	0.000*	0.329	0.000*	0.001*
Inter-echeme interval	0.955	0.937	0.093	0.003*	0.112	0.000*	0.312
Echeme period	0.694	0.692	0.597	0.000*	0.159	0.000*	0.052
Ratio echeme/interval	0.694	0.937	0.057	0.004*	0.236	0.000*	0.058

3.4. Discussion

Gross-temporal characters of the calling song in *C. orni* were found to be more variable (higher coefficients of variation) than the sound peak frequency as reported for other cicadas (*e.g.*, Sueur & Aubin, 2002). This was expected since the sound frequency characteristics are especially constrained by the physical properties of the sound producing organ. In fact, as stated for acoustic insects in general by Stumpner & von 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.

3.4.1. Temperature effect

Time variables were more consistently related to temperature than the frequency variable. Sound frequency of the songs is independent of temperature in cicadas, as expected from its sound producing system (Sanborn, 2006). In contrast, time variables, particularly syllable rate, are influenced by body temperature, since tymbal muscles contract more rapidly and with greater force as the temperature of the muscle increases (Sanborn, 1997; 2006). In this study we have measured ambient temperature and not body temperature, which are not always the same in cicadas. In fact, several cicada species are known to thermoregulate (Sanborn, 1997). Temporal properties were shown to be dependent of ambient temperature in some species but not in others (reviewed in Sanborn, 2006).

The syllable rate was found to increase significantly with the environment temperature in both *C. barbara* and *C. orni* and for both the dataset with all populations and the dataset with several recordings per individual in Crato population in 1999. Fonseca (1991) also found a temperature dependence of the period of syllables in *C. barbara*, and an increase of the syllable rate related with temperature was described for other cicada species such as *Tettigetta argentata*, *T. josei* and *Tympanistalna gastrica* (Fonseca & Revez, 2002b).

For the dataset with all populations of *C. orni* the echeme duration decreased significantly when the environment temperature increased and the inter-echeme interval showed a very low non-significant correlation with temperature. However, when analysing the dataset from Crato in 1999, with several recordings per individual at different temperatures, inter-echeme interval increased significantly with temperature. The differences observed could be due to a sampling effect.

Some of the recordings of *C. orni* made at Crato in 1999 were previously analysed for time variables using only a 10-second sample to calculate the echeme rate and 15 echemes to calculate the echeme and inter-echeme interval duration (Quartau *et al.*, 2000b). Even with such a small sample, the results were very similar to the ones obtained here (no correlation of ambient temperature with echeme rate, significant positive correlation with inter-echeme interval, negative correlation with echeme duration, even if not significant in this last case with p=0.079).

Echeme rate and echeme period in *C. orni* did not correlate with temperature but did correlate with the hour of day (controlling for temperature) which could indicate some 'motivational' difference in the singing pattern at different times of day.

3.4.2. Geographic variation of the calling song and of the amplitude modulated signal of *Cicada barbara*

The values found for the acoustic variables of the calling song of *C. barbara* in the present study are within the range found by other authors. Fonseca (1991) reported syllable rates of 162–238 per second and a peak frequency of 5.5–6.5 kHz on males of *C. barbara lusitanica*, and Boulard (1995) described the calling song of a male of *C. barbara barbara* from North Africa with a peak frequency of 6.2 kHz.

The statistical analyses carried out on both acoustic variables did not completely separate the regions or subspecies, with substantial overlap of values from both regions. 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 this study.

The syllable rate of calling songs was not significantly different among populations or among regions in this study, whereas peak frequency presented statistically significant differences. Cicada males from Ceuta showed lower peak frequency in calling songs than males from Morocco and from the Iberian Peninsula. Excluding Ceuta, peak frequency was higher in Morocco than in the Iberian Peninsula. Differences in the frequency of the sounds are generally related with the size of the sound producing organ or resonator 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). Preliminary morphologic data of these cicadas showed that Ceuta cicadas are larger on average than Iberian or Moroccan cicadas (and hence produce lower frequency sounds), and Moroccan cicadas are smaller than Iberian ones (and hence produce higher frequency sounds) (Appendix V). Body size may be pleiotropically constrained by, or adapted to, some function, such as predator avoidance or feeding. It may also be related to difference in feeding efficiency (host plant food quality or competition for resources) between the nymphs (Villet, 1995).

Ceuta population is probably isolated from the other populations by two significant barriers: the Mediterranean Sea to the north and the Rif Mountains to the south, which may be causing, at some degree, the divergence of this population from Iberian populations and from the other mainland Morocco populations, located south of the Rif Mountains. The acoustic data here presented shows a higher similarity of the calling song of Ceuta to the Iberian cicadas than to the Moroccan cicadas, but this is based on only one acoustic variable (peak frequency). This could be due to a higher genetic similarity with Iberian cicadas (higher gene flow), or, and as seen before, due to environmental factors influencing the body size of the cicadas. Microsatellite (see Chapter 2) and mitochondrial data (G. Pinto-Juma, personal communication) do not give a clear answer about the origins of this population.

3.4.3. Geographic variation of the calling song of *Cicada orni*

The values of the acoustic variables of the calling song of *C. orni* obtained in this study are within the range of variation observed in other studies. Boulard (1995) described the calling song of a male from Provence (France) with a peak frequency of 5250 Hz, which is within the range found for the French populations of this study. Fonseca (1991) described the calling

song of five males from Portugal with a syllable period of 4.1–4.8 ms (syllable rate of 208.3–243.9 per second), a duration of echemes of 45–128 ms, an interval between echemes of 43–104 ms, an echeme period of 148–155 ms, and a peak frequency of about 4.5 kHz, all within the range found for Portuguese populations of the present study. The data for the calling song of *C. orni* by Joermann & Schneider (1987) and Popov (1975), for former Yugoslavia and south USSR, respectively, are closer to the values here obtained for the Iberian Peninsula and French cicadas than to the Greek ones, especially in the inter-echeme interval duration. In fact, the range of values of this variable obtained for Yugoslavian samples was 39–127 ms (Joermann & Schneider 1987), and for south USSR was 45–87 ms (calculated from Popov (1975) by subtracting the mean echeme duration from the mean period duration of each individual). These values are 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).

Significant differences were found in some calling song properties among regions in this study, even though there was not a complete distinction. Songs of South-eastern Europe (Greece) cicadas differed from those of the Western Europe (Iberian Peninsula and France) in peak frequency, inter-echeme interval, echeme rate and echeme period. This considerable differentiation in the calling song is in agreement with the genetic differentiation found with allozymes (Quartau *et al*., 2001), microsatellites (see Chapter 2) and mitochondrial DNA (G. Pinto-Juma, personal communication) and is probably due to the considerable isolation of such populations, with the Balkan mountain ranges being probably a substantial barrier for the dispersion of these cicadas.

Inter-echeme interval was the variable that contributed most for the separation of individuals in the Principal Components Analysis. In fact, Greek male songs showed on average longer inter-echeme intervals than the songs of Western Europe, as found previously by Quartau *et al.* (1999). An exception is the Portuguese population of Sousel, which in general showed long inter-echeme intervals. In contrast, the population of Kosmas (Greece) showed shorter inter-echeme intervals and longer echemes than the typical Greek populations. As this population was the only one recorded at high altitude (>1000m), more thorough studies should be carried out on this locality as well as in Sousel to better understand these singular findings.

Songs from Greek cicadas also showed significantly higher peak frequency than cicadas from the other regions. As referred to for *C. barbara* (see 3.4.2), sound frequency is related to body size in cicadas and Greek cicadas analysed by Ribeiro (1998) were found to be

smaller than Iberian ones, which may explain the higher peak frequency found in Greek cicadas.

The only time variable that did not show any significant differences between each pair of regions, *i.e.*, which was quite constant across the studied geographic range of this cicada, was the echeme duration. Comparing *C. orni* with its sibling species *C. mordoganensis* and *C. cretensis*, echeme duration is the variable that more obviously readily distinguishes these species (Simões *et al.* 2000; Quartau & Simões, 2006). Therefore, it is probable that echeme duration might be one of the most important parameters encoding information for species recognition.

3.4.4. Calling songs in sympatric and allopatric populations of *Cicada barbara* and *C. orni* in the Iberian Peninsula

The calling songs of *Cicada barbara* and *C. orni* are easily distinguished by the human ear since the first species produces a continuous shrill and the latter a successive series of short shrills alternated with short pauses. At the song 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 higher than 1 kHz between the average for each species. However, this is not a totally diagnostic character because there is an overlap between species. In fact, some *C. orni* males produce a sound with a peak frequency above 6 kHz and some *C. barbara* males produce a peak frequency as low as 5 kHz. Also the syllable rate values show substantial overlap between species, despite the average being significantly higher in *C. orni* than in *C. barbara*. For these acoustic variables, the presence of a few specimens that were similar to the heterospecific, did not enable the Discriminant Function Analysis to show a complete species separation. As seen before, differences in the frequency between species are related to the body size, being *C. barbara* on average smaller in size than *C. orni* (Ribeiro, 1998).

The variability of the peak frequency of the calling songs among cicadas of each species was relatively low (coefficient of variation lower than 10%), as expected since frequency characteristics are constrained by physical properties of the sound-producing organ. Also the fine-temporal characteristic of the songs within both species, the syllable rate, had similar values of variability to the ones of the frequency variables. According to Gerhardt (1994), many female insects choose signals with species-typical values of fine-temporal properties, such as pulse rate or pulse duration. This would explain the stabilizing selection of these characteristics of the song. However, the variability of the characteristics of each song should

be analysed with caution, since the estimation of variability is dependent on our measuring ability and might not reflect the perceptual systems of the species (McGregor, 1991).

In acoustic insects, hybrids usually have calling songs with intermediate characteristics from those of the parental species (Walker, 1998). No evidence of hybrid cicadas with intermediate songs between *C. barbara* and *C. orni* in sympatric areas was found in this study. All male cicadas were clearly identified by the gross-temporal differences in their songs (*C. barbara* song is continuous and *C. orni* song is discontinuous). The overlap in some song variables among species is believed to be due to natural variation and not to hybridization. Moreover, no evidence of character displacement was found on the acoustic variables here studied. The discriminant analysis did not show any consistent differentiation of the calling song between sympatric and allopatric populations.

It is likely that the calling songs, which are most certainly part of the Specific-Mate Recognition Systems in these species acting as premating barriers, were already sufficiently differentiated before species came into contact in the sympatric areas, and therefore, no hybridization occurred. Since these isolating barrriers were 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 were already differentiated before the species came into contact, selection would only need to sharpen the selectivity of females, an aspect that should also be studied in these cicadas. It was already demonstrated that *C. barbara* and *C. orni* males can discriminate the frequency and temporal characteristics of their song and react preferentially to the conspecific songs (Fonseca & Revez, 2002a; Simões & Quartau, 2006), but the preferences of the females were not tested yet.

The overlap in song frequencies between the two species could produce interference in the channel of communication of the species, particularly *C. barbara* song could “mask” *C. orni* song since it is a continuous song and the temporal characteristics of *C. orni* song would be more difficult to perceive by receivers. The fact that late in the summer season, *C. orni* is not found or rarely found in sympatric areas but it remains abundant in allopatric areas, might be a result of this acoustic interference. Different ecological adaptations could eventually reduce this competition. For instance, different singing positions in the vegetation for sympatric cicada species were already described for *C. orni/Lyristes* plebejus (Claridge *et al.*, 1979) and for *C. orni/Tibicina haematodes* (Sueur & Aubin, 2003b). Some spatial segregation may also occur between *C. orni* and *C. barbara*, as observed qualitatively by Ribeiro (1998).

Boulard (1982) also noticed an opposing ecological occupation in Arrábida, with *C. barbara* singing on the sea-side of the mountain and *C. orni* on the inland side. However, in olive trees orchards of some sampling sites of the present study (*e.g.*, Crato, Portel and Sousel) this apparently does not occur, since it was common to see males of both species singing on the same trunks or branches.

In this inter-species analysis, only Iberian Peninsula populations were used to exclude any potential effect of geographic isolation. In fact, populations of these species from other areas presented some differences from the Iberian Peninsula populations, as seen above, that could be due to geographical isolation. Ecological factors, such as habitat specificities and climatic conditions could be responsible for the direct or indirect selection for certain sound frequencies, specifically for the lower frequencies found in both species in the Iberian Peninsula in both species. The influence could be indirect, for example through differences in body size.

Sousel population had longer inter-echeme intervals than the other *C. orni* populations, and this was not a temperature dependent outcome, since longer intervals would be expected at higher temperatures (as seen above) and Sousel individuals were sampled at temperatures from only 27ºC to 30ºC. The recording conditions of *C. orni* population of Sousel might have been unusual and not detected by the researchers. Nevertheless, Sousel is a sympatric area for *C. orni* and *C. barbara* and a character displacement might be occurring with longer silent pauses between echemes in *C. orni*, diverging from the song of *C. barbara* with no silent pauses. However, this hypothesis is not consistent with the findings for other sympatric areas. Furthermore, some cicadas from Piedade, an allopatric area of *C. orni*, also showed long inter-echeme intervals when compared to other populations (see Figure 3.7). In this case, however, the information about the environment temperature was not available.

3.4.5. Temporal variation of the calling song characters of *Cicada barbara* and *C. orni* at the individual and population levels

Studies in the wild on within-individual acoustic variation have been conducted in several anurans and birds (*e.g.*, Bee, 2004; Kopuchian *et al.*, 2004; Runciman *et al.*, 2005; Friedl, 2006), but in insects they are more difficult to carry out due to difficulties in marking and tracking of individuals, and therefore are usually made in controlled captivity conditions (*e.g.*, Butlin *et al.*, 1985; Fonseca & Revez, 2002b). The cicada species studied here allowed the recording of the calling songs of the same individual during one entire day since the males can sing from the same location for hours.

Variation in both frequency and time acoustic variables measured in the calling song was usually higher among individuals than within individuals. Differences among males in the premating mechanisms could affect male's relative mating success. Experimental work on the preferences of females is needed to find out if there are female preferences on the calling song properties.

Differences in the calling song properties between populations were generally higher than between years within the same population. For *C. orni* the difference between populations was mainly due to one population, Crato. Also Crato showed significant differences between years in every time variable, not explained by temperature differences. However, low sample sizes may be responsible for these differences.

Gerhardt (1991) categorised the properties of the calling sounds in tree frogs as static (within-male coefficient of variation <5%) or dynamic (within-male coefficient of variation >12%). According to this classification, several studies on insect and anuran species showed that fine-temporal properties of the songs are usually static and the gross-temporal properties are usually dynamic (Gerhardt & Huber, 2002). Moreover, female choice studies generally show that the preferences based on static properties are stabilizing or weakly directional and those based on dynamic properties are strongly directional (Gerhardt & Huber, 2002). In both *C. barbara* and *C. orni* the syllable rate (fine-temporal variable) showed median within-individual CV lower than 5% (classifying as static property) and median among-individual CVs lower than 12%. The peak frequency also showed low variation in both species, which is most likely due to functional constraints of the sound producing mechanism. In *C. orni* the gross-temporal characteristics of the song, echeme duration and inter-echeme interval, showed high median values of both within- and among-individual coefficient of variation (>12%), classifying as dynamic properties. In contrast, echeme rate and echeme period showed low levels of within-individual variation (median value < 5%) and among-individual variation (<6%). Studies of mate preference in these species are necessary to assess the biological significance of the variation found in these calling song properties.

Chapter 4 – General discussion

I have walked that long road to freedom. I have tried not to falter; I have made missteps along the way. But I have discovered the secret that after climbing a great hill, one only finds that there are many more hills to climb.

Nelson Mandela, *Long Walk to Freedom* (1995)

4. General discussion

The application of microsatellite and acoustic markers to the study of *Cicada barbara* and *C. orni* allowed us to obtain important information about the patterns of spatial and temporal variation in these species. The main results from both types of data are discussed below:

i) Genetic and acoustic analysis showed no evidence of hybridization between Cicada barbara and C. orni

Cicada barbara and *C. orni* showed highly distinct microsatellite alleles and allele frequencies and also distinct calling song properties. The occurrence of two individuals in sympatric areas with one allele typical of the other species (locus Cib03) is not sufficient evidence for hybridization or introgression, although this hypothesis cannot be entirely dismissed. Additionally, the occurrence of null alleles in some microsatellite loci may have hidden introgressed alleles. Species diverged in every calling song character analysed (frequency content and the gross- and fine-temporal patterns) and no intermediate songs (expected in hybrid individuals) were found in this study. Therefore, from this study, isolating barriers seem to be generally efficient in preventing interbreeding between these species.

ii) Sympatric and allopatric populations of C. barbara and C. orni did not differ genetically or acoustically

No genetic or acoustic differences were found between sympatric and allopatric populations in both species, similar to results from morphological characters (Ribeiro, 1998). An exception was the divergence in inter-echeme interval in one sympatric population, Sousel. However, this pattern was not observed in other areas and further investigation is needed to confirm and elucidate the reasons for it. The seasonal displacement of the calling activity of both species when in sympatric areas compared to allopatric areas may be the result of the acoustic interference between species; particularly, it could explain the earlier disappearance of *C. orni* in every sympatric area relative to allopatric situations. It is likely that the *C. orni* song is masked, at least partially, by *C. barbara*'s continuous song. Additionally, the later emergence of *C. barbara* in sympatric relative to allopatric areas may be explained by competition between nymphs underground, but these hypotheses need further testing.

The classification of which areas were sympatric and which were allopatric was made based on field observations and represent the current state of those populations. It is not possible from this study to infer whether these sympatric areas are recent or ancient and if the

allopatric populations were always allopatric. It is neither possible to infer the probable scenario for the origin of both species in the Iberian Peninsula. Mitochondrial DNA data are required but are still inconclusive (G. Pinto-Juma, personal communication).

iii) Genetic and acoustic differentiation between C. barbara subspecies, C. barbara lusitanica (Iberian Peninsula) and C. barbara barbara (northwest Africa), was higher than within subspecies

The subspecific division in *C. barbara* is supported by microsatellites (allele frequency data) and acoustic differentiation (but only in peak frequency). The sea barrier between the Iberian Peninsula and Northwestern Africa is likely the cause of this divergence, even if occasional dispersal is conceivable in these winged animals.

iv) Genetic and acoustic differentiation was weak between Iberian and Southern French populations of C. orni; it was strong between Western European and Greek populations; high levels of genetic differentiation were found among Aegean islands and between these and the Greek continental population

Both microsatellite and calling song variation indicate that the Pyrenees do not constitute a strong barrier to the dispersal of these cicadas. In contrast, Greek and Western European populations are highly differentiated at both levels, corroborating morphological results (Ribeiro, 1998), as well as mitochondrial DNA results (G. Pinto-Juma, personal communication). The hypothesis of these being different subspecies or even species may be formulated based on all these sources of information. However, more extensive sampling in areas between these two regions is needed to understand the geographic patterns of genetic and acoustic variation of intermediate populations.

The Aegean Sea is probably a major barrier to dispersal for these cicadas, as denoted in the low levels of gene flow among Aegean islands.

v) A few populations showed divergent calling song characteristics

In *C. barbara*, the Ceuta population had significantly lower peak frequency than the other populations, most probably caused by the larger body sizes in this population. The Alvor population had lower syllable rate than the other populations but temperature data was missing. *C. orni* from Sousel (Iberian Peninsula) had longer inter-echeme intervals than the other populations in the Peninsula (not explained by temperature differences) and Kosmas (Greece) had longer echemes and shorter intervals than the remaining Greek populations

(no temperature data was available). These differences may be due to isolation in divergent environmental conditions (in Ceuta and in Kosmas) and due to the presence of another species (*C. barbara*, in Sousel). These cases deserve further investigation.

vi) In the Iberian Peninsula both C. barbara and C. orni showed similar high levels of microsatellite variability and low levels of among-population differentiation; within the Iberian Peninsula, genetic differentiation among populations increased with geographical distance in C. orni but not in C. barbara

Low values of differentiation among populations may indicate high levels of current gene flow and/or a recent divergence. However, caution should be taken when analysing low values of F_{ST} calculated from microsatellite data since such patterns may be influenced by the high variability within populations. The lack of a pattern of isolation by distance in *C. barbara* in the Iberian Peninsula may indicate that this species: i) is a recent immigrant to the Iberian Peninsula and is not in demographic equilibrium; ii) it is either sedentary or highly mobile; iii) may be subject to genetic drift due to small population size; iv) may have not been sampled appropriately to detect this pattern. It was not possible to test among these hypotheses in the present study.

vii) No genetic or acoustic differences between years of emergence were detected in either species

Differences in the calling song properties among populations were generally higher than differences among years within the same population. No genetic structure was found among years in the same population. Thus, no isolation between years of emergence was detected, indicating that no fixed period of nymphal development exists in these species, which corroborates the observations in captivity by M. Boulard (personal communication).

viii) The different calling song properties showed different levels of variation

Peak frequency, being functionally constrained, showed low variability within and among individuals. Differences found between geographic regions in both species are probably explained by differences in body size. Nevertheless, the hypothesis of direct selection on this trait cannot be dismissed through the present investigations.

Syllable rate showed low variability within and among individuals, similarly to the pattern seen in many acoustic species where this fine-temporal character is subjected to stabilizing selection through female choice (Gerhardt, 1994).

Gross-temporal characteristics in *C. orni*, echeme duration and inter-echeme interval, had high variability within and among individuals and also among populations. Echeme duration was the only variable that did not differ significantly between geographic regions, being a possible candidate for a character involved in mate recognition, since, according to Paterson (1985), this type of character should be maintained relatively constant by stabilizing selection across the distribution range of the species. However, the variability level of a trait may depend on a wide variety of influencing factors: *e.g.*, the plasticity of the trait, the different kinds of selection acting directly or indirectly on the trait, or even environmental factors (such as temperature).

4.1. Final remarks

In the present study, the presence of a high frequency of null alleles in many of the loci did not allow robust conclusions from the genetic analyses. Nevertheless, the methodology adopted here showed that the patterns of genetic differentiation observed, as estimated by F_{ST}, were generally similar among datasets including or excluding these loci (as well as datasets corrected for the presence of null alleles). They were also generally in agreement with mitochondrial DNA results (G. Pinto-Juma, personal communication).

The acoustic analyses allowed the description of the spatial and temporal variation of song characters, as well as the comparison between allopatric and sympatric populations. Some of these characters are important in species recognition, at least by males, as shown by Fonseca & Revez (2002a) and Simões & Quartau (2006). However, the female responses to each of these properties of the calling song have not been studied yet. Therefore, it is not yet possible to classify them as important in the mate recognition process. In the cited experiments, syllable rate was not tested and, taking into account the low levels of variability found for this variable in the present study, it may be a potentially important character in mate recognition and, thus, future researches should include it.

The genus *Cicada* constitutes an interesting case study to analyse patterns of speciation in the Mediterranean region. It is a complex of sibling species presenting a variety of simply patterned calling songs (from continuous to discontinuous songs, these last ones presenting echemes and inter-echeme intervals of various durations depending on the species), and with a wide variety of geographic distributions (Quartau & Simões, 2006). An indication of the patterns of genetic structure have emerged with the present microsatellite work and also with the mitochondrial studies (G. Pinto-Juma, personal communication), and a fuller description

of the biogeographic patterns in the eastern part of the Mediterranean area is being carried out (P. Simões, personal communication), as this is a particularly species-rich area for this genus. Other unexplored regions could be similarly interesting, such as the Middle East and other North African regions. A study of populations of *C. orni* in the Italian Peninsula and other central Europe regions is also essential in order to understand the morphological, genetic and acoustic variation throughout the range of distribution of the species.

There is only one other species pair of this genus known to occur in sympatry, *C. lodosi* Boulard and *C. mordoganensis* in Turkey (Quartau & Simões, 2006). The first species is a continuous song producer like *C. barbara* and the second has a calling song very similar in temporal and frequency pattern to that of *C. orni*, but contrary to the pair *C. barbara*/*C. orni*, they both have very similar peak frequencies in the calling song (around 5 kHz) (Quartau & Simões, *op.cit.*). Since a continuous song with a same frequency would completely mask a discontinuous song, it would be interesting to test the existence of a reproductive character displacement in such sympatric areas in Turkey.

Further genetic studies on species of this genus would benefit from using additional markers, since the microsatellites isolated in this study presented several problems due, probably, to high rate of mutation in the microsatellite flanking regions, which supposedly should be conserved. A potentially good alternative in the future would be to develop SNPs (single nucleotide polymorphisms), when easier, faster and cheaper methods of development arise (Schlötterer, 2004).

Much work is still to be done in morphological variation and ecological requirements of these cicadas, which are essential to understand the selective pressures acting on them. A study of the micro-habitat occupation (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) would be important to test ecological adaptations. Competition between nymphs for food (at the level of the plant roots) is also a possible interaction that would be interesting to test, despite the considerable practical difficulties in implementing such study.

As stated above, female responses and preferences to the acoustic traits are essential to find out which traits are important in mate recognition and choice. Setting up female choice experiments with these cicadas is a not an easy process since they are difficult to catch in the wild and they are short-lived in captivity, but might be feasible with the adequate logistic effort. This would allow testing the efficiency of the premating isolating barriers between the

different species and subspecies of the genus. Also, it would allow testing if some of these traits are under sexual selection.

Chapter 5 – References

De tanto cantar, as cigarras rebentam pelas costas

Crença popular (Alentejo) baseada nas exúvias que se encontram no campo (exosqueleto resultante da muda de ninfa para adulto)

5. References

Allendorf FW & Luikart G. 2007. *Conservation and the Genetics of Populations*. Blackwell Publishing, Oxford.

Angers B & Bernatchez L. 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution*, 14: 230–238.

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.

Arnold ML. 1997. *Natural Hybridization and Evolution*. Oxford University Press, New York.

Astanei J, Gosling E, Wilson J & Powell E. 2005. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Molecular Ecology*, 14: 1655–1666.

Augustinos AA, Mamuris Z, Stratikopoulos EE, D'Amelio S, Zachatopoulou A & Mathiopoulos KD. 2005. Microsatellite analysis of olive fly populations in the Mediterranean indicates a westward expansion of the species. *Genetica*, 125: 231–241.

Barker JSF. 2005. Population structure and host–plant specialization in two *Scaptodrosophila* flower-breeding species. *Heredity*, 94: 129–138.

Barton NH. 2001. Speciation. *Trends in Ecology and Evolution*, 16: 325.

Beaumont MA. 2003. Conservation Genetics. *In:* Balding DJ, Bishop M & Cannings C (eds). *Handbook of Statistical Genetics*. Second edition, Volume 2, John Wiley & Sons Ltd, West Sussex, Chapter 25, pp. 751–792.

Beaumont MA. 2004. Recent developments in genetic data analysis: what can they tell us about human demographic history? *Heredity*, 92: 365–379.

Beaumont MA & Bruford MW. 1999. Microsatellites in conservation genetics. *In*: Goldstein DB & Schlotterer C (eds). *Microsatellites. Evolution and Applications*. Oxford University Press, Oxford, Chapter 13, pp. 165–182.

Beaumont MA & Rannala B. 2004. The Bayesian revolution in genetics. *Nature Reviews Genetics*, 5: 251–261.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK & Bruford MW. 2001. Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology*, 10: 319–336.

Bee MA. 2004. Within–individual variation in bullfrog vocalizations: implications for a vocally mediated social recognition system. *Journal of the Acoustical Society of America*, 116: 3770–3781.

Beeman, K. 1998 Digital signal analysis, editing, and synthesis. *In:* Hopp SL, Owren MJ & Evans CS. (eds). *Animal Acoustic Communication. Sound Analysis and Research Methods.* Springer-Verlag Berlin Heidelberg, Chapter 3, pp. 59–102.

Beerli P, Hotz H & Uzzell T. 1996. Geologically dated sea barriers calibrate a protein clock for Aegean water frogs. *Evolution*, 50: 1676–1687.

Belkhir K, Borsa P, Chikhi L, Raufaste N & Bonhomme F. 1996–2004. *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bennet-Clark HC. 1997. Tymbal mechanics and the control of song frequency in the cicada *Cyclochila australasiae*. *Journal of Experimental Biology*, 200: 1681–1694.

Bennet-Clark HC. 1998a. How cicadas make their noise. *Scientific American*, 278: 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.

Bennet-Clark HC. 1999. Resonators in insect sound production: How insects produce loud pure–tone songs. *Journal of Experimental Biology*, 202: 3347–3357.

Bennet-Clark HC & Young D. 1992. A model of the mechanism of sound production in cicadas. *Journal of Experimental Biology*, 173: 123–153.

Bennet-Clark H.C. & Young D. 1994. The scaling of song frequency in cicadas. *Journal of Experimental Biology*, 191: 291–294.

Benzécri JP. 1973. *L'Analyse des Données, Tome 2: L' Analyse des Correspondances*. Dunod, Paris.

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. 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,* 7/8: 1–72.

Boulard M. 2006. Acoustic signals, diversity and behaviour of cicadas (Cicadidae, Hemiptera). *In*: Drosopoulos S & Claridge MF (eds). *Insect sounds and communication. Physiology, Behaviour, Ecology and Evolution.* Taylor & Francis Group, Boca Raton, Chapter 25, pp: 331–349.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR & Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 368: 455–457.

Broderick D, Idaghdour Y, Korrida A & Hellmich J. 2003. Gene flow in great bustard populations across the Strait of Gibraltar as elucidated from excremental PCR and mtDNA sequencing. *Conservation Genetics*, 4: 793–800.

Brookfield JFY. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, 5: 453–455.

Broughton WB. 1976. Proposal for a new term 'echeme' to replace 'chirp' in animal acoustics. *Physiological Entomology*, 1: 103–106.

Bruford MW & Wayne RK. 1993. Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development*, 3: 939–943.

Bruford MW, Cheesman DJ, Coote T, Green AA, Haines SA, O'Ryan C & Williams TR. 1996. Microsatellites and their application to conservation genetics. *In*: Smith TB & Wayne RK (eds). *Molecular Genetic Approaches in Conservation*. Oxford University Press, New York, Chapter 17, pp. 278–297.

Buckley TR, Simon C, Shimodaira H & Chambers GK. 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (*Maoricicada*) using multiple–comparison tests of tree topology. *Molecular Biology and Evolution*, 18: 223–234.

Bues R, Eizaguirre M, Toubon JF & Albages R. 1996. Population genetic structure and ecological differences among nine populations of *Sesamia nonagrioides* Lefebre (Lepidoptera: Noctuidae) from the western basin. *Canadian Entomologist*, 128: 849–858.

Bürger R, Schneider KA & Willensdorfer M. 2006. The conditions for speciation through intraspecific competition. *Evolution*, 60: 2185–2206.

Butlin RK. 1995. Reinforcement: an idea evolving. *Trends in Ecology and Evolution*, 10: 432–434.

Butlin RK, Hewitt GM & Webb SF. 1985. Sexual selection for intermediate optimum in *Chorthippus brunneus* (Orthoptera: Acrididae). *Animal Behaviour*, 33: 1281–1292.

Callen DF, Thompson AD, Shen Y, Philips HA, Richards RI, Mulley JC & Sutherland GR. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, 52: 922–927.

Castella V, Ruedi M, Excoffier L, Ibáñez C, Arlettaz R & Hausser J. 2000. Is the Gibraltar Strait a barrier to gene flow for the bat *Myotis myotis* (Chiroptera: Vespertilionidae)? *Molecular Ecology*, 9: 1761–1772.

Chakraborty R, De Andrade M, Daiger SP & Budowle B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics*, 56: 45–57.

Chambers GK & MacAvoy ES. 1999. Molecular genetic analysis of hybridisation. *In*: *Science for Conservation*, 105. Department of Conservation, Wellington, New Zealand.

Chambers GK & MacAvoy ES. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B*, 126: 455–476.

Chikhi L & Bruford MW. 2005. Mammalian population genetics and genomics. *In*: Ruvinsky A & Graves MJ (eds). *Mammalian genomics*. CABI Publishing, Wallingford, pp. 539–584.

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

Claridge MF, Wilson MR & Singhrao JS. 1979. The songs and calling sites of two European cicadas. *Ecological Entomology*, 4: 225–229.

Claridge MF, Dawah HA & Wilson MR (eds). 1997a. *Species, the Units of Biodiversity*. Chapman & Hall, London.

Claridge MF, Dawah HA & Wilson MR. 1997b. Practical approaches to species concepts for living organisms. *In*: Claridge MF, Dawah HA & Wilson MR (eds). *Species, the Units of Biodiversity*. Chapman & Hall, London, Chapter 1, pp. 1–15.

Cocroft RB & Pogue M. 1996. Social behavior and communication in the Neotropical cicada *Fidicina mannifera* (Fabricius) (Homoptera: Cicadidae). *Journal of the Kansas Entomological Society*, 69: 85–97.

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

Cooley JR & Marshall DC. 2004. Thresholds or comparisons: mate choice criteria and sexual selection in a periodical cicada, *Magicicada septendecim* (Hemiptera: Cicadidae). *Behaviour*, 141: 647–673.

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.

Cooley JR, Marshall DC, Hill KBR & Simon C. 2006. Reconstructing asymmetrical reproductive character displacement in a periodical cicada contact zone. *Journal of Evolutionary Biology*, 19: 855–868.

Corander J, Waldmann P & Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation between populations. *Genetics*, 163: 367–374.

Coyne JA & Orr HA. 1989. Patterns of speciation in *Drosophila*. *Evolution*, 43: 362–381.

Coyne JA & Orr HA. 2004. *Speciation*. Sinauer Associates, Inc. Sunderland.

Cracraft J. 1997. Species concepts in systematics and conservation biology – an ornithological viewpoint *In*: Claridge MF, Dawah HA & Wilson MR (eds). *Species, the Units of Biodiversity*. Chapman & Hall, London, Chapter 16, pp. 325–339.

Dakin EE & Avise JC. 2004. Microsatellite null alleles in parentage analysis. *Heredity*, 93: 504–509.

Darwin C. [1859] 1964. *On the Origin of Species. A Facsimile of the First Edition with an Introduction by Ernst Mayr.* Harvard Univeristy Press, Cambridge, Massachusetts.

Daws AG, Hennig RM & Young D. 1997. Phonotaxis in the cicadas *Cystosoma saundersii* and *Cyclochila australasiae*. *Bioacoustics*, 7: 173–188.

Dawson KJ & Belkhir K. 2001. A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, 78: 59–77.

Dawson DA, Chittock J, Jehle R, Whitlock A, Nogueira D, Pellatt J, Birkhead T & Burke T. 2005. Identification of 13 polymorphic microsatellite loci in the zebra finch, *Taeniopygia guttata* (Passeridae, Aves). *Molecular Ecology Notes*, 5: 298–301.

de Boer AJ & Duffels JP. 1996. Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 124: 153–177.

Dieckmann U, Doebeli M, Metz JAJ & Tautz D (eds). 2004. *Adaptive Speciation*. Cambridge University Press, Cambridge.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M & Freimer NB. 1994. Mutational processes of simple–sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 3166–3170.

Dobzhansky T. 1951. *Genetics and the Origin of Species*. Third edition. Columbia University Press, New York.

Doolan JM & MacNally R C. 1981. Spatial dynamics and breeding ecology in the cicada *Cystosoma saundersii*: the interaction between distributions of resources and intraspecific behaviour. *Journal of Animal Ecology*, 50: 925–940.

Doolan JM & Young D. 1989. Relative importance of song parameters during flight phonotaxis and courtship in the bladder cicada *Cystosoma saundersii*. *Journal of Experimental Biology*, 141:113–31.

Dybas HS & Lloyd M. 1962. Isolation by habitat in two synchronized species of periodical cicadas (Homoptera: Cicadidae: *Magicicada*). *Ecology*, 43: 444–459.

Dytham C. 2003. *Choosing and using statistics, a biologist's guide*. Second edition. Blackwell, Science, Oxford.

Eliopoulos E. 2006. Sound and Techniques in Sound Analysis. *In*: Drosopoulos S & Claridge MF (eds). *Insect sounds and communication. Physiology, Behaviour, Ecology and Evolution*. Taylor & Francis Group, Boca Raton, Chapter 2, pp. 11–33.

Ellegren H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, 16: 551–558.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5: 435–445.

Ellegren H, Moore S, Robinson N, Byrne K, Ward W & Sheldon BC. 1997. Microsatellite evolution – a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution*, 14: 854–860.

Endsley MA, Baker MD & Krafsur ES. 2002. Microsatellite loci in the house fly, *Musca domestica* L. (Diptera: Muscidae). *Molecular Ecology Notes*, 2: 72–74.

Estoup A, Garnery L, Solignac M & Cornuet J-M. 1995. Microsatellite variation in Honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, 140: 679–695.

Estoup A, Solignac M, Cornuet J-M, Goudet J & Scholl A. 1996. Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. *Molecular Ecology*, 5: 19–31.

Estoup A, Jarne P & Cornuet J-M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11: 1591–1604.

Evanno G, Regnaut S & Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular *Ecology*, 14: 2611–2620.

Excoffier L. 2003. Analysis of population subdivision. *In:* Balding DJ, Bishop M & Cannings C (eds). *Handbook of Statistical Genetics*. Second edition, Volume 2. John Wiley & Sons Ltd, West Sussex, Chapter 24, pp. 713–750.

Excoffier L, Laval G & Schneider S. 2005. Arlequin version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1: 47–50.

Falush D, Stephens M & Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164: 1567–1587.

Feldman MW, Bergman A, Pollock DD & Goldstein DB. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics*, 145: 207–216.

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

Fonseca PJ. 1994. *Acoustic communication in cicadas (Homoptera; Cicadoidea): sound production and sound reception*. PhD thesis, Faculty of Sciences, University of Lisbon.

Fonseca PJ. 1996. Sound production in cicadas: timbal muscle activity during calling song and protest song. *Bioacoustics*, 7: 13–31.

Fonseca PJ & Bennet-Clark HC. 1998. Asymmetry of tymbal action and structure in a cicada: a possible role in the production of complex songs. *Journal of Experimental Biology*, 201: 717–730.

Fonseca PJ & Hennig RM. 1996. Phasic action of the tensor muscle modulates the calling song in cicadas. *Journal of Experimental Biology*, 199: 1535–1544.

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

Fonseca PJ & Revez MA. 2002b. Temperature dependence of cicada songs (Homoptera, Cicadoidea). *Journal of Comparative Physiology A*, 187: 971–976.

Fonseca PJ, Münch D & Hennig RM. 2000. How cicadas interpret acoustic signals. *Nature*, 405: 297–298.

Fox CW & Wolf JB (eds). 2006. *Evolutionary Genetics. Concepts and Case Studies*. Oxford University Press, New York.

Franck P, Garnery L, Solignac M & Cornuet J-M. 1998. The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, 52: 1119–1134.

Franck P, Garney L, Celebrano G, Solignac M & Cornuet J-M. 2000. Hybrid origins of honeybees from Italy (*Apis mellifera ligustica*) and Sicily (*A. m. sicula*). *Molecular Ecology,* 9: 907–921.

Friedl TWP. 2006. Individual male calling pattern and male mating success in the European treefrog (*Hyla arborea*): Is there evidence for directional or stabilizing selection on male calling behaviour? *Ethology*, 112: 116–126.

Gaggiotti OE, Lange O, Rassmann K & Gliddons C. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, 8: 1513–1520.

Garza JC, Slatkin M & Freimer NB. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution*, 12: 594–603.

Gee JM. 2005. No species barrier by call in an avian hybrid zone between California and Gambel's quail (*Callipepla californica* and *C. gambelii*). *Biological Journal of the Linnean Society*, 86: 253–264.

Gerhardt HC. 1991. Female choice in treefrogs: static and dynamic acoustic criteria. *Animal Behaviour*, 42: 615–635.

Gerhardt HC. 1994. Selective responsiveness to long–range acoustic signals in insects and anurans. *American Zoologist*, 34: 706–714.

Gerhardt HC. 1998. Acoustic signals of animals: recording, field measurements, analysis and description. *In*: Hopp SL, Owren MJ & Evans CS (eds). *Animal Acoustic Communication. Sound Analysis and Research Methods*. Springer–Verlag Berlin, Heidelberg, Chapter 1, pp: 1–25.

Gerhardt HC & Huber F. 2002. *Acoustic Communication in Insects and Anurans. Common problems and Diverse Solutions*. The University of Cicago Press, Chicago.

Giralda AA, Rodríguez JMP & Pintor FG. 1998. La cigarra del olivo o magrebí. *Boletin Fitosanitario de Avisos e informaciones*, n. 1.

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.

Goldstein D & Schlötterer C. 1999. *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford.

Goldstein DB, Linares AR, Cavalli-Sforza LL & Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139: 463–471.

Goodman SJ, Barton NH, Swanson G, Abernethy K & Pemberton J. 1999. Introgression through rare hybridisation: a genetic study of a hybrid zone between red and sike deer (Genus *Cervus*) in Argyll, Scotland. *Genetics*, 152: 355–371.

Gottelli D, Sillero-Zubiri C, Applebaum GD, Roy MS, Girman DJ, Garcia-Moreno J, Ostrander EA & Wayne RK. 1994. Molecular genetics of the most endangered canid: the Ethiopian wolf *Canis simensis*. *Molecular Ecology*, 3: 301–312.

Guillaumet A, Pons J-M, Godelle B & Crochet P-A. 2006. History of the Crested Lark in the Mediterranean region as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, 39: 645–656.

Haberl M & Tautz D. 1999. Comparative allele sizing can produce inaccurate allele size differences for microsatellites. *Molecular Ecology*, 8: 1347–1349.

Hammond RL, Saccheri IJ, Ciofi C, Coote T, Funk SM, McMillan O, Bayes MK, Taylor E & Bruford MW 1998. Isolation of microsatellite markers in animals. *In:* Karp A, Isaac PG & Ingram DS (eds). *Molecular Tools for Screening Biodiversity*. Chapman & Hall, London, pp. 279–287.

Hancock JM. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. *In*: Goldstein D & Schlötterer C (eds). *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford, pp. 1–9.

Hartl DL & Clark AG. 1997. *Principles of Population Genetics*. Third edition. Sinauer Associates.

Hedgecock D, Li G, Hubert S, Bucklin K & Ribes V. 2004. Widespread null alleles and poor cross–species amplification of microsatellite DNA loci cloned from the Pacific oysters, *Crassostrea gigas*. *Journal of Shellfish Research*, 23: 379–385.

Hedrick PW. 2000. *Genetics of Populations*. Second edition. Jones and Bartlett Publishers, Massachussetts.

Heliövaara K, Väisänen R & Simon C. 1994. Evolutionary ecology of periodical insects. *Trends in Ecology and Evolution*, 9: 475–480.

Henrique LL. 2002. *Acústica musical*. Fundação Calouste Gulbenkian.

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

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology*, 10: 537–549.

Hoare K, Hughes RN & Gliddon CJ. 1998. Polymorphic microsatellite markers isolated from the Bryozoan *Celleporella hyalina* (L.). *Molecular Ecology*, 7: 355–356.

Hoelzel AR & Green A. 1998. PCR protocols and population analysis by direct sequencing and PCR-based DNA fingerprinting. *In*: Hoelzel AR (ed). *Molecular Genetic Analysis of Populations. A Practical Approach*. Second edition, Oxford University Press, Chapter 6, pp. 201–235.

Hoffman JI & Amos W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, 14: 599–612.

Hopp SL & Morton ES. 1996. Sound playback studies. *In:* Hopp SL, Owren MJ & Evans CS (eds). *Animal Acoustic Communication. Sound Analysis and Research Methods.* Springer-Verlag Berlin, Heidelberg, Chapter 10, pp. 323–352.

Hudson RR. 2003. Linkage disequilibrium and recombination. *In:* Balding DJ, Bishop M & Cannings C. (eds). *Handbook of Statistical Genetics*. Second edition, Volume 2, John Wiley & Sons Ltd, West Sussex, Chapter 22, pp. 662–680.

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.

Hutter CM, Schug MD & Aquadro CF. 1998. Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. *Molecular Biology and Evolution*, 15: 1620–1636.

Ishibashi Y, Saitoh T, Abe S & Yoshida MC. 1996. Null microsatellite alleles due to nucleotide sequence variation in the grey–sided vole *Clethrionomys rufocanus*. *Molecular Ecology*, 5: 589–590.

Jarne P & Lagoda PJL. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11: 424–429.

Joermann G & Schneider H. 1987. The songs of four species of cicada in Yugoslavia (Homoptera: Cicadidae). *Zoologischer Anzeiger*, 219: 283–296.

Johnson NA. 2006. The evolution of reproductive isolating barriers. *In*: Fox CW & Wolf JB (eds). *Evolutionary Genetics. Concepts and Case Studies*. Oxford University Press, New York, Chapter 24, pp. 374–386.

Jones AG, Stockwell CA, Walker D & Avise JC. 1998. The molecular basis of a microsatellite null allele from the White Sands Pupfish. *The Journal of Heredity*, 89: 339–342.

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

Karhu A, Vogl C, Moran GF, Bell JC & Savolainen O. 2006. Analysis of microsatellite variation in *Pinus radiata* reveals effects of genetic drift but no recent bottlenecks. *Journal of Evolutionary Biology*, 19: 167–175.

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: 612–623.

Keller LF, Jeffery KJ, Arcese P, Beaumont MA, Hochachka WM, Smith JNM & Bruford MW. 2001. Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society of London B*, 268: 1387–1394.

Keyghobadi N, Roland J & Strobeck C. 1999. Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology*, 8: 1481–1495.

Keyghobadi N, Roland J & Strobeck C. 2005. Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology*, 14: 1897–1909.

Kopuchian C, Lijtmaer DA, Tubaro PL & Handford P. 2004. Temporal stability and change in a microgeographical pattern of song variation in the rufous-collared sparrow. *Animal Behaviou*r, 68: 551–559.

Lehman T, Hawley WA & Collins FH. 1996. An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics*, 144: 1155–1163.

Lei Z, Chou I & Li L. 1994. The sound characteristics of cicada songs and their significance in classification (Homoptera: Cicadoidea). *Entomotaxonomia*, 16: 51–59. (Abstract)

Li Y-C, Korol AB, Fahima T, Beiles A & Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11: 2453–2465.

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.

Llewellyn KS, Loxdale HD, Harrington R, Brookes CP, Clark SJ & Sunnucks P. 2003. Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Molecular Ecology*, 12: 21–34.

Loftus-Hills JJ & Littlejohn MJ. 1992. Reinforcement and reproductive character displacement in *Gastrophryne carolinensis* and *G. olivacea* (Anura: Microhylidae): a re-examination. *Evolution*, 46: 896–906.

Luikart G & England PR. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution*, 14: 253–256.

Luikart G, Allendorf FW, Cornuet J-M & Sherwin WB. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, 89: 238–247.

Lunt DH, Hutchinson WF & Carvalho GR. 1999. An efficient method for PCR-based isolation of microsatellite arrays (PIMA). *Molecular Ecology*, 8: 891–894.

Macaubas C, Jin L, Hallmayer J, Kimura A & Mignot E. 1997. The complex mutation pattern of a microsatellite. *Genome Research*, 7: 635–641.

Marcet PL, Lehman T, Groner G, Gurtler RE, Kitron U & Dotson EM. 2006. Identification and characterization of microsatellite markers in the Chagas disease vector *Triatoma infestans* (Heteroptera: Reduviidae). *Infection, Genetics and Evolution*, 6: 32–37.

Marshall DC & Cooley JR. 2000. Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year *Magicicada neotredecim*. *Evolution*, 54: 1313–1325.

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

Massonnet B & Weisser WW. 2004. Patterns of genetic differentiation between populations of the specialized herbivore *Macrosiphoniella tanacetaria* (Homoptera, Aphididae). *Heredity*, 93: 577–584.

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

Mayr E. 1970. *Populations, Species and Evolution. An abridgment of Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

McGavin GC. 2001. *Essential entomology. An Order-by-Order introduction*. Oxford University Press, Oxford.

McGoldrick DJ, Hedgecock D, English LJ, Baoprasertkul P & Ward RD. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): Selection and null alleles. *Journal of Shellfish Research*, 19: 779–788.

McGregor PK. 1991. The singer and the song: on the receiving end of bird song. *Biological Reviews,* 66: 57–81.

Michalak P & Noor MA. 2006. Genetics of reproductive isolation and species differences in model organisms. *In*: Fox CW & Wolf JB (eds). *Evolutionary Genetics. Concepts and Case Studies*. Oxford University Press, New York, Chapter 25, pp. 387–398.

Michel AP, Guelbeogo WM, Grushko O, Schemerhorn BJ, Kern M, Willard MB, Sagnon NF, Costantini C & Besansky NJ. 2005. Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*. *Insect Molecular Biology*, 14: 375–387.

Minitab Inc. 2004. MINITAB Release 14, Statistical Software.

Morin PA & Woodruf DS.1996. Noninvasive genotyping for vertebrate conservation. *In*: Smith TB & Wayne RK (eds). *Molecular Genetic Approaches in Conservation*. Oxford University Press, New York, Chapter 18, pp. 298–313.

Moulds MS. 1990. *Australian Cicadas*. New South Wales University Press, Kensington.

Moyle LC. 2006. Correlates of genetic differentiation and isolation by distance in 17 congeneric *Silene* species. *Molecular Ecology*, 15: 1067–1081.

Nast J. 1972. *Palaearctic Auchenorrhyncha (Homoptera), an annotated checklist.* Polish Scientific Publications, Warszawa.

Neigel JE. 1996. Estimation of effective population size and migration parameters from genetic data. *In*: Smith TB & Wayne RK (eds). *Molecular Genetic Approaches in Conservation*. Oxford University Press, New York, Chapter 20, pp. 329–346.

Nève G & Meglécz E. 2000. Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution*, 15: 376–377.

Newman RA & Squire T. 2001. Microsatellite variation and fine-scale population structure in the wood frog (*Rana sylvatica*). *Molecular Ecology,* 10: 1087–1100.

Noor MAF. 1999. Reinforcement and other consequences of sympatry. *Heredity*, 83: 503–508.

Onyabe DY & Conn JE. 2001. Population genetic structure of the malaria mosquito *Anopheles arabiensis* across Nigeria suggests range expansion. *Molecular Ecology*, 10: 2577–2591.

Overall ADJ & Nichols RA. 2001. A method for distinguishing consanguinity and population substructure using multilocus genotype data. *Molecular Biology and Evolution*, 18: 2048–2056.

Paetkau D & Strobeck C. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, 4: 519–520.

Paetkau D, Waits LP, Clarkson PL, Craighead L & Strobeck C. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, 147: 1943–1957.

Panhuis TM, Butlin R, Zuk M & Tregenza T. 2001. Sexual selection and speciation. *Trends in Ecology and Evolution*, 16: 364–371.

Parmakelis A, Stathi I, Chatzaki M, Simaiakis S, Spanos L, Louis C & Mylonas M. 2006. Evolution of *Mesobuthus gibbosus* (Brullé, 1832) (Scorpiones: Buthidae) in the northeastern Mediterranean region. *Molecular Ecology*, 15: 2883–2894.

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

Patterson IJ, Massei G & Genov P. 1997. The density of cicadas *Cicada orni* in mediterranean coastal habitats. *Italian Journal of Zoology*, 64: 141–146.

Pemberton JM, Slate J, Bancroft DR & Barrett JA. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, 4: 249–252.

Peterson MA & Denno RF. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *The American Naturalist*, 152: 428–446.

Pope LC, Sharp A & Moritz C. 1996. Population structure of the yellow-footed rock-wallaby *Petrogale xanthopus* (Gray, 1854) inferred from mtDNA sequences and microsatellite loci. *Molecular Ecology*, 5: 629–640.

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.

Primmer CR, Raudsepp T, Chowdhary BP, Moller AP & Ellegren H. 1997. Low frequency of microsatellites in the avian genome. *Genome Research*, 7: 471–482.

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

Pritchard JK & Wen W. 2003. *Documentation for structure software: Version 2*. URL http://pritch.bsd.uchicago.edu.

Pritchard JK, Stephens M & Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.

Puissant S & Sueur J. 2001. Contribution à l'étude des Cigales de Corse (Hemiptera, Cicadidae). *Bulletin de la Société entomologique de France*, 106: 429–436.

Quartau JA. 1988. A numerical taxonomic analysis of interspecific morphological differences in two closely related species of *Cicada* (Homoptera, Cicadidae) 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 & Fonseca PJ. 1988. An annotated check-list of the species of cicadas known to occur in Portugal (Homoptera: Cicadoidea). *In*: *Proceedings of the 6th Auchenorrhyncha Meeting,* Turin, Italy, pp. 367–375.

Quartau JA & Rebelo MT. 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 JA & Simões PC. 2003. Bioacoustic and morphological differentiation in two allopatric species of the genus *Tibicina* Amyot (Hemiptera: Cicadoidea) in Portugal. *Deutsche entomologische Zeitshrift*, 50: 113–119.

Quartau JA & Simões PC. 2004. First description of the calling song produced by *Euryphara contentei* Boulard, 1982 (Insecta : Hemiptera, Cicadoidea) in Portugal. *Arquivos do Museu Bocage*, III (22): 563–572.

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

Quartau JA & Simões PC. 2006. Acoustic evolutionary divergence in cicadas: the species of *Cicada* L. in Southern Europe. *In*: Drosopoulos S & Claridge MF (eds). *Insect sounds and communication. Physiology, Behaviour, Ecology and Evolution.* Taylor & Francis Group, Boca Raton, Chapter 17, pp. 227–237.

Quartau JA, Coelho MM, Viegas-Crespo AM, Rebelo MT, 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,* Sidney, 17–21 February 1997, pp. 47–48. (abstract)

Quartau JA, MT Rebelo, PC Simões, TM Fernandes, MF Claridge, S Drosopoulos & 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, Ribeiro M, Simões PC & Crespo A. 2000a. 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, Seabra S & Sanborn A. 2000b. Effect of ambient air temperature on the calling song of *Cicada orni* Linnaeus, 1758 (Hemiptera: Cicadidae) in Portugal. *Acta zoologica cracoviensia*, 43: 193–198.

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.

Queney G, Ferrand N, Weiss S, Mougel F & Monnerot M. 2001. Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, 18: 2169–2178.

R Development Core Team. 2003. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org.

Ragge DR & Reynolds WJ. 1998. The songs of the grasshoppers and crickets of Western Europe. Harley Books, London.

Randi E & Lucchini V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conservation Genetics*, 3: 31–45.

Rannala B & Mountain JL. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, 94: 9197–9201.

Raymond M & Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86: 248–249.

Reynolds WJ. 1988. The use of insect sounds in taxonomy. *British Journal of Entomology and Natural History*, 1: 147–152.

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.

Rieseberg LH, Church SA & Morjan CL. 2003. Integration of populations and differentiation of species. *New Phytologist*, 161: 59–69.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, 145: 1219–1228.

Rousset F. 2004. *Genetic structure and selection in subdivided populations*. Levin SA & Horn HS (eds). Princeton University Press, Princeton.

Rousset F & Raymond M. 1995. Testing heterozygote excess and deficiency. *Genetics*, 140: 1413–1419.

Roy MS, Geffen E, Smith D, Ostrander EA & Wayne RK. 1994. Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, 11: 553–570.

Runciman D, Zann RA & Murray ND. 2005. Geographic and temporal variation of the male Zebra Finch distance call. *Ethology*, 111: 367–379.

Sætre G-P, Borge T, Lindell J, Moum T, Primmer CR, Sheldon BC, Haavie J, Johnsen A & Ellegren H. 2001. Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. *Molecular Ecology*, 10: 737–749.

Sanborn AF. 1997.Thermal biology of cicadas (Homoptera: Cicadoidea). *Entomology*, 1: 89–104.

Sanborn AF. 2006. Acoustic Signals and Temperature. *In*: Drosopoulos S & Claridge MF (eds). *Insect sounds and communication. Physiology, Behaviour, Ecology and Evolution*. Taylor & Francis Group, Boca Raton, Chapter 7, pp. 111–125.

Schedl W. 1973. Zur Verbreitung, bionomie und okolgie der Singzikaden (Homoptera: Auchenorrhyncha, Cicadidae) der Ostalpen und iher benachbarten Gebiete. *Ber. Net.-med. Ver. Innsbruck*, 60: 79–94.

Schedl W. 1999. Contribution to the singing cicadas of Israel and adjacent countries (Homoptera, Auchenorrhyncha: Cicadidade et Tibicinidae). *Linzer. biol. Beitr.*, 31: 823–837.

Schlötterer C. 1998. Microsatellites. *In*: Hoelzel AR (ed). *Molecular Genetic Analysis of Populations – a pratical approach*. Oxford University Press, Oxford, Chapter 7, pp. 237–261.

Schlötterer C. 2004. The evolution of molecular markers – just a matter of fashion? *Nature Review Genetics*, 5: 63–69.

Schlötterer C, Vogl C & Tautz D. 1997. Polymorphism and locus–specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics*, 146: 309–320.

Schug MD, Hutter CM, Noor MAF & Aquadro CF. 1998. Mutation and evolution of microsatellites in *Drosophila melanogaster*. *Genetica*, 102/103: 359–367.

Scribner KT & Pearce JM. 2000. Microsatellites: evolutionary and methodological background and empirical appilcations at individual, population and phylogenetic levels. *In*: Baker AJ (ed). *Molecular Methods in Ecology*. Blackwell Science, Oxford, Chapter 10, pp. 235–273.

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. *Deutsche Entomologische Zeitschrift*, 47: 143–145.

Seddon N. 2005. Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution*, 59: 200–215.

Shriver MD, Jin L, Chakraborty R & Boerwinkle E. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics*, 134: 983–993.

Simmons LW. 2004. Genotypic variation in calling song and female preferences of the field cricket *Teleogryllus oceanicus*. *Animal Behaviour*, 68: 313–322.

Simmons AM, Popper AN & Fay RR (eds). 2003. *Acoustic communication*. Springer-Verlag, New York.

Simões PC & Quartau JA. 2006. Selective responsiveness in males of *Cicada orni* to conspecific and allospecific calling songs (Hemiptera: Cicadidae). *Entomologia Generalis*, 29: 47–60.

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: 437–440.

Simon CM. 1979. Evolution of periodical cicadas: phylogenetic inferences based on allozymic data. *Systematic Zoology*, 28: 22–39.

Simon C, Tang J, Dalwadi S, Staley G, Deniega J & Unnasch TR. 2000. Genetic evidence for assortative mating between 13-year cicadas and sympatric "17-year cicadas with 13-year life cycles" provides support for allochronic speciation. *Evolution*, 54: 1326–1336.

Slatkin M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47: 264–279.

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

Smith SW. 1997. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing. URL http://www.DSPguide.com

Sokal RR & Rohlf FJ. 1981. *Biometry. The principles and practice of statistics in biological research*. WH Freeman and Company, New York.

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

SPSS Inc. 1999. SPSS for Windows, Release 10.0.1, Chicago.

StatSoft. 2001. STATISTICA version 6.0, Tulsa.

Stölting H, Moore TE & Lakes-Harlan R. 2004. Acoustic communication in *Okanagana rimosa* (Say) (Homoptera: Cicadidae). *Zoology*, 107: 243–257.

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

Sueur J. 2003. Indirect and direct acoustic aggression in cicadas: first observations in the Paleartic genus *Tibicina* Amyot (Hemiptera: Cicadomorpha: Cicadidae). *Journal of Natural History*, 37: 2931–2948.

Sueur J & Aubin T. 2002. Acoustic communication in the Palaeartic red cicada, *Tibicina haematodes*: chorus organisation, calling–song structure, and signal recognition. *Canadian Journal of Zoology*, 80: 126–136.

Sueur J & Aubin T. 2003a. Specificity of cicada calling songs in the genus *Tibicina* (Hemiptera: Cicadidae). *Systematic Entomology*, 28: 481–492.

Sueur J & Aubin T. 2003b. 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*). *Journal of Zoology, London,* 262: 217–224.

Sueur J & Puissant S. 2002. Spatial and ecological isolation in cicadas: first data from *Tibicina* (Hemiptera: Cicadoidea) in France. *European Journal of Entomology*, 99: 477–484.

Sueur J, Puissant S, Simões PC, Seabra S, Boulard M & Quartau JA. 2004. Cicadas from Portugal: revised list of species with eco-ethological data (Hemiptera: Cicadidae). *Insect Systematics & Evolution*, 35: 177–187.

Surridge AK, Ibrahim KM, Bell DJ, Webb NJ & Hewitt GM. 1999. Fine-scale genetic structuring in a natural population of European wild rabbits (*Oryctolagus cuniculus*). *Molecular Ecology*, 8: 299–307.

Tautz D & Schlötterer C. 1994. Simple sequences. *Current Opinion in Genetics and Development*, 4: 832–837.

Tregenza T & Bridle JR. 1997. The diversity of speciation. *Trends in Ecology and Evolution*, 12: 382–383.

Tregenza T, Pritchard VL & Butlin RK. 2000. The origins of premating reproductive isolation: testing hypotheses in the grasshopper *Chorthippus parallelus*. *Evolution*, 54: 1687–1698.

Turelli M, Barton NH & Coyne JA. 2001. Theory and speciation. *Trends in Ecology and Evolution*, 16: 330–343.

Valdes AM, Slatkin M & Freimer NB. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, 133: 737–749.

Van Oosterhout C, Hutchinson WF, Wills DPM & Shipley PF. 2003. *Micro-Checker User Guide*. The University of Hull.

Van Oosterhout C, Hutchinson WF, Wills DP & Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4: 535–538.

Van Treuren R. 1998. Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haematopus ostralegus*). *Molecular Ecology*, 7: 1413–1417.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution*, 16: 381–390.

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

Villet M. 1995. Intraspecific variability in SMRS signals: some causes and implications in acoustic signalling systems. *In*: Lambert DM & Spencer HG (eds). *Speciation and the recognition concept. Theory and Applications*. The John Hopkins University Press, Baltimore, pp. 422–439.

Villet MH, Barker NP & Lunt N. 2004. Mechamisms generating biological diversity in the genus *Platypleura* Amyot & Serville, 1843 (Hemiptera: Cicadidae) in southern Africa: implications of a preliminary molecular phylogeny. *South African Journal of Science*, 100: 589–594.

Walker TJ. 1998. Trilling field crickets in a zone of overlap (Orthoptera: Gryllidae: *Gryllus*). *Entomological Society of America*, 91: 175–184.

Wang Z, Niu T, Lunetta KL, Xu X & Fang Z. 2000. Observation of null alleles of microsatellite markers in a Chinese population. *GeneScreen*, 1: 41–45.

Webster MT, Smith NGC & Ellegren H. 2002. Microsatellite evolution inferred from human – chimpanzee genomic sequence alignments. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 8748–8753.

Weir BS & Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358–1370.

Wheeler QD & Meier R (eds). 2000. *Species Concepts and Phylogenetic Theory. A debate*. Columbia University Press. New York.

Wilczynski W & Ryan MJ. 1999. Geographic variation in animal communication systems. *In:* Foster SA & Endler JA (eds). *Geographic Variation in Behavior. Perspectives on Evolutionary Mechanisms*. Oxford University Press, Chapter 11, pp. 234–261.

William J, Ballard O & Kreitman M. 1995. Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution*, 10: 485–488.

Williams KS & Simon C. 1995. The ecology, behaviour, and evolution of periodical cicadas. *Annual Review of Entomology*, 40: 269–295.

Wolda H. 1993. Diel and seasonal patterns of mating calls in some Neotropical cicadas. Acoustic interference? *Proc Kon Ned Akad v Wetensch*, 96: 369–381.

Wright S. 1943. Isolation by distance. *Genetics*, 28: 114–138.

Wu C–I. 2001 The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14: 851–865.

Wu B, Liu N & Zhao H. 2006. PSMIX: an R package for population structure inference via maximum likelihood method. *BMC Bioinformatics*, 7: 317.

Xu H & Fu Y-X. 2004. Estimating effective population size or mutation rate with microsatellites. *Genetics*, 166: 555–563.

Young D. 1972. Neuromuscular mechanism of sound production in Australian cicadas. *Journal of Comparative Physiology*, 79: 343–362.

Young D. 1990. Do cicadas radiate sound through their ear-drums? *Journal of Experimental Biology*, 151: 41–56.

Young D & Bennet-Clark HC. 1995. The role of the tymbal in cicada sound production. *Journal of Experimental Biology*, 198: 1001–1019.

Young D & Josephson RK. 1983. Pure-tone songs in cicadas with special reference to the genus *Magicicada*. *Journal of Comparative Physiology*, 152: 197–207.

Zane L, Bargelloni L & Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11: 1–16.

Zuk M, Rotenberry JT & Simmons LW. 2001. Geographical variation in calling song of the field cricket *Teleogryllus oceanicus*: the importance of spatial scale. J*ournal of Evolutionary Biology*, 14: 731–741.

Appendices

Appendix I – Values of F_{IS} calculated for each locus, population and dataset. Significance levels: *p<0.05; **p<0.01; ***p<0.001.

	Locus Cib01				Locus Cib03				Locus Cib06				Locus Cib07			
Population	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**
CbAlc95	+0.141				+0.661***	+0.174*	-0.051	-0.161	+0.134				+0.671***	+0.131**	-0.023	-0.078
CbCas96	-0.109				+0.684***	+0.148*	-0.07	+0.229	+0.282	+0.146	-0.01	-0.01	+0.357			
CbCra95	-0.029				+0.660*	+0.127	-0.083	-0.083	+0.138				+0.738***	+0.244*	-0.182	-0.182
CbCra96	-0.027				+0.597**	+0.093	+0.093	+0.093	+0.047				+0.499***	+0.192***	+0.116**	-0.041*
CbCra99	+0.067				+0.461				+0.198				+0.895***	+0.079	-0.263	-0.231
CbCra00	+0.098				+0.326				+0.309*	+0.040	+0.092	+0.067*	+0.860***	+0.200	-0.193	-0.098
CbCra01	-0.055				+0.427**	+0.164	+0.043	-0.065	-0.036				+0.791***	+0.050***	-0.269*	-0.204
CbCra02	+0.020				+0.408***	-0.102***	-0.039**	-0.152	-0.076				+0.606***	+0.148***	-0.04***	-0.142***
CbFoz99	-0.043				+0.515**	+0.088	-0.004	-0.115	+0.052				+0.770***	+0.213***	-0.029***	-0.203
CbMon95	-0.075				+0.348				+0.064*				+0.855***	+0.154***	-0.161**	-0.255*
CbMou01	+0.078				+0.494**	+0.100	+0.100	+0.100	-0.062				1***	+0.342***	-0.210	-0.150
CbPor01	-0.121				+0.482***	-0.285***	+0.012*	+0.012*	+0.155	+0.110			+0.857***	-0.368***	-0.251	-0.164
CbPor02	+0.066				+0.679***	+0138***	+0.003*	-0.094**	+0.098				+0.676***	+0.039***	-0.022**	-0.048**
CbSou01	-0.012				+0.357**	+0.076	-0.099	-0.170	+0.035				+0.951***	+0.120***	-0.247	-0.177
CbSou02	-0.025				+0.446***	+0.099*	-0.008	-0.052	-0.010				+0.594***	+0.238***	-0.053***	-0.117***
CbSev01	+0.050				+0.645***	+0.057*	-0.032*	-0.032*	+0.108				1***	+0.385***	-0.238	-0.158
CbCeu99	-0.050				+0.552***	+0.098**	-0.131*	-0.146*	+0.263**	+0.203**	+0.047	+0.053	+0.742***	+0.121***	-0.191*	-0.205*
CbFes01	+0.124	+0.064	+0.027	-0.015	+0.594***	+0136**	+0.05*	+0.05*	+0.181*	+0.081	-0.027*	-0.012***	+0.939***	-0.115***	-0.289	-0.222
CbMek01	-0.070				+0.643***	+0.131**	-0.073	-0.238	+0.057				+0.836***	+0.200**	-0.229	-0.147
CoAlt98	1*				+0.329	+0.129	+0.129	+0.129					+0.188			
CoCra01	+845***	+0.324**	+0.101***	-0.064***	-0.043								+0.488***	+0.112	-0.015	-0.015
CoMte95	-0.029				+0.054								+0.273			
CoPie96	-				+0.031								+0.252			
CoPor01	+0.782**	-0.310*	-0.237*	-0.339	+0.409**	+0.185	-0.04	-0.065					+0.256			
CoPor02	+0.280*	+0.183	-0.017	-0.102	+0.173								+0.399**	+0.133	+0.054	+0.054
CoSou02	+0.331***	+0.198*	-0.035***	-0.098**	-0.005								+0.200***	+0.100**	+0.045**	-0.013**
CoAlg01	+0.382**	+0.208*	-0.098**	-0.284***	-0.141								+0.372**	+0.071	-0.028	+0.018
CoAlg02	+0.465**	+0.242*	-0.081	-0.284*	-0.160								+0.464***	+0.191*	+0.04	+0.076
CoNar01	+0.796***	+0.092***	-0.288***	-0.275***	-0.075								+0.060			
CoNar02	+0.769***	+0.333***	-0.153	-0.104	+0.082*								+0.135			
CoStH01	+0.736***	+0.280***	-0.245***	-0.161	-0.158								+0.091			
CoStH02	+0.729***	+0.217**	-0.242	-0.141	+0.058								+0.167*			
CoAte97	+0.277*	+0.100	+0.015	+0.015	+0.032								+0.167*	+0.108	+0.052	+0.052
CoKit02	+0.085				+0.020								-			
CoNax99	+0.640**	+0.174	-0.208	-0.233	+0.439*	+0.011	-0.125	-0.06					+0.152			
CoSky02	+0.254				+0.721***	+0.303**	-0.04	-0.04					+0.144*			
CcCre00	+0.600				-0.007								+0.218*	+0.218***	+0.179***	+0.169***
CmSam97	+0.188				+0.069								+0.391***	+0.101	-0.113*	-0.093**

Appendix I (cont.)

	Locus Cio08				Locus Cib10			
Population	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**
CbAlc95	+0.127				+0.526***	+0.156	-0.05	-0.05
CbCas96	+0.300				+0.781***	+0.215**	-0.306*	-0.203*
CbCra95	+0.027				+0.855***	+0.113*	-0.297	-0.240
CbCra96	-0.015				+0.772***	+0.125*	-0.297	-0.240
CbCra99	+0.186				1***	+0.075*	-0.347	-0.283
CbCra00	-0.083				1***	+0.091***	-0.102**	-0.382
CbCra01	+0.348***	+0.124	+0.056	+0.158	+0.671***	+0.079*	-0.138	-0.192*
CbCra02	+0.138				+0.429***	+0.101	-0.091	-0.133
CbFoz99	+0.220	+0.144	+0.023	-0.078	+0.593***	+0.157*	-0.107	-0.113
CbMon95	+0.043				+0.603***	+0.035	-0.203	-0.096
CbMou01	+0.153*				+0.523***	+0.138	+0.069	+0.046*
CbPor01	-0.062				+0.616***	+0.035*	-0.170*	-0.151*
CbPor02	+0.044				+0.733***	+0.018**	-0.091*	-0.091*
CbSou01	+0.128				+0.670***	+0.021*	-0.155	-0.097
CbSou02	+0.016*				+0.798***	+0.011***	-0.113***	-0.231***
CbSev01	-0.025				+0.589***	+0.021	+0.008	+0.008
CbCeu99	+0.103*				+0.443***	+0.052	-0.021	-0.07
CbFes01	+0.061*				+0.611***	+0.123***	-0.056*	-0.181
CbMek01	+0.010				+0.754***	+0.096*	-0.245	-0.202
CoAlt98	+0.415**	-0.006	-0.132	-0.098				
CoCra01	+0.589***	+0.111	-0.093	-0.063				
CoMte95	1**	+0.294	-0.273	-0.213				
CoPie96	+0.308*							
CoPor01	+0.426*	+0.110	+0.02	-0.123				
CoPor02	+0.681***	+0.104*	-0.284	-0.224				
CoSou02	+0.605***	+0.074**	-0.054	-0.091				
CoAlg01	+0.567***	+0.128*	-0.115	-0.182				
CoAlg02	+0.102							
CoNar01	+0.544***	+0.046	-0.167	-0.158				
CoNar02	+0.659***	+0.035	-0.08	-0.022				
CoStH01	+0.797***	+0.115**	-0.154**	-0.288				
CoStH02	+0.510***	+0.063	-0.154	-0.055				
CoAte97	+0.729***	+0.119*	-0.072	-0.08				
CoKit02	+0.833***							
CoNax99	+0.457***	+0.161	+0.058	+0.146*				
CoSky02	+0.788***	+0.230**	-0.139*	-0.177				
CcCre00	+0.038							
CmSam97	+0.007							

Appendix II – Number of alleles and heterozygosity in each population per locus per dataset.

		Cib01				Cib03				Cib6				Cib07				Cio08				Cib10			
Population		**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**	**ORIG**	**MIN**	**MED**	**MAX**
CbAlc95	**Na**	7	7	7	7	4	5	5	5	9	9	9	9	8	9	9	9	7	7	7	7	6	7	7	7
	He	0.773	0.773	0.773	0.773	0.565	0.705	0.709	0.691	0.803	0.803	0.803	0.803	0.765	0.823	0.773	0.743	0.796	0.796	0.796	0.796	0.774	0.823	0.805	0.805
CbCas96	**Na**	7	7	7	7	5	6	6	6	8	9	9	9	4	4	4	4	6	6	6	6	4	5	5	5
	He	0.731	0.731	0.731	0.731	0.578	0.706	0.707	0.714	0.799	0.830	0.842	0.842	0.279	0.279	0.279	0.279	0.693	0.693	0.693	0.693	0.539	0.690	0.633	0.568
CbCra95	**Na**	5	5	5	5	3	4	4	4	10	10	10	10	5	6	6	6	5	5	5	5	4	5	5	5
	He	0.740	0.740	0.740	0.740	0.593	0.716	0.705	0.705	0.875	0.875	0.875	0.875	0.765	0.821	0.730	0.730	0.780	0.780	0.780	0.780	0.625	0.745	0.745	0.790
CbCra96	**Na**	7	7	7	7	3	4	4	4	10	10	10	10	9	10	10	10	7	7	7	7	5	6	6	6
	He	0.822	0.822	0.822	0.822	0.521	0.655	0.655	0.655	0.847	0.847	0.847	0.847	0.825	0.859	0.852	0.818	0.788	0.788	0.788	0.788	0.659	0.759	0.686	0.655
CbCra99	**Na**	6	6	6	6	3	3	3	3	7	7	7	7	5	6	6	6	7	7	7	7	4	5	5	5
	He	0.758	0.758	0.758	0.758	0.269	0.269	0.269	0.269	0.798	0.798	0.798	0.798	0.733	0.778	0.661	0.633	0.787	0.787	0.787	0.787	0.694	0.748	0.612	0.558
CbCra00	**Na**	6	6	6	6	3	3	3	3	8	8	9	9	4	5	5	5	5	5	5	5	5	6	6	6
	He	0.736	0.736	0.736	0.736	0.541	0.541	0.541	0.541	0.799	0.799	0.832	0.777	0.710	0.778	0.608	0.498	0.776	0.776	0.776	0.776	0.674	0.755	0.704	0.621
CbCra01	**Na**	7	7	7	7	5	6	6	6	10	10	10	10	6	7	7	7	7	8	8	8	5	6	6	6
	He	0.770	0.770	0.770	0.770	0.414	0.534	0.605	0.636	0.862	0.862	0.862	0.862	0.685	0.765	0.679	0.634	0.774	0.816	0.828	0.827	0.669	0.757	0.736	0.711
CbCra02	**Na**	7	7	7	7	5	6	6	6	11	11	11	11	8	9	9	9	7	7	7	7	8	9	9	9
	He	0.802	0.802	0.802	0.802	0.526	0.679	0.686	0.684	0.875	0.875	0.875	0.875	0.620	0.738	0.736	0.690	0.807	0.807	0.807	0.807	0.672	0.750	0.760	0.755
CbFoz99	**Na**	8	8	8	8	4	5	5	5	10	10	10	10	9	10	10	10	9	10	10	10	6	7	7	7
	He	0.797	0.797	0.797	0.797	0.539	0.661	0.688	0.693	0.805	0.805	0.805	0.805	0.727	0.807	0.762	0.673	0.819	0.835	0.855	0.848	0.693	0.772	0.752	0.726
CbMon95	**Na**	4	4	4	4	3	3	3	3	7	7	7	7	6	7	7	7	7	7	7	7	4	5	5	5
	He	0.726	0.726	0.726	0.726	0.517	0.517	0.517	0.517	0.687	0.687	0.687	0.687	0.602	0.732	0.678	0.597	0.749	0.749	0.749	0.749	0.693	0.770	0.715	0.650
CbMou01	**Na**	8	8	8	8	3	4	4	4	10	10	10	10	7	8	8	8	7	7	7	7	8	9	9	9
	He	0.795	0.795	0.795	0.795	0.537	0.652	0.652	0.652	0.842	0.842	0.842	0.842	0.792	0.837	0.511	0.410	0.768	0.768	0.768	0.768	0.794	0.837	0.832	0.821
CbPor01	**Na**	10	10	10	10	4	5	5	5	12	13	13	13	8	9	9	9	6	6	6	6	6	7	7	7
	He	0.778	0.778	0.778	0.778	0.555	0.696	0.690	0.690	0.861	0.871	0.871	0.871	0.722	0.723	0.651	0.567	0.732	0.732	0.732	0.732	0.720	0.786	0.763	0.757
CbPor02	**Na**	8	8	8	8	4	5	5	5	10	10	10	10	11	12	12	12	7	7	7	7	7	8	8	8
	He	0.723	0.723	0.723	0.723	0.553	0.691	0.698	0.696	0.811	0.811	0.811	0.811	0.820	0.838	0.822	0.807	0.795	0.795	0.795	0.795	0.753	0.797	0.776	0.776
CbSou01	**Na**	7	7	7	7	4	5	5	5	10	10	10	10	9	10	10	10	7	7	7	7	7	8	8	8
	He	0.732	0.732	0.732	0.732	0.523	0.618	0.675	0.681	0.767	0.767	0.767	0.767	0.778	0.800	0.601	0.513	0.679	0.679	0.679	0.679	0.812	0.831	0.751	0.663
CbSou02	**Na**	5	5	5	5	3	4	4	4	8	8	8	8	15	16	16	16	7	7	7	7	10	11	11	11
	He	0.752	0.752	0.752	0.752	0.445	0.570	0.612	0.628	0.781	0.781	0.781	0.781	0.731	0.813	0.780	0.707	0.671	0.671	0.671	0.671	0.707	0.772	0.748	0.702
CbSev01	**Na**	8	8	8	8	4	5	5	5	10	10	10	10	8	9	9	9	6	6	6	6	7	8	8	8
	He	0.777	0.777	0.777	0.777	0.530	0.672	0.686	0.686	0.861	0.861	0.861	0.861	0.738	0.827	0.518	0.399	0.751	0.751	0.751	0.751	0.765	0.810	0.806	0.806
CbCeu99	**Na**	12	12	12	12	5	6	6	6	12	13	13	13	7	8	8	8	10	10	10	10	9	10	10	10
	He	0.828	0.828	0.828	0.828	0.613	0.715	0.719	0.714	0.858	0.874	0.876	0.862	0.628	0.744	0.703	0.680	0.839	0.839	0.839	0.839	0.849	0.869	0.848	0.817
CbFes01	**Na**	15	16	16	16	4	5	5	5	14	15	15	15	10	11	11	11	8	8	8	8	8	9	9	9
	He	0.864	0.871	0.881	0.884	0.420	0.572	0.609	0.609	0.813	0.835	0.851	0.850	0.820	0.773	0.667	0.598	0.771	0.771	0.771	0.771	0.580	0.704	0.717	0.698
CbMek01	**Na**	12	12	12	12	5	6	6	6	9	9	9	9	7	8	8	8	9	9	9	9	5	6	6	6
	He	0.882	0.882	0.882	0.882	0.563	0.702	0.704	0.678	0.830	0.830	0.830	0.830	0.680	0.780	0.618	0.541	0.824	0.824	0.824	0.824	0.710	0.784	0.696	0.668

Appendix II (cont.)

Population		Cib01				Cib03				Cib07				Cio08			
		ORIG	MIN	MED	MAX	ORIG	MIN	MED	MAX	ORIG	MIN	MED	MAX	ORIG	MIN	MED	MAX
CoAlt98	**Na**	2	2	2	2	5	6	6	6	7	7	7	7	8	9	9	9
	He	0.111	0.111	0.111	0.111	0.756	0.798	0.798	0.798	0.658	0.658	0.658	0.658	0.730	0.806	0.778	0.764
CoCra01	**Na**	5	6	6	6	5	5	5	5	6	7	7	7	7	8	8	8
	He	0.341	0.554	0.649	0.627	0.759	0.759	0.759	0.759	0.718	0.783	0.788	0.788	0.736	0.805	0.768	0.753
CoMte95	**Na**	4	4	4	4	3	3	3	3	3	3	3	3	3	4	4	4
	He	0.591	0.591	0.591	0.591	0.641	0.641	0.641	0.641	0.515	0.515	0.515	0.515	0.612	0.735	0.530	0.465
CoPie96	**Na**	1	1	1	1	4	4	4	4	6	6	6	6	3	3	3	3
	He	0.000	0.000	0.000	0.000	0.685	0.685	0.685	0.685	0.691	0.691	0.691	0.691	0.405	0.405	0.405	0.405
CoPor01	**Na**	3	4	4	4	5	6	6	6	6	6	6	6	5	6	6	6
	He	0.328	0.530	0.592	0.529	0.689	0.755	0.758	0.711	0.743	0.743	0.743	0.743	0.709	0.770	0.758	0.728
CoPor02	**Na**	6	7	7	7	4	4	4	4	8	9	9	9	5	6	6	6
	He	0.515	0.579	0.687	0.705	0.668	0.668	0.668	0.668	0.712	0.779	0.788	0.788	0.583	0.715	0.667	0.636
CoSou02	**Na**	7	8	8	8	6	6	6	6	9	10	10	10	7	8	8	8
	He	0.504	0.589	0.708	0.637	0.695	0.695	0.695	0.695	0.789	0.815	0.827	0.833	0.778	0.820	0.789	0.751
CoAlg01	**Na**	5	6	6	6	4	4	4	4	8	9	9	9	7	8	8	8
	He	0.285	0.391	0.620	0.564	0.633	0.633	0.633	0.633	0.806	0.842	0.820	0.801	0.689	0.777	0.756	0.723
CoAlg02	**Na**	6	7	7	7	4	4	4	4	12	13	13	13	4	4	4	4
	He	0.419	0.545	0.659	0.626	0.615	0.615	0.615	0.615	0.875	0.896	0.853	0.832	0.404	0.404	0.404	0.404
CoNar01	**Na**	6	7	7	7	5	5	5	5	6	6	6	6	5	6	6	6
	He	0.477	0.649	0.637	0.598	0.540	0.540	0.540	0.540	0.694	0.694	0.694	0.694	0.714	0.782	0.753	0.723
CoNar02	**Na**	5	6	6	6	4	4	4	4	8	8	8	8	6	7	7	7
	He	0.457	0.642	0.467	0.413	0.596	0.596	0.596	0.596	0.730	0.730	0.730	0.730	0.737	0.793	0.636	0.570
CoStH01	**Na**	8	9	9	9	6	6	6	6	6	6	6	6	5	6	6	6
	He	0.600	0.744	0.646	0.587	0.627	0.627	0.627	0.627	0.598	0.598	0.598	0.598	0.646	0.751	0.709	0.648
CoStH02	**Na**	6	7	7	7	6	6	6	6	7	7	7	7	5	6	6	6
	He	0.463	0.648	0.555	0.475	0.615	0.615	0.615	0.615	0.693	0.693	0.693	0.693	0.686	0.759	0.709	0.647
CoAte97	**Na**	8	9	9	9	7	7	7	7	14	15	15	15	13	14	14	14
	He	0.660	0.718	0.751	0.751	0.792	0.792	0.792	0.792	0.877	0.882	0.895	0.895	0.894	0.874	0.776	0.711
CoKit02	**Na**	14	14	14	14	8	8	8	8	2	2	2	2	7	7	7	7
	He	0.782	0.782	0.782	0.782	0.778	0.778	0.778	0.778	0.219	0.278	0.219	0.219	0.847	0.847	0.847	0.847
CoNax99	**Na**	4	5	5	5	5	6	6	6	11	11	11	11	10	10	11	11
	He	0.295	0.455	0.609	0.607	0.639	0.734	0.731	0.703	0.858	0.858	0.858	0.858	0.865	0.865	0.770	0.651
CoSky02	**Na**	2	2	2	2	3	4	4	4	9	9	9	9	7	8	8	8
	He	0.454	0.454	0.454	0.454	0.288	0.465	0.566	0.566	0.730	0.730	0.730	0.730	0.709	0.794	0.705	0.635
CcCre00	**Na**	3	3	3	3	3	3	3	3	23	23	24	24	13	13	13	13
	He	0.611	0.611	0.611	0.611	0.054	0.054	0.054	0.054	0.933	0.933	0.939	0.935	0.882	0.882	0.882	0.882
CmSam97	**Na**	5	5	5	5	11	11	11	11	6	7	7	7	8	8	8	8
	He	0.450	0.450	0.450	0.450	0.832	0.832	0.832	0.832	0.622	0.705	0.735	0.714	0.732	0.732	0.732	0.732

Appendix III – F_{ST} average values ± standard deviations for each locus and each dataset calculated between *C. barbara* populations and between *C. orni* populations – for all populations, within each region and among regions.

		Locus Cib01	Locus Cib03	Locus Cib06	Locus Cib07	Locus Cio08	Locus Cib10
C. barbara							
All populations	**ORIG**	0.042±0.0425	0.098±0.1282	0.05±0.0463	0.088±0.0957	0.035±0.0358	0.045±0.065
	MIN	0.042±0.0424	0.078±0.0953	0.048±0.0458	0.062±0.0954	0.035±0.0353	0.021±0.0312
	MED	0.042±0.0423	0.081±0.0934	0.053±0.0442	0.074±0.1215	0.037±0.0346	0.022±0.0236
	MAX	0.042±0.0425	0.085±0.0953	0.059±0.0479	0.081±0.1354	0.041±0.0354	0.028±0.0273
Iberian Peninsula	**ORIG**	0.020±0.0246	0.068±0.1047	0.048±0.0488	0.088±0.1046	0.037±0.0389	0.031±0.0578
	MIN	0.020±0.0246	0.064±0.0824	0.047±0.0487	0.065±0.1064	0.037±0.0383	0.013±0.025
	MED	0.020±0.0246	0.070±0.0846	0.051±0.0477	0.083±0.1331	0.04±0.0374	0.016±0.0201
	MAX	0.020±0.0246	0.074±0.0869	0.056±0.0525	0.091±0.1477	0.044±0.0382	0.026±0.0278
Africa	**ORIG**	0.094±0.0294	0.170±0.1498	0.055±0.0396	0.086±0.0713	0.031±0.0271	0.078±0.0695
	MIN	0.094±0.0293	0.112±0.1141	0.052±0.0382	0.054±0.0624	0.031±0.027	0.041±0.0356
	MED	0.093±0.0289	0.108±0.1079	0.058±0.0345	0.054±0.0864	0.032±0.0266	0.035±0.0258
	MAX	0.094±0.0282	0.112±0.109	0.065±0.0348	0.057±0.0981	0.034±0.0269	0.034±0.0255
Iberian Peninsula vs. Africa	**ORIG**	0.098±0.0245	0.178±0.1505	0.057±0.04	0.089±0.0717	0.031±0.0276	0.075±0.0682
	MIN	0.098±0.0244	0.118±0.1152	0.054±0.0386	0.055±0.0635	0.031±0.0276	0.039±0.0348
	MED	0.098±0.0241	0.113±0.1093	0.059±0.035	0.057±0.0885	0.032±0.0272	0.034±0.0252
	MAX	0.098±0.0237	0.116±0.111	0.066±0.0352	0.059±0.1007	0.034±0.0274	0.033±0.0255
C. orni							
All populations	**ORIG**	0.252±0.2542	0.119±0.1298		0.134±0.1374	0.099±0.1073	
	MIN	0.229±0.2132	0.106±0.1041		0.125±0.1196	0.087±0.0967	
	MED	0.254±0.1927	0.112±0.0952		0.135±0.1234	0.113±0.1262	
	MAX	0.270±0.207	0.119±0.0987		0.139±0.1245	0.127±0.1418	
Iberian Peninsula	**ORIG**	0.149±0.229	0.036±0.0522		0.096±0.1116	0.081±0.116	
	MIN	0.154±0.2114	0.039±0.0492		0.088±0.0885	0.094±0.1143	
	MED	0.219±0.2043	0.055±0.0504		0.094±0.0788	0.142±0.1444	
	MAX	0.258±0.2206	0.066±0.0623		0.097±0.0832	0.157±0.1566	
France	**ORIG**	-0.011±0.015	-0.007±0.003		0±0.0144	0.021±0.0372	
	MIN	-0.013±0.0062	-0.007±0.003		0±0.0144	0.008±0.0154	
	MED	0.010±0.0167	-0.007±0.003		0±0.0144	0.011±0.0121	
	MAX	0.009±0.0152	-0.007±0.003		0±0.0144	0.013±0.0133	
Greece	**ORIG**	0.232±0.1473	0.212±0.1776		0.23±0.1447	0.061±0.0525	
	MIN	0.202±0.1338	0.147±0.1304		0.209±0.1233	0.056±0.0316	
	MED	0.206±0.1046	0.133±0.0921		0.228±0.1457	0.075±0.0696	
	MAX	0.215±0.103	0.140±0.0966		0.228±0.1457	0.106±0.1027	
Iberian Peninsula vs. France	**ORIG**	0.021±0.0476	0.040±0.0405		0.028±0.0394	0.04±0.0629	
	MIN	0.058±0.0832	0.044±0.0389		0.037±0.0402	0.042±0.0777	
	MED	0.173±0.196	0.055±0.0461		0.055±0.0547	0.073±0.1222	
	MAX	0.181±0.2285	0.062±0.0588		0.059±0.061	0.086±0.1419	
Iberian Peninsula vs. Greece	**ORIG**	0.522±0.1296	0.214±0.1163		0.21±0.1349	0.159±0.1106	
	MIN	0.450±0.123	0.179±0.0833		0.192±0.1198	0.12±0.1005	
	MED	0.361±0.1465	0.176±0.074		0.204±0.1321	0.129±0.1278	
	MAX	0.362±0.1523	0.183±0.0767		0.207±0.131	0.142±0.1441	
France vs. Greece	**ORIG**	0.451±0.1051	0.296±0.0975		0.243±0.1261	0.134±0.0857	
	MIN	0.334±0.0828	0.267±0.0654		0.234±0.1132	0.087±0.0367	
	MED	0.311±0.1075	0.264±0.0568		0.241±0.1269	0.076±0.0585	
	MAX	0.330±0.1234	0.268±0.0599		0.241±0.1269	0.083±0.0786	

Appendix IV – Multilocus F_{ST} values (loci in Hardy-Weinberg equilibrium) for each pair of populations of *C. barbara* (a) and of *C. orni*, *C. cretensis* and *C. mordoganensis* (b). Significant values (p<0.05) are shown in bold (results from 1000 permutations).

a)	CbCas96	CbCra95	CbCra96	CbCra99	CbCra00	CbCra01	CbCra02	CbFoz99	CbMon95	CbMou01	CbPor01	CbPor02	CbSou01	CbSou02	CbSev01	CbCeu99	CbFes01	CbMek01
CbAlc95	**0.03511**	**0.03075**	0.01203	0.00577	**0.03817**	**0.02893**	**0.03234**	**0.01725**	**0.0601**	0.00591	**0.01755**	0.01179	0.04987	0.0608	0.0247	**0.05938**	**0.0534**	**0.06019**
CbCas96		**0.05861**	**0.05486**	0.01799	**0.03656**	**0.02631**	**0.03859**	**0.04765**	**0.10943**	**0.04744**	**0.08085**	**0.06004**	**0.02288**	**0.01861**	**0.03969**	**0.0595**	**0.05543**	**0.06749**
CbCra95			0.01019	0.01863	0.01339	0.00351	-0.00584	0.01451	**0.04901**	0.00839	**0.04783**	**0.05115**	**0.04395**	**0.04183**	0.01286	**0.03127**	**0.04852**	**0.04122**
CbCra96				0.0131	**0.02954**	0.00912	**0.01598**	**0.02256**	**0.05758**	**0.02914**	**0.041**	**0.0433**	**0.0639**	**0.07266**	**0.03353**	**0.02895**	**0.05575**	**0.03929**
CbCra99					-0.00196	0.00815	0.00839	0.01514	**0.09396**	0.00725	**0.03898**	**0.04294**	0.01191	**0.01852**	0.01778	**0.02566**	**0.05495**	**0.04812**
CbCra00						0.00458	-0.00098	**0.02317**	**0.09463**	**0.03223**	**0.04561**	**0.06171**	**0.04122**	**0.04353**	**0.02558**	**0.03926**	**0.08533**	**0.05981**
CbCra01							0.0044	**0.02935**	**0.08912**	**0.02968**	**0.06639**	**0.06876**	**0.03997**	**0.04731**	**0.03656**	**0.04764**	**0.05898**	**0.05064**
CbCra02								**0.02001**	**0.06414**	**0.01942**	**0.04587**	**0.05482**	**0.03546**	**0.03497**	**0.01411**	**0.03356**	**0.0683**	**0.05389**
CbFoz99									**0.03688**	0.00671	**0.01911**	**0.02993**	**0.03493**	**0.03502**	0.00844	**0.03839**	**0.06482**	**0.05235**
CbMon95										**0.06032**	**0.04416**	**0.04945**	**0.10919**	**0.10549**	**0.02792**	**0.1004**	**0.1262**	**0.11166**
CbMou01											**0.01648**	**0.02879**	**0.03788**	**0.03866**	**0.01376**	**0.04977**	**0.05522**	**0.06146**
CbPor01												0.00528	**0.08363**	**0.08156**	**0.0237**	**0.07279**	**0.09628**	**0.08567**
CbPor02													**0.0763**	**0.07444**	**0.02495**	**0.08082**	**0.08557**	**0.08751**
CbSou01														0.00721	**0.03745**	**0.06532**	**0.08868**	**0.08625**
CbSou02															**0.02681**	**0.05349**	**0.07707**	**0.0907**
CbSev01																**0.04061**	**0.06999**	**0.07388**
CbCeu99																	**0.03792**	**0.01645**
CbFes01																		**0.02424**

Appendix IV (cont.)

b)	CoCra01	CoMte95	CoPie96	CoPor01	CoPor02	CoSou02	CoAlg01	CoAlg02	CoNar01	CoNar02	CoStH01	CoStH02	CoAte97	CoKit02	CoNax99	CoSky02	CcCre00	CmSam97
CoAlt98	0.00526	0.01545	-0.01561	0.03438	-0.00201	0.01751	**0.0676**	**0.06621**	0.02568	0.0115	0.00929	0.01663	**0.16245**	**0.28505**	**0.17272**	**0.2832**	**0.39002**	**0.22121**
CoCra01		0.04422	0.01805	0.01327	0.02019	**0.02991**	**0.07192**	**0.06686**	**0.06649**	**0.04491**	**0.06561**	**0.04199**	**0.143**	**0.25979**	**0.1317**	**0.27026**	**0.37271**	**0.21049**
CoMte95			-0.02601	0.04266	0.0186	**0.06849**	**0.08552**	**0.07564**	0.0113	0.00636	-0.00108	0.00701	**0.19569**	**0.35782**	**0.20105**	**0.3076**	**0.44365**	**0.26309**
CoPie96				0.02179	0.00126	0.02766	0.04225	0.03531	0.01379	-0.00132	0.00066	0.00996	**0.13682**	**0.2915**	**0.13895**	**0.2742**	**0.39948**	**0.18787**
CoPor01					**0.04576**	0.00171	0.03177	-0.00196	0.0255	0.01052	**0.05541**	0.00592	**0.15619**	**0.28503**	**0.11518**	**0.30517**	**0.38171**	**0.19622**
CoPor02						**0.02617**	**0.08897**	**0.08278**	**0.04177**	**0.0197**	**0.04269**	**0.02167**	**0.14877**	**0.26576**	**0.19257**	**0.20498**	**0.39221**	**0.23632**
CoSou02							**0.03203**	**0.02103**	**0.036**	**0.01898**	**0.05513**	**0.02381**	**0.14939**	**0.26897**	**0.14655**	**0.26579**	**0.35362**	**0.19934**
CoAlg01								0.0031	**0.08743**	**0.07084**	**0.09346**	**0.08398**	**0.15631**	**0.28885**	**0.08405**	**0.31205**	**0.37225**	**0.21366**
CoAlg02									**0.05652**	**0.04476**	**0.07476**	**0.05457**	**0.14666**	**0.27317**	**0.09625**	**0.29149**	**0.35994**	**0.17489**
CoNar01										-0.0102	-0.00331	-0.00909	**0.22066**	**0.36389**	**0.2368**	**0.33107**	**0.42127**	**0.24641**
CoNar02											0.00323	-0.01031	**0.19128**	**0.31985**	**0.20501**	**0.29577**	**0.4044**	**0.22993**
CoStH01												0.00904	**0.21828**	**0.37156**	**0.23252**	**0.35548**	**0.42503**	**0.2495**
CoStH02													**0.19847**	**0.32466**	**0.21404**	**0.2966**	**0.4101**	**0.23603**
CoAte97														**0.18529**	**0.11845**	**0.16913**	**0.31416**	**0.14065**
CoKit02															**0.27804**	**0.33535**	**0.44464**	**0.27651**
CoNax99																**0.32987**	**0.35474**	**0.23156**
CoSky02																	**0.43943**	**0.33336**
CcCre00																		**0.36397**

Appendix V – Total body length of cicadas (from tip of head to tip of the right wing) from Iberian Peninsula (N=248), Ceuta (N=25) and Fès (Morocco) (N=47). The rectangular box is delimited by the quartiles 25% and 75%, with the median value shown as a horizontal line; the whiskers indicate the non-outlier maximum and minimum, the circles are outliers, and the stars are extremes. All three regions differed significantly (Mann-Whitney tests, $p<0.003$).