

DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae)

T. Malausa¹, A. Fenis¹, S. Warot², J.-F. Germain³, N. Ris², E. Prado⁴, M. Botton⁵, F. Vanlerberghe-Masutti¹, R. Sforza⁶, C. Cruaud⁷, A. Couloux⁷ & P. Kreiter²

1 Institut National de la Recherche Agronomique, UMR 1301, INRA/UNSA/CNRS, Sophia-Antipolis, France

2 Institut National de la Recherche Agronomique, Unité Expérimentale de Lutte Biologique, Sophia-Antipolis, France

3 LNPV Montpellier, Campus International de Baillarguet, Montferrier-sur-Lez, France

4 Empresa de Pesquisa Agropecuária de Minas Gerais, EPAMIG, Lavras, Brazil

5 Embrapa Uva e Vinho, Bento Gonçalves, Brazil

6 European Biological Control Laboratory, USDA-ARS, Campus International de Baillarguet, Montferrier-sur-Lez, France

7 Génoscope, Centre National de Séquencage, Evry, France

Keywords

Planococcus, *Pseudococcus*, Coccoidea, mitochondrial DNA, molecular characterization, ribosomal DNA

Correspondence

Thibaut Malausa (corresponding author), Institut National de la Recherche Agronomique, UMR 1301, Equipe BPI, 400, Route des Chappes, BP 167, 06903 Sophia-Antipolis Cedex, France.
E-mail: tmalausa@sophia.inra.fr

Received: August 12, 2009; accepted: November 30, 2009.

doi: 10.1111/j.1439-0418.2009.01495.x

Abstract

Mealybugs (Hemiptera: Pseudococcidae) are major pests of a wide range of crops and ornamental plants worldwide. Their high degree of morphological similarity makes them difficult to identify and limits their study and management. We aimed to identify a set of markers for the genetic characterization and identification of complexes of taxa in the Pseudococcidae. We surveyed and tested the genetic markers used in previous studies and then identified new markers for particularly relevant genomic regions for which no satisfactory markers were available. We tested all markers on a subset of four taxa distributed worldwide. Five markers were retained after this first screening: two regions of the mitochondrial cytochrome oxidase I gene, 28S-D2, the entire internal transcriber space 2 locus and the rpS15-16S region of the primary mealybug endosymbiont *Tremblaya princeps*. We then assessed the utility of these markers for the characterization and identification of 239 samples from 43 sites in France and Brazil. The five markers studied (i) successfully distinguished all species identified by morphological examination, (ii) disentangled complexes of species by revealing intra-specific genetic variation and identified a set of closely related taxa for which taxonomic status requires clarification through further studies, and (iii) facilitated the inference of phylogenetic relationships between the characterized taxa.

Introduction

Mealybugs (Hemiptera: Pseudococcidae) are major pests infesting a wide range of crops and ornamental plants worldwide (Ben-Dov 1994; Miller and Rossman 1995; Abate et al. 2000; Millar 2002; Miller et al. 2002, 2005). These sap-sucking insects have been studied intensively for decades because of the economic losses they cause to agriculture, through

direct damages to crops, virus transmission and decreases in yield quality (Sforza et al. 2003; Charles et al. 2006; Kuniyuki et al. 2006; Douglas and Kruger 2008; Meyer et al. 2008; Nakaune et al. 2008).

The high degree of morphological similarity between different taxa of mealybugs is a major problem for the management of these pests and for studies of their systematic or population biology. Indeed, morphological examination is time-consuming, can only be carried

out at appropriate developmental stages (mostly in adult females), requires a high level of expertise and cannot always distinguish between taxa, particularly if the taxa concerned are closely related. This situation has motivated studies on the molecular characterization of mealybugs and the starting of 'barcoding of life' projects (e.g. the Scale Insect Barcoding Initiative in Africa). Several studies have focused on the use of DNA markers to distinguish between closely related species of mealybugs for fundamental research or agronomic applications (Beuning et al. 1999; Demontis et al. 2007; Cavalieri et al. 2008; Rung et al. 2008; Saccaggi et al. 2008). A few DNA-based phylogenetic studies have also led to a re-examination and revision of the classification of Pseudococcidae (Downie and Gullan 2004; Hardy et al. 2008). A few molecular studies have also focused on particular aspects of mealybug biology, such as the co-evolution of these insects with symbionts (von Dohlen et al. 2001; Baumann 2005; Baumann and Baumann 2005) and epigenetic mechanisms (Nur and Brett 1988; Bongiorno et al. 1999, 2004, 2007; Field et al. 2004; Khosla et al. 2006).

These studies have contributed to the development of new tools for mealybug characterization, but it remains difficult to evaluate the efficacy of different markers for distinguishing between closely related Pseudococcidae taxa and for intraspecific studies. Indeed, the various markers identified to date have been used in very different contexts, in studies with different aims, using different ranges and numbers of taxa. For example, Downie and Gullan (2004) and Hardy et al. (2008) used three genes (18S, 28S and EF1- α) in two large-scale phylogenetic studies of the Pseudococcidae. Thao et al. (2002) and Baumann and Baumann (2005) focused on the co-evolution of mealybugs and their endosymbionts. They studied the genes of the primary endosymbiont *Tremblaya princeps* (16S-23SrDNA) and a few regions of mealybug mitochondrial DNA (CytB-16SrDNA), but used only a few samples. Other genes, such as those encoding internal transcriber space 1 and 2 (ITS1, ITS2) (Beuning et al. 1999), triose phosphate isomerase, dynamin (Hardy 2007) and cytochrome oxidase I (COI) (Rung et al. 2008), have also occasionally been used. The heterogeneity of the genes or parts of genes used in different studies of the Pseudococcidae strongly limits the sharing of data or results within the scientific community and assessments of the relative value of different markers for various applications. For example, the efficacy of markers already validated for phylogeny applications (Downie and Gullan 2004; Hardy et al. 2008) for disentangling complexes of closely

related taxa remains unclear. Conversely, the various markers used to approach specific questions – including the study of complexes of taxa (Rung et al. 2008) – have not been evaluated for studies of the phylogenetic relationships between mealybugs.

In this study, we aimed at selecting a set of markers for fast and reliable routine characterization of Pseudococcidae and fine-scale study of complexes of cryptic taxa within this family. We tested markers identified in a survey of previous molecular studies of the Pseudococcidae or designed on the basis of data collected from international databases. All the available polymerase chain reaction (PCR) primers were first tested on a subset of common taxa. We selected those markers for which amplification was entirely successful and taxonomic discrimination efficiency was satisfactory. We then assessed the discrimination efficiency and utility for phylogeny applications of the markers retained for analysis, by characterizing 239 unidentified samples collected from various crops in France and Brazil. These markers (i) clearly distinguished closely related species in several genera; (ii) revealed the existence of intraspecific variation and, possibly, of unknown complexes of species; and (iii) facilitated the investigation of phylogenetic relationships between identified taxa.

Materials and Methods

Sample collection

We studied 37 populations of mealybugs collected from crops or ornamental plants throughout France or received by mail from French producers (for further details, see Table S1, Supporting Information). Six populations were also collected in Brazil. Samples were conserved in ethanol (70% or 95%) and stored at -20°C .

DNA extraction, amplification and sequencing

Genomic DNA was extracted with the DNeasy Tissue Kit (QIAGEN, Hilden, Germany). Mealybugs were not crushed before extraction. Instead, we extended the cell lysis time beyond the manufacturer's recommendations (~ 6 – 8 h). Two elution steps ($2 \times 50 \mu\text{l}$ of AE buffer) were performed to increase the amount of DNA extracted.

All PCRs were performed with Phusion High-Fidelity DNA polymerase 530L (FINNZYMES, Espoo, Finland). We chose to work with this enzyme because of its high fidelity and its high speed,

making it possible to perform routine PCR within about an hour on standard thermocyclers. All PCRs were performed in a total volume of 25 μ l: 23 μ l of mix + 2 μ l of diluted DNA (between 1 and 20 ng). Reagent concentrations were: 1x Phusion HF buffer, 0.01 U/ μ l Phusion enzyme, 200 μ M dNTPs and 0.5 μ M of each primer. All the primers used are given in table 1. PCR conditions were as specified in the Phusion High-Fidelity DNA polymerase 530L kit (FINNZYMES): initial denaturation at 98°C for 30 s, followed by 35 cycles of (i) denaturation at 98°C for 10 s, (ii) annealing at temperatures of 48–60°C for 15 s (see table 2), (iii) elongation at 72°C for 15 s

and by a final extension period at 72°C for 5 min. PCR products were separated by electrophoresis in agarose gels (1.5–2.5% agarose, depending on the size of the PCR products), to check their quality.

Polymerase chain reaction products for sequencing were sent to Genoscreen (Lille, France) or to the French National Genoscope for capillary electrophoresis on ABI 3130XL automatic sequencers (Applied Biosystems, Foster City, CA, USA). PCR products were sequenced on both strands. Consensus sequences were generated and checked with SEQSCAPE v2.5 (ABI). Alignments were manually edited with BIOEDIT 7.01 (Hall 1999).

Table 1 List of all the markers tested in this work: name of the region targeted, name of primers, 5'–3' sequence of primers, reference from which primer sequences were collected

Locus name	Primer name	F/R	Primer sequence	References
18S	18S2880	F	CTGGTTGATCCTGCCAGTAG	Downie and Gullan (2004)
	18SB	R	CCGCGGCTGCTGGCAACCGA	
28S	None (D2)	F	AGAGAGAGTTCAAGAGTACGTG	Belshaw and Quicke (1997)
	None (D2)	R	TTGGTCCGTGTTTCAAGACGGG	
	None (D10)	F	GTAGCCAAATGCCTCGTCA	
EF-1 α	None (D10)	R	CACAATGATAGGAAGAGCC	Dietrich et al. (2001)
	M3	F	CACATYAACATTGTCGTSATYGG	
	RcM44.9	R	CTTGATGAAATCYCTGTGTCC	
COI (HCO–LCO region)	M51.9	F	CARGACGTATACAAAATCGG	Downie and Gullan (2004)
	RcM53.2	R	GCAATGTGRGCTGTGGCA	
	LCO-1490	F	GGTCAACAAATCATAAAGATATTGG	
	HCO-2198	R	TAAACTTCAGGGTGACCAAAAAATCA	
COI (region 2)	LCO-M-2d-F	F	ATAACTATACCTATYATTATTGGAG	New
	LCO-M-2d-R	R	AATAAATGTTGATATAAAATTGG	
	C1-J-2183	F	CAACATTTATTTGATTTTTTGG	
	C1-N-2568	R	GCWACWACRTAATAKGTATCATG	
ITS	ITS5	F	CGCGCGGATCCGGAAGTAAAGTCGTAACAAGG	Beuning et al. (1999)
	MS6	R	GACGAGAAGCCYGTTCGATC	
	MS1	F	GATCGAACRGGCTTCTCGTC	
	MS4	R	CGACCCTCAGACAGGCGTGG	
	MS3	F	CCACGCCTGTCTGAGGGTCTG	
	MS2	R	ACGGCCGGAATAACGACGCG	
	MS5	F	CGCGTCGTTATTCCGGCCGT	
	ITS26	R	TATATGGATCCATATGCTTAAGTTCAGCGGGT	
ITS2 (complete)	ITS2-M-F	F	CTCGTGACCAAGAGTCTCTG	New
	ITS2-M-R	R	TGCTTAAGTTCAGCGGGTAG	
Tpi	197fin (Tpi) 1F	F	AAYTGGAARATGAAYGG	Hardy (2007)
	197fin (Tpi) 2R	R	GCCCANACNGGYTCRTA	
Dynamin	3006fin (dynamin) 1F	F	CCNGAYATGGCNTTYGA	
	3006fin (dynamin) 2R	R	TCYTCRTGRTTNGTRTTCATRTA	
leuA-16S	leuA	F	GTATCTAGAGGNATHCAYCARGAYGGNG	Baumann et al. (2002)
	U16S	R	GCCGTMCGACTWGCATGTG	
CytB-nd1-16S	CytB-nd1-16S F	F	GAGGATCCGGAATAATATTATTAATTTGAATTTGAGG	Baumann and Baumann (2005)
	CytB-nd1-16S R	R	GAGGTACCATTACTTTAGGGATAACAGG	
16S-12S	16S-12S-F	F	GAGGTACCGATAGAAACCAACCTGGCTTACACCGG	
	16S-12S-R1	R	GTGGATCCGTGCCAGCAGTTGCGGTTAAAC	
	16S-12S-R2	R	CAGTAATAAATTTTAAGGGGA	

Table 2 Annealing temperature and PCR results for each pair of primers and for each species tested. '+' indicates PCR generating a clear single band, 'mb' PCR resulting in multiple bands or smears, '-' PCR failing to produce a product. Lines in bold characters correspond to the markers and PCR conditions retained for routine protocols

Region	Primers	Anneal. temp. (°C)	Species tested							
			Screening step				Optimization step			
			<i>Ps. viburni</i>	<i>Ps. comstocki</i>	<i>H. bohemicus</i>	<i>Pl. citri</i>	<i>D. brevipennis</i>	<i>Pl. minor</i>	<i>Pl. ficus</i>	<i>Ps. longispinus</i>
18S	18S2880/18SB	52	mb	mb	–	mb				
		55	mb	mb	–	+				
		58	+	+	+	+				
28S	D2	52	mb	mb	+	mb				
		58	+	+	+	+	+	+	+	+
		60	+	+	+	+	+	+	+	+
	D10	52	mb	mb	mb	mb				
		58	mb	mb	mb	mb				
Elongation factor	M3/RcM44.9	48	+	–	–	+				
	M51.9/RcM53–2	52	+	–	–	+				
		48	+	–	–	+				
		52	+	–	–	+				
COI	1490/2198	46	–	mb	mb	mb				
		48	–	–	–	+				
	New (LCO)	48	–	+	+	+	–	+	+	+
	C1-J-2183/C1-N-2568	50	–	+	+	+	–	+	+	+
52		mb	+	+	+					
54		+	+	+	+	+	+	+	+	
		56	+	+	+	+	+	+	+	+
ITS	ITS5/MS6	52	mb	mb	+	+				
		55	mb	mb	+	+				
		58	+	+	mb	+				
	MS1/MS4	55	mb	mb	mb	mb				
58		mb	+	+	+					
	MS3/MS2	52	mb	mb	mb	mb				
56		+	mb	mb	+					
	MS5/ITS26	52	+	+	+	+				
New (ITS2)		58	+	+	+	+	+	+	+	+
		60	+	+	+	+	+	+	+	+
Tpi	197fin 1F/2R	52	mb	mb	mb	mb				
		56	+	+	–	+				
Dynammin	3006fin 1F/2R	52	mb	mb	mb	mb				
		56	+	mb	+	+				
Tremblaya princeps	leuA - U16S	58	+	+		+	+	+	+	+
		60	+	+		+	+	+	+	+
CytB-nd1	CytB-nd1–16S F/R	45	mb	mb	mb	mb				
		52	mb	–	–	mb				
16S-12S	16S-12S-F	45	+	–	–	+				
	16S-12S-R1	52	–	–	–	–				
	16S-12S-F	45	mb	mb	mb	+				
	16S-12S-R2	52	–	–	+	+				

Morphological examination

For samples with sequences different from those published in international databases, a morphological examination was carried out for taxonomic identification. Mealybugs were prepared by a five-stage process before mounting: (1) the specimen was heated

gently ($\leq 40^{\circ}\text{C}$) in 10% KOH for 20 min; (ii) the specimen was washed in distilled water for 20 min; (iii) the specimen was stained by incubation for 1 h in a saturated solution of fuschin in a 1 : 1 : 1 mixture of distilled water, lactic acid and glycerol; (iv) the specimen was washed in glacial acetic acid for 1 h to stabilize the staining; and (v) the specimen was

transferred to lavender oil for at least 1 h, placed in a drop of Canada balsam on a slide and covered with a coverslip. The slide was then labelled and observed immediately under a microscope or after drying (40°C for 4–6 weeks). In most cases, identification was based on the keys of Beardsley (1966), Cox (1989), Williams and Watson (1988), Williams and Granara de Willink (1992) and Williams (2004).

When possible, taxonomic identification was carried out for at least two sequenced individuals with identical sequences. In most cases, a reference specimen was stored in 70% ethanol at –20°C or mounted on a slide. Slide-mounted specimens have been deposited to the national collection of Laboratoire National de la Protection des Végétaux (Montpellier, France). Samples stored in ethanol are available at INRA, Sophia Antipolis, France.

Testing and selection of markers

The PCR primers used in previous studies for the sequencing of Pseudococcidae were surveyed and ordered (primer sequences and associated references are given in table 1). All primer pairs were first tested on a subset of samples consisting of three to four individuals from four common species – *Planococcus citri*, *Pseudococcus viburni*, *Pseudococcus comstocki* and *Heliooccus bohemicus* – collected from populations previously identified on the basis of morphology. If PCR failed for all the individuals of a given species, it was repeated at least three times [at least once at low temperature (48°C)] for confirmation of this result.

We selected only markers amplified in all four species, plus one marker that we expected to fail in *H. bohemicus* (primary symbiont *T. princeps*’ rpS15–16S) because *T. princeps* is restricted to the Pseudococcinae and is thus absent from the Phenacoccinae and Putoidae (Hardy et al. 2008).

Optimization of selected primers, identification of new primers and sequencing tests

For the selected markers, we increased the range of species for amplification reactions by adding available identified samples from several species: *Planococcus minor*, *Planococcus ficus*, *Pseudococcus longispinus* and *Dysmicoccus brevipes* (3–4 individuals per species).

Polymerase chain reaction annealing temperatures, reagent concentrations and elongation times were adjusted to optimize PCR product quality on the basis of the appearance of the products on agarose gels. PCR products giving a single band on agarose gels were sent for DNA sequencing. We assessed the read-

ability of the resulting sequences and selected a few primer pairs amplifying genes for which the sequences obtained distinguished between the different species and revealed intraspecific polymorphism.

We designed new primers for genes thought to be important for ‘barcoding of life’ applications [e.g. the LCO region of COI (Hebert et al. 2003) and ITS2] for which no appropriate primers had previously been found (i.e. the primers available gave few amplifications or amplified only very short regions). We aligned the sequences generated in this study and those already available for closely related organisms in international databases. We then looked for conserved regions and designed several combinations of primers. The online version of PRIMER3 (<http://frodo.wi.mit.edu/primer3/input.htm>) was used to design primers. We then tested these primers by following the procedure described above.

Generic characterization of unidentified samples

We tested five selected pairs of primers for the routine characterization of unidentified samples collected in France and Brazil.

DNA was extracted, amplified and sequenced and morphological examinations were carried out as described above. The PCR conditions used for the five targeted genes (see Results) were identical except for annealing temperature (table 2).

Phylogenetic analyses

We assessed the value of using a combination of all five markers for phylogenetic studies, by using a mixture model, as implemented in BayesPhylogenies (Pagel and Meade 2004) to infer the phylogenetic relationships between all taxa sampled and characterized at more than two loci. Analyses with nQ+ Γ mixture models, with n varying between one and six independent rate matrices (Qs), were performed and compared, using Bayes factors. As suggested by Pagel and Meade (2004), we also used a GTR model on our dataset. The Markov chain Monte Carlo method was used, with four chains of 10 000 000 iterations and a print frequency of 1000 iterations. The length of the burn-in period was assessed by plotting likelihood across iterations. All iterations corresponding to the burn-in period were removed from the outputs of BayesPhylogenies before subsequent analyses. We used the sump command of MRBAYES (Ronquist and Huelsenbeck 2003) to obtain a summary of BayesPhylogenies outputs and to calculate Bayes factors. Majority rule consensus trees

were then drawn with PAUP 4.0b10 (Swofford 2003) ('contree/Majrule' command) from the output of the BayesPhylogenies analysis (9000 trees), using the best model selected. Trees were rooted on the most divergent taxa, based on the most recent phylogeny of Pseudococcidae available (Hardy et al. 2008).

Results

Testing and selection of markers

Five of the 17 pairs of primers tested gave single PCR products in all species used in the first screening step (table 2) and amplified the following regions: 18S, 28S-D2, ITS, COI and the rpS15-16S region of *T. princeps*.

Only small parts of the ITS region were amplified with previously reported combinations of primers, so we designed a new combination of primers for amplification of the entire ITS2 region (~800 bp). This new combination gave an amplicon for 100% of the samples tested (table 2).

Amplification of the 'barcoding' region of COI (LCO-HCO) was successful for only a minority of samples. We designed several new combinations of primers, using alignments of hemipteran sequences available from GenBank (data not shown), and tested them on the problematic taxa, with a view to identify the most efficient primers. One of these new combinations amplified a 491-bp fragment in a few more species, but could not amplify LCO in all the species tested.

We selected five markers from the initial screening for further DNA sequencing. Four of these markers gave amplicons with all DNA samples tested: 28S-D2 (~320 bp), the 2183-2568 region of COI (385 bp), rpS15-16S from *T. princeps* (~1050 bp), and the new ITS2 (~800 bp), referred to hereafter as '28SD2', 'COI-region2', '*T. princeps*' and 'ITS2', respectively. The new primers for the LCO region of COI (491 bp, referred to as 'COI-LCO') were not 100% successful, but were chosen nonetheless, because of the importance of this region in the 'barcoding of life' procedures applied to a wide range of organisms. Updated primers and protocols will be kept available at <http://bpi.sophia.inra.fr/dnabarcoding/>.

The number of fully unambiguous sequences (consensus of forward/reverse sequences) obtained for the five markers in the second stage of testing was 75 for COI-LCO, 131 for *T. princeps*, 157 for ITS2, 166 for COI-region2 and 172 for 28SD2. Most technical failures were due to the co-amplification of parasitoid DNA, because the whole body of the mealybugs, including

any parasitoids present, was used for DNA extraction. We identified 11 haplotypes for *T. princeps*, 14 for 28SD2, 16 for COI-LCO, 16 for ITS2 and 17 for COI-region2 (table 3, figs 2 and 3, Tables S2 and S3). In total, we obtained 22 multilocus haplotypes (table 3). All sequences have been deposited in GenBank under accession numbers GU134641–GU134711 and GU168801–GU168803.

Identification of mealybug species

The morphological characterization of samples collected from the 43 populations led to the identification of 11 different species: *Planococcus citri* (Risso), *Planococcus minor* (Maskell), *Planococcus ficus* (Signoret), *Pseudococcus viburni* (Signoret), *Pseudococcus comstocki* (Kuwana), *Pseudococcus longispinus* (Targioni Tozzetti), *Pseudococcus calceolariae* (Maskell), *Dysmicoccus brevipes* (Cockerell), *Phenacoccus parvus* Morrison, *Vryburgia rimariae* Tranfaglia, *Heliococcus bohemicus* Šulc. All the markers distinguished between the species identified on morphological examination, even the most closely related species, such as *Planococcus citri* and *Planococcus minor* (table 4).

They also identified two taxa closely related to *Ps. longispinus* (referred to as *Pseudococcus* nr *microadonidum*) and *Pseudococcus maritimus* (referred to as *Ps.* nr *maritimus*), leading to morphological re-examination of the specimens concerned.

Pseudococcus nr *microadonidum* is morphologically similar to *Ps. longispinus* and *Ps. microadonidum*. The principal characteristics of *Ps. longispinus* are the presence of one large and two smaller oral-rim tubular ducts close on the dorsal side to each abdominal cerarius, anterior to the antepenultimate and nearest thoracic cerarii, strongly sclerotized anal lobe and penultimate cerarii with highly concentrated trilocular pores and a labium frequently about 150 µm in length, but shorter in some specimens. *Ps. microadonidum*, a Micronesian species also present in the Seychelles, has two oral-rim tubular ducts, one large and one small, on the lateral margins of the dorsum. The penultimate cerarii and anal lobe are highly sclerotized, but with only about 40 trilocular pores in each anal lobe cerarius, resulting in a lower concentration of these pores than in *Ps. longispinus*, which has at least 60 such pores. The labium is short, at about 120 µm. The detected taxon has two oral-rim tubular ducts on the lateral margins of dorsum, a short labium and strongly sclerotized penultimate and anal lobe as in *Ps. microadonidum*, but with a large concentration of trilocular pores on the anal lobe.

Table 3 Multilocus haplotypes discovered for each taxon identified by morphological examination. GenBank accession numbers are given for each sequence. Codes of slide-mounted specimens (deposited at national collection of LNPV, Montpellier) are given for each multilocus haplotype. '–' designates unavailable sequences or samples for which no voucher specimen could be prepared (e.g. when only small larvae were available)

Morphological identification	Haplo-type no.	GenBank accession number of haplotype					Voucher no.
		28SD2	ITS2	COI-region2	COI-LCO	<i>Tprinceps</i>	
<i>Phenacoccus parvus</i>	1	GU134663	GU134679	GU134696	GU134711	–	0803245, 0803246, 0803247
<i>Heliooccus bohemicus</i>	1	GU134664	–	GU134695	GU134709	–	0900073, 0900074
	2	GU134664	–	GU134695	GU134710	–	–
<i>Vryburgia rimariae</i>	1	GU134659	GU134674	GU134680	GU134697	GU134641	0803233, 0803234, 0803235
<i>Pseudococcus nr microadonidum</i>	1	GU134656	GU134670	GU134681	GU134698	GU134642	0803248
<i>Pseudococcus longispinus</i>	1	GU134657	GU134671	GU134682	GU134699	GU134643	08024inra
<i>Pseudococcus calceolariae</i>	1	GU168802	GU168803	GU134684	–	GU134645	0801500, 0801121
<i>Pseudococcus comstocki</i>	1	GU134654	GU134668	GU134688	GU134701	GU134648	0900067, 0900068
<i>Pseudococcus nr maritimus</i>	2	GU134654	GU134669	GU134688	GU134701	GU134648	–
	1	GU134655	GU134672	GU134683	GU134700	GU134644	0803225, 0803228
<i>Pseudococcus viburni</i>	1	GU134653	GU134667	GU134685	–	GU134646	0803239, 0803240, 0803241, 0803242, 0803243, 0803244
	2	GU134652	GU134665	GU134686	–	GU134646	0900069, 0900070
	3	GU134652	GU134666	GU134686	–	GU134646	–
<i>Dysmicoccus brevipes</i>	1	GU134658	GU134673	GU134690	GU134702	GU134650	0803226, 0803227, 0803229, 0803230, 0803231, 0803232
<i>Planococcus ficus</i>	1	GU134662	GU134677	GU134687	GU134703	GU134647	08258inra, 08260inra
	2	GU134662	GU134677	GU134693	GU134704	GU134647	08259inra
<i>Planococcus minor</i>	1	GU134661	GU134676	GU134694	GU134708	GU134649	0803223, 0803224
	2	GU134661	GU134676	GU134689	GU134708	GU134649	–
<i>Planococcus citri</i>	1	GU134660	GU134678	GU134691	GU134705	GU134651	08089inra
	2	GU134660	GU134675	GU134692	GU134706	GU134651	08002inra, 08014inra
	3	GU134660	GU134675	GU134692	GU134707	GU134651	0830220, 0830222
	4	–	–	–	GU168801	GU134651	08122inra

We identified two samples as *Ps. nr maritimus*. This taxon could be distinguished from *Ps. viburni* on the basis of (i) the higher density of dorsal oral-rim ducts (particularly in the median abdominal area), (ii) the density of ventral oral collar tubular ducts in the ventral inter-antennal area, (iii) the presence of translucent pores on the hind coxas, and (iv) the absence of associated discoidal pores near the eyes (these last two characters are used to distinguish *Ps. comstocki* from *Ps. viburni*). The limited number of specimens currently available makes more accurate identification impossible.

Intraspecific variation

Substantial genetic variation was observed between specimens identified as the same species on the basis of morphological analysis, for all markers except

Tprinceps. At ITS2, three haplotypes were observed in populations morphologically identified as *Ps. viburni* and two haplotypes were observed in both *Ps. comstocki* and *Pl. citri*. At COI-LCO, four haplotypes were observed for *Pl. citri* and two for both *Pl. ficus* and *H. bohemicus*. Finally, at COI-region2, two haplotypes were associated with *Ps. viburni*, *Pl. minor*, *Pl. ficus* and *Pl. citri*. Finally, at 28SD2, two haplotypes were found in the only population of *H. bohemicus* available. At ITS2 and COI-region2, considerable divergence was observed between the haplotypes of French and Brazilian *Ps. viburni* populations.

Phylogenetic trees

The variability of ITS2 sequences between taxa was so great that it was impossible to align these sequences. We therefore generated topologies from a

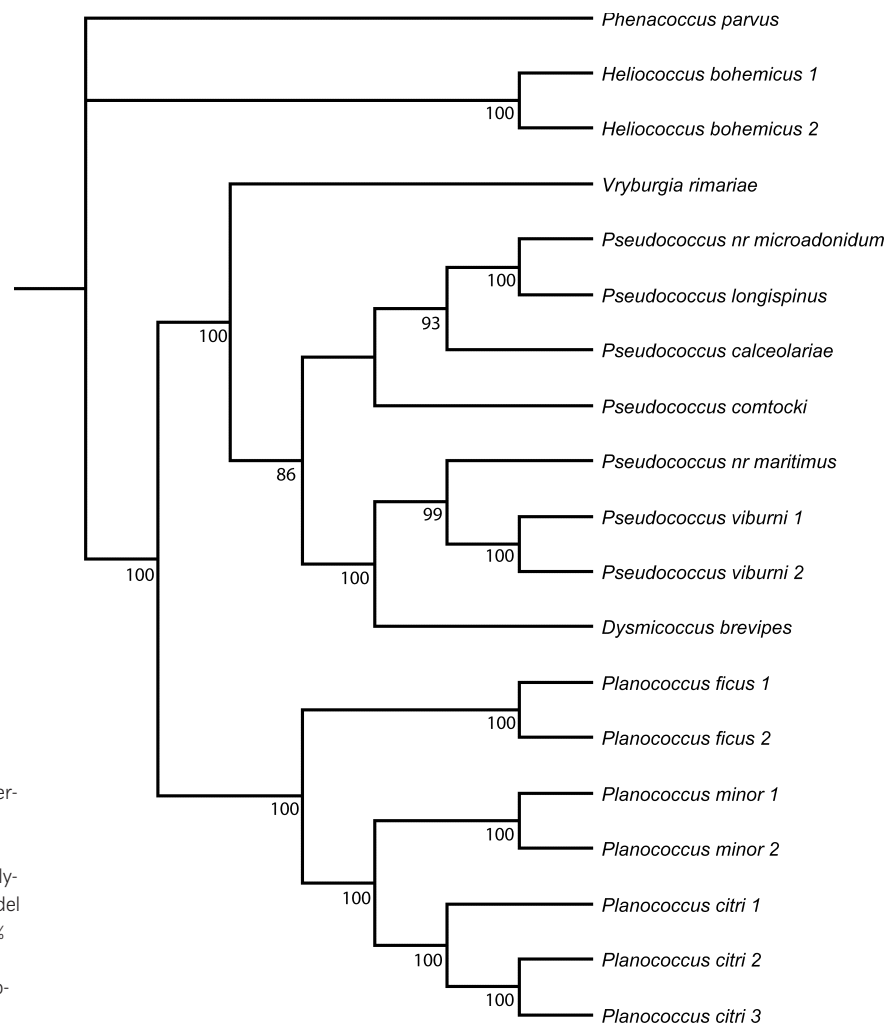


Fig. 1 Phylogenetic relationships between the taxa identified in this study and characterized at three or more DNA regions. The majority rule consensus tree was obtained from 9000 trees generated by Bayesian analysis, based on the best selected mixture model (three matrices). Posterior probabilities >70% (calculated from 9000 trees) are displayed under the nodes. Names of multilocus haplotypes correspond to table 3.

dataset consisting of DNA sequences for COI-LCO, COI-region2, 28SD2 and *Tprinceps*. Relationships between the haplotypes of the various genes taken individually are shown in figs 2 and 3. The topologies obtained on the basis of all loci and for individual loci analysed separately were similar (figs 1–3), but with different degrees of resolution at internal nodes and overall support. The genus *Planococcus* appeared monophyletic in all cases. Within this genus, *Pl. citri* and *Pl. minor* clustered together for all genes, but were separated from *Pl. ficus* in all cases. The genus *Pseudococcus* displayed three distinct clusters: (i) *Ps. longispinus* and *Ps. nr microadonidum*, (ii) the two divergent *Ps. viburni* (from Brazil and France), *Ps. nr maritimus* and *D. brevipes*, and (iii) *Ps. comstocki* and *Ps. calceolariae* were positioned with variable support between these two clusters. *Vryburgia* sp., *D. brevipes* and the genus *Pseudococcus* formed a monophyletic group with respect to *Planococcus* and

Heliococcus. *Ph. parvus* was used as the outgroup for the rooting of trees, following (Hardy et al. 2008).

Discussion

We propose here a set of five molecular markers suitable for the reliable characterization of *Pseudococcidae*. These markers distinguished between taxa collected from crops and ornamental plants at a large number of sites, previously identified morphologically. They also separated closely related taxa that are difficult to differentiate morphologically (e.g. *Ps. longispinus* and *Ps. nr microadonidum*, *Pl. minor* and *Pl. citri*). The *Tprinceps* marker displayed polymorphism patterns consistent with morphological identification (except for *Ps. viburni*, for which differences were observed between the Brazilian and French populations), and the other markers were variable enough to reveal intraspecific polymorphism

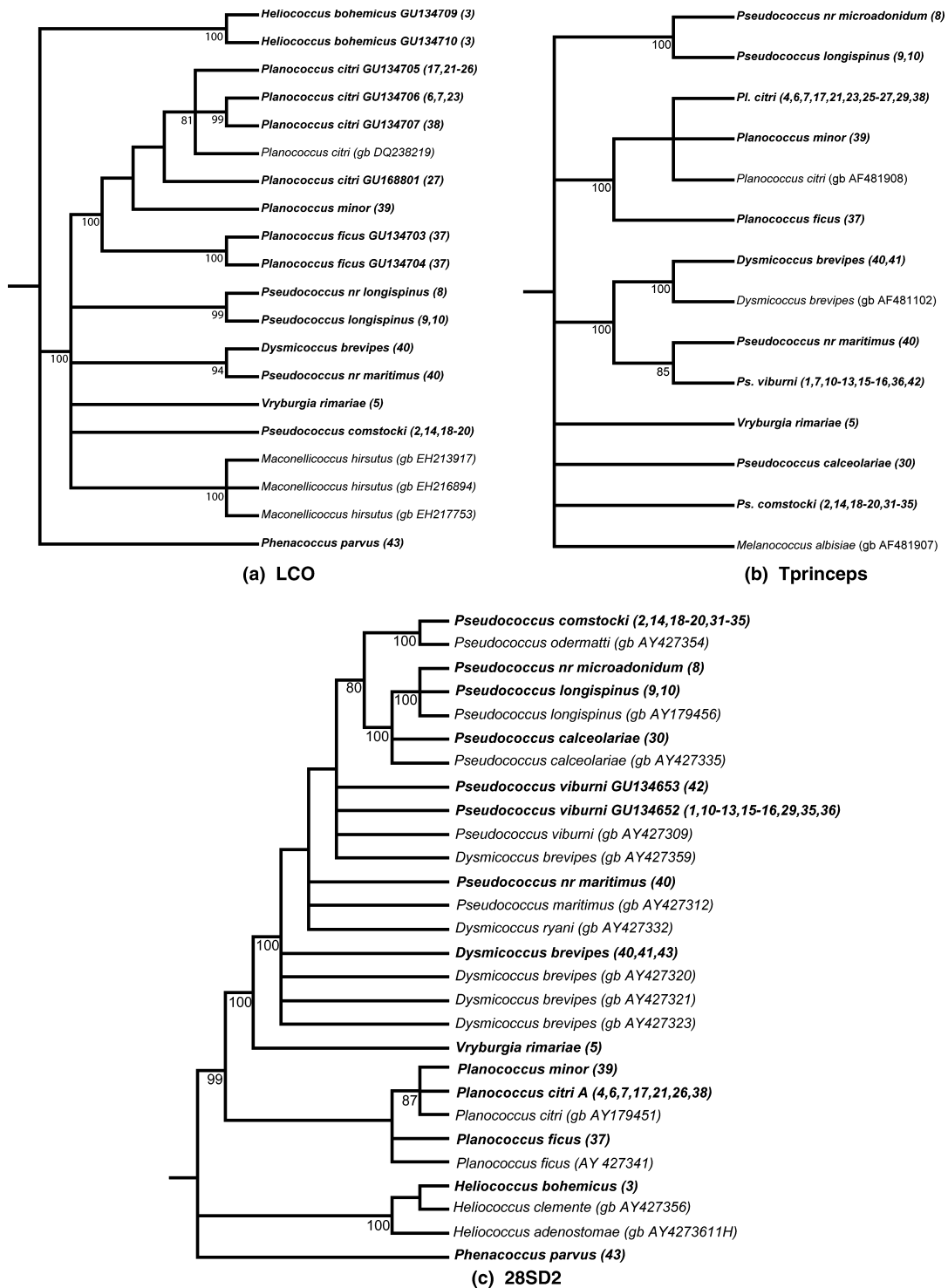


Fig. 2 Phylogenetic relationships between haplotypes at loci LCO (a), *Tprinceps* (b) and 28SD2 (c). For each locus, the majority rule consensus tree was obtained from 9000 trees produced by Bayesian analyses on the basis of mixture models. The mixture models selected consist of one matrix for 28SD2 and two matrices for LCO and *Tprinceps*. Haplotypes obtained from the GenBank database are followed by 'gb' and their accession number. Haplotypes discovered within a same morphologically identified taxon are distinguished using their GenBank accession. The occurrence of each haplotype in the populations collected (see codes in Table S1) is shown in brackets. Posterior probabilities >70% (calculated from 9000 trees) are displayed under the nodes.

Table 4 Number of differences between sequences obtained from closely related identified species or between different populations of a single identified species. In each comparison, the number of sequences from each taxon used in the comparison is given. The number of differences is the number of sites (insertions/deletions are considered as a single site) varying between species but not within species (in inter-species comparisons), or within species (for intra-species-variations)

	Inter-taxon variations					Intra-taxon variations			
	sp1	<i>Pl. citri</i>	<i>Pl. citri</i>	<i>Ps. longisp.</i>	<i>Ps. viburni</i> Fr	<i>Ps. viburni</i> (France)	<i>Pl. ficus</i>	<i>Ps. comstocki</i>	<i>Pl. citri</i>
	sp2	<i>Pl. minor</i>	<i>Pl. ficus</i>	<i>Ps. nr microad.</i>	<i>Ps. viburni</i> Br				
COI-LCO	N sp1	33	33	4	–	–	4	15	33
	N sp2	5	4	2	–	–			
	Fix diff.	7	26	20	–	–	2	0	4
COI-region2	N sp1	34	34	5	35	35	4	55	34
	N sp2	5	4	2	6				
	Fix diff.	7	25	13	1	0	1	0	3
ITS2	N sp1	24	24	4	49	49	4	52	24
	N sp2	4	4	2	5				
	Fix diff.	4	>50	15	6	1	0	1	1
28S	N sp1	34	34	6	44	44	4	55	34
	N sp2	4	4	2	6				
	Fix diff.	1	1	3	1	0	0	0	0
<i>Tprinceps</i>	N sp1	30	30	4	39	39	3	35	30
	N sp2	4	3	2	3				
	Fix diff.	1	21	12	0	0	0	0	0

(table 4). Indeed, several haplotypes were observed at COI-LCO, COI-region2, ITS2 and, to a lesser extent, at 28SD2, in most of the commonly collected taxa (e.g. *Pl. citri*, *Ps. viburni*, *Ps. comstocki*). The existence of this fine-scale genetic variability and the haplotype divergence observed between populations suggest that these markers may prove useful for disentangling complexes of taxa and the population genetics patterns driven by selective or geographic factors. For example, the populations morphologically identified as *Pl. citri* and *Ps. viburni* displayed a range of differentiated taxa as a function of the region and host considered. These 'species' are therefore probably a set of taxa currently diverging.

Given the extent of the variation observed, these genes could also be used to design identification tools based on species-specific DNA amplification. The tools developed to date have been based mainly on COI (Demontis et al. 2007; Saccaggi et al. 2008). However, the structure of the polymorphism of this locus is not ideal for primer design: (i) this is an AT-rich region making it difficult to design primers with high annealing temperatures; (ii) high levels of homoplasy are observed; and (iii) point mutations are predominant and make it difficult to design robust species-specific primers. 28SD2 appears to be insufficiently variable for the design of species-specific primers, but the ITS2 and *Tprinceps* regions are

useful, because they are highly variable between taxa, with large 'blocks' of variable sites or insertions/deletions, making them highly suitable for such applications.

Finally, the new data obtained (including data for some previously unsequenced taxa, such as *Ps. comstocki*, *Ps. nr maritimus*, *Ps. nr microadonidum*, *V. rima-riae* and *Ph. parvus*) made it possible to assess the usefulness of these markers for phylogenetic studies, individually, or pooled together (~2000 bp) with analysis on the basis of mixture models. Congruent phylogenetic signals were detected based on the similar patterns of the topologies obtained (figs 1–3) and the identification of mixture models with small numbers of partitions as the best models in BayesPhylogenies even when four loci were used. Similar results were obtained for the loci of the primary endosymbiont, *Tremblaya princeps*, consistent with the tight co-evolution observed in previous studies (Spaulding and von Dohlen 1998; von Dohlen et al. 2001; Thao et al. 2002; Baumann 2005; Baumann and Baumann 2005). All topologies obtained are fully consistent with the phylogenetic results of Downie and Gullan (2004) and Hardy et al. (2008). In particular, they clearly confirm the monophyly of the genus *Planococcus* and the absence of monophyly for *Pseudococcus*. Indeed, a clear divergence was observed between *Ps. viburni*, *Ps. nr maritimus* on the one hand and

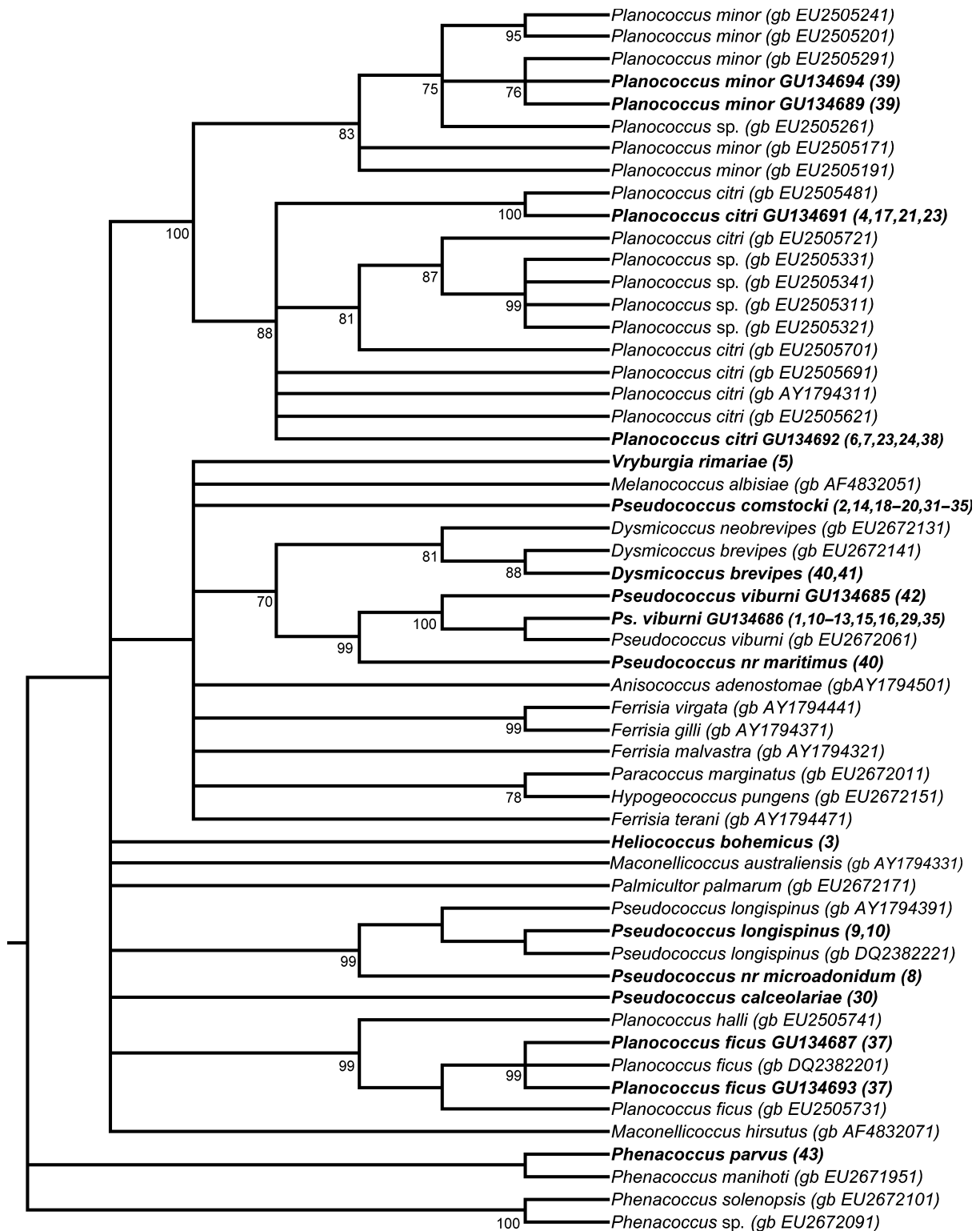


Fig. 3 Phylogenetic relationships between haplotypes at the COI-region2 locus. The majority rule consensus tree was obtained from 9000 trees generated by Bayesian analysis on the basis of the best selected mixture model (three matrices). Haplotypes obtained from the GenBank database are followed by 'gb' and their accession number. Haplotypes discovered within a same morphologically identified taxon are distinguished using their GenBank accession. The occurrence of each taxon in the populations collected (see codes in Table S1) is given in brackets. Posterior probabilities >70% (calculated from 9000 trees) are displayed under the nodes.

Ps. longispinus, *Ps. nr microadonidum*, *Ps. comstocki* and *Ps. calceolariae* on the other, with the monophyly of the genus disrupted by the presence of *D. brevipes*, which is closely related to *Ps. viburni* and *Ps. nr maritimus*. In addition, the genus *Vryburgia* was positioned, with a high level of support, between the genera *Pseudococcus* and *Planococcus*, as suggested by Downie and Gullan (2004) and Hardy et al. (2008). One limitation of the current set of markers is the failure of the COI-LCO marker to amplify a significant proportion of taxa, particularly in the genera *Pseudococcus* and *Dysmicoccus*. Additional studies are required to establish a fully satisfactory marker for this barcoding region. For other genes, failures of DNA amplification or sequencing resulted mostly from problems related to the DNA extracts: insufficient DNA or, in many cases, the presence of DNA from a parasitoid taxon infesting the mealybug and amplified with the primers used.

In conclusion, the markers tested and selected in this work are promising tools for future studies of complexes of cryptic taxa in mealybugs. The results obtained in this study suggest that different sets of markers could be chosen according to the questions addressed: (i) for studies in systematics and phylogeny, all genes except ITS2 are useful (ITS2 sequences cannot be aligned when divergence between taxa is high); (ii) for 'Barcoding of Life' as well as population genetics, ITS2 and COI regions (LCO and COI-region2) are the most informative; and (iii) for fine-scale study of complexes of cryptic taxa and delimitation of species, all markers but 28SD2 may prove useful (few differences between closely related taxa appear at 28SD2). In addition, all protocols selected in this study are particularly suitable for use in routine work as they require no gene cloning and make use of rapid, cost-efficient procedures (based on the use of a particularly rapid and cheap PCR enzyme).

Acknowledgements

We thank two anonymous referees who significantly improved the quality of this manuscript. Many thanks to Lenira V. C. Santa-Cecília and Brígida Souza for collecting samples in Brazil and for providing us with a warm welcome at EPAMIG and 'Universidade Federal de Lavras'. We also thank Dr Jon Martin, for lending us *Pseudococcus microadonidum* slides from Gilbert Island and the Seychelles, and our doubtful colleagues Pouaz'rlk for intellectual support. We are also greatly indebted to Paula Ceoto for the help with phylogenetic analyses. This

work was supported by the 'Santé des Plantes et Environnement' Department of INRA.

References

- Abate T, van Huis A, Ampofo JKO, 2000. Pest management strategies in traditional agriculture: an African perspective. *Annu. Rev. Entomol.* 45, 631–659.
- Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P, 2002. The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. *Appl. Environ. Microb.* 68, 3198–3205.
- Baumann P, 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59, 155–189.
- Baumann L, Baumann P, 2005. Cospeciation between the primary endosymbionts of mealybugs and their hosts. *Curr. Microbiol.* 50, 84–87.
- Beardsley J, 1966. Insects of Micronesia. Homoptera: Coccoidea. *Insects Micronesia* 6, 377–460.
- Belshaw R, Quicke DLJ, 1997. A molecular phylogeny of the aphidiinae (Hymenoptera: Braconidae). *Mol. Phylogenet. Evol.* 7, 281–293.
- Ben-Dov Y, 1994. A systematic catalogue of the mealybugs of the world (Insecta: Homoptera: Coccoidea: Pseudococcidae and Putoidae) with data on geographical distribution, host plants, biology and economic importance. Intercept, Andover.
- Beuning L, Murphy P, Wu E, Batchelor T, Morris B, 1999. Molecular-based approach to the differentiation of mealybug (Hemiptera: Pseudococcidae) species. *J. Econ. Entomol.* 92, 463–472.
- Bongiorni S, Cintio O, Prantera G, 1999. The relationship between DNA methylation and chromosome imprinting in the coccid *Planococcus citri*. *Genetics* 151, 1471–1478.
- Bongiorni S, Fiorenzo P, Pippoletti D, Prantera G, 2004. Inverted meiosis and meiotic drive in mealybugs. *Chromosoma* 112, 331–341.
- Bongiorni S, Pasqualini B, Taranta M, Singh PB, Prantera G, 2007. Epigenetic regulation of facultative heterochromatinisation in *Planococcus citri* via the me(3)k9h3-hp1-me(3)k20h4 pathway. *J. Cell Sci.* 120, 1072–1080.
- Cavaliere V, Mazzeo G, Garzia GT, Buonocore E, Russo A, 2008. Identification of *Planococcus ficus* and *Planococcus citri* (Hemiptera: Pseudococcidae) by PCR-RFLP of coi gene. *Zootaxa* 1816, 65–68.
- Charles JG, Cohen D, Walker JTS, Forgie SA, Bell VA, Breen KC, 2006. A review of the ecology of grapevine leafroll associated virus type 3 (GLRaV-3). *N. Z. Plant Prot.* 59, 330–337.
- Cox J, 1989. The mealybug genus *Planococcus* (Homoptera: Pseudococcidae). *Bull. Br. Mus. Nat. Hist. (Entomol.)* 58, 1–78.

- Demontis MA, Ortu S, Cocco A, Lentini A, Migheli Q, 2007. Diagnostic markers for *Planococcus ficus* (Signoret) and *Planococcus citri* (Risso) by random amplification of polymorphic DNA-polymerase chain reaction and species-specific mitochondrial DNA primers. *J. Appl. Entomol.* 131, 59–64.
- Dietrich CH, Rakitov RA, Holmes JL, Black WC, 2001. Phylogeny of the major lineages of Membracoidea (Insecta : Hemiptera : Cicadomorpha) based on 28S rDNA sequences. *Mol. Phylogenet. Evol.* 18, 293–305.
- von Dohlen CD, Kohler S, Alsop ST, McManus WR, 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* 412, 433–436.
- Douglas N, Kruger K, 2008. Transmission efficiency of grapevine leafroll-associated virus 3 (glrav-3) by the mealybugs *Planococcus ficus* and *Pseudococcus longispinus* (Hemiptera : Pseudococcidae). *Eur. J. Plant Pathol.* 122, 207–212.
- Downie DA, Gullan PJ, 2004. Phylogenetic analysis of mealybugs (Hemiptera : Coccoidea : Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher classification. *Syst. Entomol.* 29, 238–259.
- Field LM, Lyko F, Mandrioli M, Prantera G, 2004. DNA methylation in insects. *Insect Mol. Biol.* 13, 109–115.
- Gullan PJ, Downie DA, Steffan SA, 2003. A new pest species of the mealybug genus *Ferrisia* Fullaway (Hemiptera : Pseudococcidae) from the united states. *Ann. Entomol. Soc. Am.* 96, 723–737.
- Hall T, 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hardy NB, 2007. Phylogenetic utility of dynamin and triose phosphate isomerase. *Syst. Entomol.* 32, 396–403.
- Hardy NB, Gullan PJ, Hodgson CJ, 2008. A subfamily-level classification of mealybugs (Hemiptera : Pseudococcidae) based on integrated molecular and morphological data. *Syst. Entomol.* 33, 51–71.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR, 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 313–321.
- Khosla S, Mendiratta G, Brahmachari V, 2006. Genomic imprinting in the mealybugs. *Cytogenet. Genome Res.* 113, 41–52.
- Kuniyuki H, Gioria R, Rezende JAM, Willink CGd, Novo JPS, Yuki VA, 2006. Transmission of the grapevine virus b by the mealybug *Pseudococcus longispinus* Targioni-Tozzetti (Hemiptera:Pseudococcidae) in Brazil. *Summa Phytopathol.* 32, 151–155.
- Meyer JB, Kasdorf GGF, Nel LH, Pietersen G, 2008. Transmission of activated-episomal banana streak ol (badna)virus (bsolv) to cv. Williams banana (musa sp.) by three mealybug species. *Plant Dis.* 92, 1158–1163.
- Millar IM, 2002. Mealybug genera (Hemiptera : Pseudococcidae) of South Africa: identification and review. *Afr. Entomol.* 10, 185–233.
- Miller DR, Rossman AY, 1995. Systematics, biodiversity, and agriculture – systematic analyses of small but important organisms provide crucial information for improvement of agriculture. *Bioscience* 45, 680–686.
- Miller DR, Miller GL, Watson GW, 2002. Invasive species of mealybugs (Hemiptera: Pseudococcidae) and their threat to US agriculture. *Proc. Entomol. Soc. Wash.* 104, 825–836.
- Miller DR, Miller GL, Hodges GS, Davidson JA, 2005. Introduced scale insects (Hemiptera: Coccoidea) of the united states and their impact on us agriculture. *Proc. Entomol. Soc. Wash.* 107, 123–158.
- Nakaune R, Toda S, Mochizuki M, Nakano M, 2008. Identification and characterization of a new vitivirus from grapevine. *Arch. Virol.* 153, 1827–1832.
- Nur U, Brett BLH, 1988. Genotypes affecting the condensation and transmission of heterochromatic-b chromosomes in the mealybug *Pseudococcus affinis*. *Chromosoma* 96, 205–212.
- Pagel M, Meade A, 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 55, 571–581.
- Ronquist F, Huelsenbeck JP, 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rung A, Scheffer SJ, Evans G, Miller D, 2008. Molecular identification of two closely related species of mealybugs of the genus *Planococcus* (Homoptera: Pseudococcidae). *Ann. Entomol. Soc. Am.* 101, 525–532.
- Saccaggi DL, Kruger K, Pietersen G, 2008. A multiplex PCR assay for the simultaneous identification of three mealybug species (Hemiptera: Pseudococcidae). *Bull. Entomol. Res.* 98, 27–33.
- Sforza R, Boudon-Padieu E, Greif C, 2003. New mealybug species vectoring grapevine leafroll associated viruses 1 and 3 (GLRaV-1 and 3). *Eur. J. Plant Pathol.* 109, 975–981.
- Spaulding AW, von Dohlen CD, 1998. Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha). *Mol. Biol. Evol.* 15, 1506–1513.
- Swofford D, 2003. Paup*: phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA.
- Thao M, Gullan PJ, Baumann P, 2002. Secondary (γ -proteobacteria) endosymbionts infect the primary (β -proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl. Environ. Microbiol.* 68, 3190–3197.
- Williams D, 2004. Mealybugs of southern Asia. The Natural History Museum, Southdene.

Williams D, Granara de Willink M, 1992. Mealybugs of central and southern America. CAB International, Wallingford, 635pp.

Williams D, Watson G, 1988. The scale insects of the tropical south pacific region part 2 the mealybugs (Pseudococcidae). C.A.B. International, London, 260pp.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Population codes, origin of samples and number of individuals collected and used for DNA extraction and sequencing.

Table S2. ITS2 haplotypes discovered and the taxon identified and associated with each haplotype. For each haplotype, the populations (see Table S1 for correspondence of codes) in which the haplotype was observed are given.

Table S3. Complete list of samples with corresponding haplotypes: code of individual, GenBank accession numbers for haplotypes, codes of voucher slide-mounted specimens, population code (see Table S1).

Figure S1. DNA matrix for 28S-D2 used for phylogenetic analyses. This alignment differs from the raw alignment because some parts of the raw sequences could not be aligned and were removed.

Figure S2. Comparison of morphological characters in *Pseudococcus longispinus*, *Pseudococcus microadonidum* and *Pseudococcus nr microadonidum*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.