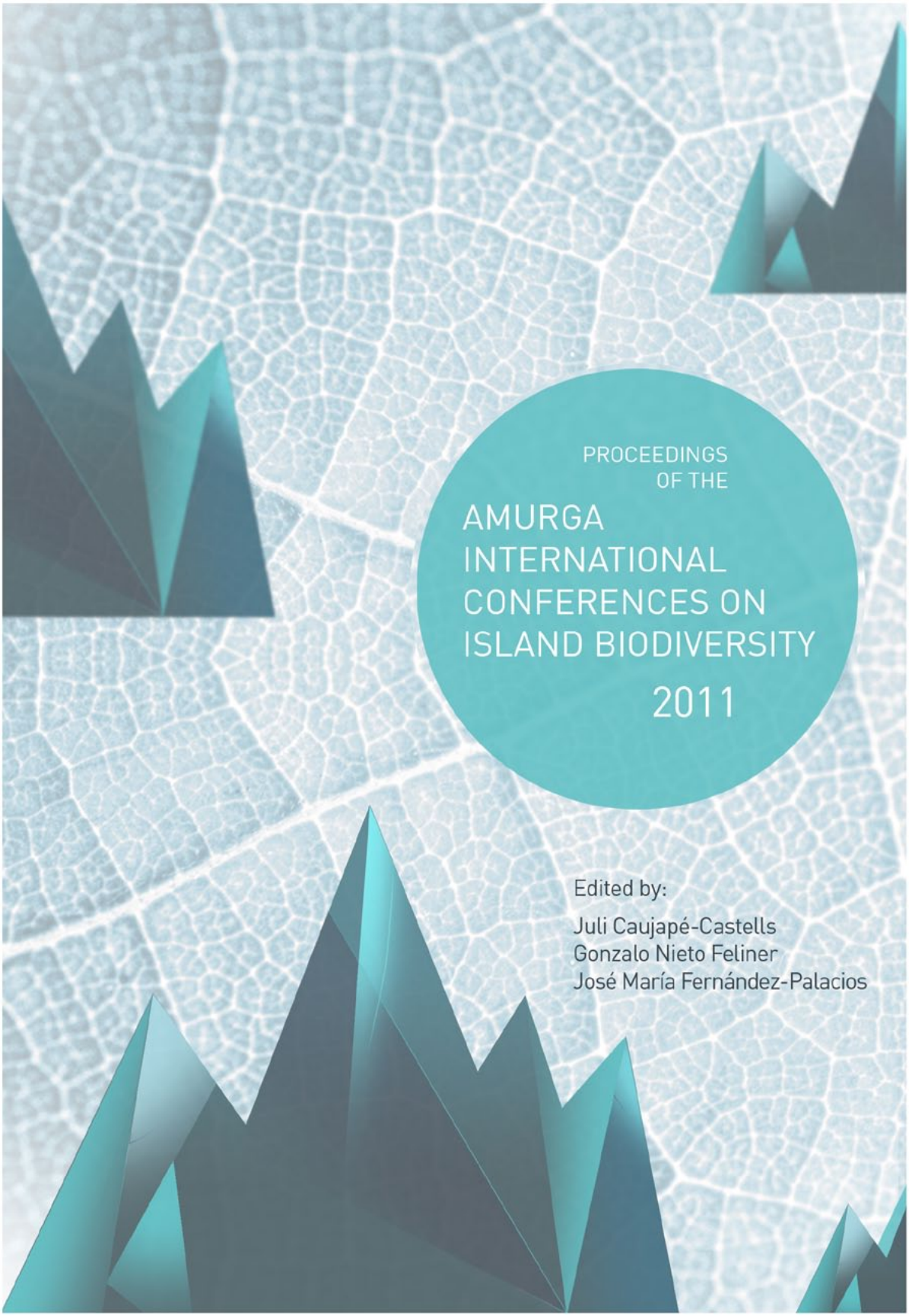
PROCEEDINGS
OF THE
AMURGA
INTERNATIONAL
CONFERENCES ON
ISLAND BIODIVERSITY
2011
Edited by:
Juli Caujapé-Castells
Gonzalo Nieto Feliner
José María Fernández-Palacios

A "Tree of Life" for the Canarian endemic trees: a first step toward the Phylogenetic Diversity of the archipelago's flora.

Ruth Jaén-Molina[1]*, Águedo Marrero[1]*, José María Fernández-Palacios[2], Alain Franc[3], Juli Caujapé- Castells[1]

[1]Jardín Botánico Canario "Viera y Clavijo"-Unidad Asociada CSIC (Cabildo de Gran Canaria, Spain);

[2]Universidad de La Laguna, Santa Cruz de Tenerife (Spain);

[3]INRA-UMR Biodiversité, Gènes et Communautés, Arcachon (France).

E-mails:
ruthjaen@gmail.com
aguedomarrero@gmail.com
jmferpal@gmail.com
julicaujape@gmail.com

Introduction

A DNA barcode is a short DNA sequence taken from standardized portions of the genome, normally used to identify species (Kress et al, 2005; Savolainen et al, 2005). Lately, DNA barcodes have been employed for a number of different purposes in addition to the identification such as the construction of phylogenies (Kress et al, 2009), the assistance to classical taxonomy in the elaboration of censuses of plant biodiversity (Janzen 2009), and the application of these data to conservation strategies. Many studies have shown that DNA barcoding is very efficient in the discrimination of species and in helping to flag possible new species by using one or a combination of several DNA regions (Hebert et al. 2003, 2009). The application of the mitochondrial marker *cytochrome c oxidase I* (COI) has been succesfull and it usually offers a high discrimination rate (>90%) for species identification in a wide range of animal taxa (see references Ran et al, 2010); however, this marker was found inadequate for barcoding plants. A multi-locus approach based on the chloroplast genome was increasingly accepted as an effective strategy for species identification and recognition in land plants (Newmaster et al. 2008; Chase et al. 2007; Kress and Erickson 2007; Fazekas et al. 2008; Lahaye et al. 2008a). In 2009, the Consortium for the Barcoding of Life (CBOL) Plant Working Group (PWG from now on) recommended the two-locus combination *rbcL+matK* as the universal DNA barcode for land plants (CBOL Plant Working Group 2009). The discrimination among highly speciose genera, or cases of recently evolved species are the most complicated challenges when barcoding plants. Compared with the successful identification rates of over 90% in animal, fungi and algae species, the average rate of successful species discrimination in land plants was found to rarely exceed 70%. (e.g. Newmaster et al. 2008; CBOL Plant Working Group 2009). Another issue to overcome in plants is that the discrimination rate varies greatly among different lineages, and it gets still lower the more congeneric species are sampled, or when processes such as hybridization and/or introgression are taking place.

To use the two accepted Angiosperm "barcode" cpDNA sequences (*rbcL* and *matK*) to assess their usefulness to

identify any Canarian endemic tree species should be greatly challenging for various reasons: i) in oceanic islands the incidence of evolutionary processes, namely radiation and hybridization, can be much higher than in other insular or mainland regions and ii) it has been shown that woody plant lineages show consistently lower rates of molecular evolution as compared with herbaceous plant lineages, suggesting the application of DNA barcoding concepts should be more difficult for tree floras than for non-woody floras. Significantly, trees in general are perhaps the organisms with a lower capacity of adaptation to the current global changes, due to their long generation times. Hence, testing the applicability of DNA-barcodes for trees of an oceanic archipelago such as the Canary Islands seems to be of utmost importance in elaborating more accurate censuses that support world-wide conservation efforts.

The project "arBOLcan" has as the main goal to barcode all 46 Canarian endemic tree taxa. The objectives, methodologies, and results of this study ("arBOLcan") will be shared with other international projects under way in Europe and other continents (such as the AR-BOL project in Mesoamerica) because Dr. Juli Caujapé Castells (Jardín Botánico Canario "Viera y Clavijo"-Unidad Asociada CSIC) and Dr. Félix Forest (Jardín Botánico de Kew) are co-coordinators of Tree-BOL in Europe.

We examined more extensively samples of a Macaronesian endemic tree *Heberdenia excelsa* (Sol. in Aiton) Banks ex A. DC. *(sensu lato*) that is distributed in Canary Islands and Madeira. Notable morphological and ecological differences have been found between the forms that grow associated to the laurel trees ("saquitero") and the ones that grow in the termophilous woodlands ("aderno") although, they have not been reflected in the taxonomy of this specie (Jaén Molina et al, 2007). Currently, this taxon is not included in any threatened status but if the molecular markers confirm the morphological divergency found between its populations and betwen the islands, than *Heberdenia excelsa* could be divided into two taxa, both of which should be under a threatened category.

The use of the DNA barcodes sequences to estimate a region's Phylogenetic Diversity (PD, an essential parameter to define *in situ* conservation strategies and priorities) is relatively new. Basically, this methodology uses molecular information to suggest conservationist priorities based on phylogenetic singularities of the territory (Faith et al. 2005, Forest et al. 2007, Schaefer et al. in press), thereby complementing estimates based on taxonomic diversity per unit area (e.g., Reyes-Betancort et al. 2008). To estimate a "tree of life" for the Macaronesian Flora and the PD of the Canarian Flora have been two of the aims of the DNA Bank of the Canarian Flora (JBCVC-CSIC) since its inception (Caujapé-Castells et al. 2006, and see also http://www.bioclimac.com/mbdna/). The DNA bank at the JBCVC-CSIC contains at present more than 7,000 samples that represent ca. 70% of the Canarian endemic flora (i.e., at least one population per taxon, though not every island of occurrence is sampled in all cases). Ongoing strategic sampling aims at presenting the first robust PD estimates for at least 85% of the endemic flora, in two years.

In this communication, based on preliminary results from project arBOLcan and others, we use a multi-locus approach (*rbcL*, *trnH-psbA* and *trnK-rps16*) to detect possibly cases of cryptic species in a focal case of the Myrsinaceae (*Myrsine-Pleiomeris-Heberdenia)*. We also use the *rbcL* and *matK* sequences to (i) to assess the genetic diversity of endemic trees relative to non-tree endemics in the Canarian Flora, (ii) build a preliminary tree of life for the Canarian endemic trees (including the available outgroups from other geographical regions) and provide a baseline data to estimate Phylogenetic Diversity for the Canarian endemic trees by the end of this year. We cross-compared the topologies obtained by Parsimony, ML and Bayesian inferences with the results of the application of the "automatic barcoding" (*Declic*, Franc et al, in prep.) that to our knowledge is a pipeline for data analysis that has not been performed yet with the Canarian Flora.

Material and methods

Sampling

For the focal case (Myrsinaceae) we sampled in total 77 individuals, 49 of them correspond to Heberdenia excelsa (13 from populations in La Palma; 11 in La Gomera; 7 in Tenerife; 5 in el Hierro; 4 in Gran Canaria and 9 from populations in Madeira) plus 11 samples of Pleiomeris canariensis (5 from populations in Gran Canaria and 6 in Tenerife) and 17 accesions of Myrsine africana (8 from different populations of Sao Miguel; 7 from different populations of Faial). Further details of the localities where the samples were collected are included in Table 1. For the Canarian endemic trees our sampling includes 138 accessions, 35 of which are sequences downloaded from GenBank (the majority of the continental representatives used as outgroup samples) that represent 16 Angiosperm families and 2 Gymnosperm families (Cupressaceae and Pinaceae), encompassing 26 genera and 33 taxa of the 46 Canarian endemic trees. With the exception of 11 taxa that are from Madeira, the rest of the ingroup´s taxa are from the Canary Islands. As for outgroups we have samples from Morocco, Cape Verde, Azores and other continental areas, as well. All the Macaronesian endemic taxa are represented by a minimum of one individual (*Arbutus canariensis, Limonium dendroides, Pinus canariensis* and *Sideroxylon mirmulano*) and a maximum of 9-10 individuals (*Dracaena draco, Rubus bollei* and *Sambucus nigra*) (see Table 2 for more details). For Myrsinaceae vouchers were collected for each population and, in most cases of the Canarian trees, herbarium sheets corresponding to the sampled accessions are deposited in the LPA herbarium.

Table 1. Myrsinaceae Samples analyzed as part of the focal case (vouchers in Gran Canaria: Marrero Á & Caujapé-Castells J; in El Hierro: Fdez-Palacios JM & Dámaso Perera López, P; in Tenerife: Fdez-Palacios JM, Rüdiger O & Fernández Lugo S; in La Palma: Fdez-Palacios JM & Romero P; in La Gomera: Fdez-Palacios JM, Romero P, Marrero Á, Naranjo J, Caujapé-Castells J & Jaén-Molina R; in Madeira: Barone Tosco R; in Azores: Fdez-Palacios JM, Caujapé-Castells J & Rodrigues N.)

DNABANK VIALCODE	BOTANICAL TAXONOMY	LOCALITY/ISLANDS	Nº seq. rbcL	Nº seq. psb-trnH	Nº seq trnK-rps16	Nº seq All 3-reg.
6764-6769	*Myrsine africana*	Pico da Vara / Sao Miguel	3	2	3	2
6770-6779	***Myrsine africana***	Lagoa /Sao Miguel	3	2	3	2
6780-6788	***Myrsine africana***	Caldeira / Faial	3	3	2	2
6789-6793	***Myrsine africana***	Capelo/ Faial	2	3	3	2
6794-6799	***Myrsine africana***	Freguesia/Faial	2	3	3	2
3724-3733	***Myrsine africana***	Sierra de Tronqueira /Sao Miguel	2	2	2	2
		TOTAL Individuals/Locality/ Islands =	17/6/2	15/6/2	16/6/2	**12/6/2**
3709, 3711,3712	*Pleiomeris canariensis*	Teno/ Tenerife	3	2	1	1
3715,3716	***Pleiomeris canariensis***	Pista Las Hiedras/ Tenerife	2	2	0	0
3714	***Pleiomeris canariensis***	Mirador Pico del Inglés/ Tenerife	1	1	1	1
7038	***Pleiomeris canariensis***	Albercón de la Virgen/ Gran Canaria	1	1	0	0
7042-7044	***Pleiomeris canariensis***	Bco. de Azuaje (Andén)/ Gran Canaria	2	2	0	0
7048-7050	*Pleiomeris canariensis*	Bco. de Azuaje (Telesf.)/ Gran Canaria	1	3	0	0
3803	*Pleiomeris canariensis*	Bco. de Azuaje (Cornisa)/ Gran Canaria	1	1	1	1
		TOTAL Individuals/Locality/ Islands	**11/7/2**	**12/7/2**	**3/3/2**	**3/3/2**
3754-3758	*Heberdenia excelsa*	Barranco seco / Palma	1	3	2	1
3759-3763	***Heberdenia excelsa***	Ladera Tagoja/ Palma	4	3	4	3
3764-3768	***Heberdenia excelsa***	Los Tilos/ Palma	4	4	3	3
3769-3773	***Heberdenia excelsa***	Tajadre/ Palma	4	2	2	1
3774-3778	***Heberdenia excelsa***	Bajada a Jinama/ Hierro	5	1	2	1
3734-3738	*Heberdenia excelsa*	El Cedro/ Gomera	0	3	2	0
3739-3743	***Heberdenia excelsa***	Chorros de Epina /Gomera	3	4	3	3
3744-3748	***Heberdenia excelsa***	Bco. de Majona/ Gomera	5	2	1	1
3749-3753	***Heberdenia excelsa***	Altos de Hermigua/ Gomera	5	2	4	2

3942-3944	**Heberdenia excelsa**	Valle Gran Rey/ Gomera	2	1	2	1
3642-3646	**Heberdenia excelsa**	El Palmar/ Tenerife	2	2	2	0
3705-3708	**Heberdenia excelsa**	Teno/ Tenerife	2	2	1	1
7045-7046	**Heberdenia excelsa**	Bco.de Azuaje (Andén)/ Gran Canaria	1	1	1	1
7047	**Heberdenia excelsa**	Bco. de los propios/ Gran Canaria	1	1	1	1
3795	**Heberdenia excelsa**	Barranco Oscuro/Gran Canaria	1	0	0	0
3796	**Heberdenia excelsa**	Bco. de Palo Blanco/ Gran Canaria	1	1	0	0
3779-3783	**Heberdenia excelsa**	Laguna de Dona Beija/ Madeira	1	1	1	1
3784-3788	**Heberdenia excelsa**	Lagoa do Vento/ Madeira	4	0	0	0
3789-3793	**Heberdenia excelsa**	Levada de Dos Fontes/ Madeira	4	3	2	2
		TOTAL Individuals/Locality/ Islands	**49/18/6**	**36/17/6**	**35/15/6**	**21/14/6**

Table 2 (below). Macaronesian endemic trees (distribution following Acebes-Ginovés et al. 2004). In black are highlighted the islands were samples have been collected and, with an asterisk the taxa that are not represented, yet.

	ESPECIE	FAMILIA	DISTRIBUCIÓN INSULAR						
			H	P	G	T	C	F	L
M	*Ilex canariensis Poir*	*Aquifoliaceae*	*H*	*P*	**G**	*T*	*C*		
	Ilex perado Ait. ssp. lopezlilloi (G. Kunkel) A. Hansen & Sunding	*Aquifoliaceae*			*G*				
	Ilex perado Ait. ssp. platyphylla (Webb et Berth.) Tutin	*Aquifoliaceae*		*P*	**G**	**T**			
	Phoenix canariensis Chab	*Arecaceae*	*H*	*P*	*G*	*T*	**C**	*F*	*L*
	Viburnum rigidum Vent.	*Caprifoliaceae*	*H*	*P*	*G*	*T*	**C**		
	Sambucus palmensis Link.	*Caprifoliaceae*		*P*	*G*	*T*	*C*		
	Maytenus canariensis (Loes.) Kunk. et Sund	*Celastraceae*	*H*	*P*	*G*	**T**	**C**	*F*	
	*Juniperus turbinata Guss. ssp canariensis **	*Cupressaceae*	*H*	*P*	*G*	*T*	*C*		
M	*Juniperus cedrus Webb & Berthel.*	*Cupressaceae*		*P*	*G*	*T*	*C*		
M	*Dracaena draco L. ssp. draco*	*Dracaenaceae*	*H*	*P*	*G*	*T*	**C**		
	Dracaena tamaranae Marrero Rodr., Almeida-Pérez & González-Martín	*Dracaenaceae*					*C*		
	Arbutus canariensis Veill.	*Ericaceae*	*H*	*P*	*G*	*T*	**C**		
	Erica platycodon (Webb et Berth.) Rivas-Mart. & al.	*Ericaceae*	*H*		*G*	*T*			
M	*Euphorbia mellifera Aiton var. canariensis Boiss.*	*Euphorbiaceae*		*P*	**G**	**T**			

M	*Apollonias barbujana (Cav.) Bornm. ssp. barbujana*	*Lauraceae*	*H*	*P*	**G**	*T*	*C*	*F*	
	Apollonias barbujana (Cav.) Bornm. ssp. ceballosi (Svent.) G. Kunkel	*Lauraceae*			*G*				
M	*Laurus azorica (Seub.) Franco*	*Lauraceae*	*H*	*P*	**G**	*T*	*C*	*F*	*L*
	Laurus novocanariensis (Seub.) Franco	*Lauraceae*			*G*				
M	*Ocotea foetens (Ait.) Berth. et Hook. f.*	*Lauraceae*	*H*	*P*	**G**	*T*	*C*		
M	*Persea indica (L.) Spreng. Rkse.*	*Lauraceae*	*H*	*P*	**G**	*T*	*C*		
M	*Morella faya (Aiton) Wilbur y Morella rivas-martinezii A. Santos & J. Herber*	*Myricaceae*	*H*	*P*	**G**	*T*	*C*	*F*	*L*
	Morella faya (Aiton) Wilbur y Morella rivas-martinezii A. Santos & J. Herber	*Myricaceae*	*H*	*P*	**G**				
M	*Heberdenia excelsa (Ait.) Banks et DC.*	*Myrsinaceae*	**H**	**P**	**G**	**T**	**C**	*F*	
	Pleiomeris canariensis (Willd.) A. DC.	*Myrsinaceae*		*P*	*G*	*T*	*C*		
	Olea europaea L. ssp. cerasiformis (Webb et Berth.) Kunk. et Sund.	*Oleaceae*	*H*	*P*	*G*	*T*	*C*	*F*	*L*
M	*Jasminum odoratissimum L.*	*Oleaceae*	*H*	*P*	*G*	*T*	**C**	*F*	
M	*Picconia excelsa (Ait.) DC.*	*Oleaceae*	*H*	*P*	*G*	*T*	*C*	*F*	
	Pinus canariensis Chr. Sm. ex DC.	*Pinaceae*	*H*	*P*	*G*	*T*	**C**		
	*Limonium dendroides Svent.**	*Plumbaginaceae*			*G*				
M	*Rhamnus glandulosa Aiton*	*Rhamnaceae*		*P*	*G*	*T*	**C**		
	Rhamnus crenulata Aiton	*Rhamnaceae*	*H*	*P*	*G*	*T*	*C*	**F**	*L*
	Rhamnus integrifolia DC.	*Rhamnaceae*				*T*			
M	*Prunus lusitanica L. ssp. hixa [Willd.] Franco*	*Rosaceae*	*H*	*P*	**G**	**T**	*C*		
	Marcetella moquiniana (Webb & Berthel.) Svent.	*Rosaceae*			*G*	*T*	**C**		
	Dendriopoterium menendezii Svent.	*Rosaceae*					*C*		
	Dendriopoterium pulidoi Svent. ex Bramwell	*Rosaceae*					*C*		
	Bencomia brachystachia Svent. ex Nordborg	*Rosaceae*					*C*		
	Bencomia exstipulata Svent.	*Rosaceae*		*P*		*T*			
	Bencomia sphaerocarpa Svent.	*Rosaceae*	*H*						
	*Bencomia caudata (Aiton) Webb & Berthel.**	*Rosaceae*	*H*	*P*		*T*	*C*		
	Rubus palmensis A. Hansen	*Rosaceae*		*P*		*T*	*C*		
	Rubus bollei Focke	*Rosaceae*	*H*	*P*	**G**	*T*	**C**		
M	*Salix canariensis Chr. Sm. ex Link*	*Salicaceae*	*H*	**P**	*G*	*T*	*C*		
M	*Sideroxylon mirmulano Banks ex Lowe*	*Sapotaceae*	*H*	*P*	*G*	*T*	**C**	**F**	
	Tamarix canariensis Willd.	*Tamaricaceae*		*P*	*G*	*T*	**C**	*F*	**L**
M	*Visnea mocanera L. f.*	*Theaceae*	*H*	*P*	**G**	*T*	*C*	**F**	

DNA extraction, amplification and sequencing

DNA extractions were performed from both silica-gel dried or fresh material using the CTAB 2X method. The quality of total extracted DNA was checked on 1% agarose gels, and concentrations were measured in an Eppendorf biophotometer. Subsequently, aliquots of the DNA extracts were deposited in the DNA Bank of the Canarian Flora at the Jardín Botánico Canario "Viera y Clavijo" (JBCVC-Unidad CSIC) (see "DNABank" codes in Table 1 and 2). PCR amplifications were performed for *rbcL*, *psbA-trnH* and *trnK-rps16* (for the focal case) and *rbcL* and *matK* (general case), with primers and conditions recommended by CBOL and in Shaw et al. (2007), for *trnK-rps16*. The PCR products were sent to Macrogen in Korea or to the INRA in France, for sequencing. Sequence alignment was performed using the accessory application CLUSTAL W implemented in BIOEDIT with manual adjustments as needed. The concatenated alignment matrices (3-loci for the focal case and 2-loci for the general case) were obtained using GENEIOUS. Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI) analyses were conducted for each marker separately and combined. Parsimony analyses were carried out in PAUP* v4.010b, ML analyses with PhyML and Bayesian analyses were conducted with MrBayes v3.0b4. The obtained topologies were crossed-compared with the output file of pipeline *declic* (Franc et al, in prep.).

Results and discussion

1. Focal case:

   A) Diagnostic characters

   We explored the matrix containing the three-locus aligned sequences to detect diagnostic characters (one or a combination of nucleotides in a position of aligned sequences that is unique to all individuals available for a specific taxa) that will discriminate the species of Myrsinaceae and also to determine whether or not the morphological divergence found, so far, in *Heberdenia excelsa* ("saquitero" vs. "aderno") was supported by molecular data. From the three locus analyzed, the one that found more diagnostic characters for each taxa was *trnK-rps16*, the second was *psbA-trnH* and, interestingly the third was *rbcL* that only found one unique nucleotide for *Myrsine africana*. When we estimated, the total number of variable characters (diagnostic and polymorphic) detected by each locus, we found out that *rps16* detected twice of diagnostic characters than *psbA* and 20 diagnostic more than *rbcL*. We only examined the *three-locus* alignment for DNA barcode positions between populations, because was the only that presented enough variability to show some intraspecific polymorphism. No diagnostic characters were found to distinguish between all populations examined. One character distinguished one sequence of Caldeira do Faial, from all the sequences of the same and different populations of *Myrsine africana*.

One character was shared by all samples from the three different populations of Sao Miguel. Four characters are exclusive of one sample of *Pleiomeris* from Teno and two characters were found that distinguished one sequence (from Tenerife-Pico del Inglés) from the remaining *Pleiomeris* samples (Gran Canaria). Within *Heberdenia*, a group of samples from La Gomera (Chorros de Epina) shared a nucleotide with samples from La Palma (Los Tilos and Tajadre). One sample from Gran Canaria has an exclusive character. The remaining samples of the different populations even some of Madeira were 100% identical.

B) Phylogenetic reconstructions

MP (Paup), ML (PhyML) and BI (MrBayes) analyses were carried out to evaluate whether or not *rbcL* ("official barcode") recovered the three Myrsinaceae species as monophyletic and whether or not the "non-official barcodes regions" recovered the same clades as *rbcL*.

The consensus bayesian tree obtained for all the 77 *rbcL* sequences (427chs.) grouped each genus in three main clades, one for each taxa. For *psbA-trnH* (63 taxa/ 418 chs.) again, all sequences for each taxa were clustered together in three main clades but now with a higher resolution within *Myrsine* (Faial´s samples are in a clade separated from samples of Sao Miguel) and within *Heberdenia* (although populations of the different Canary Islands and Madeira are mixed). There is no discrimination between samples of *Pleiomeris* from Gran Canaria and Tenerife. The Bayesian consensus tree for *rps16-trnk* (58 taxa/ 589 chs.) showed a less resolved topology between *Myrsine*. *Pleiomeris* from Pico del Inglés (T) was slightly different than the others from Tenerife and Gran Canaria.

The resolution within *Heberdenia* is better, although populations of different islands are not completely segregated from each other. The three-locus combination (37 taxa and 1,427 characters) provided the most fully resolved phylogeny with the higher support values recovered (Fig. 1). The Bayesian consensus tree topology resulted, in practically the same clades as for *rps*16 alone, for *Heberdenia and Myrsine* with a mix of populations in them. There are identical sequences for *Heberdenia excelsa* samples from different islands and populations**,** so a clear discrimination between populations and islands was not possible and neither between the two morphological forms ("saquitero" and "aderno"). It was the position of Pleiomeris what, notably distinguished these results with respect to those from other analyses. This discrepancy is not completely unexpected because in the past, sometimes species of *Heberdenia* have been named after *Ardisia (Heberdenia excelsa=Ardisia teneriffae*) or species of *Pleiomeris* after *Myrsine* (*Pleiomeris canariensis=Myrsine canariensis*). The morphological and molecular variability of the Macaronesian Myrsinaceae need further analyses, in order to bring more light to the taxonomy of the group. One important aspect to consider is the difficulties to collect fruits and flowers of *"Heberdenia excelsa"* that make it very difficult a precise study of the presence of the different forms (saquitero and aderno) in each island.

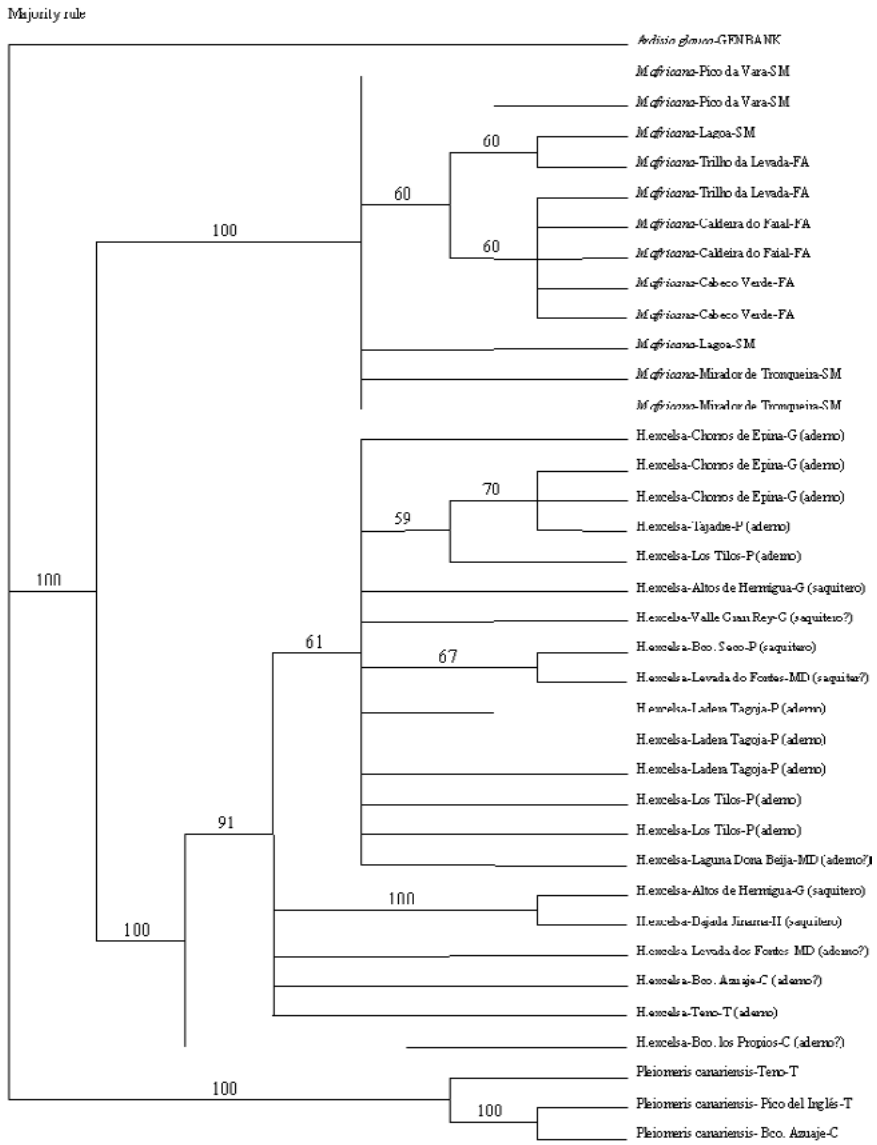

Figure.1. Bayesian consensus tree for 36 taxa of Macaronesian Myrsinaceae based on three locus combined analyses (rbcL+ psbA-trnH + rps16-trnK) with each genus coloured differently (Heberdenia excelsa-blue, Pleiomeris canariensis-green, Myrsine africana-orange)

C) Declic with Myrsinaceae data

Most of barcoding or phylogeny approaches rely on multiple alignments. *Declic* is an alignment free method. Its first step is to build a pairwise edit distance matrix (the minimum number of edits needed to transform one string into the other, with

the allowable edit operations being, insertion, deletion, or substitution of a single character), the one implemented here is Levenshtein distance. Second step in pipeline *declic* is to transform genetic distances between observed sequences into estimates of branch length using an evolution model. Third step in the pipeline is to translate the updated edit distance matrix into a graph. A graph is a list of nodes, some of them being linked by an edge. A connex component (cc) in a graph is a maximum set of nodes such that any two of them can be linked by a path. A clique in a graph is a set of nodes such that any pair of nodes is linked by an edge. The graph is a set of cliques. If taxonomy would be perfect, a cc would be a clique: any pair in the clique is linked by an edge. Each clique is a taxon. Pipeline *declic* allows a link between phylogenies, barcodes, and resolutive and variable markers. No alignments, no block selection, no phylogenies, no BLAST, are needed, just graph on edit distances!! We run declic on a *.fasta file of 63 *psbA-trnH* sequences representing the three genus of Myrsinaceae. In Table 3 are included the composition of connex components with a gap = 10. The results showed that *declic* was able to find cliques for *Heberdenia excelsa* and *Myrsine africana* (with very few exceptions) with *Pleiomeris canariensis* in a clear position, too.

CONNEX COMPONENTS	
gap = 0 field = Genus	
1 ->	*Heberdenia : 36 / 46 Pleiomeris : 2 / 10*
2->	*Myrsine : 13 / 14*
3 ->	*Pleiomeris : 8 / 10*
4 ->	*Heberdenia : 3 / 46*
5 ->	*Heberdenia : 1 / 46*
6 ->	*Heberdenia : 1 / 46*
7 ->	*Heberdenia : 1 / 46*
8 ->	*Heberdenia : 1 / 46*
10 ->	*Heberdenia : 1 / 46*
11 ->	*Heberdenia : 1 / 46*
12 ->	*Myrsine : 1 / 14*
9 ->	*Heberdenia : 1 / 46*

Table 3. *Declic* connex components for *trnH-psbA*

2. General case

Our sampling includes more sequences for *rbcL* (138) than for *matK* (85), because we encountered technical problems not only in he amplification of this region but also in the obtention of high quality sequences. We analyzed three matrices: one per locus and a two-loci combination. We did look for diagnostic characters only in conflictive cases as *Ilex* (data not shown).

A) Phylogenetic reconstructions

For each dataset MP, ML and BI analyses were performed. The MP consensus tree for rbcL distinguished well between families (color shaped) and also between genera, with the exceptions of Lauraceae (although the genera that conform this family are easily recognizable at the field). For genus represented by a larger number of individuals, the assignment of the individuals to the correct specie was 100% with the exception of Ilex (spread across different parts of the tree). The only two Gymnosperm families represented (Cupressaceae and Pinaceae) are placed together. The limited taxon sampling and the lack of species per genus in some cases, may be the reason of not much congruence in the recovered of some clades, as the one that grouped Dracaena and Phoenix. The clades recovered by parsimony analyses for *matK* were very similar as the ones obtained for *rbcL*, with the exceptions of Ilex (these time all sequences are grouped together) and within Lauraceae, species of Apollonias and Laurus are distinguished.

Within some genera as *Rubus* and *Morella* the resolution was higher than for *rbcL*. The two-loci combination (85 taxa/ 599 chs.) analyses resulted in a better resolved tree (Fig.2), where all families are clearly differenciated and more clades were supported with higher values.

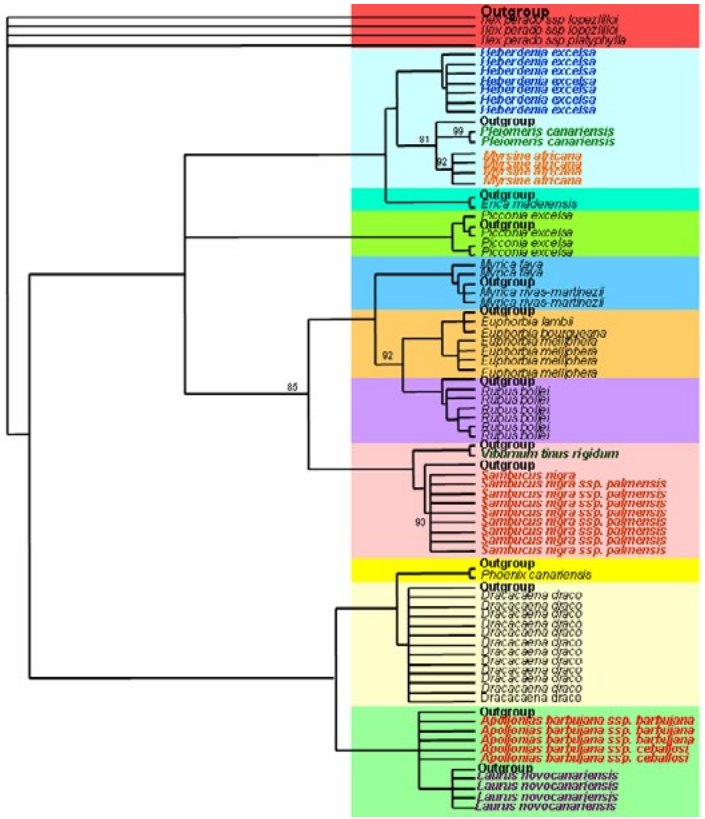

Figure.2. Bayesian consensus tree for 85 Macaronesian endemic taxa and based on two locus combined analyses (*rbcL*+ *matK*) with families color shaded and genera within the no- monotypic families colored differently.

B) Declic with Canarian tree data

We run declic (gap=10) on a *.fasta file for the 138 *rbcL* sequences representing the Canarian endemic trees. The results showed (Fig.3) that *declic* was able to find cliques for almost all the genera, with very few exceptions, as for the genera within Lauraceae and Myrsinaceae that were not well distinguished.

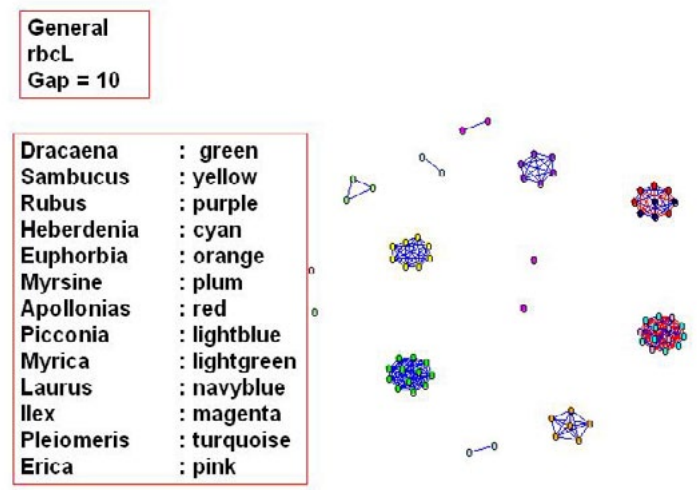

Figure.3. Decli´s graph output showing cliques for the majority of the genera (see legend box for family colors, one genera- one color)

3. Phylogenetic Diversity and "Tree of Life" for the Canarian flora.

With the data generated for the arBOLcan and Garajonay´s projects (Jaén Molina et al, 2010), we began to have some extensive data sets (especially for *rbcL*) of the Angiosperm endemic Canarian flora, that should allow us to construct a preliminary "Tree of Life" for the Canarian Flora (Fig.4) and to estimate the PD of the islands that are more represented (Gomera and Gran Canaria), once some strategic additional sampling is completed.

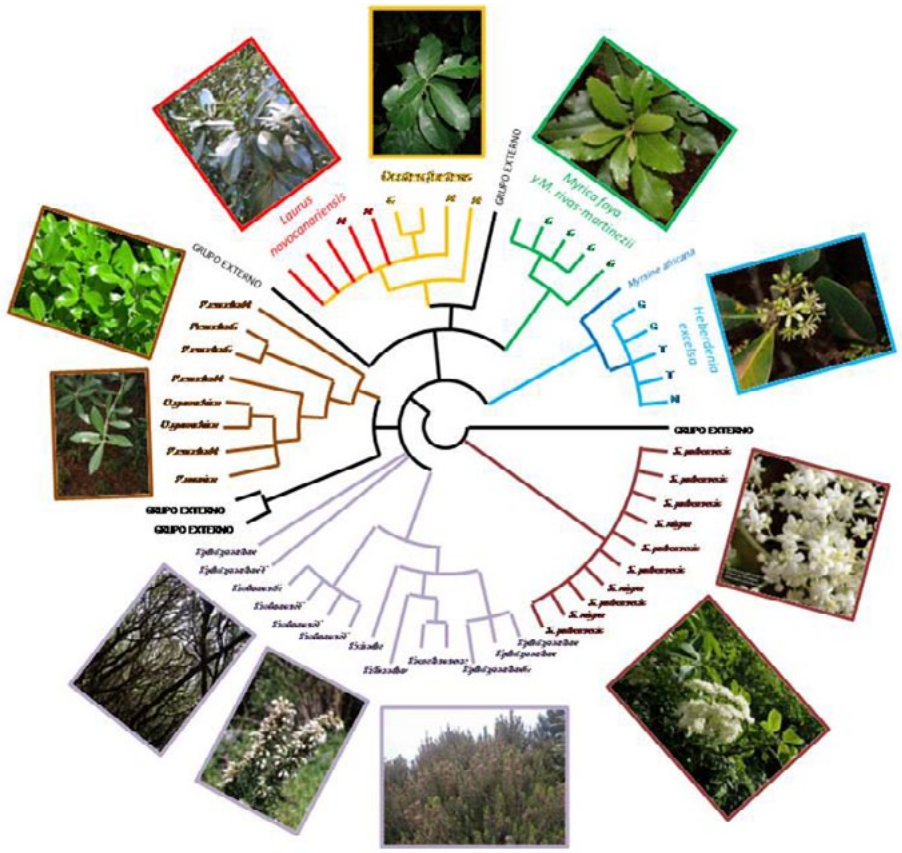

Figure 4. "Tree of Life" for some Macaronesian endemic trees based on *rbcL* sequences.

Acknowledgements

We thank to our extensive network of collaborators of other Spanish and International Research Centres (Botanic Gardens, National Parks, Museums, and Universities). To the team of Alain Franc for laboratory assistance (INRA, France). To the the Canarian Research Agency and Autonomous Organisation of National Parks for funding the "arBOLcan" and "Garajonay" Projects (C200801000239 and 129/2006), respectively.

References

Caujapé-Castells J, Jaén Molina R, Cabrera García N (2006) El banco de ADN de la flora canaria: creación, progresos y líneas futuras de desarrollo. *Botánica Macaronésica* 26: 3-16.

Chase MW, Cowan RS, Hollingsworth PM, van der Berg C, Madriñán S, et al. (2007) A proposal for standardised protocol to barcode all land plants. *Taxon* 56: 295-299.

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc. Nat Acad. Sci. USA* 106: 12794-12797.

Faith DP, Williams KJ (2005) How Large-scale DNA Barcoding Programs Can Boost Biodiversity Conservation Planning: Linking Phylogenetic Diversity (PD) Analyses to the Barcode of Life Database (BoLD). In: Australian Entomological Society's 36th AGM and Scientific Conference/7th Invertebrate Biodiversity and Conservation Conference/ Australian Systematics Society. Canberra, Australia, 4-9 December 2005, pp. 83-84.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M and Barrett SC (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3 (7), E2802

Forest F, Grenyer R, Rouget M, Davies J, Cowling RM (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445: 757-760.

Jaén-Molina R, Caujapé-Castells J, Fernández-Palacios JM, Marrero Á (2007)¿Pueden las Técnicas Moleculares ayudar a la Taxonomía Clásica? el caso de Heberdenia excelsa (Myrsinaceae) en Canarias. III Congreso de Biología de Conservación de Plantas, 25-28 de Septiembre. Puerto de La Cruz (Tenerife).

Jaén-Molina R, Marrero Á, Reyes-Betancort JA, Naranjo Suárez J, Santos Guerra A, Caujapé-Castells J (2010) La Flora Endémica Del Parque Nacional De Garajonay Bajo La Perspectiva Molecular: El Código De Barras Molecular Como Herramienta Taxonómica. In: L. Ramírez Sanz & B. Asensio Nistal (eds.). Proyectos de investigación en parques nacionales: 2006-2000. Naturaleza y Parques Nacionales. Serie investigación en la red. Ministerio de Medio Ambiente y Medio Rural y Marino. Madrid.

Janzen, DH and 44 authors (2009) 'Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity', *Molecular Ecology Resources* 9(s1): 1-26.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identification through DNA barcodes. Proc. Roy. Soc. Lond. B: 270:313-321.

Hebert PDN, DeWaard JR, Landry JF (2009) DNA barcodes for 1/1000 of the animal kingdom. - Biology Letters, doi: 10.1098/rsbl.2009.0848

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non coding trnH-psbA spacer region. PloSONE.

Kress WJ, Erickson DL, Jones,FA, Swenson NG, Perez, R, Sanjur O and Bermingham E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc. Natl. Acad. Sci. U.S.A. 106 (44): 18621-18626.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions with species of recent origin in Myristicaceae. Mol Ecol Notes 8: 480-490.

Ran JH, Wang PP, Zhao HJ, Wang XQ (2010) A Test of Seven Candidate Barcode Regions from the Plastome in Picea (Pinaceae). Journal of Integrative Plant Biology, 52 (12): 1109–1126

Reyes-Betancort, JA, Santos Guerra, A, Guma IR, Humphries CJ, Carine MA (2008) Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. Anales del Jardín Botánico de Madrid 65: 25-45

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans, Ser B360: 1850–1811. (doi:10.1098/rstb.2005.1730).

Shaw J, Lickey EB, Schilling EE, Small RI (2007) Comparison Of Whole Chloroplast Genome Sequences To Choose Noncoding Regions For Phylogenetic Studies In Angiosperms: The Tortoise And The Hare III. American Journal of Botany 94(3): 275–288.

Vane-Wright, RI, Humphries CJ, Williams PH (1991) What to Protect? Systematics and the agony of choice. Biological Conservation 55: 235.

Ruth Jaén-Molina

Ruth Jaén-Molina has been combining different activities as a researcher of the Department of Molecular Biodiversity in addition to managing the DNA Bank of the Canarian Flora at the Jardín Botánico Canario "Viera y Clavijo"-Unidad Asociada CSIC (since May 2005). Her research interests are in molecular systematics, molecular phylogeny, molecular taxonomy, biogeography, and diversification of island plants, particularly from Macaronesia. The majority of the projects in which she participates have a multidisciplinary approach and the conservation of the species as a priority aim. Some examples of the emblematic groups that she has been studying are *Matthiola R.BR, Parolinia Webb, Dracaena Vand. ex L., Minuartia L., Heberdenia Banks ex A.DC., Myrsine L., Echium L.*, etc. The main goals of her research are to use molecular tools (DNA sequencing but also ISSR and microsatellites) to better understand the origin, evolution and biogeographical patterns of mainly, but not only, the species of the Canary Islands. For the past few years, she has been involved in projects to explore and develop the potential of DNA barcoding (*matK* & *rbcL*) as a practical tool for species identification as complementary to the classical taxonomy (Projects "Garajonay", "Tree Bol" and "arBOLcan").

Águedo Marrero Rodríguez

Águedo Marrero Rodríguez is curator of the Herbarium of the JBCVC (LPA). He devoted his first research period to cytogenetics, generating contributions on the Macaronesian Sideritis. Later on, he moved his research focus to the areas of nomenclature, taxonomy, systematics and biogeography of the endemic Canarian and Macaronesian floras. He has developed an intense botanical prospection task across the entire Canarian Archipelago, discovering several hundred new populations of threatened or endangered taxa, re-discovering taxa that were feared extinct, and describing about 20 new taxa. He has contributed to numerous projects on Conservation Biology and Natural Spaces, and is an advisor for numerous initiatives related to recovery Plans of threatened species. He has published ca. 45 scientific papers and participated in many R+D projects funded by national and international agencies. He has been a most decisive collaborator in the DNA Bank of the JBCVC since its inception in 2005.