Molecular characterization of *Echium portosanctensis* and a reappraisal of Macaronesian *Echium* phylogeny

Silva, N.[1], Gouveia, M.[1,2], Jardim, R.[1], Jesus, J.[1,2] & Menezes de Sequeira, M.[1,2]

[1] Madeira Botanical Group (GBM), Faculty of Life Sciences, University of Madeira, Funchal, Portugal
[2] InBio, Research Network in Biodiversity and Evolutionary Biology, CIBIO-Azores, Portugal

INTRODUCTION

The genus *Echium* L. (Boraginaceae) is represented in the Macaronesian Islands (Azores, Madeira, Selvages, Canary and Cape Verde) by 35 endemic taxa (species and subspecies). Phylogenetic studies for these taxa were carried out in the past by several authors (Böhle *et al*., 1996; García-Maroto *et al*., 2009; Romeiras *et al*., 2011), suggesting that the Macaronesia taxa derived from a single colonization event. This event occurred in the Canary Islands, through the invasion of a continental herbaceous ancestor species. A secondary colonization event led to the dispersion of the taxa from the Canary Islands to Cape Verde and to Madeira Archipelagos. The specialization and diversification of the taxa might have occurred concurrently with the colonization of the Islands, giving rise to a high number of endemic taxa (Böhle *et al*. 1996). Following this hypothesis *Echium webbi* Coincy, an endemic species of the Canary Islands, is pointed out as the ancestral species of the genus in the Madeira Archipelago and its radiation will have originated three endemic woody taxa: *E. nervosum* W.T.Aiton (Fig. 1), *E. candicans* L.f (Fig. 2) and *E. portosanctensis* J.A.Carvalho, Pontes, Bat.-Marques & R.Jardim (Fig. 3).

Since *E. portosanctensis* was described recently (Carvalho *et al.*, 2010) as endemic to Porto Santo, an island older than the island of Madeira, this study aims to obtain molecular markers from both nuclear and chloroplast genomes in order to reappraise the Macaronesian *Echium* phylogeny.

Fig. 1. *Echium nervosum*, Madeira Island.

Fig. 2. *Echium candicans* L.f , Madeira Island.

Fig. 3. *Echium portosanctensis*, Porto Santo Island.

MATERIAL AND METHODS

Genomic DNA of *E. portosanctensis* (up to three individuals) was extracted from leaves dried in silica gel (ca. 30 mg) by the Doyle & Doyle (1990) procedure with minor changes. Polymerase chain reactions (PCR) were performed by Advantage *Taq* DNA Polimerase (Clontech) to amplify five molecular markers using primers published previously: ITS region (O´Kane *et al*., 1996; White *et al*., 1990), intergenic spacer $trnT^{UGU}$-$trnL^{UAA}$, $trnL^{UAA}$ intron, intergenic spacer $trnL^{UAA}$-$trnF^{GAA}$ (Taberlet et al., 1991) and *matK* (Ford *et al*., 2009; Dunning & Savolainen, 2010). The PCR products with the expected sizes were purified by using SpinPrep PCR clean-up kit (Novagen) and sequenced by Macrogen Spain. The sequences obtained were compared to homologous sequences present in GenBank. Sequences were aligned using ClustalW (Thompson *et al*., 1994) as implemented in MEGA version X (Kumar *et al*., 2018) sequence alignment editor and subjected to visual inspection when necessary. True evolutionary relationships may be obscured or skewed in DNA sequence data sets if sites have become saturated by multiple substitutions (Swofford *et al*., 1996). To test for saturation, observed pairwise proportions of transitions and transversions were plotted against sequence divergence and calculated using DAMBE version 5 (Xia, 2013). The combined dataset of ITS1, $trnL^{UAA}$ intron, intergenic spacer $trnT^{UGU}$-$trnL^{UAA}$, intergenic spacer $trnL^{UAA}$-$trnF^{GAA}$) fragments revealed some saturation, although the high proportion of invariable sites. Our fragments were compared to homologous sequences presented at GeneBank to construct the trees using Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian Analysis.

The Maximum Likelihood (ML) analysis was performed using MEGA X (Kumar *et al*., 2018), based on the General Time Reversible (GTR) model (Rodríguez *et al*., 1990). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1000)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 45.70% sites). Before doing ML analysis in MEGA X we search for best models of evolution using JModeltest 2.1.2 (Darriba *et al*., 2012). Using the Akaiké information criterion (Posada & Buckley, 2004), the best model was the GTR +I+G. The –LnL= 3.526.58 was very similar to that obtained with mega X, P-inv and gamma parameter were slightly different. Maximum Parsimony analysis a was performed using MEGA X (Kumar *et al*., 2018). The best tree have a length = 209. The consistency index was (0.680556), the retention index (0.878307). The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The estimates of bootstraps values involved 500 replicates in both analysis. The Bayesian analysis was implemented using Mr-Bayes version 3.1.2 (Huelsenbeck & Ronquist, 2001), which calculates Bayesian posterior probabilities using a Metropolis-coupled, Markov chain Monte Carlo (MC-MCMC) analysis. Bayesian analysis was conducted with random starting trees, four MCMC chains (one cold, three heated), run 2.0×10^6 generations, and sampled every 100 generations. The burn-in value was 2000. All analysis involved 34 nucleotide sequences and a total of 1628 positions in the final dataset.

Fig. 4. *Echium hypertropicum* Webb, Santiago Island (Cabo Verde).

Fig. 5. *Echium vulcanorum Chev.* Fogo Island, Cabo Verde.

Fig. 6. *Echium webbii* (http://commons.hortipedia.com/images/6/6c/Echium_webbii_PDB.jpg).

```
E.portosanctensis  .....ATGCCTTGCSCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCCCCCCGGCGCGAGAA
E.nervosum         .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.candicans        .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.webbii           .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.hierrense        .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.simplex          .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.virescens        .....ATGCCTTGCGCTCGACTGTCTCCTGACGTTTCC............CAAACAAACCC--CCGGCGCGAGAA
E.hypertropicum    .....ATACCTTGCGCTCGACTGTCTCCTCACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
E.vulcanorum       .....ATACCTTGCGCTCGACTGTCTCCTCACGTTGCC............CAAACAAACCC--CCGGCGCGAGAA
                          *       *                *      *                        **
```

Figure 7. Partial alignment of the ITS1 sequences with some Macaronesian taxa from Madeira Island, Canary and Cape Verde islands, showing heterozygosity (S = C or G) and an indel unique to *E. portosanctensis* (highlight in yellow). Asterisks indicate nucleotide polymorphisms. (GenBank accessions numbers: *E. nervosum* L43256, *E. candicans* L43192, *E. webbii* EU048854, *E. simplex* EU048851, *E. hierrense* L43216, *E. virescens* EU048850, *E. hypertropicum* EU048858, *E. vulcanorum* L43304).

```
E. portosantensis   TCAA-TTCAGAGAAACC  16    CCAAAATATATTTTTTTT-GAAAAAAAAAAGAAGAGT  270
E. nervosum         TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
E. candicans        TCAAATTCAGAGAAACC  60    CCAAAATATATTTTTTTT-GAAA-AAAAAAGAAGAGT  313
E. webbii           TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
E. hierrense        TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
E. simplex          TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
E. virescens        TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAA--AAAAAAGAAGAGT  313
E. hypertropicum    TCAAATTCAGAGAAACC  53    CCAAAATCTATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
E. vulcanorum       TCAAATTCAGAGAAACC  53    CCAAAATATATTTTTTTTTGAAA-AAAAAAGAAGAGT  307
```

Figure 8. Partial alignment of the sequences obtained for the $trnL^{UAA}$ intron marker of *E. portosanctensis* with some endemic *Echium* taxa. Nucleotide polymorphisms of *E. portosanctensis* are highlighted in yellow. (GenBank accessions numbers: *E. nervosum* L43254, *E. candicans* EU433607, *E. webbii* L43306, *E. hierrense* L43214, *E. simplex* L33359, *E. virescens* EU433601, *E. hypertropicum* L43230, *E. vulcanorum* L43302).

RESULTS AND DISCUSSION

Sequence lengths for the noncoding DNA regions either from nuclear or chloroplast genomes (ITS, $trnL^{UAA}$ intron, *matK*, intergenic spacer $trnT^{UGU}$-$trnL^{UAA}$, intergenic spacer $trnL^{UAA}$-$trnF^{GAA}$) of *E. portosanctensis* were similar to other Macaronesian *Echium* species. The alignment of all molecular markers revealed polymorphic positions for ITS1 (Fig. 7). To a lesser extent nucleotide variability was detected for $trnL^{UAA}$ intron (Fig. 8), whereas no genetic diversity was observed for the remainder markers.

All markers revealed many invariant sites, but those that span are often transversions. Trees obtained with Maximum Likelihood, Maximum Parsimony and Bayesian analysis gave similar topologies (Fig. 9). In all analyses, the phylogenetic trees suggest that the Macaronesian *Echium* derives from a single continental herbaceous ancestor that initially dispersed to the Canary Islands, supporting the hypothesis advanced by Böhle *et al.* (1996). A secondary radiation event led to the rapid dispersion of the taxa in Macaronesia. Our results also revealed that Cape Verde endemic taxa always appeared in a defined cluster. Moreover, the results do not seem to support the idea that *E. webbii* originated indeed the taxa in the archipelago of Madeira, as suggested by Böhle *et al.* (1996). Also, the hypothesis of the origin of Madeira *Echium* species in Porto Santo must not be discarded as Porto Santo Island is the oldest in the archipelago of Madeira.

Although morphologically the Madeiran and Canary endemic taxa are distinct, due to the lack of genetic divergence between them, these species are frequently grouped in the same cluster, suggesting a recent specialization and diversification event.

The phylogenetic history of *Echium* remains undefined in Macaronesia. Many questions persist unanswered that only a more detailed study can enlighten. A larger sample of individuals would be necessary as well as more discriminant molecular markers.

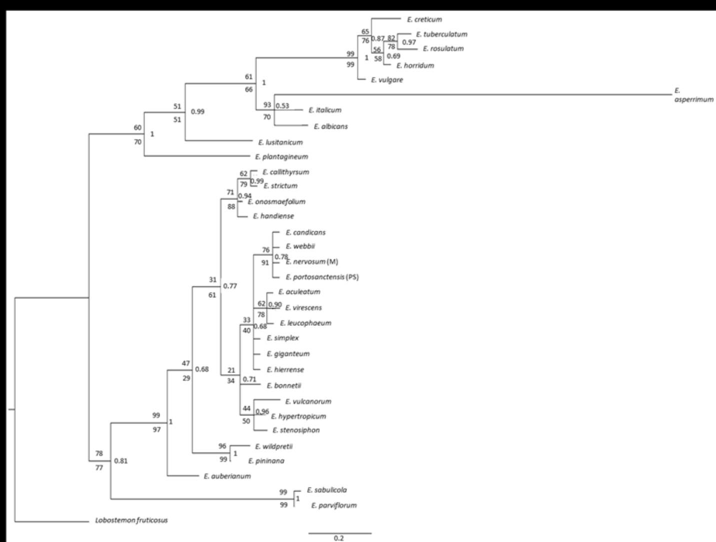

Figure 9. Tree derived from Bayesian analysis of combined of ITS1, $trnL^{UAA}$ intron, intergenic spacer $trnT^{UGU}$-$trnL^{UAA}$, intergenic spacer $trnL^{UAA}$-$trnF^{GAA}$) fragments. Average posterior probabilities are shown in the right side of nodes. The analysis involved 34 taxa, and 1628 positions. The tree was rooted using *Lobostomon fruticosus*. The tree derived from ML analysis, obtained by MEGA X and with a model of sequence evolution GTR + I + G and the trees derived from MP analysis, shows a similar pattern, except for being in general less well resolved (for the major groups the bootstrap values for ML and MP are shown, respectively, above and below the nodes).

REFERENCES

Böhle UR, Hilger HH & Martin WF (1996). Island colonization and evolution of the insular woody habit in Echium L. (Boraginaceae). PNAS 93: 11740-11745.
Carvalho JA, Pontes T, Batista-Marques MI & Jardim R (2010). A new species of Echium (Boraginaceae) from the island of Porto Santo (Madeira Archipelago). Anales Jard. Bot. Madrid 67(2): 87-96.
Darriba D, Taboada GL, Doallo R & Posada D (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature Meth. 9:772.
Doyle J & Doyle J (1990). Isolation of plant DNA from fresh tissue.Focus 12: 13-15.
Dunning LT & Savolainen V (2010). Broad-scale amplification of matK for DNA barcoding plants, a technical note. Bot. J. Linn. Soc. 164: 1-9.
Ford CS, Ayres KL, Toomey N, Haider N et al. (2009). Selection of candidate coding DNA barcoding regions for use on land plants. Bot. J. Linn. Soc. 159: 1-10.
García-Maroto F, Mañas-Férnandez A, Garrido-Cárdenas JA, Alonso DL et al (2009). D6-Desaturase sequence evidence for explosive Pliocene radiations within the adaptive radiation of Macaronesian Echium (Boraginaceae). Mol. Phylogenetics Evol. 52: 563-574.
Huelsenbeck JP & Ronquist F (2001). MR-BAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.
Kumar S, Stecher G, Li M, Knyaz C & Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms.Mol. Biol. Evol. 35:1547-1549.
O´Kane SL, Schall BA & Al-Shehbaz IA (1996). The origins of Arabidopsis suecica (Brassicaceae) as indicated by nuclear rDNA sequences. Syst. Bot 21: 559-556.
Posada D & Buckley T (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches Over Likelihood Ratio Tests. Syst. Biol., 53: 793–808.
Rodrıíguez F, Oliver JF., Marıín A & Medina JR (1990). The generar stochastic model of nucleotide substitutions. J.Theor.Biol. 142: 485–501.
Romeiras MM, Paulo OS, Duarte MC, Pina-Martins F, Coutrim HC, Carine MA & Pais MS (2011).Origin and diversification of the genus Echium (Boraginaceae) in the Cape Verde archipelago. Taxon 60: 1375-1385.
Swofford DL, Olsen GJ, Waddell PJ & Hillis DM (1996). Phylogenetic Inference. In: Hiillis DM, Moritz D, and Mable BK, editors, Molecular Systematics, pp. 407-514. Sinauer Associates, Sunderland, Massachusetts.
Taberlet P, Gielly L, Pauyou G & Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol 17: 1105-1109.
Thompson JD, Higgins DG & Gibson TJ (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
White T J, Bruns T, Lee S & Taylor J (1990). Amplification and direct sequencing of fungal ribossomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ & White TJ (eds.). PCR protocols. pp. 315-322. New York: Academic Press.
Xia X (2013). DAMBE5: A comprehensive software package for data analysis in in molecular biology and evolution. Mol. Biol. Evol. 30: 1720–1728.

ACKNOWLEDMENTS

This work was funded by: FEDER funds through the Operational Programme for Competitiveness Factors – COMPETE; National Funds through FCT - Foundation for Science and Technology under the UID/BIA/50027/2013 and POCI-01-0145-FEDER-006821; MACFLOR (MAC/4.6d/190) Project under the Cooperation Programme INTERREG MAC 2014-2020.

floramac 2018 Madeira Cabo Verde Selvagens Açores CANÁRIAS

UNIVERSIDADE da MADEIRA

GBM Grupo de Botânica da Madeira

macflor Atlas Macaronésico de Biología Reproductiva. Marcadores de vigor y de amenaza con aplicaciones a la conservación. MAC/4.6d/190

CIBIO

InBIO Research Network Biodiversity and Evolutionary Biology