

Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus *Himantoglossum* s.l. (Orchidaceae)

Gábor Sramkó^{1,2,*}, Attila Molnár V.², Julie A. Hawkins³ and Richard M. Bateman⁴

¹MTA-ELTE-MTM Ecology Research Group, Pázmány P. sétány 1/C, Budapest, 1117, Hungary, ²Department of Botany, University of Debrecen, Egyetem ter 1., Debrecen, 4032, Hungary, ³School of Biological Sciences, University of Reading, Reading RG6 6AS, UK and ⁴Jodrell Laboratory, Royal Botanical Gardens Kew, Richmond, Surrey TW9 3DS, UK

* For correspondence. E-mail sramko.gabor@science.unideb.hu

Received: 27 February 2014 Returned for revision: 1 May 2014 Accepted: 18 July 2014 Published electronically: 7 October 2014

- **Background and Aims** Lizard orchids of the genus *Himantoglossum* include many of Eurasia's most spectacular orchids, producing substantial spikes of showy flowers. However, until recently the genus had received only limited, and entirely traditional, systematic study. The aim of the current work was to provide a more robust molecular phylogeny in order to better understand the evolutionary relationships among species of particular conservation concern.
- **Methods** All putative species of *Himantoglossum* s.l. were sampled across its geographical range. A large subsample of the 153 populations studied contributed to an initial survey of nuclear ribosomal internal transcribed spacer (nrITS) ribotypes. Smaller subsets were then sequenced for four plastid regions and the first intron of the low-copy-number nuclear gene *LEAFY*. Rooted using *Stenienella* as outgroup, phylogenetic trees were generated using parsimony and Bayesian methods from each of the three datasets, supplemented with a ribotype network.
- **Key Results** The resulting trees collectively determined the order of branching of the early divergent taxa as *Himantoglossum comperianum* > *H. robertianum* group > *H. formosum*, events that also involved significant morphological divergence. Relaxed molecular clock dating suggested that these divergences preceded the Pleistocene glaciations (the origin of the *H. robertianum* group may have coincided with the Messinian salinity crisis) and occurred in Asia Minor and/or the Caucasus. Among more controversial taxa of the *H. hircinum-jankae* clade, which are only subtly morphologically divergent, topological resolution was poorer and topological incongruence between datasets was consequently greater.
- **Conclusions** Plastid sequence divergence is broadly consistent with prior, morphologically circumscribed taxa and indicates a division between *H. hircinum-adriaticum* to the west of the Carpathians and *H. jankae-caprinum* (plus local endemics) to the east, a distinction also suggested by nrITS ribotypes. *LEAFY* phylogenies are less congruent with prior taxonomic arrangements and include one likely example of paralogy. *Himantoglossum metlesicsianum* fully merits its IUCN Endangered status. Potentially significant genetic variation was detected within *Stenienella satyrioides*, *H. robertianum* and *H. hircinum*. However, confident circumscription of the more derived species of *Himantoglossum* s.s., including local endemics of hybrid origin, must await future morphometric and population-genetic analyses.

Key words: Dating, evolution, *Himantoglossum* s.l., hybridization, incomplete lineage sorting, internal transcribed spacer, nrITS, *LEAFY*, molecular phylogeny, orchid, Orchidaceae, molecular phylogeography, topological incongruence.

INTRODUCTION

Lizard orchids of the genus *Himantoglossum* W.D.J. Koch s.l. (i.e. as circumscribed by Delforge, 1999; Bateman *et al.*, 2003; Table 1) include many of Eurasia's most spectacular orchids, producing substantial spikes of showy flowers characterized by large, complex labella that in the majority of species are highly contorted and follow a remarkable floral ontogeny (Bateman *et al.*, 2013a). The flowers of all species indulge in food deceit, typically attracting a wide range of pollinators – most commonly solitary bees (Claessens and Kleynen, 2011). The group has long attracted interest from ecologists, particularly since an early study (Good, 1936) revealed the apparent utility of the type species, *Himantoglossum hircinum*, as a biogeographical indicator of subtle changes in European climate (see also Carey, 1998, 1999; Pfeifer *et al.*, 2006, 2009; Bateman *et al.*, 2013a). Kropf

and Renner (2008) used this species in a detailed study of geitonogamy within inflorescences, while students of pollination biology have been attracted by the unusual biochemistry of its scent (Kaiser, 1993). Moreover, an unusually large proportion of the taxa within the genus have achieved high conservation status in many European countries; the geographically widespread species tend to be very local across much of their ranges (and to be allopatric rather than sympatric), whereas several putative species are local endemics and therefore, by definition, rare. Indeed, three species – *H. metlesicsianum*, *H. adriaticum* and *H. jankae* – are presently listed on the Habitats Directive of the European Union, partly reflecting threats from habitat degradation and/or harvesting of tubers for saleg production. Contrasts in the extent and location of the species' distributions also encourage biogeographical study, not least because closely related species appear largely allopatric.

TABLE 1. Pilot comparison of DNA regions used in this study. For the purposes of this exercise, *H. robertianum* is treated as the outgroup (O) relative to two ingroup members (Is), *H. hircinum* and *H. caprinum* – two closely related species from the *H. hircinum* aggregate. Gaps are treated as missing data

DNA-region	Genomic location	Length range	Aligned length	Variability in terms of			
				Absolute number		Percentage	
				O–Is	Is only	O–Is	Is only
<i>accD-psaI</i>	Plastid	938–941	951	14	3	1.47 %	0.32 %
<i>atpF-atpH</i>	Plastid	278–280	281	2	0	0.71 %	0 %
<i>rps16</i>	Plastid	818–863	864	18	0	2.08 %	0 %
<i>trnH-psbA</i>	Plastid	812–819	826	7	0	0.85 %	0 %
<i>trnL-rpl32-ndhF</i>	Plastid	1098–1285	1368	40	2	2.92 %	0.15 %
<i>ycf1</i>	Plastid	1429–1472	1472	23	1	1.56 %	0.07 %
nrITS	Nucleus	638–639	640	54	6	8.44 %	0.94 %
LFY	Nucleus	2139–2245	2325	170	28	7.31 %	1.20 %

However, until recently, *Himantoglossum s.l.* had received limited and entirely traditional systematic study (e.g. Nelson, 1968; Sundermann, 1973; Delforge, 1999). The first molecular phylogenetic study to represent the genus generated parsimony trees from sequences derived from the nuclear ribosomal internal transcribed spacer (nrITS) region and included four species of *Himantoglossum s.l.* (Pridgeon et al., 1997). The genus was shown to be the earliest-divergent of four major lineages within a clade that also contains *Ophrys* and *Serapias* + *Anacamptis s.l.* and is delimited by a reduction in typical chromosome number from $2n = 42$ to $2n = 36$. Unsurprisingly, *H. robertianum* (at the time widely referred to as *Barlia robertiana*) was shown to be sister to, and substantially divergent from, three species of *Himantoglossum*: *H. hircinum*, *H. adriaticum* and *H. ‘caprinum’*. Bateman et al. (2003) performed an analysis using a similar approach but based on twice as many species of tribe Orchideae. Their study showed that (1) *H. calcaratum* was closely related to *H. ‘caprinum’*, (2) the Canarian endemic *H. metlesicsianum* was sister to, but moderately divergent from, *H. robertianum*, and (3) *H. comperianum* (formerly *Comperia comperiana*) was sister to, and strongly divergent from, all other *Himantoglossum* species. These results were fortunate, as in the interim Delforge (1999) had used traditional morphological taxonomic arguments to transfer ‘*Comperia*’ *comperiana*, ‘*Barlia*’ *robertiana* and ‘*B.*’ *metlesicsiana* to an expanded genus *Himantoglossum s.l.*

The nrITS study of Bateman et al. (2003) also usefully identified a tentative sister genus to *Himantoglossum s.l.* in *Steveniella satyrioides*, presently the only species in a genus of much smaller and highly morphologically divergent plants with a distribution centred on Turkey. Both morphological comparison and previous molecular phylogenetic studies (e.g. Bateman et al., 2003) demonstrate that this species constitutes an ideal outgroup for the present study. As species delimitation within the expanded genus *Himantoglossum* remains controversial, and some species have also recently been subject to purely nomenclatural changes (Molnár V. et al., 2012; Sramkó et al., 2012), we have summarized our current taxonomy, authorities, clade names and geographical distributions in Fig. 1, together with a synthesis of chromosome counts that indicates the expected dominant base number of $2n = 36$ (Bateman et al., 2013a). It is especially noteworthy that, setting aside highly localized endemics, species of *Himantoglossum s.s.* align geographically west to east from the

Atlantic coasts of Europe and North Africa to the Caucasus: *H. hircinum*, *H. adriaticum*, the *H. jankae* group (formerly *H. ‘caprinum’*, and including *H. ‘robustissimum’*), *H. caprinum s.s.* (formerly named *H. affine*; see Discussion) and *H. formosum*. The eastern Mediterranean *H. jankae* group also contains four taxonomically controversial local endemics whose possible origins through recent hybridity or more ancient gene exchange in glacial refugia have been discussed periodically (e.g. Alibertis and Alibertis, 1989; Delforge, 1999; Vakhrameeva and Tatarenko, 2008). Listed from west to east they are *H. calcaratum*, *H. × samariense*, *H. montis-tauri* and *H. galilaeum*.

Here we provide a more robust molecular phylogeny and evolutionary context for *Himantoglossum s.l.* by (1) sampling multiple individuals in multiple populations of all putative species, and (2) adding to the biparentally inherited nrITS a wider range of molecular markers to include low-copy-number nuclear genes and maternally inherited plastid regions. In addition to the molecular studies described here, a wide range of morphometric characters were sampled from the same individuals; these results will be reported elsewhere, in a paper that will also discuss in detail the complex taxonomic and nomenclatural issues.

In the present paper we ask the following questions: (1) Can the substantial molecular divergence of the Canarian endemic *H. metlesicsianum* from *H. robertianum* reported by Bateman et al. (2003) be confirmed? (2) What is the phylogenetic position of the enigmatic Caucasian endemic *H. formosum*, which is morphologically intermediate between the *H. robertianum* and *H. hircinum*-*H. jankae* groups? (3) What are the phylogenetic relationships among the subtly morphologically distinct, taxonomically problematic members of the *H. hircinum*-*H. jankae* group? (4) Is there evidence for the stabilized hybrid origins for *H. adriaticum*, *H. montis-tauri* and/or *H. × samariense* postulated by some previous workers (e.g. Delforge, 1999)? (5) What are the most likely approximate dates of key events in the evolution of *Himantoglossum s.l.*? (6) In which geographical regions are these events most likely to have taken place?

MATERIALS AND METHODS

Plant material and DNA extraction

Samples for DNA extraction, either portions of leaves or entire flowers, were collected in the field. Some were placed in silica

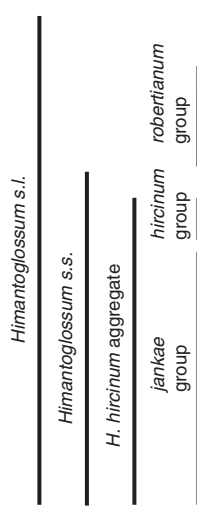
Clades	Taxon/authority	Geographic distribution	2n
	<i>Steeniella satyrioides</i> (Sprengel) Schlechter [outgroup]	Turkey, Crimea, Caucasus, Iran	?38 ¹
	<i>Himantoglossum comperianum</i> (Steven) P. Delforge	Asia Minor, Caucasus, Crimea, Near East	?30 ²
	[= <i>Comperia comperiana</i> (Steven) Ascherson & Graebner]		
	<i>H. robertianum</i> (Loiseleur) P. Delforge	Circum-Mediterranean (near-coastal)	36, ?38 ³
	[= <i>Barlia robertiana</i> (Loiseleur) Greuter		
	<i>H. metlesicsianum</i> (W.P. Teschner) P. Delforge	Tenerife (Canary Islands)	36 ⁴
	[= <i>Barlia metlesicsiana</i> W.P. Teschner]		
	<i>H. formosum</i> (Steven) K. Koch	E Caucasus	
	<i>H. hircinum</i> (L.) Sprengel	N Africa, Iberia, NW Europe, S Italy/Sicily	36, ?24, ?18 ³
	<i>H. adriaticum</i> H. Baumann	N Italy, Slovenia, Hungary, Austria, Slovakia	36 ^{3,5}
	<i>H. galilaeum</i> Schiffman	Israel, S Jordan	NA
	<i>H. jankae</i> Somlyay, Kreutz & Óvári	Slovakia, Hungary, Balkans, N Turkey	NA
	[= <i>H. 'caprinum'</i> auct. mult. non (M.Bieb.) Spreng., between 1897 and 2012 ⁶]		
	incl. <i>H. caprinum</i> ssp. <i>robustissimum</i> Kreutz	N Turkey	NA
	<i>H. calcaratum</i> (G. Beck) Schlechter	Bosnia-Herzegovina, Serbia	NA
	<i>H. montis-tauri</i> Kreutz & W. Lueders	SW Turkey	NA
	<i>H. x samariense</i> C. & A. Alibertis	Crete, S Greece	NA
	<i>H. caprinum</i> (M. Bieb.) Sprengel	Greece, Turkey, Syria, Lebanon, Crimea	NA
	[= <i>H. affine</i> (Boissier) Schlechter ⁶]		

FIG. 1. Details of species sampled: nomenclature, distribution and chromosome numbers, together with names of major clades (left). More detailed taxonomic reviews were presented by Delforge (1999) and Bateman *et al.* (2003). ¹Sundermann and von der Bank, 1977; ²Ströhlein and Sundermann, 1972; ³D'Emérico *et al.*, 1992; ⁴Bernardos *et al.*, 2006; ⁵Löve, 1976; ⁶Sramkó *et al.*, 2012 (chromosome numbers reviewed by Bateman *et al.*, 2003, 2013a). NA, not available; ?, uncertain data.

gel and stored at room temperature, others were placed in 96 % ethanol and stored at 4 °C until extraction. Our sampling covered almost the entire distribution of the genus, omitting only (due to political unrest) an eastward projection into the Kurdish regions of Iran and Iraq (Fig. 2). Altogether, we analysed samples from 130 populations of *Himantoglossum* plus six populations of the outgroup, *S. satyrioides* (Supplementary Data Table S1). The highly endangered nature of several species allowed only preparation of mounted flowers as vouchers: one flower was removed from 30–40 % of the distance from the base to the apex of the inflorescence, dissected and mounted onto double-sided adhesive tape. Together with scaled images of representative sampled plants, these flowers [the majority stored in the herbarium at the University of Debrecen (DE) and the remainder in R.M.B.'s private collection] constitute vouchers of the studied species (Supplementary Data Table S1). All samples were examined for nrITS variability, whereas only a representative subset of 35 populations (Fig. 2) provided samples for inclusion in the phylogeny reconstruction analyses.

DNA extraction followed a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). Between 1 and 30 mg of dried leaves were thoroughly ground in liquid nitrogen and then resuspended in lysis buffer (2 % CTAB, 20 mM EDTA, pH 8, 100 mM Tris–HCl, pH 9, and 1.4 mM NaCl). After incubation at 65 °C for at least 60 min, samples were centrifuged at 20 000 g for 5 min. The supernatant was then treated with an equal volume of chloroform:isoamyl alcohol (24:1), thoroughly mixed in a rotary shaker, and subsequently centrifuged for 5 min at 20 000 g. This procedure was repeated once. The DNA was precipitated from the supernatant by adding an equal volume of isopropanol and 0.08 volume of

7.5 M ammonium acetate, followed by storage at –20 °C for at least 1 h. Subsequent centrifugation at 20 000 g for 15 min yielded a DNA pellet that was washed twice with 70 % ethanol, dried at room temperature and finally redissolved in 40 µl of 10 mM Tris (pH 8.0).

Choice of genic regions for phylogeny reconstruction

When using DNA regions for species-level molecular phylogenetics, relatively rapidly evolving regions (such as introns or intergenic spacers) are needed; they are sampled from the plastid, the nuclear genome or, ideally, both (Zimmer and Wen, 2012). The nrITS region is one of the most powerful DNA regions in molecular phylogenetic work (Álvarez and Wendel, 2003) but it should be interpreted with caution, especially given the extensive evidence of paralogy (Bailey *et al.*, 2003; Nieto-Feliner and Rosselló, 2007). When divergent nrITS copies are found within an individual, International Union of Pure and Applied Chemistry (IUPAC) ambiguity symbols (Cornish-Bowden, 1985) can be used to describe polymorphism. Although this coding significantly decreases the resolving power of the nrITS, it is the most efficient way to analyse directly sequenced nrITS data in the presence of extensive intra-individual paralogy.

The usual laboratory-based solution for the interpretation of nrITS paralogy is cloning to separate diverged nrITS haplotypes, commonly termed ribotypes. The molecular mechanism acting to homogenize nrITS arrays, termed 'concerted evolution' (Baldwin *et al.*, 1995), is seldom complete (Álvarez and Wendel, 2003). Instead, as a by-product of these molecular mechanisms, many intermediate ribotypes between parental

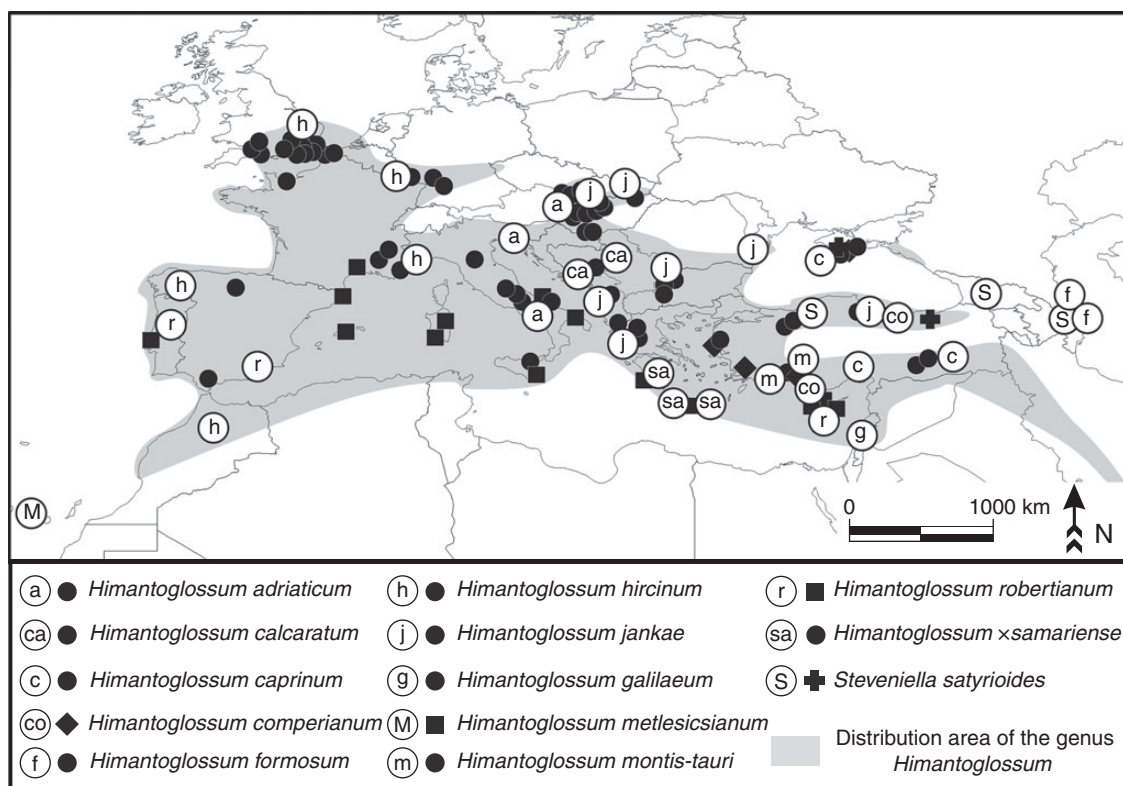


FIG. 2. Sampling sites of *Himantoglossum* s.l. and *Steveniella* (outgroup) for this study. Open circles represent populations included in the phylogenetic analyses, taxa being indicated by the enclosed letters. Solid symbols represent additional sample sites used to assess variability in the nrITS region.

copies can arise due to interactions of divergent copies (i.e. various degrees of recombination). Therefore, cloning usually recovers multiple ribotypes, among which we should find the orthologous copies but also intra-individual recombinants that weaken phylogenetic inference between closely related taxa. LaJeunesse and Pinzón (2007) argued that the more frequent copies are likely to be orthologues and so should be prioritized when selecting clones for inclusion in the resulting matrix.

Low-copy-number nuclear regions are an increasingly popular alternative to nrITS (Hughes *et al.*, 2006; Calonje *et al.*, 2009; Zimmer and Wen, 2012). In European orchids, one of the most promising low-copy-number nuclear regions is *LEAFY* (*LFY*, equivalent to *FLORICULA* in *Antirrhinum*), which regulates flower development. The ~2.5-kb-long gene consists of three exons separated by two introns, and was found to be present as a single copy in the diploid orchid species *Orchis simia* using Southern blotting (Montieri *et al.*, 2004). The phylogenetic utility of the first intron at low taxonomic levels was demonstrated by Schlüter *et al.* (2007), who also provided universal primers to amplify this intron by PCR in European orchids.

Hybridization often plays an important role in plant speciation (Rieseberg, 1997). At the molecular level, it is most readily detected by comparing data from biparentally inherited nuclear markers with those from a uniparentally inherited organelle, in plants most importantly plastid markers. Initial screening of plastid regions commonly used in plant phylogenetics allowed selection of the most variable regions for further analysis. Six plastid regions (Table 1) were tested for sequence variability in

single samples of three species (*H. robertianum*, *H. hircinum* and *H. caprinum*) using the primers described in the original papers: *accD-psaI* (Small *et al.*, 1998), *atpF-atpH* (Ki-Joong in Lahaye *et al.*, 2008), gene *rps16* (Oxelman *et al.*, 1997), *trnH-psbA* (Sang *et al.*, 1997), *trnL-ndhF* including gene *rpl32* (Shaw *et al.*, 2007) and gene *ycf1* (Neubig *et al.*, 2009).

Topologies between contrastingly inherited regions were then compared to check for incongruence as the result of recent gene flow via hybridization (Wendel and Doyle, 1998).

PCR amplification and cloning of nrITS

The nrITS of one to four individuals (1.3 ± 0.5 individuals, mean \pm s.d.) randomly chosen from each population was amplified using the plant-specific ITS1A (5'-GACGTCGCGAGAAG TCCA-3') primer and the universal primer ITS4 (White *et al.*, 1990). The PCR mixture contained 0.1 volume of $10 \times$ Taq buffer with $(\text{NH}_4)_2\text{SO}_4$ (Thermo Scientific, USA), 200 μM of each dNTP (Thermo Scientific, USA), 2 mM MgCl_2 , 0.2 μM of each primer, 1.25 U Taq DNA polymerase (Thermo Scientific, USA) and $\sim 5 \text{ ng } \mu\text{l}^{-1}$ genomic DNA extract. Amplifications were performed on a GeneAmp PCR System 2400 (Perkin Elmer), programmed for a denaturation step at 94°C for 4 min 30 s, followed by 33 cycles of denaturation for 30 s at 94°C , annealing for 30 s at 51°C and extension for 30 s at 72°C , the extension time being increased by 1 s in each successive cycle; thermal cycling was ended by a final extension for 7 min at 72°C .

For direct sequencing, the PCR products were directly submitted to a commercially available service provided by Macrogen Inc. (South Korea) using ITS1A as sequencing primer. Cloning was performed on a subset of 14 samples, selecting one individual per population that had displayed overlapping peaks indicative of additive polymorphic sites (APSs) during direct sequencing. The purified nrITS PCR products were ligated and transformed into the pGEM-T Easy Vector System II (Promega Corp., USA) following the manufacturer's instructions. The plant nrITS region was amplified and cycle-sequenced from five or six clones (5.8 ± 0.4) per individual under the PCR and sequencing conditions described above.

Study of nrITS variability

Direct nrITS sequences were checked for the presence of APSs and such sites were coded using IUPAC ambiguity symbols (Cornish-Bowden, 1985). APSs were identified by comparing the relative signal heights of the primary and secondary peaks; when the secondary peak was considerably higher than those observed in the neighbouring bases, the site was labelled as an APS (note the different usage of the same term by Fuertes Aguilar and Nieto Feliner, 2003). Paralogous copies were classified as either length-identical or length-polymorphic. Because length-polymorphic nrITS copies were impossible to read to full length, length mutations were differentiated based on the location of the site at which readability collapsed. Once categorized, these length-polymorphic sequences were excluded from further analyses. In addition, further nrITS analyses were confined to samples of *Himantoglossum* s.s., as the phylogenetic positions of the earlier-divergent lineages were already clear. The remaining direct nrITS sequences that lacked length polymorphisms were collapsed into ribotypes by using Collapse v1.2 (David Posada, University of Vigo). Relationships among the resulting ribotypes were inferred using TCS v1.21 (Clement et al., 2000) under default parameters. The ribotype groups were then mapped to reveal any geographical patterns, an analytical procedure that was later repeated with the cloned nrITS sequences.

PCR amplification of plastid regions

The three most informative plastid regions (*accD-psaI*, *trnH-psbA* and *trnL-rpl32-ndhF*) were sequenced for a much wider representative set of accessions (Fig. 2). All regions were amplified under the same touchdown PCR regime: denaturation at 94 °C for 3 min, followed by 16 cycles of denaturation for 30 s at 94 °C, annealing for 30 s starting at 56 °C and decreasing by 0.5 °C in each cycle, extension for 90 s at 72 °C, followed by 20 cycles with the same settings but using the same 48 °C temperature for annealing; thermal cycling concluded with a final extension for 7 min at 72 °C. The PCR mixture contained the same amounts of ingredients for all regions: 0.1 volume of 10× reaction buffer (Thermo Scientific, USA), 200 μM of each dNTP (Thermo Scientific, USA), 2 mM MgCl₂, 1 mg μl⁻¹ bovine serum albumin, 0.3 μM of each primer, 0.03 U Dream Taq DNA polymerase (Thermo Scientific, USA) and 1 μl of unquantified genomic DNA extract. Successfully amplified samples were submitted to Macrogen Inc. (South Korea) for purification and sequencing. Sequencing was performed in

both forward and reverse reads using the original primers as sequencing primers, though in the case of *trnL-ndhF* the whole region was sequenced in four reads, using the two internal primers (rpl32-F and rpl32-R) of Shaw et al. (2007).

PCR amplification and cloning of the LFY region

The first intron of the *LFY/FLO* gene was targeted for this study as it reputedly occurs as a single copy in European orchids (Montieri et al., 2004). We used the exon-anchoring primers (E1Cf, E2Gr) of Schlüter et al. (2007) to amplify this region in a hot-start PCR using JumpStart REDAccuTaq LA DNA polymerase (Sigma-Aldrich, USA). The PCR mixture and regime followed the manufacturer's recommendations. Successfully amplified samples were submitted to Macrogen Inc. (South Korea) for purification and direct sequencing. Based on the partial reads, we designed new internal sequencing primers to obtain a whole read of this ~2.3 kb region. Altogether, two new internal primers, one forward and one reverse, spanned the whole region: LFY-If2 (5'-CTGGG CCATTAAATTGGAT-3') and LFY-Ir3 (5'-ACTCTCTCT AAGAAATCAACG-3').

As we found ~30 APSs in the central part of the *LFY* intron of nine accessions (Supplementary Data Table S1), additional internal primers that specifically amplified the variable region were designed [LFY-INTf (5'-ACACCCTAGACGCACT-3') and LFY-INTR3 (5'-GCCGTGGAATATTGGAGTC-3')], and PCR products of the variable region were cloned and sequenced. Cloning of the partial intron of *LFY* followed the same procedure as described above for nrITS. Because subsequent comparison of the samples (see Results) indicated the origin of APSs in intra-individual paralogy, we randomly selected a single clone to represent each individual in further analyses.

Phylogeny reconstruction

For each region, sequences were aligned by eye using BioEdit v7.1.3 (Hall, 1999) and phylogenetic tree reconstruction used two contrasting search philosophies: heuristic search under the maximum parsimony (MP) criterion as implemented in PAUP* v4.0b10 (Swofford, 2003), and Bayesian phylogeny estimation as implemented in MrBayes v3.2.1 (Ronquist et al., 2012). One nrITS region that was especially rich in APSs at phylogenetically informative nucleotide positions was excluded from the phylogenetic analysis. A subset of 35 accessions of *Himantoglossum* and three accessions of *Steveniella*, designed to best represent the geographical distribution of each taxon (Fig. 2), was analysed separately for the plastid DNA regions selected (*accD-psaI*, *trnH-psbA* and *trnL-ndhF*) and for the nuclear gene *LFY*. Plastid sequences of the three sequenced regions were checked for combinability via an incongruence length difference (ILD) test in PAUP* v4.0b10, implemented as a partition homogeneity test with 100 replicates. All analyses were run on Bioportal (Kumar et al., 2009).

For the MP heuristic search, we coded each unambiguous indel in the alignments as a single mutation event following the 'simple' indel-coding approach *sensu* Simmons and Ochoterena (2000). The heuristic search settings were: starting trees were obtained using 1000 random sequence replicates, holding ten trees at each step during stepwise addition with the

steepest descent option not in effect, and applying a TBR branch-swapping algorithm. Branch-swapping was conducted with the MulTrees option on and MaxBrLen plus AccTran applied, and without any topological or tree-buffer constraints. For the Bayesian estimate we first identified the most appropriate evolutionary model using MrModeltest v2.3 (Nylander, 2004), and implemented that model in the MrBayes analysis. We ran two independent chains for 25 million generations, sampling every 2500th generation. The resulting log-likelihood values of the runs were checked for convergence using Tracer v1.5 (Drummond and Rambaut, 2007). Summary trees were calculated after discarding the first 25 % of trees as burn-in.

Molecular dating

In order to provide an estimate of divergence times within the *Stenieniella*–*Himantoglossum* clade, we applied a Bayesian relaxed molecular clock approach using the uncorrelated log-normal distribution model (Drummond et al., 2006) as implemented in BEAST v1.7.4 (Drummond and Rambaut, 2007). The dataset used for this analysis was based on the nrITS dataset for Orchidinae compiled by Bateman et al. (2003), though the original sequences of *Himantoglossum* s.l. were replaced by sequences generated during the current study. In the case of *Himantoglossum* s.s., where nrITS showed intra-individual polymorphism (see below), only the two dominant cloned ribotypes were selected to represent the aggregate (following the recommendations of Lajeunesse and Pinzón, 2007). This dataset spans taxonomically the whole of European Orchidinae plus representatives of their putative sister group Habenariinae (possibly paraphyletic, but the two subtribes together forming the undoubtedly monophyletic tribe Orchideae), together with single species of Satyriinae and Disinae serving as outgroups.

As no unequivocal fossil remains exist for Orchideae, we used the secondary calibration points inferred by Gustafsson et al. (2010). As secondary dating points, we applied a normal prior age of 24 ± 6.3 Ma to the split between Habenariinae and Orchidinae, plus 17 ± 5.65 Ma to the putative split between *Orchis* and *Platanthera* inferred by Inda et al. (2012) (but see Bateman, 2012a). Two separate Bayesian runs were conducted for 50 million generations, sampling every 50 000th generation. All runs had the same prior settings: the ingroup (i.e. Orchideae) was kept monophyletic during the run, the GTR + G + I model was specified, a log-normal relaxed clock (uncorrelated) was applied, and the Yule process was set as the speciation model. The resulting tree files and log files were combined, and the latter were analysed using Tracer v1.5 (Drummond and Rambaut, 2007) to check for reliably large effective sample sizes. The tree judged to be most credible on external criteria (notably morphometrics) was visualized as a maximum clade credibility tree with mean node heights using TreeAnnotator, after conservatively discarding the first 20 % as burn-in. The substitution rate of the nrITS region in angiosperms was reviewed by Kay et al. (2006), who reported an average rate of the molecular clock to be 4.13 substitutions/site/year $\times 10^{-9}$ in herbaceous plants (extremes reported were 1.72 and 8.34 substitutions/site/year $\times 10^{-9}$), offering the opportunity to compare the mutation rate obtained from our study with those found in other plant groups. We therefore used the inferred mean rate of substitutions

estimated using BEAST under the parameter meanRate, expressed as the total number of substitutions per site divided by the total period of time represented by the tree (because time in the tree is measured in millions of years, the rates are given as unit substitutions/site/Myr).

Ancestral geographical area reconstruction analysis

The plastid dataset was used for ancestral area reconstruction in a Bayesian DIVA (dispersal–vicariance) analysis (Nylander et al., 2008) as implemented in RASP v2.1 (Yu et al., 2010). Six geographical regions (Fig. 9B) were defined for the samples: Mediterranean basin (A), western sub-Mediterranean (B), central sub-Mediterranean (C), Balkan sub-Mediterranean (D), eastern sub-Mediterranean (E = Asia Minor plus the Crimea) and Caucasus (F). Circumscription of these geographical areas largely followed Meusel et al. (1965), though we used the narrower circumscription of the Mediterranean basin (A) advocated by Thompson (2005), thus leaving an expanded sub-Mediterranean region (B–E). Note that samples from high mountains in the Mediterranean (e.g. from the Moroccan mid-Atlas Mountains, the Cretan uplands and the Taurus Mountains in Turkey) were classified into the corresponding sub-Mediterranean region, because the high-altitude habitats in these regions constitute non-Mediterranean ecological conditions (higher precipitation and lower mean temperatures) that are broadly comparable with sub-Mediterranean environments. Bayesian trees of the MrBayes run were used as input trees. The Bayesian analysis of ancestral areas was run for 250 000 cycles, using ten chains and sampling every 100th generation; the first 10 % of the resulting data were discarded as burn-in. The sequence of *S. satyrioides* from Azerbaijan was used as the outgroup.

RESULTS

Information content of chosen DNA regions

Each of the chosen DNA regions was successfully amplified for all accessions, except for the *LFY* region. As this could not be amplified in either accession of *H. formosum*, we could only obtain a partial read of the first *LFY* intron for this species.

Initial comparison of sequence variability provided an important yardstick for assessing the resolving power of the chosen DNA regions (Table 1). The nuclear regions outperformed the plastid regions in terms of both absolute number and percentage of variable positions within the aligned length. Although the first intron of *LFY* provided more variable positions than the more widely used nrITS, the two regions provided similar amounts of information relative to sequencing effort required. However, given the presence of extensive intra-individual paralogy in nrITS (discussed below in the section nrITS variability), the single-copy *LFY* arguably appeared a better choice for phylogeny reconstruction.

Of the six plastid regions initially analysed, three displayed no variation among the selected ingroup members. Among the three variable regions, *trnL-ndhF* provided by far the most phylogenetic information, echoing the conclusions of the classic study by Shaw et al. (2007). The second most informative region was *accD-psaI*, followed by *ycf1*. However, as the *ycf1* region is

reported to be under positive selective pressure in some plant families (e.g. Pinaceae; Parks *et al.*, 2009), we eventually preferred *trnH-psbA*, a region widely used in phylogenetic studies. These three regions were sequenced for all selected samples (Fig. 2), and an ILD test of topological incongruence among single-region trees indicated maximal combinability of these regions ($P = 1.0$). We therefore performed all subsequent analyses of plastid sequences using this combined matrix.

nrITS variability

Direct sequencing of the nrITS of 136 populations produced 176 sequences of the entire region, but 99 of these accessions showed APSs in their sequence data. A further 35 samples showed length polymorphism: *H. robertianum* was characterized by a length-polymorphic site (LPS) located 566 bp from the 5' end of the nrITS, all *H. comperianum* accessions by an LPS at 483 bp, three accessions of *H. hircinum* by an LPS at 41 bp and one accession of *H. × samariense* by an LPS at 483 bp. A further LPS at 216 bp united *H. hircinum*, *H. jankae* and *H. calcaratum*. Two accessions originating from the same population (*H. hircinum* from Saint-Paul-en-Forêt, France) showed

an unusually large number of point mutations; these putative pseudogenes were excluded from further analysis. All species examined other than *S. satyrioides* and *H. metlesicsianum* (represented by six and three plants, respectively) yielded sequences exhibiting both categories of APS.

Within *Himantoglossum* s.s., an alignment of the direct nrITS sequences lacking length polymorphisms yielded a matrix of 647 positions, 51 displaying APSs. These 101 accessions were collapsed into 75 ribotypes, which were then reduced using TCS to 14 ribotype groups (Fig. 3A). These were dispersed around two dominant ribotype groups, here labelled 'RGa1' and 'RGb1', which contained 72 and 18 ribotypes, respectively. The remaining ribotypes formed satellites of between one and four accessions. Most notably, the *H. formosum* ribotypes were separated from the remainder by a minimum of five mutational steps.

The ribotype groups (Fig. 3B) showed surprisingly limited phylogeographical patterns. The most frequent ribotype group (RGa1) occurred throughout the sampled area and so encompassed all taxa of the *H. hircinum* and *H. jankae* clades, whereas the second most frequent group (RGb1) was confined to the region east of the Adriatic and taxonomically restricted to the *H. jankae* clade, occurring in *H. jankae*, *H. calcaratum*

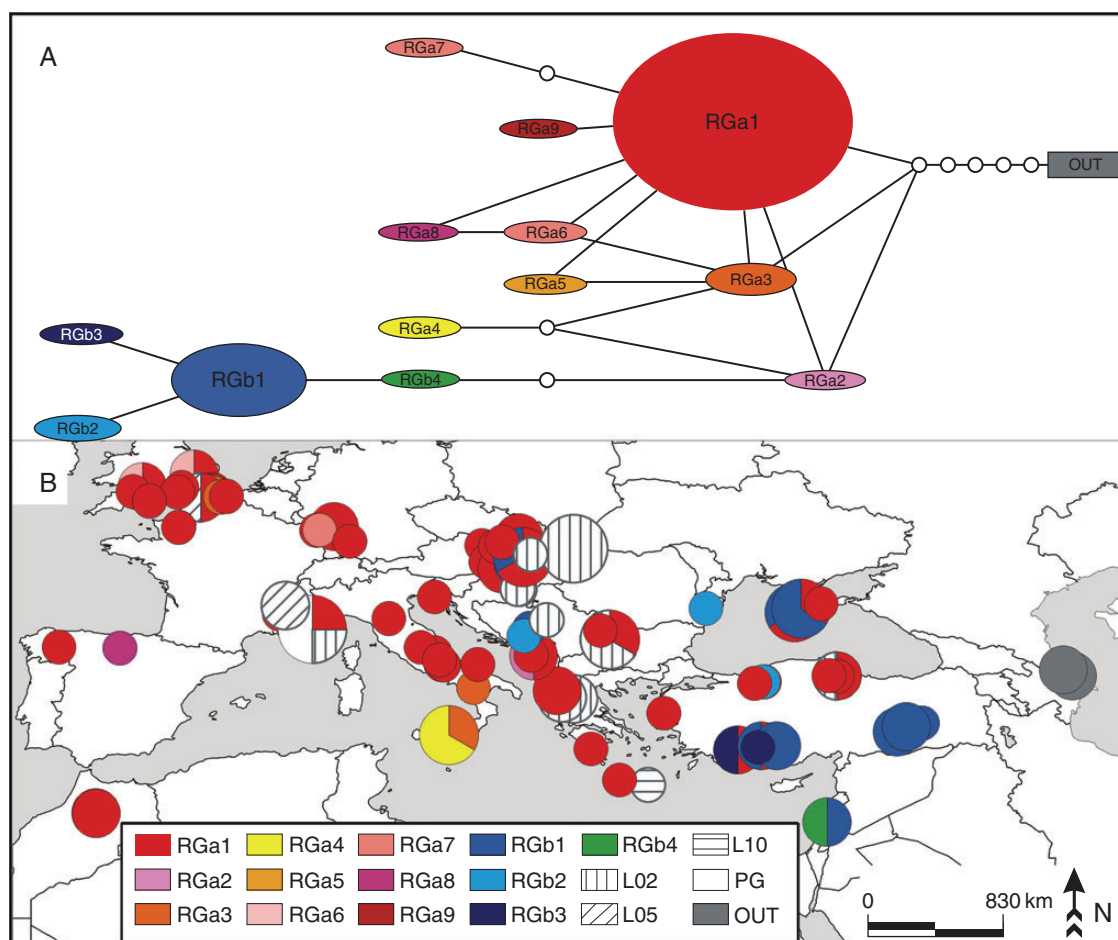


FIG. 3. Parsimony network of ribotype groups of *Himantoglossum* s.s. identified by TCS (A) and their geographical distribution (B), together with three length-polymorphic variants (L02, L05 and L10), a putative pseudogene (PG) and the ribotype group of the outgroup, *H. formosum* (OUT). Blank circles (A) represent undiscovered (i.e. hypothetical) ribotypes. The size of pie charts (B) is proportional to the number of samples analysed from the same population (maximum of four).

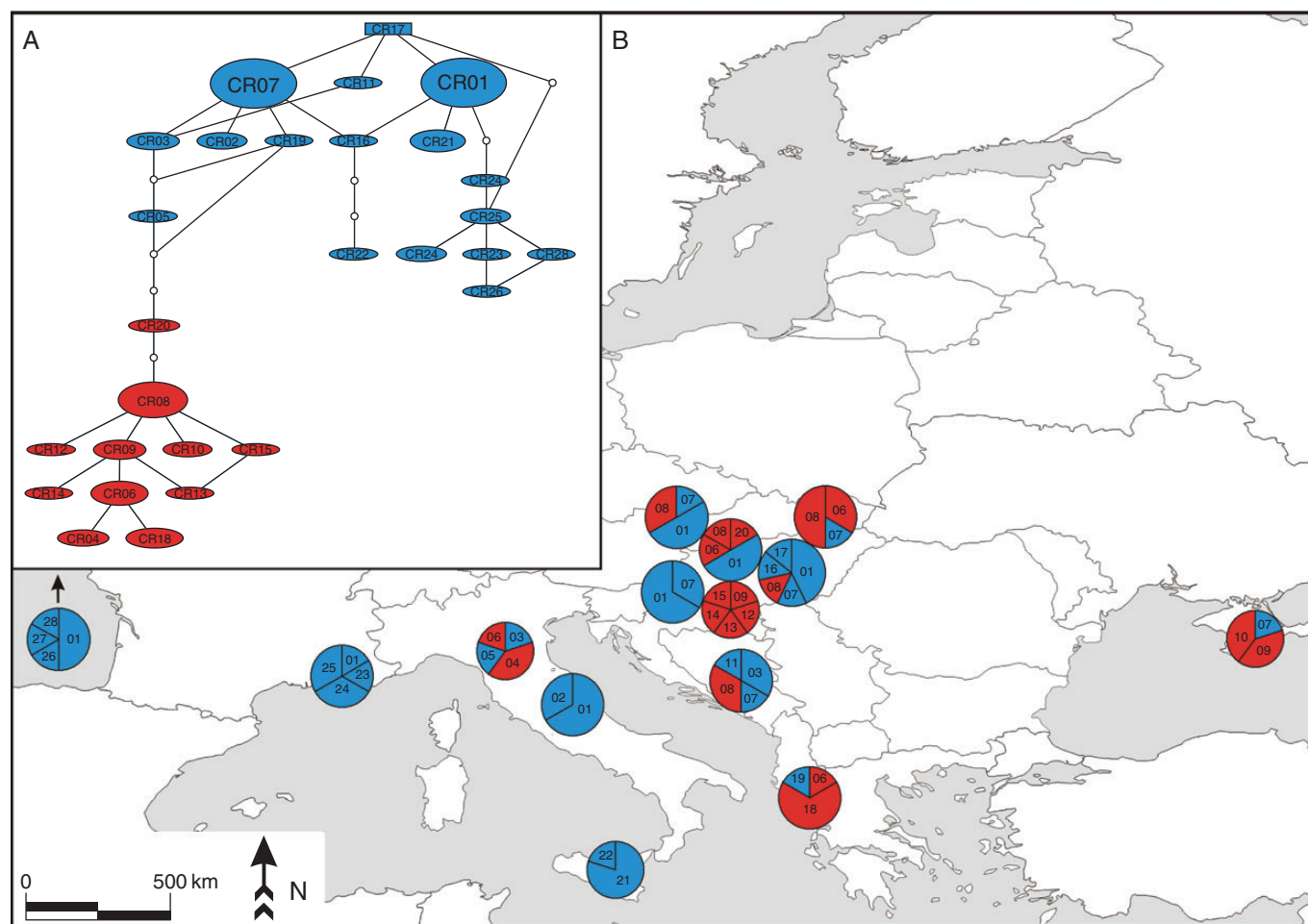


FIG. 4. Parsimony network of cloned nrITS sequences of the *Himantoglossum hircinum-jankae* clade identified by TCS (A) and their geographical distribution (B). The TCS network (A) is coloured according to the two main groups identified; each cloned ribotype is coded as 'CR' followed by a number. Blank circles (A) represent undiscovered (i.e. hypothetical) ribotypes. The pie charts on the map (B) show the ribotype composition of a cloned sample (i.e. ribotypes cloned from a given specimen); their size is proportional to the number of clones sequenced. Colouring and numbering (without the acronym 'CR') of ribotypes on the pie charts follow those of the TCS network; the arrowed pie chart is located in south-west England.

and *H. caprinum*. The less frequent ribotype groups were more geographically localized, being confined either to small regions (e.g. RGA6 in southern England and western Germany and RGA2 in Bosnia-Herzegovina, Romania and Turkey) or even to a single population (e.g. RGA2, RGA8).

The 81 cloned nrITS sequences were aligned into a 638-bp-long matrix, and were subsequently collapsed into 28 ribotypes. The distributions of these ribotypes were uneven; two codominant ribotypes (denoted CR01 and CR07) were found to be equally frequent (14 occurrences), followed by CR08 (9), whereas the remaining ribotypes only occurred between one and five times (mean 1.76, median 1.0), most being found in just one accession (Fig. 4A).

LEAFY variability

The central part of the *LFY* intron showed APSs in nine of 36 analysed samples, so the polymorphic samples were subjected to cloning. The alignment yielded a 1472-bp-long matrix of 29 sequences. A heuristic MP search found 16 most-parsimonious

unrooted trees, topological differences being confined to the distal nodes. Because sequences cloned from the same species clustered together [in most cases sequences from the same individual clustered together (Supplementary Data Fig. S1)] we treated the cloned sequence variants as intra-individual paralogues.

Phylogenetic tree reconstructions

The MP phylogenetic analysis of the concatenated plastid sequences used 348 parsimony-informative characters and yielded two most-parsimonious trees of 406 steps (consistency index = 0.94, retention index = 0.98) on the same optimal parsimony island, the two trees differing only in the topology among the three outgroup samples (Fig. 5). The Bayesian tree (not shown) had the same topology, but some branches received no statistical support [see posterior probabilities (PPs) on the branches of the MP tree presented as Fig. 5]. The plastid phylogeny provided an important insight into the phylogeny of the group. It entirely corroborated the previous tree based on nrITS (Bateman *et al.*, 2003) by placing *H. comperianum* as sister to

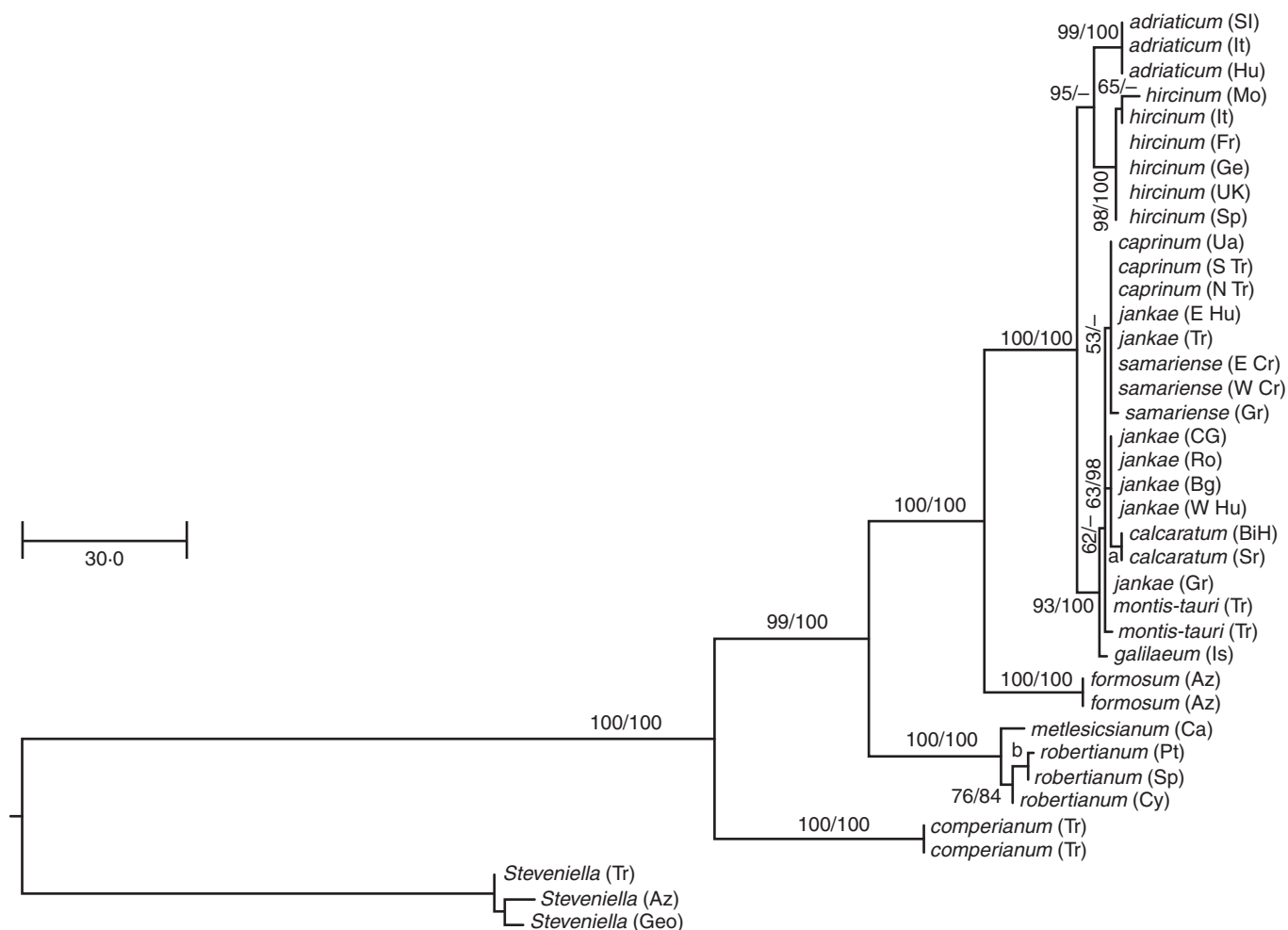


FIG. 5. Phylogram of *Himantoglossum* s.l. based on four concatenated plastid intergenic spacer sequences. Region of origin (displayed as country code) is shown in parentheses. Numbers above branches represent statistical support (bootstrap value/Bayesian posterior probability; values for nodes a and b are 87/- and 95/100, respectively).

the remaining species and the *H. robertianum* group as sister to *Himantoglossum* s.s. As predicted from morphological examination, *H. formosum* was placed as sister to the rest of *Himantoglossum* s.s., and within this group two main clades were found: a western clade consisting of *H. hircinum* plus *H. adriaticum* (both clearly separate) and an eastern clade epitomized by *H. jankae* and *H. caprinum*, together with several more localized named segregates (Fig. 1). Unfortunately, no resolution was observed within the eastern clade other than the weakly supported separation of the Israeli species *H. galilaeum* as sister to the rest.

The MP search of the first intron of *LFY* used 598 parsimony-informative sites and found two most-parsimonious trees (consistency index = 0.95, retention index = 0.98) of length 894 on the same parsimony island. The two trees differ from each other only in the relationships inferred among the *H. hircinum* samples: one tree tentatively shows the Spanish sample as sister to the rest (Fig. 6) whereas in the other tree it collapses into the general polytomy among *hircinum* samples. This second topology was found by Bayesian phylogenetic tree

inference (not shown), and all bootstrap-supported branches were also supported (i.e. PP > 95 %) by posterior probability values (shown in Fig. 6). Although the *LFY* tree ostensibly gave useful resolution (Fig. 6), it also showed significant topological differences from the plastid tree (Fig. 7). Contrary to the plastid phylogeny, *LFY* resolved two main clades within the eastern *H. jankae*-*H. caprinum* clade separating *H. caprinum* (and an east Cretan sample of *H. × samariense*) from the rest of the samples (*H. jankae*, *H. calcaratum*, *H. galilaeum*, *H. montis-tauri* and the remaining two samples of *H. × samariense*). In addition, *H. montis-tauri* was found to be sister to all other samples within the *H. jankae* group, though this separation was only weakly supported.

The most striking topological difference between *LFY* and the remaining trees (Fig. 7) concerns *H. formosum*, which is placed as sister to the rest of *Himantoglossum* s.l. Within the *H. hircinum*-*H. jankae* clade, central European samples of *H. jankae* were grouped together with *H. adriaticum* on the *LFY* tree, whereas these samples were grouped into the unresolved polytomy of the eastern group on the plastid tree, together with all other

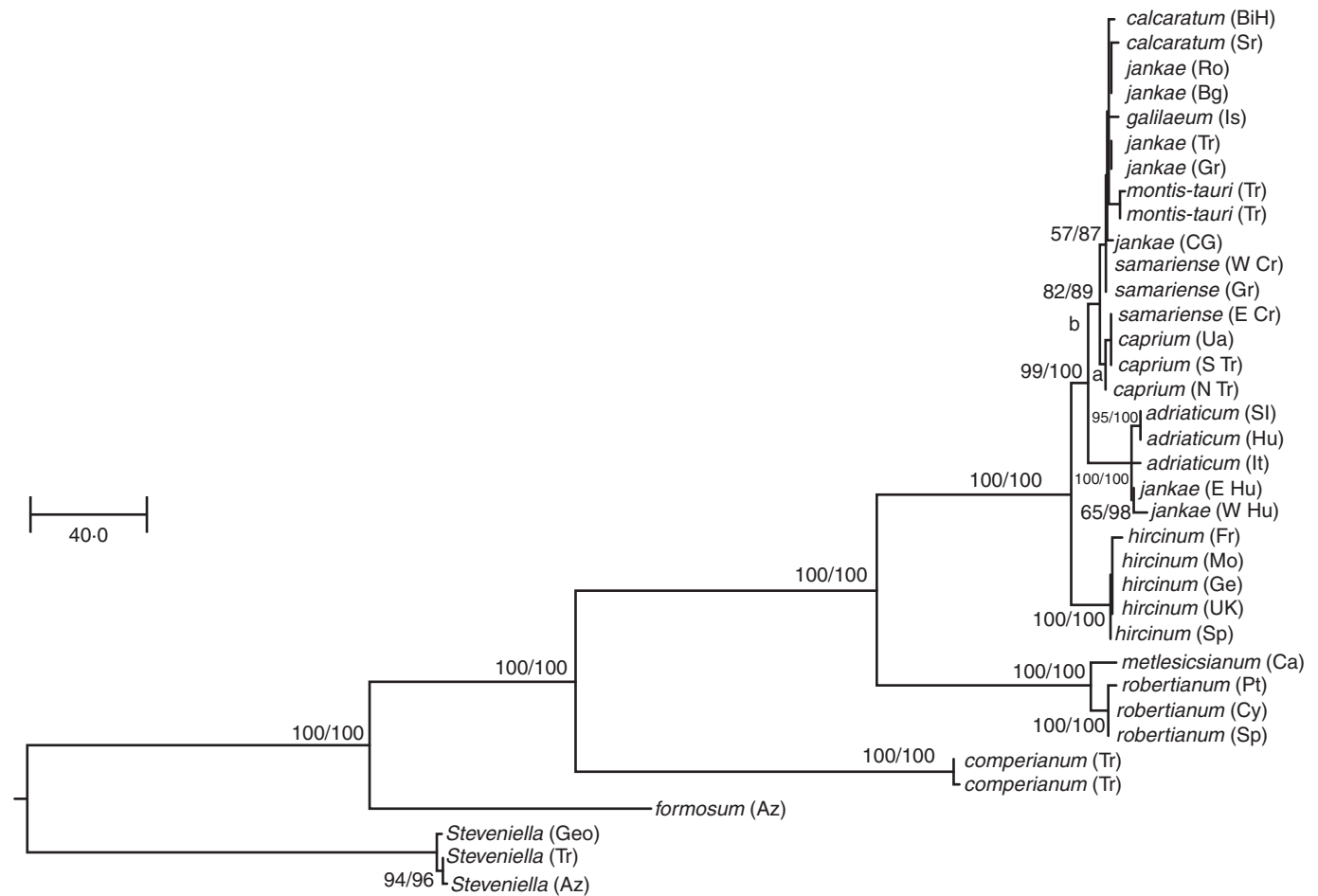


FIG. 6. Nuclear DNA phylogram of the genus *Himantoglossum* based on the first intron of the single-copy *LFY* gene. Region of origin (displayed as the country code) is shown in parentheses. Numbers above branches represent statistical support (bootstrap value/Bayesian posterior probability; values for nodes a and b are 86/100 and 87/100, respectively).

H. jankae samples, both relationships receiving high statistical support (Fig. 6). Similarly, samples of *H. × samariense* were separated on the *LFY* tree; the eastern Cretan sample was confidently grouped with all three *H. caprium* samples, whereas the western Cretan and southern Peloponnesian samples were included in the taxonomically diverse clade epitomized by *H. jankae*.

Finally, we checked the potential combinability of the plastid and *LFY* datasets via an ILD test, as implemented in PAUP. The results indicated extensive topological incongruence between these two datasets ($P = 0.01$) – incongruence that is also evident in the ‘hard incongruences’ (Wendel and Doyle, 1998) identifiable by eye between the trees obtained from contrasting genomes. We therefore abandoned further analyses of the concatenated matrix, as such matrices perform poorly in such circumstances (Heled and Drummond, 2010).

Molecular dating

Our key goal for the molecular dating analysis was to provide a temporal framework for the evolution of *Himantoglossum s.l.*, using far better taxon sampling than that of Inda et al. (2012) for *Himantoglossum s.l.* – a clade that they did not discuss in detail. Given the analytical similarities, it is not surprising that

the chronogram resulting from our analysis (Fig. 8) generally yielded dates similar to those reported by Inda et al.

The placement of *S. satyrioides* as sister to *Himantoglossum s.l.* receives no statistical support (posterior probability 0.64), as in the analysis by Bateman et al. (2003), in which bootstrap support was <50 %. In contrast, *Himantoglossum s.l.* is clearly monophyletic. The remainder of the chronogram reflects the topologies obtained from both the plastid and the *LFY* matrix. Nevertheless, the paucity of species-specific nrITS ribotypes prevented us from dating speciation events within the *H. hircinum-jankae* clade; only the split between the two codominant ribotypes was supported by sufficient data to allow statistically credible dating.

Our results indicated the age of the most recent common ancestor (MRCA) of *Himantoglossum s.l.* to be 9.15 Ma [95 % highest posterior density interval (HPD) 4.64–14.2 Ma (lower–upper)]; this was also the date inferred for the separation of *H. comperianum*. The split between the *H. robertianum* group and *Himantoglossum s.s.* was dated to 5.7 Ma (HPD 2.41–9.46 Ma), whereas isolation of the Canary Island endemic *H. metlesicsianum* was estimated at 1.8 Ma (HPD 0.56–3.52 Ma). Within *H. robertianum*, a split shown by our plastid dataset (Fig. 5) between western and eastern Mediterranean samples (Iberian versus Cypriot samples) is also

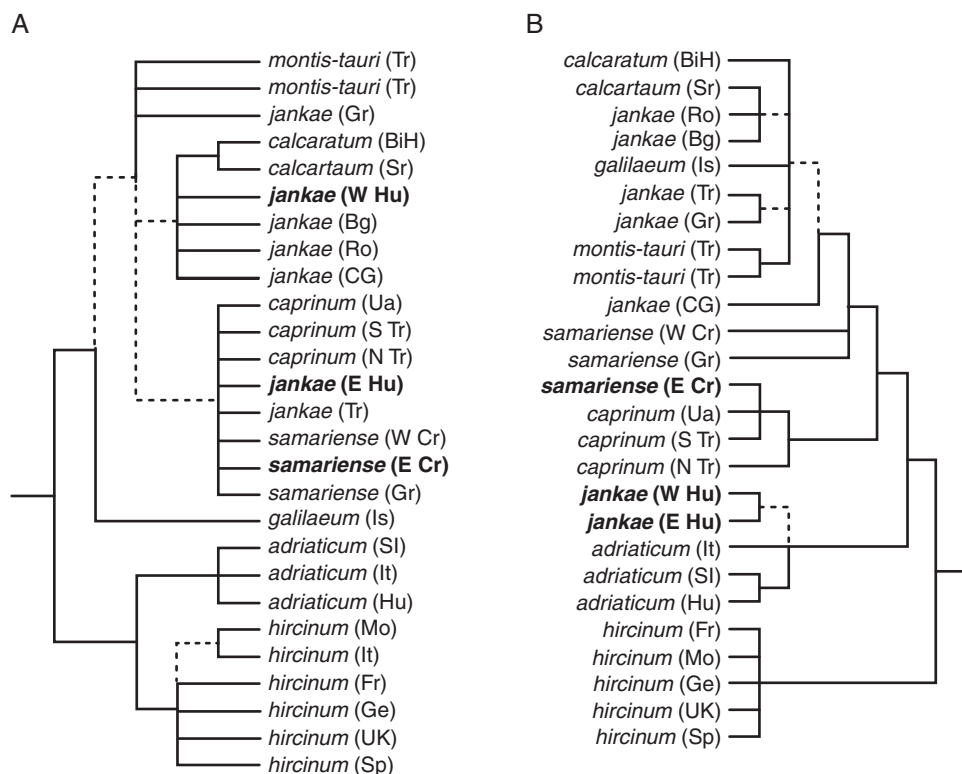


FIG. 7. Topological incongruences between the contrastingly inherited genomes within the as shown by the cladogram of MP trees of four combined plastid spacers (A) and the nuclear *LFY* first intron (B). Statistically weakly supported branches (i.e. those with <74 % bootstrap support) are dashed, and the most conspicuous topological incongruences are in bold.

evident in the nrITS chronogram and was dated at 1.0 Ma (HPD 0.24–2.22 Ma). Within *Himantoglossum* s.s., separation of the Caucasian endemic *H. formosum* from later-divergent species was estimated at 1.5 Ma (HPD 0.39–3.09 Ma), whereas the subsequent separation of the two dominant clone ribotypes within the *H. hircinum*–*H. jankae* clade was tentatively dated at 0.59 Ma (HPD 0.07–1.58 Ma). The mean substitution rate of nrITS was estimated to be 7.2 substitutions/site/year $\times 10^{-9}$.

Ancestral geographical area reconstruction

The Bayesian DIVA analysis of plastid data identified an unambiguous (PP 96 %) ancestral range for the MRCA of the *Steveniella*–*Himantoglossum* clade in the Caucasus Mountains (Fig. 9). The ancestral range of *Himantoglossum* s.l. most likely lies either in the Caucasus (PP 67 %) or, with less confidence, in Asia Minor (PP 18 %). Distal to this node, all backbone nodes on the tree apparently represent a dispersal event that led to vicariance between daughter lineages. The first dispersal event is the separation of the Mediterranean lineage of the *H. robertianum* group from *Himantoglossum* s.s., an event that was followed by the dispersal of *Himantoglossum* s.s. to sub-Mediterranean Europe from a confidently identified (PP 88 %) ancestral area in the Caucasus Mountains, a region where today the genus is represented by the putatively relict species *H. formosum*.

The MRCA of the *H. hircinum*–*jankae* clade, represented by the next backbone node, had the most ambiguous ancestral area on the tree (Fig. 9): an origin is approximately equally likely in

Asia Minor (PP 29 %), the Caucasus (PP 27 %) or the Balkans (PP 18 %), and an origin is even possible in western or central European (PP 7 % for both). The next node, representing the MRCA of *H. jankae* plus *H. caprinum*, received equal support for an origin in either the Balkans (PP 37 %) or Asia Minor (PP 36 %), though a shared ancestral area is also quite likely (PP 24 %). Similarly, the MRCA of *H. hircinum* plus *H. adriaticum* had an ancestral area with similar probabilities in either western Europe (PP 43 %) or central Europe (PP 39 %), but in this case a shared area was deemed much less likely (PP 8 %).

DISCUSSION

Phylogenetics of *Himantoglossum* s.l.

Our three molecular phylogenetic tools (nrITS, *LFY* and plastids) have provided contrasting degrees of resolution at two contrasting taxonomic levels: *Himantoglossum* s.l. and the *H. hircinum*–*jankae* clade (Figs 3–6). The two levels are therefore best discussed separately.

Despite the discordant placement of *H. formosum* in the *LFY* tree (see below, section Incongruence between gene trees), we are confident from the morphometric (not shown), nrITS and plastid data that species diverged in the order *comperianum* > *robertianum* + *metlesicsianum* > *formosum* > *hircinum*–*jankae* clade. Multiple samples of these taxa clustered tightly together, and the robust topologies are consistent with previous, albeit less well sampled, nrITS phylogenies (Pridgeon *et al.*, 1997; Bateman

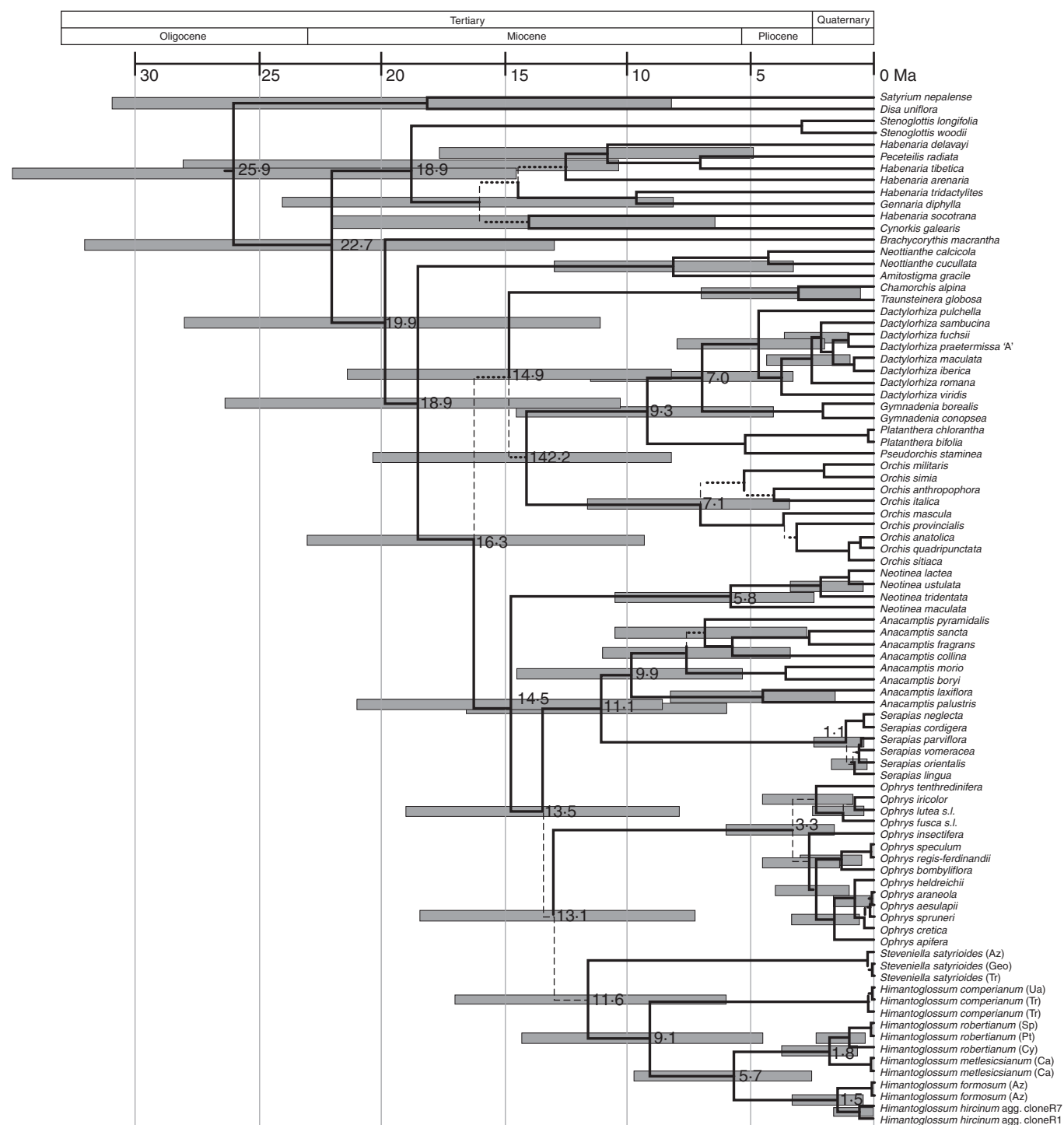


FIG. 8. Bayesian chronogram of nrITS sequences of a representative set of European orchids with enhanced sampling of the *Steveniella*-*Himantoglossum* clade. The tree is the maximum clade credibility tree obtained in a Bayesian analysis with a log-normal relaxed clock (uncorrelated) after 10 % burn-in and secondary calibration points derived from Gustafsson *et al.* (2010). The numbers at nodes represents mean divergence dates of key clades in million years before present (Ma); dashed branches have low (<0.83) posterior support.

et al., 2003). Authors indulging in topological speculation have been less successful: Delforge (1999) misplaced *H. comperianum* as sister to *H. robertianum* + *metlesicsianum* and within the *H. hircinum-jankae* clade, he was joined by Bateman *et al.* (2003)

in speculatively misplacing *H. caprinum* (formerly *H. affine*) as diverging earlier than the remaining taxa within the group.

This sequence of lineage divergence suggests that laterally fused sepals, long filiform lateral labellar lobes, a long spur

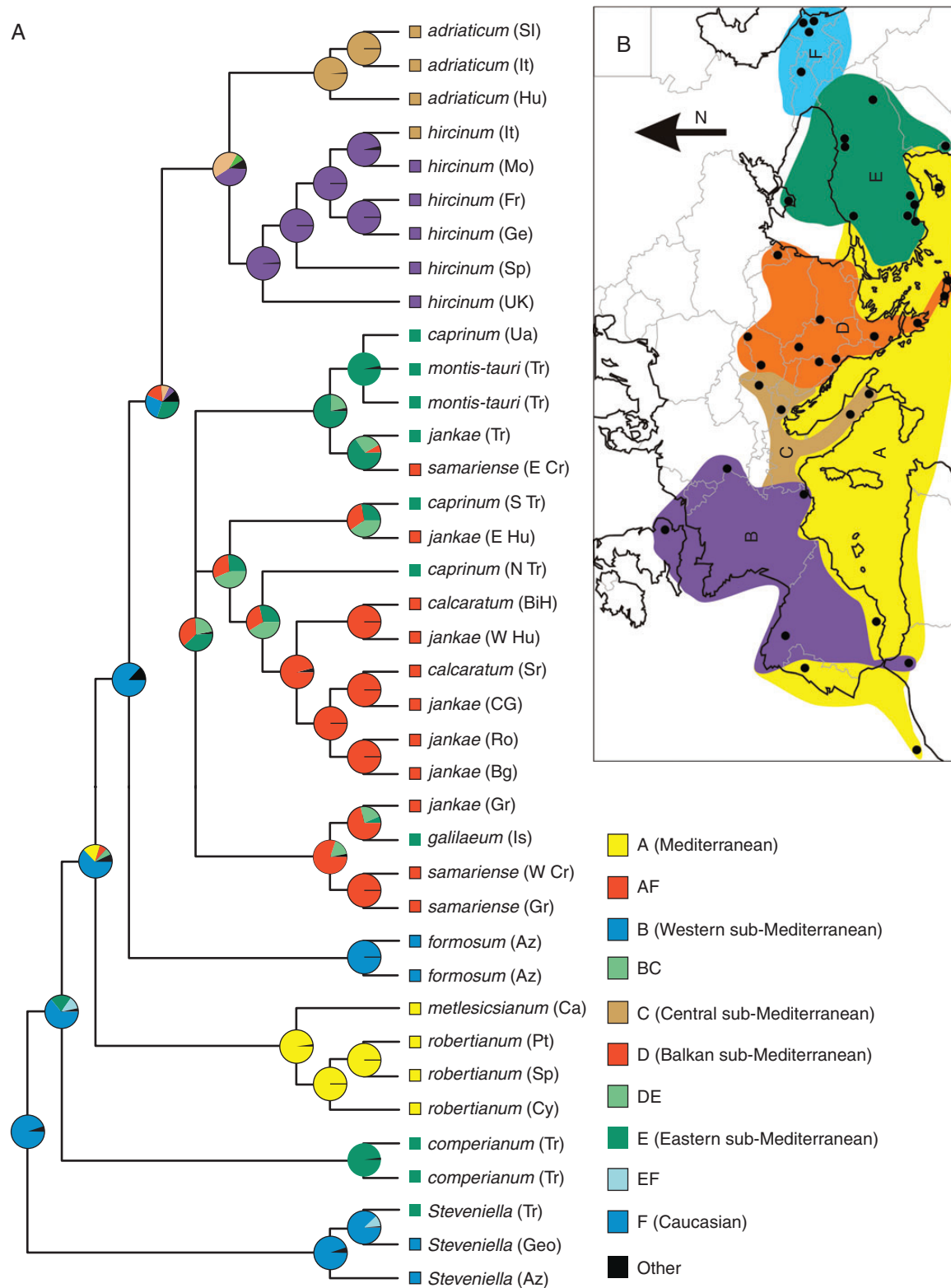


FIG. 9. Ancestral area reconstruction of the *Steveniella*-*Himantoglossum* clade based on the concatenated plastid regions and using a Bayesian DIVA approach. (A) The resulting tree has nodes displayed as pie charts denoting possible ancestral areas proportional to their Bayesian posterior probability, drawn after 250 000 cycles using ten chains and sampling every 100th generation with 10 % burn-in. Region of origin (displayed as country code) is shown in parentheses. (B) Initial classification of the geographical origin of samples (dots) used in this analysis.

and possibly a chromosome number of $2n = 30$ are apomorphies of *H. comperianum*, whereas early flowering, loss of lateral petal teeth, shortened central labellar lobes and a triangular stigma are apomorphies of the *H. robertianum* group; this group shares with *Himantoglossum* s.s. the synapomorphy of laterally fused viscidia (see Delforge, 1999; Bateman et al., 2003; Claessens and Kleynen, 2011).

Within *Himantoglossum* s.s., the Caucasian endemic *H. formosum* branches first, with good statistical support in both the plastid (Fig. 5) and nrITS (Fig. 8) trees. However, our DNA data permit far less confidence when discussing the circumscription of, and relationships among, taxa within the succeeding *H. hircinum*–*H. jankae* clade. Within this group, the separation of a western clade (*H. hircinum* plus *H. adriaticum*) from an eastern clade (*H. jankae* s.l. plus *H. caprinum* plus local endemics) is suggested by the plastid tree (Fig. 5) with high bootstrap support (95 %) but no Bayesian support. In contrast, the *LFY* tree (Fig. 6) shows with high statistical robustness (the relevant branches have 82–100 % bootstrap values and 0.89–1.00 PP values) *H. hircinum* alone branching earliest, followed by *H. adriaticum* (plus two accessions of *H. jankae*) and *H. caprinum* (plus one accession of *H. × samariense*). The *LFY* tree leaves only a poorly resolved clade containing most of the accessions of *H. jankae* plus its segregates (or hybrids?) *H. calcaratum*, *H. montis-tauri* and *H. × samariense*. If interpreted literally, this topology indicates a reversal of the east-to-west direction of speciation events through time that characterizes the early history of not only *Himantoglossum* s.l. but also many other Eurasian orchid genera (cf. Bateman et al., 2003); rather, eastward events are implied in *Himantoglossum* s.s., beginning in western Europe and ending in the local endemics of the eastern Mediterranean.

Morphological variation within the *H. jankae* group is similarly more subtle than that among earlier divergent lineages (this topic will be discussed in greater detail in a future paper). Much of the variation stems from floral pigmentation characters; as a sweeping generalization, variously placed spots and/or stripes tend to decrease in frequency eastwards, and flower colour also is often paler in the Balkans and Asia Minor. Labellum shape is also highly variable – again, most strongly in the Balkans and Asia Minor – and has recently been attributed to heterochronic shifts (Bateman et al., 2013a). This variation in floral characters such as spur length, which differentiates some species of *Himantoglossum* s.l. (e.g. Sramkó et al., 2012), may be associated with pollinator guilds characteristic of different geographical areas, as demonstrated for some other European orchids (Boberg et al., 2014; Sun et al., 2014; but see Bateman and Sexton, 2008).

nrITS variability within the *H. hircinum*–*jankae* clade

We placed particular emphasis on the archetypally variable nrITS region (Table 1) in search of phylogeographical patterns. However, neither our direct sequencing efforts (Fig. 3) nor the subsequent cloning exercise (Fig. 4) provided more than a basic insight, uncovering one exceptionally widespread ribotype group (RGa) and a second most frequent ribotype group (RGb) that is confined to the area east of the Adriatic. This paucity of phylogenetic resolution in a fast-changing nuclear marker can be explained either by recent gene flow between taxa or by

incomplete lineage sorting during speciation events (Sang, 2002; Álvarez and Wendel, 2003). However, we found no sign of widespread hybridization in *LFY*, the other nuclear marker included in this study; instead, in most cases multiple accessions of the same species grouped together, thus corroborating the morphologically based circumscriptions of species without offering evidence of substantial genetic exchange.

The exceptions to this generalization are central European samples of *H. jankae* and the apparently hybridogenous *H. × samariense*. A more likely explanation for the pattern observed in the nrITS is incomplete lineage sorting of ribotypes between the comparatively recently separated species of the *H. hircinum*–*adriaticum* and *H. jankae*–*caprinum* groups (i.e. we suspect that the ribotypes are shared between species and infraspecific taxa within these groups due to recent common ancestry).

The most surprising result of the nrITS-based phylogeographical survey was evidence (albeit equivocal) that the western groups (*H. hircinum* and *H. adriaticum*) are older than the eastern clade (*H. jankae* plus *H. caprinum* plus local endemics). Ribotype group RGa occurs throughout the geographical distribution of the clade, whereas group RGb is confined to the region east of the Carpathians/Adriatic and is therefore tentatively viewed as being younger (an interpretation that assumes similar rates of migration in the two groups). This inference receives further support from the *LFY* tree (Fig. 6), which suggests divergence of species in the order *hircinum* > *adriaticum* > *jankae*–*caprinum* group. Lastly, Bateman et al. (2013a) noted deviant ribotypes in the two populations of *H. hircinum* sampled from the supposed geographical outlier of this species in southern Italy and Sicily, suggesting that detailed morphological and karyological examinations should be pursued in this region.

Incongruence between gene trees

Topological incongruence between phylogenetic trees obtained from regions of the typically maternally inherited plastid genome and biparentally inherited nuclear genome usually hints at past hybridization (Rieseberg, 1997). In the case of *LFY*, where the presence of a single copy has been verified by Southern blotting in several European orchids (Montieri et al., 2004), this process can easily replace the maternal gene with the paternal gene, thereby introducing reticulation into the gene tree.

Comparing the results of our three molecular datasets, the most prominent incongruence is the placement of *H. formosum* improbably close to the root of the *LFY* tree. Although this topology could in theory reflect an ancient hybridization event, we believe that loss of an orthologous copy, leading to incongruent placement of a taxon with a non-orthologous copy, provides a more likely explanation (Sang, 2002; Small et al., 2004; Ilut and Doyle, 2012; Zimmer and Wen, 2012). As recently demonstrated in the family Lamiaceae (Curto et al., 2012), the post-duplication fate of paralogous copies can be sorting into different lineages, or different copies may be lost from different species. In the present case we suspect that, unlike all other members of the genus, *H. formosum* retained a more ancient paralogous copy of *LFY*. This interpretation is suggested by difficulties experienced during amplification of *LFY* in *H. formosum* and the failure of internal primers to obtain reads of the intron. *Himantoglossum formosum* is sister to the morphologically similar species of *Himantoglossum* s.s. in both the plastid tree (Fig. 5) and the

nrITS chronogram (Fig. 8), as well as the better-sampled nrITS tree recently published by Bateman *et al.* (2013a).

In contrast, the different placement of central European *H. jankae* samples on the *LFY* tree (Figs 5 and 6) may indeed indicate hybridization. The central European samples are found together with *H. jankae* samples from other geographical regions on the plastid tree, potentially reflecting genetic exchange between *H. adriaticum* and central European *H. jankae*. This observation raises the possibility of a hybrid origin for *H. adriaticum*, because this interchange could have caused the topological contrast evident between the two trees (i.e. *H. hircinum* and *H. adriaticum* are represented as monophyletic sister species in the plastid tree but not in the *LFY* tree). The conflicting placements of *H. adriaticum* suggest it is an ancient hybrid between *H. hircinum* and *H. jankae* that underwent speciation after its formation. Our conclusion of hybridity is supported by flower morphology, which combines character states of both putative parents (slightly larger and more colourful flowers compared with those of *H. hircinum*).

The placement of different *H. × samariense* samples into contrasting clades can again be explained by hybridization if the maternal parent of the eastern Cretan sample was *H. jankae* whereas the paternal parent was *H. caprinum*. This explanation fits well into the hybrid origin postulated by the author of the species, who described this taxon as a nothospecies derived from *H. jankae* and *H. caprinum* [albeit then named as *H. caprinum* and *H. affine*, respectively (Alibertis and Alibertis, 1989; see also Delforge, 1999)]. Here, too, the flower morphology of the supposed hybrid also shows a combination of parental characteristics; some flowers of *H. × samariense* exhibit the labellum markings typical of *H. jankae* whereas others lack them, as does *H. caprinum*.

Finally, the weakly supported (bootstrap value 64 %) placement on the plastid tree of *H. montis-tauri* as sister to all other samples of the *H. jankae*–*H. caprinum* clade, together with its possession of a subtly distinct but unique nrITS ribotype (R**G**b3) and its relative morphological distinctiveness (not shown), corroborates the view expressed by many authors who regard this taxon as an entity separate from *H. jankae* s.s. In contrast, the lack of molecular distinctiveness of *H. calcaratum* from *H. jankae* s.s. and the morphological intermediacy between them of *H. ‘robustissimum’* in all characters but spur length (not shown) call into question its taxonomic separation as a full species.

Key events in the evolution of *Himantoglossum* s.l.

When considered together with the ancestral area reconstruction, molecular dating sheds light on key events in the evolution of the genus. Admittedly, the broad error bars incurred by the molecularly inferred dates urge caution when indulging in causal interpretation.

Our finding of an average nrITS mutational rate of 7.2 substitutions/site/year $\times 10^{-9}$ clearly places *Himantoglossum* s.l. towards the higher end of the spectrum of substitution rates reported by Kay *et al.* (2006). However, their survey did not include any Orchidaceae, a family that is characterized by relatively high neutral substitution rates (Barracough and Savolainen, 2001).

Given the similar dataset and taxonomic sampling, dates obtained by us for the separation of the main lineages are, not

surprisingly, close to those reported by Inda *et al.* (2012). The first split within *Himantoglossum* s.l., marking the separation and dispersal of *H. comperianum* from the Caucasus to Asia Minor, is dated to 9.15 Ma (HPD 4.64–14.2 Ma), a point in time that offers no obvious environmental driver. However, the next split, the separation of the *H. robertianum* group at 5.7 Ma (HPD 2.4–9.46 Ma) and its dispersal to the Mediterranean, apparently coincided with the Messinian salinity crisis, the dramatic wholesale desiccation of the Mediterranean Basin that occurred between 5.96 and 5.33 Ma (Rouchy and Caruso, 2006). Similar events have been reported in other Mediterranean plant groups (e.g. Rosselló *et al.*, 2002; Pimentel *et al.*, 2007; Font *et al.*, 2009; Escudero *et al.*, 2010; Emadzade *et al.*, 2011).

The Mediterranean region was largely wooded prior to the Messinian crisis, which triggered a transition to more arid habitats and more open plant communities. The modified environment may have suited *Himantoglossum*. All members of the *Himantoglossum* s.l. lineage have large tubers and fleshy stems, and most have wintergreen leaves that wither during anthesis, well before seed-set. The one exception – the *H. robertianum* group – flowers unusually early in spring and bears fleshy xeromorphic leaves. This adaptation could be a key character that enabled this lineage to disperse and diversify across the arid region of the Mediterranean basin created by the Messinian event.

The rarest and most endangered species examined in the present study was the Canarian endemic *H. metlesicsianum*; a few small populations cling to the middle to lower slopes west of the 3700-m-high volcanic caldera of Mount Teide on Tenerife (Stierli-Schneider, 2004; Kropf *et al.*, 2012) and of the 2400-m-high volcanic peak of Mount Taburiente on La Palma (Acevedo Rodríguez and Mesa Coello, 2013). The estimated 1.8 Ma (HPD 0.56–3.52 Ma) divergence of *H. metlesicsianum* post-dates the emergence above the sea of the three volcanic cores upon which the present island of Tenerife was built (~ 7 Ma) but falls within the period when the three cores were joined volcanically to eventually form a single composite island (between ~ 3 and 0.8 Ma), as well as the estimated maximum age of La Palma (~ 3 Ma). It also corresponds to the divergence time of several other Canary Island endemics (Carine, 2005). The origin of this species on the Canary Islands most likely reflects a colonization event by an ancestor resembling *H. robertianum*, westwards from Morocco or Iberia, followed (most likely rapidly) by anagenetic speciation (cf. Bateman, 2012b). This biogeographical scenario is shared with almost all Macaronesian orchids (Bateman *et al.*, 2014) and many other Canarian endemic plants and animals (Juan *et al.*, 2000).

The rise of the taxonomically problematic *H. hircinum*–*jankae* group was relatively recent; this lineage separated from *H. formosum*, the only lineage to remain in the ancestral Caucasian area, at 1.5 Ma (HPD 0.4–3.1 Ma) – that is, within the Calabrian stage of the Pleistocene epoch that preceded the major northern hemisphere glaciations. The drivers behind their separation and the subsequent divergence of the two main nrITS lineages within the group at 0.6 Ma (HPD 0.07–1.58 Ma) remain unclear. However, the dramatic impact of Quaternary glaciations on the European biota (Hewitt, 1999) constitutes a credible driver of the more recent speciation within the members of the genus *Himantoglossum*, as well as potentially explaining a degree of ongoing gene flow between them.

Hybridization remains possible across the entire genus, as was demonstrated by the recent discovery of natural hybrids between the two most divergent species, *H. comperianum* and (putative) *H. montis-tauri*, on the Aegean island of Lesbos (Karatzas, 2004). We also note that the geographical divide between the largely allopatric *H. hircinum-adriaticum* group to the west and *H. jankae-caprinum* group to the east passes through the Adriatic and Carpathians, coinciding with a similar genetic barrier that was inferred from DNA data in the related orchid genus *Ophrys* by Devey et al. (2009).

Implications for conservation

Our results strongly suggest that *H. metlesicsianum* is a *bona fide* species of significant age. Rare, highly localized and constantly threatened by various human impacts, it merits the highest conservation attention, being justly categorized as Endangered on the 2012 IUCN Red List (Bateman et al., 2013b). In addition, our nrITS and plastid datasets both show that there is potentially significant molecular diversity within the widespread Mediterranean species *H. robertianum*, which evidently merits deeper Europe-wide study using molecular and morphological methods. *Himantoglossum formosum* is even more morphologically and molecularly distinct than *H. metlesicsianum*, as well as offering a potential ancestral morphology for the *H. hircinum-jankae* clade. Although its distribution is poorly known, this species appears to us sufficiently distinct and uncommon to warrant high-level legal protection.

We are presently less confident when making conservation recommendations within the given the limited resolution uncovered by each of the molecular approaches that we have applied. Confirmation of the circumscription of, and most appropriate taxonomic rank for, taxa within this clade must await integration of the molecular data discussed in this paper with our detailed morphometric survey (this topic will form the core of a future systematics paper). Included in both the IUCN Red List and European Habitats Directive, *H. adriaticum* can confidently be distinguished from *H. hircinum* s.s. on both morphological and molecular criteria. *Himantoglossum jankae* (still listed incorrectly as *H. caprinum*) is also included in the Habitats Directive. Both species are confined to slightly disturbed habitat patches within oakwoods of high conservation value and hence are valuable indicators of important European habitats that require legal protection. Further east, *H. caprinum* (formerly named *H. affine*) is not presently included in international conservation lists but appears to merit high conservation status. In Asia Minor and the Crimea, all members of the *Himantoglossum* clade are threatened by the collection of tubers to make ‘salep’, a traditional nutritious foodstuff used to produce beverages and ice cream (e.g. Kasperek and Grimm, 1999). Consequently, in Turkey, both *H. caprinum* and *H. jankae* have become confined to small habitat patches, most commonly cemeteries.

Establishing the evolutionary origins of, and the most appropriate taxonomic rank for, the four named local endemics (*calcaratum*, ‘*robustissimum*’, *samariense* and *montis-tauri*) is more challenging; it requires not only morphometric comparison but also more focused, well-sampled investigation using population genetic markers. Indeed, such work could give greater meaning to the genetic variation observed by us within *H. hircinum* (nrITS), *H. robertianum* (all genic regions) and even within the

outgroup, *S. satyrioides* (all genic regions). Until such work is conducted, it would perhaps be premature to award these taxa the highest conservation status.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following: Table S1: a list of specimens sampled, their geographical origin and abbreviations used in the study, together with GenBank accession numbers for the resulting DNA sequences (provided as an Excel file). Figure S1: maximum parsimony phylogenetic tree of cloned *LFY* sequences.

ACKNOWLEDGEMENTS

We thank the following people for assisting in extensive sampling of the genus: A. Alibertis, E. Alirzayeva, V. Ali-Zade, Z. Barina, L. Bartha, J. Bódis, H.-J. Esser, V. Farzaliyev, M. Fiedler, R. Hall, R. Kish, K. Könyves, C. A. J. Kreutz, B. A. Lukács, G. Magos, G. Matus, M. Óvári, H. Presser, P. Rudall, I. Sramkóné Gáspár, H. Saslis-Laoudakis, D. Schmidt, A. Shifman, V. Virók and A. V. Yena. This work was supported by a NKTH-OTKA-EU FP7 (Marie Curie action) co-funded MOBILITY grant (OTKA-MB08-A 80332) and a Hungarian “National Excellence Program” European Social Fund co-funded grant (TÁMOP 4.2.4.A/2-11-1-2012-0001). We also acknowledge the support of the Hungarian National Scientific Fund (OTKA-K 108992 and OTKA-PD 109686).

LITERATURE CITED

- Acevedo Rodríguez A, Mesa Coello R. 2013. 152: Adiciones corológicas de *Himantoglossum metlesicsianum* (W. P. Tschner) P. Delforge (Orchidaceae): primera cita para la isla de La Palma (Islas Canarias). *Botanica Macaronésica* 28: 123–128.
- Alibertis A, Alibertis C. 1989. Crète (la) n’a pas fini de livrer ses secrets (2). *L’Orchidophile* 20: 110.
- Álvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.
- Bailey CD, Carr TG, Harris SA, Hughes CE. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435–455.
- Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- Barracough TG, Savolainen V. 2001. Evolutionary rates and species diversity in flowering plants. *Evolution* 55: 677–683.
- Bateman RM. 2012a. Circumscribing genera in the European orchid flora: a subjective critique of recent contributions. *Berichte aus den Arbeitskreisen heimische Orchideen Beiheft* 8: 94–126.
- Bateman RM. 2012b. Circumscribing species in the European orchid flora: multiple datasets interpreted in the context of speciation mechanisms. *Berichte aus den Arbeitskreisen Heimische Orchideen Beiheft* 8: 160–212.
- Bateman RM, Sexton R. 2008. Is spur length of *Platanthera* species in the British Isles adaptively optimized or an evolutionary red herring? *Watsonia* 28: 1–21.
- Bateman RM, Hollingsworth PM, Preston J, Yi-Bo L, Pridgeon AM, Chase MW. 2003. Molecular phylogenetics and evolution of Orchidaceae and selected Habenariaceae (Orchidaceae). *Botanical Journal of the Linnean Society* 142: 1–40.
- Bateman RM, Rudall PJ, Bidartondo MI, et al. 2014. Speciation via floral heterochrony and apparent mycorrhizal host-switching of endemic butterfly

- orchids on the Azorean archipelago. *American Journal of Botany* **101**: 1–23.
- Bateman RM, Rudall PJ, Hawkins JA, Sramkó G. 2013a. Morphometric, molecular, ontogenetic and demographic observations on selected populations of the lizard orchid, *Himantoglossum hircinum*. *New Journal of Botany* **3**: 122–140.
- Bateman RM, Rudall PJ, Moura M. 2013b. Systematic revision of *Platanthera* in the Azorean archipelago: not one but three species, including one of Europe's rarest orchids. *PeerJ* **1**: e218 <http://dx.doi.org/10.7717/peerj.218>.
- Bernardos S, García-Barriuso M, León-Arencibia MC, et al. 2006. A cytotoxic study of three endemic orchids of the Canary Islands. *Annales Botanici Fennici* **43**: 161–166.
- Boberg E, Alexandersson R, Jonsson M, Maad J, Ågren J, Nilsson LA. 2014. Pollinator shifts and the evolution of spur length in the moth-pollinated orchid *Platanthera bifolia*. *Annals of Botany* **113**: 267–275.
- Calonje M, Martín-Bravo S, Dobeš C, et al. 2009. Non-coding nuclear DNA markers in phylogenetic reconstruction. *Plant Systematics and Evolution* **282**: 257–280.
- Carey PD. 1998. Modelling the spread of *Himantoglossum hircinum* (L.) Spreng. at a site in the south of England. *Botanical Journal of the Linnean Society* **126**: 159–172.
- Carey PD. 1999. Changes in the distribution and abundance of *Himantoglossum hircinum* (L.) Sprengel (Orchidaceae) over the last 100 years. *Watsonia* **22**: 353–364.
- Carine MA. 2005. Spatio-temporal relationships of the Macaronesian endemic flora: a relic series or window of opportunity? *Taxon* **54**: 895–903.
- Claessens J, Kleynen J. 2011. *The flower of the European orchid: form and function*. Published by the authors.
- Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 487–494.
- Cornish-Bowden A. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Research* **13**: 3021–3030.
- Curto MA, Puppo P, Ferreira D, Nogueira M, Meimberg H. 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). *Molecular Phylogenetics and Evolution* **63**: 758–767.
- Delforge P. 1999. Contribution taxonomique et nomenclaturale au genre *Himantoglossum* (Orchidaceae). *Naturalistes Belges* **80**: 387–408.
- D'Emerico S, Bianco P, Medagli P. 1992. Karyological studies on Orchidaceae. Tribe Ophrydeae, subtribe Serapiadinae. *Caryologia* **45**: 301–311.
- Devey DS, Bateman RM, Fay MF, Hawkins JA. 2009. Genetic structure and systematic relationships within the *Ophrys fuciflora* aggregate (Orchidaceae: Orchidinae): high diversity in Kent and a wind-induced discontinuity bisecting the Adriatic. *Annals of Botany* **104**: 483–495.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin* **19**: 11–15.
- Drummond A, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.
- Emadzade K, Gehrke B, Peter Linder H, Hörandl E. 2011. The biogeographical history of the cosmopolitan genus *Ranunculus* L. (Ranunculaceae) in the temperate to meridional zones. *Molecular Phylogenetics and Evolution* **58**: 4–21.
- Escudero M, Vargas P, Arens P, Ouborg NJ, Luceno M. 2010. The east-west-north colonization history of the Mediterranean and Europe by the coastal plant *Carex extensa* (Cyperaceae). *Molecular Ecology* **19**: 352–370.
- Font M, Garcia-Jacas N, Vilatersana R, Roquet C, Susanna A. 2009. Evolution and biogeography of *Centaurea* section *Acrocentron* inferred from nuclear and plastid DNA sequence analyses. *Annals of Botany* **103**: 985–997.
- Fuertes Aguilar J, Nieto Feliner G. 2003. Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). *Molecular Phylogenetics and Evolution* **28**: 430–447.
- Good R. 1936. On the distribution of lizard orchid (*Himantoglossum hircinum* Koch). *New Phytologist* **35**: 142–170.
- Gustafsson AL, Verola C, Antonelli A. 2010. Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae: Epidendroideae). *BMC Evolutionary Biology* **10**: 177.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.
- Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multi-locus data. *Molecular Biology and Evolution* **27**: 570–580.
- Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**: 87–112.
- Hughes CE, Eastwood RJ, Bailey CD. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **361**: 211–225.
- Ilut DC, Doyle JJ. 2012. Selecting nuclear sequences for fine detail molecular phylogenetic studies in plants: a computational approach and sequence repository. *Systematic Botany* **37**: 7–14.
- Inda LA, Pimentel M, Chase MW. 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Annals of Botany* **110**: 71–90.
- Juan C, Emerson BC, Oromí P, Hewitt GM. 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology and Evolution* **15**: 104–109.
- Kaiser R. 1993. *The scent of orchids: olfactory and chemical investigations*. Basel: Editiones Roche.
- Karatzas I. 2004. *×Comptoglossum agiasense*, eine seltene Hybride von *Lesvos*. *Journal Europäischer Orchideen* **36**: 951–956.
- Kasperek M, Grimm U. 1999. European trade in Turkish saleg with special reference to Germany. *Economic Botany* **53**: 396–406.
- Kay KM, Whittall JB, Hodges SA. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* **6**: 36.
- Kropf M, Renner S. 2008. Pollinator-mediated selfing in two deceptive orchids and a review of pollinium tracking studies addressing geitonogamy. *Oecologia* **155**: 497–508.
- Kropf M, Sommerkamp E, Bernhardt K-G. 2012. Population dynamics of *Barlia metlesicsiana* (Orchidaceae) endemic to Tenerife (Canary Islands). *Bocconea* **24**: 271–276.
- Kumar S, Skjaeveland A, Orr R, et al. 2009. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* **10**: 357.
- Lahaye RRY, Savolainen V, Duthoit S, Maurin O, van der Bank M. 2008. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. Available from *Nature Precedings*. <http://hdl.handle.net/10101/npre.2008.1896.1>.
- LaJeunesse TC, Pinzón JH. 2007. Screening intragenomic rDNA for dominant variants can provide a consistent retrieval of evolutionary persistent ITS (rDNA) sequences. *Molecular Phylogenetics and Evolution* **45**: 417–422.
- Löve Á. 1976. IOPB chromosome number reports LIII. *Taxon* **25**: 483–500.
- Meusel H, Jäger E, Weinert E. 1965. *Vergleichende Chorologie der Zentraleuropäischen Flora—Karten*. Jena: Gustav Fischer Verlag.
- Molnár V. A, Kreutz CAJ, Óvári M, et al. 2012. *Himantoglossum jankae* (Orchidaceae: Orchidoideae), a new name for a long-misnamed lizard orchid. *Phytotaxa* **73**: 8–12.
- Montieri S, Gaudio L, Aceto S. 2004. Isolation of the LFY/FLO homologue in *Orchis italica* and evolutionary analysis in some European orchids. *Gene* **333**: 101–109.
- Nelson E. 1968. *Monographie und Ikongraphie der Orchidaceen-Gattungen Serapias, Aceras, Loroglossum, Barlia*. Chernerx-Montreux: published by the author.
- Neubig KM, Whitten WM, Carlswald BS, et al. 2009. Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Plant Systematics and Evolution* **277**: 75–84.
- Nieto-Feliner G, Rosselló JA. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* **44**: 911–919.
- Nylander JAA. 2004. MrModeltest 2.3. Program distributed by the author. Uppsala: Evolutionary Biology Centre, Uppsala University.
- Nylander JAA, Olsson U, Alström P, Sanmartín I. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-variance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology* **57**: 257–268.
- Oxelmann B, Lidén M, Berglund D. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* **206**: 393–410.

- Parks M, Cronn RC, Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.
- Pfeifer M, Heinrich W, Jetschke G. 2006. Climate, size and flowering history determine flowering pattern of an orchid. *Botanical Journal of the Linnean Society* 151: 511–526.
- Pfeifer M, Schatz B, Pico FX, et al. 2009. Phylogeography and genetic structure of the orchid *Himantoglossum hircinum* (L.) Spreng. across its European central-marginal gradient. *Journal of Biogeography* 36: 2353–2365.
- Pimentel M, Sahuquillo E, Catalán P. 2007. Genetic diversity and spatial correlation patterns unravel the biogeographical history of the European sweet vernal grasses (*Anthoxanthum* L., Poaceae). *Molecular Phylogenetics and Evolution* 44: 667–684.
- Pridgeon AM, Bateman RM, Cox AV, Hapeman JR, Chase MW. 1997. Phylogenetics of the subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. I. Intergeneric relationships and polyphyly of *Orchis sensu lato*. *Lindleyana* 12: 89–109.
- Rieseberg LH. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28: 359–389.
- Ronquist F, Teslenko M, van der Mark P, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Rosselló JA, Cebrián MC, Mayol M. 2002. Testing taxonomic and biogeographical relationships in a narrow mediterranean endemic complex (*Hippocrepis balearica*) using RAPD markers. *Annals of Botany* 89: 321–327.
- Rouchy JM, Caruso A. 2006. The Messinian Salinity Crisis in the Mediterranean basin: a reassessment of the data and an integrated scenario. *Sedimentary Geology* 188: 35–67.
- Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.
- Sang T, Crawford D, Stuessy T. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.
- Schlüter PM, Kohl G, Stuessy TF, Paulus HF. 2007. A screen of low-copy nuclear genes reveals the LFY gene as phylogenetically informative in closely related species of orchids (*Ophrys*). *Taxon* 56: 493–504.
- Shaw J, Lickey EB, Schilling EE, Small RL. 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: 275–288.
- Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.
- Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *ADH* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301–1315.
- Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.
- Sramkó G, Óvári M, Yena AV, et al. 2012. Unravelling a century of misuse: typification of the name *Himantoglossum caprinum* (Orchidaceae: Orchideae). *Phytotaxa* 66: 21–26.
- Stierli-Schneider J. 2004. Beitrag zur *Barlia metlesicsiana* Teschner auf Teneriffa. *Journal Europäischer Orchideen* 36: 735–744.
- Ströhlein H, Sundermann H. 1972. Chromosomenzählungen bei Erdorchideen. *Jahresbericht naturwissenschaftlichen Vereins Wuppertal* 25: 75–80.
- Sun M, Gross K, Schiestl FP. 2014. Floral adaptation to local pollinator guilds in a terrestrial orchid. *Annals of Botany* 113: 289–300.
- Sundermann H. 1973. *Himantoglossum* (Loroglossum) *hircinum*-*caprinum*-*calcaratum*-affine. *Acta Botanica Academiae Scientiarum Hungaricae* 19: 367–374.
- Sundermann H, Von der Bank C. 1977. Neue Beiträge zur Zytotaxonomie der Erdorchideen II (*Ophrys*, *Orchis*, *Neotinea*, *Steveniella* und *Gennaria*). *Orchideen* 28: 167–175.
- Swofford DL. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland, Massachusetts: Sinauer Associates.
- Thompson JD. 2005. *Plant evolution in the Mediterranean*. Oxford: Oxford University Press.
- Vakhrameeva MG, Tatarenko IV. 2008. *Himantoglossum caprinum* (M. Bieb.) C. Koch. In: Vakhrameeva MG, Tatarenko IV, Varlygina TI, Torosyan GK, Zagulski MN. eds. *Orchids of Russia and adjacent countries (within the borders of the former USSR)*. Ruggell: A. R. Gantner, 274–276.
- Wendel J, Doyle J. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis D, Soltis P, Doyle J. eds. *Molecular systematics of plants II*. London: Chapman & Hall, 265–296.
- White TJ, Bruns TD, Lee S, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ. eds. *PCR protocols: a guide to methods and applications*. San-Diego: Academic Press, 315–322.
- Yu Y, Harris AJ, He X. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution* 56: 848–850.
- Zimmer EA, Wen J. 2012. Using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774–785.