Plant Systematics and Evolution (2020) 306:49
https://doi.org/10.1007/s00606-020-01676-6

ORIGINAL ARTICLE

Extremely low genetic diversity in the European clade of the model bryophyte *Anthoceros agrestis*

Thomas N. Dawes[1,2] · Juan Carlos Villarreal A.[3,4] · Péter Szövényi[5] · Irene Bisang[6] · Fay-Wei Li[7,8] · Duncan A. Hauser[7,8] · Dietmar Quandt[9] · D. Christine Cargill[10] · Laura L. Forrest[1]

Received: 2 May 2019 / Accepted: 13 March 2020 / Published online: 4 April 2020

Abstract

The hornwort *Anthoceros agrestis* is emerging as a model system for the study of symbiotic interactions and carbon fixation processes. It is an annual species with a remarkably small and compact genome. Single accessions of the plant have been shown to be related to the cosmopolitan perennial hornwort *Anthoceros punctatus*. We provide the first detailed insight into the evolutionary history of the two species. Due to the rather conserved nature of organellar loci, we sequenced multiple accessions in the *Anthoceros agrestis–A. punctatus* complex using three nuclear regions: the ribosomal spacer ITS2, and exon and intron regions from the single-copy coding genes *rbc*S and phytochrome. We used phylogenetic and dating analyses to uncover the relationships between these two taxa. Our analyses resolve a lineage of genetically near-uniform European *A. agrestis* accessions and two non-European *A. agrestis* lineages. In addition, the cosmopolitan species *Anthoceros punctatus* forms two lineages, one of mostly European accessions, and another from India. All studied European *A. agrestis* accessions have a single origin, radiated relatively recently (less than 1 million years ago), and are currently strictly associated with agroecosystem habitats.

Keywords Genome streamlining · Hornworts · Phylogenetics · Phytochrome

Handling Editor: Andreas Tribsch.

Thomas N. Dawes and Juan Carlos Villarreal A. have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00606-020-01676-6) contains supplementary material, which is available to authorized users.

✉ Laura L. Forrest
l.forrest@rbge.ac.uk

1 Royal Botanic Garden Edinburgh, Edinburgh, Scotland

2 Victoria University of Wellington, Wellington, New Zealand

3 Department of Biology, Université Laval, Quebec, Canada

4 Smithsonian Tropical Research Institute, Panama City, Panama

5 University of Zurich and Zurich-Basel Plant Science Center, Zurich, Switzerland

6 Swedish Museum of Natural History, Stockholm, Sweden

7 Boyce Thompson Institute, Ithaca, NY, USA

8 Plant Biology Section, Cornell University, Ithaca, NY, USA

9 Nees-Institut für Biodiversität Der Pflanzen, Bonn, Germany

10 Australian National Botanic Gardens, Canberra, Australia

Introduction

Large-scale projects have implemented model organisms in each of the major green plant phyla, including algae and bryophytes (Chang et al. 2016). One of the latest model organisms to be sequenced is the bryophyte *Anthoceros agrestis* (Szövényi 2016). This hornwort is emerging as a model for studies of plant N_2-fixing bacterial interactions (Villarreal A and Renzaglia 2015), carbon concentrating mechanisms (Li et al. 2017) and stomatal evolution (Renzaglia et al. 2017).

The choice of a plant species as a model organism requires balancing economical and evolutionary significance with pragmatism: the ideal model species should have a small genome size, be easy to grow and genetically tractable. Firstly, a small genome facilitates whole genome sequencing. In this regard, *Anthoceros agrestis* is ideal. It has one of the smallest genome sizes among plant model systems and the smallest reported bryophyte genome, with a genome size of 70–110 Mbp, and only 6.98% transposable elements (Szövényi 2016). Easy laboratory and genetic manipulation are practical necessities for the study of organ development, gene function and genome architecture. *Anthoceros*

agrestis also fits this criterion: its life cycle can be completed in 1–3 months in field and laboratory conditions (Bisang 2003), and *Agrobacterium*-mediated genetic transformation is currently being developed for the species (E. Frangedakis unpublished data). Finally, hornworts are at centre stage in the debate on early land plant diversification (Puttick et al. 2018), and thus are highly evolutionarily significant.

Instigating a novel plant species as a "model system" has frequently prompted detailed investigations of its evolutionary history, placing it into a broader biological context. For example, the moss model system *Physcomitrella patens* (473 Mbp) shows a complex evolutionary history, with the genus *Physcomitrella* apparently polyphyletic (McDaniel et al. 2010; Beike et al. 2014; Medina et al. 2019). In contrast, the liverwort model system, *Marchantia polymorpha* ssp *ruderalis* (226 Mbp; Bowman et al. 2017) has a relatively recent origin relative to ssp *monitvagans* and ssp *polymorpha* (~ 5 Ma) (Villarreal A et al. 2016). The most recently chosen bryophyte model organisms include the fire moss *Ceratodon purpureus* (340 Mbp; McDaniel and Shaw 2005), which has a history of introgression, recurrent intercontinental dispersal and selective sweeps and the peat mosses *Sphagnum fallax* (473 Mbp; Shaw et al. 2016) and *S. magellanicum,* which also have complex histories, with hybridisation and the presence of geographically structured species complexes.

The taxonomic history of *Anthoceros agrestis* is difficult to separate from that of the widespread perennial hornwort, *Anthoceros punctatus*. Both species are found in or near man-made habitats (near roads, paths, ditches, in agricultural fields and waste land) and are widely distributed in Europe, North America and Asia (Proskauer 1958; Paton 1979; Schuster 1992). *Anthoceros agrestis* has been recognised informally for decades, and was formally described by Paton (1979), based on British plant specimens; however, due to their identical spore architecture, *A. agrestis* has often been considered a subspecies or variety of *A. punctatus* (Proskauer 1958; Schuster 1992). Conversely, there is a suggestion that *A. punctatus* could be a neopolyploid form of *A. agrestis* (Proskauer 1957): A collection of the species from Portugal was reported to have $n = 10$ chromosomes, instead of the five chromosomes that are typical for monoicous hornworts (Proskauer 1957). The larger reported genome size of *Anthoceros puctatus* (129–178 Mbp; Li unpublished data; Bainard and Villarreal A 2013) compared to a size of 70–110 Mbp for *A. agrestis* (Szövényi 2016; Szövényi unpublished data) also suggests that polyploidy may be involved in this lineage. The major morphological differences between the two species are the size of their antheridia, and to a lesser degree their thallus and sporophyte size (*A. agrestis* being the smaller of the two). This also seems to tie in with the hypothesis of a polyploid origin for *A. punctatus*, given the links between cell/spore size and genome size in other plant groups. Additionally, *A. punctatus* is a rather longer lived and more robust species (Proskauer 1958; Paton 1979; Schuster 1992).

Phylogenetic analyses using single accessions of both species confirm their close relationship (Duff et al. 2007, using plastid *rbc*L, mitochondrial *nad*5 and nuclear 18S markers), with a more recent work suggesting that the Ascension Island neoendemic *Anthoceros cristatus* is also involved in the species complex (Villarreal A et al. 2017, using plastid *rbc*L and *mat*K and mitochondrial *nad*5 markers).

There are some ecological differences across the geographic range of *Anthoceros agrestis*. In North America, for example, *Anthoceros agrestis* is found associated with crops and in recently abandoned fields, whereas in Europe, it is found almost entirely in arable fields that are regularly tilled (Bisang 1992).

In many parts of Europe, *Anthoceros agrestis* is a short lived summer annual, with the plants overwintering as spores (Paton 1979; Bisang 1995). Sporophytes of the European *A. agrestis* mature from late July/August to December, depending on climate, weather conditions and management (Bisang 1995, 1998), doing best when the fields are stubble rather than growing crops and/or spring cultivated (Bisang 1998; Blockeel et al. 2014). The populations that grow up in these fields are dynamic, with a fast growth rate, reproductive maturity and death occurring more or less continuously throughout the season (Bisang 1995). Unfortunately, the population size is dwindling in many parts of its European distribution range due to changes in farm practices and habitat loss (Schnyder et al. 2019); the species was recently assessed as near-threatened in the Red List of European bryophytes.

Adaptations to agricultural environments seem to be an intrinsic part of the biology and life cycle of *Anthoceros agrestis,* giving rise to the hyopthesis that the species has evolved alongside the domestication and spread of agricultural crops in Eurasia and America, which occurred in the last 9–12,000 years (Matsuoka et al. 2002; Purugganan and Fuller 2009). The annual life cycle of the species, potentially imposed by farming practices, could then have triggered streamlining of its genome (Villarreal A et al. 2017).

Based on the natural history and preliminary phylogenetic information, there are several distinct scenarios for the origin of *A. agrestis*. There may have been a single origin of *A. agrestis,* accessions of which would resolve monophyletically within a clade of *A. punctatus* accessions. *Anthoceros agrestis*, as a ruderal species associated with early human settlements, could have spread rapidly around the world along with the Holocene agricultural revolution, in which case we would also expect a molecular signature of rapid, recent population expansion (an agricultural ecotype). In contrast*, Anthoceros agrestis* may have been the progenitor species for *Anthoceros punctatus.* In this case, we would

expect to see *A. punctatus* phylogenetically nested within an *A. agrestis* complex (essentially the reverse of the previous scenario). The phylogenetic analyses for such scenarios will assume that markers used have coalesced for the taxa and that they now form monophyletic entities. It is also possible that the *Anthoceros agrestis* lifestyle may have arisen multiple times as a response to similar ecological conditions, making the morphological taxon polyphyletic.

In the present study, to assess the phylogenetic affinities of *A. agrestis,* we have sequenced three nuclear genes from multiple accessions of *A. agrestis* and *A. punctatus*, focusing mostly on Europe and North America.

Materials and methods

Molecular methods

We sequenced 51 accessions representing *Anthoceros punctatus* and *A. agrestis* from Europe, the Americas, Australia and Asia (Online Resource 1, Fig. 1). Given some uncertainty as to what the closest extant relative of *A. agrestis* and *A. punctatus* is, we included several related species of *Anthoceros* as outgroup taxa, based on previous studies (Duff et al. 2007; Villarreal A et al. 2017). We included five accessions of *A. neesii*, three accessions of *A. cristatus* and one accession each of *A. venosus, A. lamellatus* and *A. fragilis*. All those species share similar spore morphology and have been associated with the *A. punctatus* complex (Schuster 1992; Villarreal A et al. 2015). Voucher information and GenBank accession numbers are available in Online Resource 1. The DNA extraction and PCR amplification followed standard protocols (available upon request). The data set includes nucleotide sequences from two single-copy nuclear loci: part of the RuBisCO small subunit (*rbc*S) that includes sequence data from three exons (1, 2 and 3) and the two introns between them; part of the phytochrome gene that includes sequence data from exons 2 and 3 and the intron between them; and the internal transcribed spacer 2 region (ITS2). Primers to amplify the *rbc*S and phytochrome loci were newly generated as part of this study (Online Resource 2). In previous work and as part of a pilot study for this paper, we amplified several standard plastid intron and spacer regions; we ruled out inclusion of these loci (e.g. *psb*A-trnH; trnL-trnF) due to low genetic variability between accessions.

We used Geneious 9.0.5 (Biomatters Limited) to generate an initial alignment of the nucleotide sequences, with subsequent manual augmentation. Due to some failed amplifications, the final data matrix (hereafter called dataset I) comprises forty-seven accessions that have sequence data for at least two loci, including thirty accessions of *A. agrestis*, twelve accessions of *A. punctatus*, two accessions of *A. neesii* and one each of *A. cristatus, A. lamellatus, A. venosus* and *A. fragilis.* A subset of the matrix used for the dating analyses (data set II) contains 20 accessions and includes

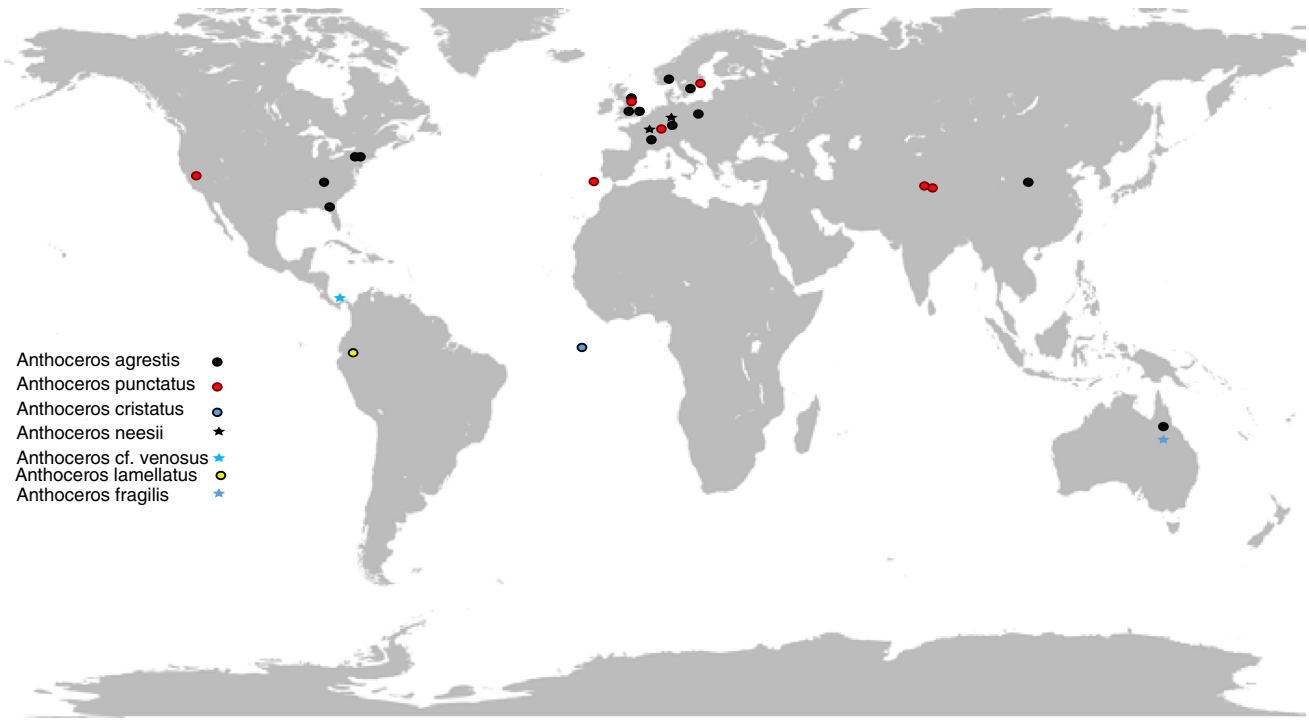

Fig. 1 Geographic location of samples of *Anthoceros* included in this study. The provenance of each sample is found in Table S1

representatives of *A. agrestis*, *A. punctatus* and the outgroup species *A. neesii, A. cristatus* and *A. fragilis*.

Phylogenetic analyses

We analysed each locus separately, and also as a single partitioned data set with models suggested by PartitionFinder (Lanfear et al. 2012), under the maximum likelihood criterion (ML), using RAxML black box (Stamatakis 2014) with 500 bootstrap replicates (MLB). Basic statistics on each data set were retrieved using PAUP* (Swofford 2002). Bayesian analyses on the combined data set were conducted in MrBayes 3.2 (Ronquist et al. 2012), using the default two runs and four chains, with default priors on most parameters. To assess burn-in and convergence, we compared the bipartitions across the two runs. Convergence was usually achieved in MrBayes after 2×10^7 generations, with trees sampled every 10,000th generation for a total length of 10×10^6 generations; we discarded 25% of each run and then pooled runs. The final matrix is available at Dryad https://doi.org/10.5061/dryad.qnk98sfcb. All analyses were run under the CIPRES platform (Warnow 2010).

Molecular clock dating

We used a root prior from a previous study (Villarreal A et al. 2015) to date a phylogeny produced using data set II, with *A. cristatus*, *A. fragilis* and *A. neessii* as outgroup taxa. There is no reliable fossil described from *Anthoceros*, although spores similar to those of *A. punctatus* and referred to as *Rudolphisporis rudolfii* have been reported from Late Miocene (Machnín, Bohemia), Pliocene and Pleiostocene deposits from Europe (Krutzsch 1963). We have decided to use a root prior because of the uncertainty around this fossil spore and the lack of precise dating of the stratum. We gave the root a normal prior, with a mean of 20 Ma and sd of 3, to account for the highest posterior density (HPD) from a previous study (Villarreal A et al. 2015). In addition, we used the crown age of the *A. punctatus–agrestis* group from this previous study, applying it to the ingroup with a mean of 6 Ma and sd of 1. The last calibration used was the age of Ascension Island, as a proxy age for the neoendemic *A. cristatus*. Ascension Island is the tip of an undersea volcano that is thought to have emerged from the ocean one million years ago (Ashmole and Ashmole 2000). We used a normal prior with a mean of 1 and sd of 0.25. To explore the effective priors, we ran analyses with either one or two calibrations at a time, and one on an empty alignment, to compare the frequency distribution of age estimates for each calibrated node with the prior. Bayesian divergence time estimation used a Yule process tree prior with unlinked data partitions, using same substitution models suggested by PartitionFinder. The analyses were done using an uncorrelated lognormal (UCLN) relaxed clock model. The MCMC chains were run for 900 million generations, with parameters sampled every 10,000th generation using BEAST 1.8.3 (Drummond et al. 2012). Tracer 1.6 (Rambaut et al. 2014) was used to assess effective sample sizes (ESS) for all estimated parameters and to decide the appropriate percentages of burn-in. We verified that all ESS values were > 200. Trees were combined in TreeAnnotator 1.8 (part of the BEAST package), and maximum clade credibility trees with mean node heights were visualised using FigTree 1.4.0. We report the HPD intervals (the interval containing 95% of the sampled values).

Results

Phylogenetic analyses

Nucleotide sequences were generated for the three nuclear loci (aligned length 1,965 bp): part of the small RuBisCO subunit gene including exons and spacers (*rbc*S), part of the phytochrome gene including exons and a spacer and the internal transcribed spacer 2 region (ITS2). Sequencing was most successful for the ITS2 locus, however, there are only 14 parsimony informative characters (3.0%). The phytochrome locus comprises a similar proportion of parsimony informative characters (2.9%), while the *rbc*S locus contains most of the variation in the data, comprising 17% parsimony informative characters (Online Resource 3). Single locus analyses produce different levels of resolution and support. The phytochrome and *rbc*S analyses support few monophyletic groups and resolve the backbone of the phylogeny poorly (Online Resource 4). Plants with *Anthoceros agrestis* morphology are placed in three distinct and statistically supported lineages: a highly supported European *A. agrestis* clade, a Chinese/Australian clade and an American *A. agrestis* clade (only supported in MrBayes run, data not shown). A monophyletic *Anthoceros punctatus* is not recovered in any analysis.

The three-locus concatenated data set (data set I) recovers a monophyletic ingroup (79% MLB, 1.00 PP) (Fig. 2a). The Ascension Island endemic *Anthoceros cristatus* is sister to a clade that includes multiple accessions of *A. agrestis* and *A. punctatus, A. venosus* from Panama and *A. lamellatus* from Ecuador. The relationships between these lineages are not supported in our phylogenetic analyses; they are separated by extremely short branches (Fig. 2a). The ingroup segregates into (Fig. 2a):

A. An unsupported clade/grade containing *A. punctatus* from India, *A. lamellatus* and *A. venosus* (62% MLB, 0.74 PP);

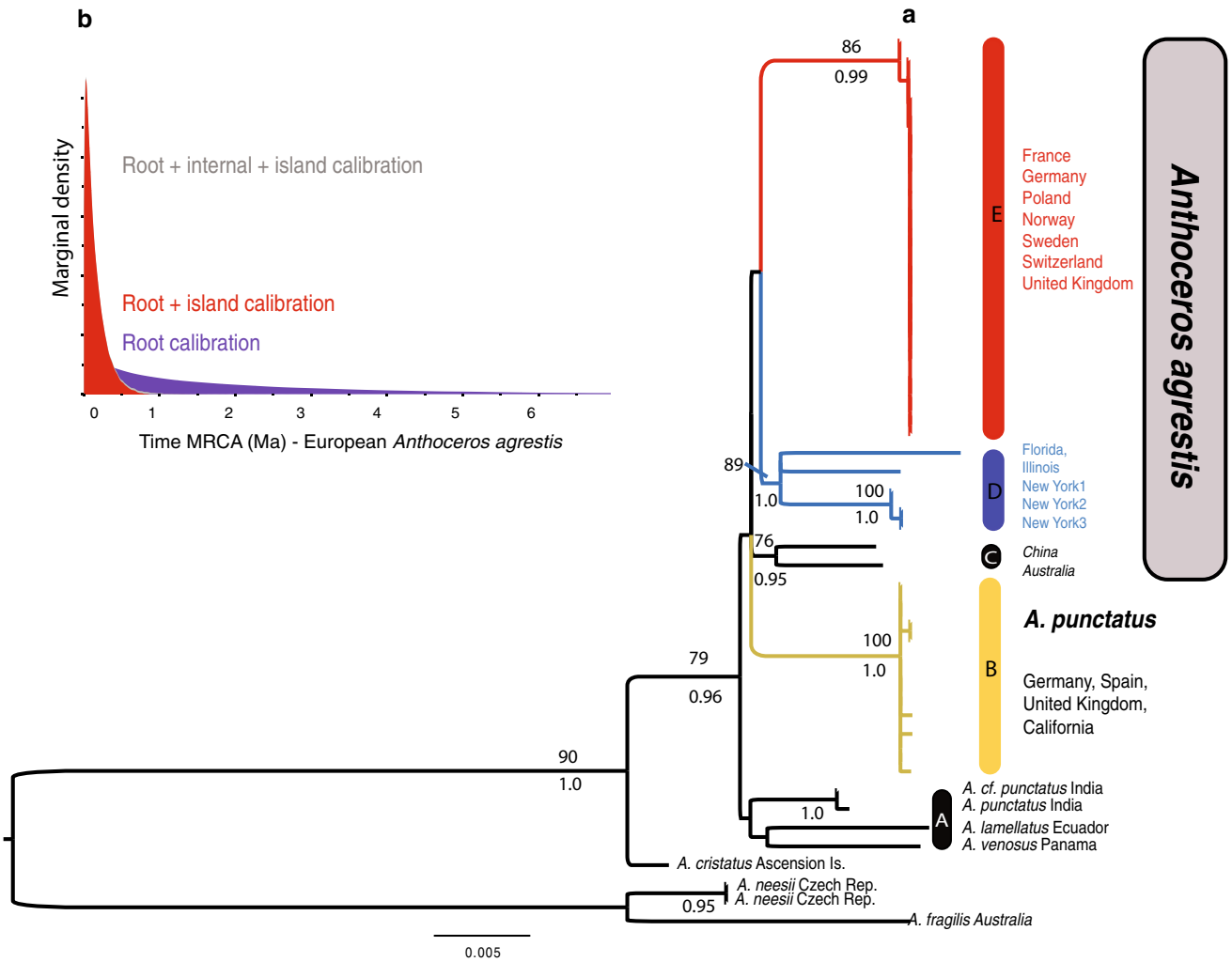

Fig. 2 **a** Maximum likelihood (ML) phylogram from the combined phytochrome, *rbc*S and ITS2 DNA sequence analyses of *Anthoceros*. A single origin of European *A. agrestis* accessions is highlighted (red), as is the US clade of *A. agrestis* (blue) and a Europe–US clade of *A. punctatus* (yellow); ML bootstrap and posterior probability values are annotated as symbols on the trees. Clades A-E are annotated in the text (**b**). Age of diversification of European *A. agrestis* using three calibration schemes and their marginal density: (1) a root calibration (purple), (2) root and island calibration (red) and, (3) root, internal and island calibration. Two of them (2, 3) are overlapping, and therefore, only the root + island calibration is visible. The node ages and HPD intervals are shown in Online Resource 5

B. The *A. punctatus* accessions from Europe and the USA (California) (100% MLB, 1.00 PP), with nearly invariable sequences;
C. The Chinese and Australian accessions of *Anthoceros agrestis* (76% MLB; 0.95 PP);
D. All the *Anthoceros agrestis* samples from the USA (New York, Florida and Illinois) (89% MLB, 1.00 PP);
E. All European accessions of *Anthoceros agrestis*, from France, Germany, Norway, Poland, Sweden, Switzerland and the United Kingdom (86% MLB, 1.00 PP), with nearly invariable sequences.

Molecular clock dating

Using the root prior and the island calibration, the crown age of the European *A. agrestis* lineage is 160,000 years (HPD—0.000001–0.4 Ma) (Fig. 2b, Online Resource 5, Online Resource 6). Under this calibration scheme, the crown age of the European and the US *A. punctatus* lineage is 0.24 Ma (HPD—0.000003–0.6 Ma). Other calibration schemes give slightly older estimates with overlapping HPD; however, the root calibration resulted in an older age

for the crown European *A. agrestis* with a broader HPD (Fig. 2b, Online Resource 5; Online Resource 6).

Discussion

A single and recent origin of the European *Anthoceros agrestis*

Our analyses recover a single recent origin of the European lineage of the model plant *Anthoceros agrestis* (Figs. 2, 3). However, overall, the concatenated gene analyses support several distinct *A. agrestis* lineages: the European clade and a clade from the USA, while accessions of *A. agrestis* from China and Australia are genetically distinct from both these lineages. Although the exact relationships of these lineages to each other cannot be ascertained with our data, this is likely due to a relatively rapid diversification event in the complex (Online Resource 6a-c). Multiple origins of either species are possible, concurrent with this rapid diversification.

Accessions of *Anthoceros punctatus* from Europe and the USA (California) form a monophyletic group, while *Anthoceros punctatus* from India, *A. lamellatus* from Ecuador and *A. venosus* from Panama form another clade (albeit with low MLB/PP support). *Anthoceros lamellatus* and *A. venosus* are morphologically similar to *A. punctatus*, with slight variation in the spore architecture (Schuster 1992; Villarreal A et al. 2015). The surprising finding that the Californian accession of *A. punctatus* resolves with the European *A. punctatus* clade with little genetic divergence suggests either recent long-distance or human-mediated dispersal. Long-distance dispersal is commonplace among bryophytes species: Recently biotic (e.g. birds) and abiotic vectors have been shown to be responsible for even bipolar distributions (Cargill et al. 2013; Lewis et al. 2014).

The genetically distinct lineages found within *A. agrestis*– *A. punctatus* are not hierarchically congruent with the morphological concepts of either species. This seems to be a recurrent theme in bryophyte systematics (Cargill et al. 2013; Lewis et al. 2014; Biersma et al. 2017).

The short branch lengths and low levels of support for the branching pattern between lineages of *A. agrestis* and *A. punctatus* in the phylogenetic analyses make it difficult to rule out a single origin of *A. agrestis*. However, the lineages of *A. agrestis* have deep levels of genetic divergence between Europe and America, indicative of a long separation. Proskauer's suggestion (1957) that *A. punctatus* may have evolved from *A. agrestis* is not supported by our data, given that the *A. punctatus* lineages occur in several places,

Fig. 3 Known European distribution of *Anthoceros agrestis* shown in yellow (10 March 2017), with permission from LIFE IUCN European Red List project (contract number LIFE14PREBE001). Map sources: Esri, HERE, deLorme, Intermap increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors and the GIS user community. Inset: a field photograph of Czech *A. agrestis*, courtesy of L. Hedenäs

interspersed with lineages of *A. agrestis*. Furthermore, we did not pick up alleles or paralogues of the reportedly single-copy nuclear loci in our sequence data, as we might find in a polyploid plant, although amplicon-based Sanger sequencing is not the best way to explore gene copy number. Our data instead show several genetically distinct lineages with *A. agrestis* life history types, in China, Australia, the Americas and Europe (Online Resource 4). Further molecular and morphological study is required to establish the taxonomic status of these lineages—given that *A. agrestis* is typified by plants from near Oxton, Nottinghamshire, England (Paton 1979), it may be that other names will be required for the non-European lineages.

One of most important implications of this current study, though, is that any data relating to the "model" organism *Anthoceros agrestis* needs to be qualified with the origin of the material. For example, the plants that have been high-throughput sequenced for genome assembly came from near Hirschbach, Germany and Berwickshire, Scotland (Szövényi et al. 2015). The genome size estimate of 71 Mbp is based on Kmers analysis of Illumina sequencing data of material from near Hirschbach, Germany (Szövényi et al. 2015), while an estimate of 83 Mbp is based on flow cytometry of material from an unknown location (Greihhuber 2005; Leitch and Bennett 2007). A further estimate of 110 Mbp is based on Kmer analysis of Illumina sequencing data of material from Berwickshire, Scotland (Li unpublished data). There is, as yet, no information about the genome size or chromosome number of plants belonging to the American or Asian *A. agrestis* lineages. An estimate of 178 Mbp for *Anthoceros punctatus* is based on flow cytometry of a sample from Australia (Bainard and Villarreal A 2013), while an estimate of 129 Mbp is based on Kmer analysis of Illumina sequencing data from a sample from Humboldt County, California, USA (Li unpublished data). Samples of the related species *Anthoceros lamellatus* from Colombia and Panama, measured using flow cytometry, have genomes of 185 Mbp and 193 Mbp, respectively (Bainard and Villarreal A 2013), although there are no chromosome number counts for these plants, to correlate polyploidy with geographical origin or phylogenetic affinities.

The low level of nucleotide diversity and correspondingly short branch lengths in the crown European *A. agrestis* lineage reflect a recent divergence of extant populations, and the dated phylogeny supports this, placing the divergence of these accessions in the last 1 million years (Fig. 2b), depending on the analysis. However, given the low amount of genetic variation that we have observed within European *A. agrestis*, it is not possible to determine with certainty whether the divergence actually took place as recently as the last 10,000–20,000 years, or whether the low observed genetic diversity within the European *A. agrestis* lineage could be due to recolonisation after glaciation.

Alternatively, we propose that *A. agrestis* may have been a ruderal species, effectively pre-adapted, by its occurrence in disturbed habitats, to human settlements; the species may subsequently have become "co-opted" more recently to its current rotating crop-field habitats as a by-product of the crop domestication process. In this case, the *A. agrestis* lineage itself may substantially predate the establishment of agricultural cultivation in Europe, surviving in refugia in interglacial periods, with the species only latterly becoming associated with agricultural practices.

Estimation of trajectories of changes in effective population size over time, using pairwise sequentially Markovian coalescent (PSMC), could cast more light on the population history of European *A. agrestis*. A recent pilot SNP analysis of two individuals of *A. agrestis*, one collected near Hirschbach, Germany and one collected in Berwickshire, Scotland, reported nearly 2 SNPs/kbp (Szövényi et al. 2015). This level of polymorphism is similar to reports from crop and weedy accessions of the cereal *Sorghum bicolor* (L.) Moench (2.7 SNPs/kbp) (Morris et al. 2012) and suggests that despite the low genetic diversity and lack of phylogenetic resolution in analyses using the three loci we have sampled, there is promising information to be harvested from a genome-wide study across the European *A. agrestis* lineage.

The other two *A. agrestis* lineages seem to differ from the European populations in several respects. From our phylogeny, it is evident that the American *A. agrestis* clade is more structured than the European clade, with longer branch lengths suggesting genetic isolation. The annual *A. agrestis* plants from Florida and Illinois are not strictly associated with agricultural fields (Schuster 1992), and they produce sporophytes in the spring instead of late summer/autumn. *Anthoceros agrestis* from New York is associated with maize and soybean fields and, like the European plants, the sporophytes are found in the late summer and autumn. In general, there is limited data about the natural history of *A. agrestis* in North America and Asia, so the tight association of these populations with agricultural practices, as seen in Europe, cannot be sufficiently confirmed. Since historic agricultural practices in eastern Asia are very different from those seen in Europe, it is entirely conceivable that ecological and evolutionary constraints on the plants will not be the same.

Agriculture management and land use have undergone immense changes since the Second World War, with very negative impacts on the wildlife of agricultural ecosystems (e.g. Tscharntke et al. 2005; BirdLife International 2015). The intensification of crop farming practices has resulted in an alarming decline of, and threat to, many arable species that are adapted to regular disturbance through tillage (e.g. Donald et al. 2001; Meyer et al. 2013). Likewise, hornwort populations have decreased in many parts of Europe (Bisang et al. 2009; Schnyder et al. 2019). Hornworts are not known from any places that would be considered "primary" habitats

in Europe today, and meanwhile, the negative impact of agricultural intensification on biodiversity is predicted to continue (Reidsma et al. 2006). The low level of sequence variation among European accessions of *A. agrestis,* as shown in this study, indicates limited genetic variation within the lineage. The corollary of this is that the species may have low adaptability to environmental changes, leaving the European clade of *A. agrestis* at risk of extinction in the absence of targeted conservation actions (Bisang et al. 2009; Schnyder et al. 2019).

Conclusion

We have studied the evolutionary history of the model hornwort species, *Anthoceros agrestis*. Our data suggest that all European accessions of the species have a single origin, with little genetic divergence between them. However, lineages of *A. agrestis* from other geographic areas may have independent origins. We suggest that the European *A. agrestis* may have been a ruderal species, pre-adapted to disturbed settings, that was "co-opted" to the rotating field habitats that resulted from crop domestication. This hypothesis could be tested in other bryophyte species that are strongly associated with arable field habitats (Porley 2008), as well as in vascular plants that are associated with rotating crops in Europe and North America, and compared with patterns of evolution in rice-paddy-associated bryophytes. Our study prompts consideration of secondary implications of agricultural practices on non-crop plant population biology.

Acknowledgements We would like to acknowledge the LIFE IUCN European Red List team and Norbert Schnyder for access to population size data for European *Anthoceros agrestis* from the Red List work (contract number LIFE14PREBE001). Thanks to Ilia Leitch for information about the *Anthoceros agrestis* genome size, and to the many colleagues who improved our sampling by sending us their *Anthoceros* collections. We are also grateful to the anonymous reviewers of a previous version of this manuscript, whose comments have hopefully led to some beneficial changes. This study formed the research thesis element of Tom Dawes' MSc degree at the University of Edinburgh and Royal Botanic Garden, Edinburgh.

Author contribution LLF, JCVA and TD contributed to the study conception and design. Material preparation, data collection and analyses were performed by TD, JCVA and LLF. The first draft of the manuscript was written by LLF, JCVA and TD, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding JCVA acknowledges support from the Earl S. Tupper Fellowship (Smithsonian Tropical Research Institute) and le Conseil de recherches en sciences naturelles et en genie du Canada RGPIN/05967-2016; TD and LLF acknowledge funding from the Royal Botanic Garden, Edinburgh, which is supported by the Scottish Government's Rural and Environment Science and Analytical Services Division. PS acknowledges funding from Swiss National Science Foundation grant nos 160004 and 131726, the Georges and Antoine Claraz Foundation, The Forschungskredit and the University Research Priority Program.

Availability of data and materials The data set generated and analysed during the current study is available from GenBank and DRYAD [https://datadryad.org/stash/dataset/doi:10.5061/dryad.qnk98sfcb].

Information on Electronic Supplementary Material

Online Resource 1. Voucher table for the specimens included in the multigene analyses.
Online Resource 2. PCR and Sanger sequencing primers used in this study.
Online Resource 3. Basic statistics for the three loci.
Online Resource 4. Phylogeny produced using phytochrome (**a**) and *rbc*S (**b**) nucleotide sequence data, and annotated with ML bootstrap values and posterior probabilities as symbols on the trees.
Online Resource 5. *Anthoceros* divergence dates (in Ma) from the three calibration schemes (see text). (**a**) UCLN "relaxed" clock with a single calibration (root calibration) of 20 Ma and sd of 3; (**b**) UCLN "relaxed" clock with the root calibration and the island calibration on the node of the neoendemic *A. cristatus,* with a normal prior with a mean of 1 and sd of 0.25; (**c**) UCLN "relaxed" clock (Online Resource 4) with three calibrations (Online Resource 6): a normal root calibration, the *A. cristatus* calibration, and the crown age of the *A. punctatus–agrestis–lamellatus–venosus* group from the previous study applied to the ingroup, with a mean of 6 Ma and sd of 1.
Online Resource 6. (**a**) Uncorrelated lognormal (UCLN) relaxed clock dated tree produced using the root calibration scheme (normal prior, mean 20 Ma, sd 3) based on values in Villarreal A et al. 2015. (**b**). UCLN dated tree produced using a 2-calibration scheme, based on the root calibration and a crown age for *A. punctatus–agrestis* (mean 6 Ma, sd 1) based on values in Villarreal A et al. 2015. (**c**). UCLN dated tree produced using a 3-calibration scheme, based on the root calibration, a crown age for *A. punctatus–agrestis* and the age of Ascension Island as a proxy for *A. cristatus* (normal prior, mean 1 Ma, sd 0.25).

References

Ashmole P, Ashmole M (2000) St. Helena and Ascension Island: a natural history. Anthony Nelson, Oswestry

Bainard JD, Villarreal A JC (2013) Genome size increases in recently diverged hornwort clades. Genome 56:431–435. https://doi.org/10.1139/gen-2013-0041

Beike AK, von Stackelberg M, Schallenberg-Rüdinger M, Hanke ST, Follo M, Quandt D, McDaniel SF, Reski R, Tan BC, Rensing SA (2014) Molecular evidence for convergent evolution and allopolyploid speciation within the *Physcomitrium-Physcomitrella* species complex. BMC Evol Biol 14:158. https://doi.org/10.1186/1471-2148-14-158

Biersma EM, Jackson JA, Hyvönen J, Koskinen S, Linse K, Griffiths H, Convey P (2017) Global biogeographic patterns in bipolar moss species. Roy Soc Open Sci 4:170147. https://doi.org/10.1098/rsos.170147

Bird Life International (2015) Why agriculture is the greatest threat to European biodiversity. Available at: https://www.birdlife.org/europe-and-central-asia/news/why-agriculture-greatest-threat-european-biodiversity. Accessed 23 Mar 2018

Bisang I (1992) Hornworts in Switzerland—endangered? Biol Conservation 59:145–149. https://doi.org/10.1016/0006-3207(92)90574-7

Bisang I (1995) On the phenology of *Anthoceros agrestis* [Anthocerotae, Anthocerotaceae], with special reference to Central Europe. Fragm Florist Geobot 40:513–518

Bisang I (1998) The occurrence of hornwort populations (Anthocerotales, Anthocerotopsida) in the Swiss Plateau: the role of management, weather conditions and soil characteristics. Lindbergia 23:94–104

Bisang I (2003) Population development, demographic structure, and life cycle aspects of two hornworts in Switzerland. Lindbergia 28:105–112

Bisang I, Bergamini A, Lienhard L (2009) Environmental-friendly farming in Switzerland is not hornwort-friendly. Biol Conservation 142:2104–2113. https://doi.org/10.1016/j.biocon.2009.04.006

Blockeel TL, Bosanquet SDS, Preston P (eds) (2014) Atlas of British and Irish bryophytes, vol. 1. Pisces Publications and British Bryological Society, Newbury

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, Adam C, Aki SS, Althoff F, Araki T, Arteaga-Vazquez AM, Balasubrmanian S, Barry K, Bauer D, Boehm CR, Briginshaw L, Caballero-Perez J, Catarino B, Chen F, Chiyoda S, Chovatia M, Davies KM, Delmans M, Demura T, Dierschke T, Dolan L, Dorantes-Acosta AE, Eklund DM, Florent SN, Flores-Sandoval E, Fujiyama A, Fukuzawa H, Galik B, Grimanelli D, Grimwood GU, Hamada T, Haseloff J, Hetherington AJ, Higo A, Hirakawa Y, Hundley HN, Ikeda Y, Inoue K, Inoue SI, Ishida S, Jia Q, Kakita M, Kanazawa T, Kawai Y, Kawashima T, Kennedy M, Kinose K, Kinoshita T, Kohara Y, Koide E, Komatsu K, Kopischke S, Kubo M, Kyozuka J, Lagercrantz U, Lin SS, Lindquist E, Lipzen AM, Lu CW, De Luna E, Martienssen RA, Minamino N, Mizutani M, Mizutani M, Mochizuki N, Monte I, Mosher R, Nagasaki H, Nakagami H, Naramoto S, Nishitani K, Ohtani M, Okamoto T, Okumura M, Phillips J, Pollak B, Reinders A, Rövekamp M, Sano R, Sawa S, Schmid MW, Shirakawa M, Solano R, Spunde A, Suetsugu N, Sugano S, Sugiyama A, Sun R, Suzuki Y, Takenaka M, Takezawa D, Tomogane H, Tsuzuki M, Ueda T, Umeda M, Ward JM, Watanabe Y, Yazaki K, Yokoyama R, Yoshitake Y, Yotsui I, Zachgo S, Schmutz J (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell 171:287–304.e15. https://doi.org/10.1016/j.cell.2017.09.030

Cargill DC, Vella NGF, Sharma I, Miller JT (2013) Cryptic speciation and species diversity among Australian and New Zealand hornwort taxa of *Megaceros* (Dendrocerotaceae). Austral Syst Bot 26:356. https://doi.org/10.1071/sb13030

Chang C, Bowman JL, Meyerowitz EM (2016) Field guide to plant model systems. Cell 167:325–339. https://doi.org/10.1016/j.cell.2016.08.031

Donald PF, Green RE, Heath MF (2001) Agricultural intensification and the collapse of Europe's farmland bird populations. Proc Roy Soc London, Ser B, Biol Sci 268:25–29. https://doi.org/10.1098/rspb.2000.1325

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molec Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075

Duff RJ, Villarreal A JC, Cargill DC, Renzaglia KS (2007) Progress and challenges toward developing a phylogeny and classification of the hornworts. Bryologist 110:214–243

Greilhuber J (2005) (iv) Bryophytes. In Genome size: a research discipline in development. Report on the International Botanical Congress official workshop held at the Institute of Botany, University of Vienna 22nd July 2005. Available at: https://data.kew.org/cvalues/vienna05_report.pdf. Accessed Feb 2019

Krutzsch W (1963) Atlas der Mittel- und Jungtertiaren dispersen Sporen- und Pollen-sowie der Mikroplanktonformen des nordlichen Mitteleuropas. Lieferung 2: Die Sporen der Anthocerotaceae und der Lycopodiaceae. Gustav Fischer Verlag, Jena and Berlin

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molec Biol Evol 29:1695–1701. https://doi.org/10.1093/molbev/mss020

Leitch IJ, Bennett MD (2007) Genome size and its uses: the impact of flow cytometry. In: Doležel J, Greilhuber J, Suda J (eds) Flow cytometry with plant cells. Wiley, Weinheim, pp 153–176

Lewis LR, Rozzi R, Goffinet B (2014) Direct long-distance dispersal shapes a New World amphitropical disjunction in the dispersal-limited dung moss *Tetraplodon* (Bryopsida: Splachnaceae). J Biogeogr 41:2385–2395. https://doi.org/10.1111/jbi.12385

Li F-W, Villarreal A JC, Szövényi P (2017) Hornworts: an overlooked window into carbon-concentrating mechanisms. Trends Pl Sci 22:275–277. https://doi.org/10.1016/j.tplants.2017.02.002

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084. https://doi.org/10.1073/pnas.052125199

McDaniel SF, Shaw AJ (2005) Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. Molec Ecol 14:1121–1132. https://doi.org/10.1111/j.1365-294x.2005.02484.x

McDaniel SF, von Stackelberg M, Richardt R, Quatrano RS, Reski R, Rensing SA (2010) The speciation history of the *Physcomitrium-Physcomitrella* species complex. Evolution 64:217–231. https://doi.org/10.1111/j.1558-5646.2009.00797.x

Medina R, Johnson MG, Liu Y, Wickett NJ, Shaw AJ, Goffinet B (2019) Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorrhegma*. J Syst Evol 57:404–417. https://doi.org/10.1111/jse.12516

Meyer S, Wesche K, Krause B, Leuschner C (1950s) Dramatic losses of specialist arable plants in Central Germany since the 1950s/60s—a cross-regional analysis. Diversity Distrib 19:1175–1187. https://doi.org/10.1111/ddi.12102

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2012) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA 110:453–458. https://doi.org/10.1073/pnas.1215985110

Paton JA (1979) *Anthoceros agrestis*, a new name for *A. punctatus* var. *cavernosus* sensu Prosk. 1958, non (Nees) Gottsche et al. J Bryol 10: 257–261. http://dx.doi.org/10.1179/jbr.1979.10.3.257

Porley R (2008) Arable bryophytes. A field guide to the mosses, liverworts, and hornworts of cultivated land in Britain and Ireland. Princeton University Press, Princeton

Proskauer J (1957) Studies on Anthocerotales V. Phytomorphology 7:113–135

Proskauer J (1958) Nachtrag zur Familie Anthocerotaceae. In: Müller K (ed) Die Lebermoose Europas, Rabenhorst's Kryptogamen-Flora, 3rd edn. Akademische Verlagsgesellschaft, Leipzig, pp 1303–1319

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457:843–848. https://doi.org/10.1038/nature07895

Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, Donoghue PCJ (2018) The interrelationships of land plants and the nature of the ancestral embryophyte. Current Biol 28:733–745.e2. https://doi.org/10.1016/j.cub.2018.01.063

Rambaut A, Suchard M, Drummond A (2014) Tracer, version 1.6.0. Available at: https://tree.bio.ed.ac.uk/software/tracer/. Accessed Feb 2019

Reidsma P, Tekelenburg T, van den Berg M, Alkemade R (2006) Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union. Agric Ecosyst Environm 114:86–102. https://doi.org/10.1016/j.agee.2005.11.026

Renzaglia KS, Villarreal A JC, Piatkowski BT, Lucas JR, Merced A (2017) Hornwort stomata: Architecture and fate shared with 400-million-year-old fossil plants without leaves. Pl Physiol (Lancaster) 174:788–797. https://doi.org/10.1104/pp.17.00156

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenback JP (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Schnyder N, Bisang I, Caspari S, Hedenäs L, Hodgetts N, Kiebacher T, Kučera J, Ştefănuţ S, Vana J (2019) *Anthoceros agrestis.* The IUCN Red List of Threatened Species 2019. Available at: https://www.iucnredlist.org/species/83658361/87732544. Accessed Oct 2019

Schuster RM (1992) The Hepaticae and Anthocerotae of North America, vol. 6. Field Museum, Chicago

Shaw AJ, Schmutz J, Devos N, Shu S, Carrell AA, Weston DJ (2016) The *Sphagnum* genome project. In: Rensing SA (ed) Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns. Advances Bot Res 78:167–187. https://dx.doi.org/10.1016/bs.abr.2016.01.003

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swofford D (2002) PAUP* (Phylogenetic Analysis Using Parsimony) (*and Other Methods). Version 4. Sinauer Associates, Sunderland

Szövényi P, Frangedakis E, Ricca M, Quandt D, Wicke S, Langdale JA (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. BMC Pl Biol 15:98. https://doi.org/10.1186/s12870-015-0481-x

Szövényi, P (2016) The genome of the model species *Anthoceros agrestis*. In: Rensing SA (ed) Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns. Advances Bot Res 78:189–211. https://dx.doi.org/10.1016/bs.abr.2015.12.001

Tscharntke T, Klein AM, Kruess A, Steffan-Dewenter I, Thies C (2005) Landscape perspectives on agricultural intensification and biodiversity—ecosystem service management. Ecol Lett 8:857–874. https://doi.org/10.1111/j.1461-0248.2005.00782.x

Villarreal A JC, Renzaglia KS (2015) The hornworts: important advancements in early land plant evolution. J Bryol 37:157–170. https://doi.org/10.1179/1743282015y.0000000016

Villarreal A JC, Cusimano N, Renner SS (2015) Biogeography and diversification rates in hornworts—the limitations of diversification modeling. Taxon 64:229–238

Villarreal A JC, Crandall-Stotler BJ, Hart ML, Long DG, Forrest LL (2016) Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytol 209:1734–1746. https://doi.org/10.1111/nph.13716

Villarreal A JC, Duckett JG, Pressel S (2017) Morphology, ultrastructure and phylogenetic affinities of the single-island endemic *Anthoceros cristatus* Steph. (Ascension Island). J Bryol 39:226–234. https://doi.org/10.1080/03736687.2017.1302153

Warnow T (2010) The CIPRES Portals. Available at: https://www.phylo.org. Accessed Mar 2018