Phycologia Volume 56 (1), 54–62 Published 2 November 2016

DNA barcodes and morphometric data indicate that *Codium fragile* (Bryopsidales, Chlorophyta) may consist of two species

Heroen Verbruggen*, Margaret J. L. Brookes and Joana F. Costa

School of BioSciences, University of Melbourne, Victoria 3010, Australia

Abstract: The siphonous green alga *Codium fragile* currently has 10 accepted subspecies, among which is the problematic invasive subsp. *fragile*. The subspecies are defined on the basis of five polymorphic nucleotides in the plastid *rps*3–*rpl*16 spacer region, and this subdivision into subspecies has not been investigated in detail with other DNA markers. Here, we compiled a data set of the green algal DNA barcoding marker *tuf*A that includes 110 *C. fragile* specimens and an additional 80 specimens of eight related *Codium* species. Our *tuf*A haplotype tree showed two distinct clades within *C. fragile*, and the three algorithmic species delimitation methods that were applied agreed that these are likely to represent separate species. The first clade consisted of the invasive subspecies *fragile*, whereas the second contained the remaining subspecies present among our sequences. Using a newly developed morphometric method to capture fine details of utricle morphology, we show that the two molecular clades also have subtle morphological differences. Despite all this new evidence in favour of two species being present in *C. fragile*, we refrain from making formal taxonomic changes, chiefly because our *tuf*A data do not agree with previous *rps*3–*rpl*16 results. We propose four complementary approaches to resolve these conflicts.

Key Words: *Codium fragile*, Chlorophyta, DNA barcodes, Green algae, Invasive species, Morphometrics, Species delimitation, Taxonomy, *tuf*A

INTRODUCTION

Codium fragile (Suringar) Hariot is a branched, siphonous green alga (order Bryopsidales) present in many temperate coastal areas around the world. Its thallus is composed of a central medullar region of intertwined siphons and a thick cortex adjoining siphon endings called utricles (Schmidt 1923). The utricles of *C. fragile* have pointed tips called mucros. One of the subspecies, *C. fragile* subsp. *fragile* (Suringar) Hariot, is a marine pest species (Trowbridge 1998) in many parts of the world including the United Kingdom and mainland Europe (Silva 1955), Australia (Campbell 1999; McDonald *et al*. 2015), New Zealand (Dromgoole 1975), US East Coast and Canada (Begin & Schiebling 2003) and Chile (Neill *et al*. 2006). Its native range includes Japan and Korea (Silva 1955; Provan *et al*. 2005). This subspecies has parthenogenetic reproduction via swarmer cells and wide ecological tolerances, which have enabled it to successfully and rapidly invade new areas (Watanabe *et al*. 2009).

Recognition of regional subspecies within *Codium fragile* has a long history. Agardh (1887) described three geographic forms of the species as the varieties *californicum, tasmanicum* and *novae-zelandiae*. They were described as varieties of *C. mucronatum* J.Agardh, but this species has since been synonymized with *C. fragile* and the varieties are now recognized as subspecies (Silva 1955; Silva & Womersley 1956). *Codium fragile* subsp. *atlanticum* (A.D.Cotton) P.C.Silva was described from Northern Ireland (Cotton 1912), and the collections from the type locality (Japan) were placed in *C. fragile* subsp. *tomentosoides* (Van Goor) P.C.Silva (Van Goor 1923).

Geographical origins played a major role in the original recognition of subspecies, but these boundaries have been blurred with the introduction of subsp. *fragile* in many areas outside its native range. Particularly in areas where native subspecies now co-occur with introduced subsp. *fragile*, it is critical for decision-makers to have good identification tools. Both Agardh (1887) and Silva (1955) noted minor differences in the shape of the utricle and its mucro among subspecies, but added that intergrades were present and that morphology was not reliable in separating the subspecies. It is also known that with age and at different stages in thallus development, the shape, length, width and composition of both the mucro and the utricle can differ, even on the same plant (Provan *et al*. 2008; McDonald *et al*. 2015).

Because of the difficulty of recognizing subspecies on the basis of morphology, significant efforts have been made to redefine them on the basis of molecular data (Maggs & Kelly 2007; Provan *et al*. 2008). These studies were based on the spacer between the *rps*3 and *rpl*16 genes of the chloroplast genome. This region, also known as UCP6 (universal chloroplast primers 6, Provan *et al*. 2004), provided polymorphisms at five locations. The resulting haplotypes were used to redefine the existing subspecies and recognize five additional subspecies based on DNA descriptions, but no morphological characteristics were given (Maggs & Kelly 2007; Provan *et al*. 2008). The boundaries between subspecies have not been comprehensively verified with other DNA markers, which we consider a useful exercise considering the very subtle differences between subspecies in the *rps*3–*rpl*16 region and the importance of being able to accurately identify specimens as native or introduced.

* Corresponding author (heroen.verbruggen@unimelb.edu.au).
DOI: 10.2216/16-54.1

This study aims to revisit the species boundaries of *Codium fragile* as well as subspecies delimitation on the basis of *tuf*A DNA barcodes. Our approach consists of applying species delimitation algorithms on a large data set that includes samples of *C. fragile* and related species. We also assess morphological variation within the species using statistical analyses of morphometric data sets.

MATERIAL AND METHODS

Samples identified as *Codium fragile* were preserved in silica-gel for molecular work and in ethanol or formalin for morphological examination. The samples originated from many parts of the world where populations of native or introduced populations are known. We also used 80 samples of species related to *C. fragile* to provide the species delimitation algorithms with more data for parameter optimisation. These species are *C. yezoense* (Tokida) K.L.Vinogradova, *C. galeatum* J.Agardh, *C. tomentosum* Stackhouse, *C. duthieae* P.C.Silva, *C. decorticatum* (Woodward) M.Howe, *C. cylindricum* Holmes, *C. papenfussii* P.C.Silva and "*C.* sp.6" from Verbruggen *et al.* (2007).

DNA was extracted following Saunders & McDevit (2012). The *tuf*A gene and *rps*3-*rpl*16 spacer (UCP6 region) were amplified and sequenced as previously described (Lee & Kim 2015; Verbruggen & Costa 2015). Alignments were made by hand in Geneious 8.0.4 (http://www.geneious.com/). The alignments were 1224 and 445 nucleotides long for *tuf*A and UCP6, respectively. A haplotype tree of *tuf*A sequences was inferred with BEAST v.1.8.2 (Drummond *et al.* 2012) using a GTR+Γ_4 model of sequence evolution, a log-normal relaxed clock model and a coalescent tree prior (constant population size). The analysis was run for 10 million generations, storing a tree every 10 thousand generations. A maximum likelihood (ML) phylogeny was inferred with RAxML v.7.7.1 (Stamatakis 2006) using a GTR+Γ_4 model and 100 bootstrap replicates.

We applied three single-marker species delimitation algorithms to the *tuf*A data: automatic barcode gap detection (ABGD; Puillandre *et al.* 2013), generalized mixed Yule coalescent (GMYC; Pons *et al.* 2006) and Poisson tree processes (PTP; Zhang *et al.* 2013). All programs were run with default settings from online servers: ABGD on the sequence alignments at http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html, GMYC on the BEAST tree at http://species.h-its.org/gmyc/ and PTP on the RAxML tree at http://species.h-its.org/ptp/.

A data set of UCP6 sequences was also analysed to assess the subspecies to which our samples belonged. Subspecies are defined by the reference UCP6 sequences given in Maggs & Kelly (2007). Comparing our samples, which are also present in the *tuf*A analyses, with these reference sequences permits assigning subspecies names to clades in the *tuf*A tree. The UCP6 data were analysed with unweighted pair-group method with arithmetic mean (UPGMA) using a Tamura–Nei model in Geneious 8.0.4. Our choice for this method is justified by the fact that it clearly shows which sequences are identical in a data set, which in this case indicates subspecies membership. We did not perform bootstrapping because there were only five polymorphic sites, so resampling would naturally result in some bootstrap alignments not having all the characters needed to distinguish between subspecies. Furthermore, our goal with this analysis was not to resolve the branching order, making branch support unnecessary.

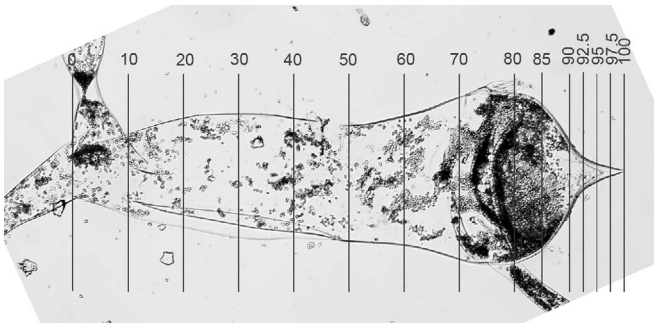


Fig. 1. Illustration of the grid used to digitise utricle morphology. Microscope images of utricles were rotated to be horizontal, and an image of vertical lines was overlaid so that line 0 was at the base and line 100 at the tip of the mucro. Landmarks were then placed at each point where the vertical lines intersected the utricle wall, and metrics were derived from the landmark coordinates. The density of vertical lines increases toward the utricle tip to better capture shape differences occurring over short distances.

Morphometric data were gathered from a subset of 17 samples representing the clades observed in the *tuf*A data (5 of the 17 samples were subsp. *fragile*, see Results). A piece of the cortex was pinched out of the thallus halfway along the branch below an apical branch to reduce the age variation of utricles (i.e. avoiding the very young utricles from apical branches and very old utricles closer to the base; see Verbruggen *et al.* 2005a). The tissue was spread out on a microscope slide and 10 randomly selected utricles were photographed using a Leica DM750 compound microscope equipped with a Canon EOS 600D digital camera. Using photo-editing software, the utricles were aligned horizontally and a grid of vertical lines was placed over them (Fig. 1). The vertical lines were at the base of the utricle (0%), at the tip of the utricle (100%) and at various points in between (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92.5%, 95% and 97.5%). The density of lines was increased near the tip of the utricle to capture fine details of the tip and mucro shape, as these were considered important characters for distinguishing between subspecies (e.g. Silva 1957). The size and shape of the utricles was digitised by placing landmarks where the lines of the grid intersected the utricle cell wall using PointPicker in ImageJ (http://imagej.nih.gov/ij), and the relative width of the utricle was calculated at each point (cf. fig. 1 of Verbruggen *et al.* 2005b). This resulted in a data matrix of 16 variables (length, maximal width, position of maximal width, relative width at 13 points listed above) and 167 utricles belonging to 17 samples. This deviates slightly from the 10 utricles per sample because a handful of strongly curved utricles was excluded.

The morphometric data matrix was analysed with multivariate statistics in R v.3.2.4 (R Core Team 2016). First, we performed a principal components analysis (PCA, FactoMineR package) to examine the overall structure in the data set as a biplot. For this analysis, measurements were averaged across all utricles of a specimen to yield a data set

with one entry per specimen. Discriminant analysis (DA) was used to identify the differences between subspecies *fragile* and the remaining subspecies. The DAs were performed on data from individual utricles. The candisc package was used to produce a plot of correlations between morphometric predictor variables and the canonical axes of the DA and the MASS package was used to calculate classification success both for the training data and test data (cross-validation function).

It is important to note that the DA-based classification approach above predicted group membership (*fragile* vs native) for individual utricles. In reality, of course, one is interested in identifying specimens, not utricles. For this study, we collected *c*. 10 utricles per specimen, and the question should be how well the data of these 10 utricles combined can predict group membership. To answer this question, we averaged the group membership probabilities across the 10 utricles of each specimen. These average group membership probabilities were then used to predict the most likely identity of the specimen.

RESULTS

The haplotype tree inferred from our 190 *tuf*A sequences (Fig. 2, Supplement 1) showed distinct groups corresponding to described species with two exceptions. First, *Codium duthieae* separated into three allopatric clusters (called *duthieae*.1, *duthieae*.2, *duthieae*.3 – see under *C. decorticatum* complex in Fig. 2). This was also shown in previous studies and interpreted as cryptic diversity (Verbruggen *et al*. 2007; Verbruggen & Costa 2015). Second, *C. fragile* separated into two distinct clusters, one corresponding to subspecies *fragile* (at bottom of Fig. 2), the other being made up of the remaining subspecies. We will refer to these clades as the '*fragile* clade' and the 'native clade'. Within the latter clade, three groupings were observed. The first, which had high support [posterior probability (PP) = 0.99], consisted of specimens from the northeast Pacific (subspecies *mexicanum*). The second, which was not supported (PP = 0.38), consisted of specimens from South Africa (subspecies *bonnespei*) and a sample from Tasmania (possibly subspecies *tasmanicum*). The third, which had high support (PP = 0.95), consisted of specimens from Australia and New Zealand (subspecies *novae-zelandiae*).

PTP suggested that there were 12 species-level groups in the *tuf*A data (Fig. 2). A histogram of pairwise sequence distances did not show a clear barcode gap but three distinct peaks (Supplement 2). However, ABGD inferred the presence of eight species in our *tuf*A data set across a wide range of prior intraspecific divergence settings (Supplement 2), and these are shown on the haplotype tree (Fig. 2). GMYC indicated a wide range of possibilities (confidence interval ranging from 3 to 45 species), with the ML solution being eight species (Fig. 2). The likelihood-ratio test of the GMYC analysis indicated that the mixed model did not provide a significantly better fit than the null model with a single branching rate ($P = 0.09$). The different species delimitation algorithms agreed fairly well on species boundaries, disagreeing only about species limits in the *Codium decorticatum* complex. PTP separated this complex into the five species that it is thought to contain (see Verbruggen *et al*. 2007), but ABGD and GMYC indicated that the entire complex is a single species. More important for this study, all the methods suggested that *C. fragile* consisted of two species, corresponding to the *fragile* clade and the native clade.

The UPGMA tree of UCP6 sequences showed a much shallower haplotype structure within *Codium fragile*, without the split into *fragile* and native clades seen in *tuf*A (Supplement 3). The UCP6 data also had more subspecies represented. The number of sequences available for UCP6 was substantially smaller than for *tuf*A (often a single sequence per species). Because algorithmic species delimitation algorithms rely on large numbers of sequences per species to detect suitable thresholds, we have not used them with the UCP6 data. The main purpose of the UCP6 tree was to allow comparison of intraspecific clades in *C. fragile* between the two markers. The sequences considered to be diagnostic for the different subspecies were indicated in boldface, and samples for which we have corresponding *tuf*A data are in an oblique font type in the figure. Samples in the *fragile* clade of our *tuf*A data clearly matched with the reference sequence, confirming their identity as subspecies *fragile*. Similarly, our specimens of subspecies *novae-zelandiae*, *mexicanum* and *bonnespei* matched the corresponding UCP6 reference sequences. Our Tasmanian sample HV04038 did not match up with the reference sequence for subspecies *tasmanicum*, so we remain uncertain of its identity and have added a question mark to its subspecies assignment.

The morphometric data of a subset of sequenced specimens suggested that utricles of the *fragile* and native clades are sufficiently different to be distinguished statistically. The PCA showed *fragile* clade specimens toward the lower right corner of the biplot, separated from specimens from the native clade, which covered a wider range of utricle morphologies (Fig. 3). The variable structure of the PCA showed that the area of the biplot where *fragile* clade samples were located had a positive correlation with the maximum utricle width (maxW) and the relative width at 70–80% of the utricle (Fig. 4; refer to Fig. 1 for positions) and negative correlations with utricle length and the relative widths near the tip of the utricle (relW85, 90, 92.5, 95, 97.5).

The DA clearly separated the utricles of the two predefined groups along the DA axis (Fig. 5) and showed the same correlations between variables that were indicated by the PCA, i.e. *fragile* group was positively correlated with maxW and relW70, and negatively with relW85–97.5 (Fig. 5). The linear discriminant model assigned utricles to the correct group with very high confidence. For the cross-validation analysis, where utricles are left out of the analysis one by one and then assigned to the *fragile* or native group on the basis only of their morphology, the discriminant model correctly assigned 93.4% of the utricles. Using the approach that uses information from 10 utricles of a specimen to jointly predict to which species the specimen belongs, all but one of the specimens were correctly identified (Supplement 4).

To translate the morphometric results back to a familiar form, we created models depicting the clade averages for utricle morphology. These models show that *fragile* clade

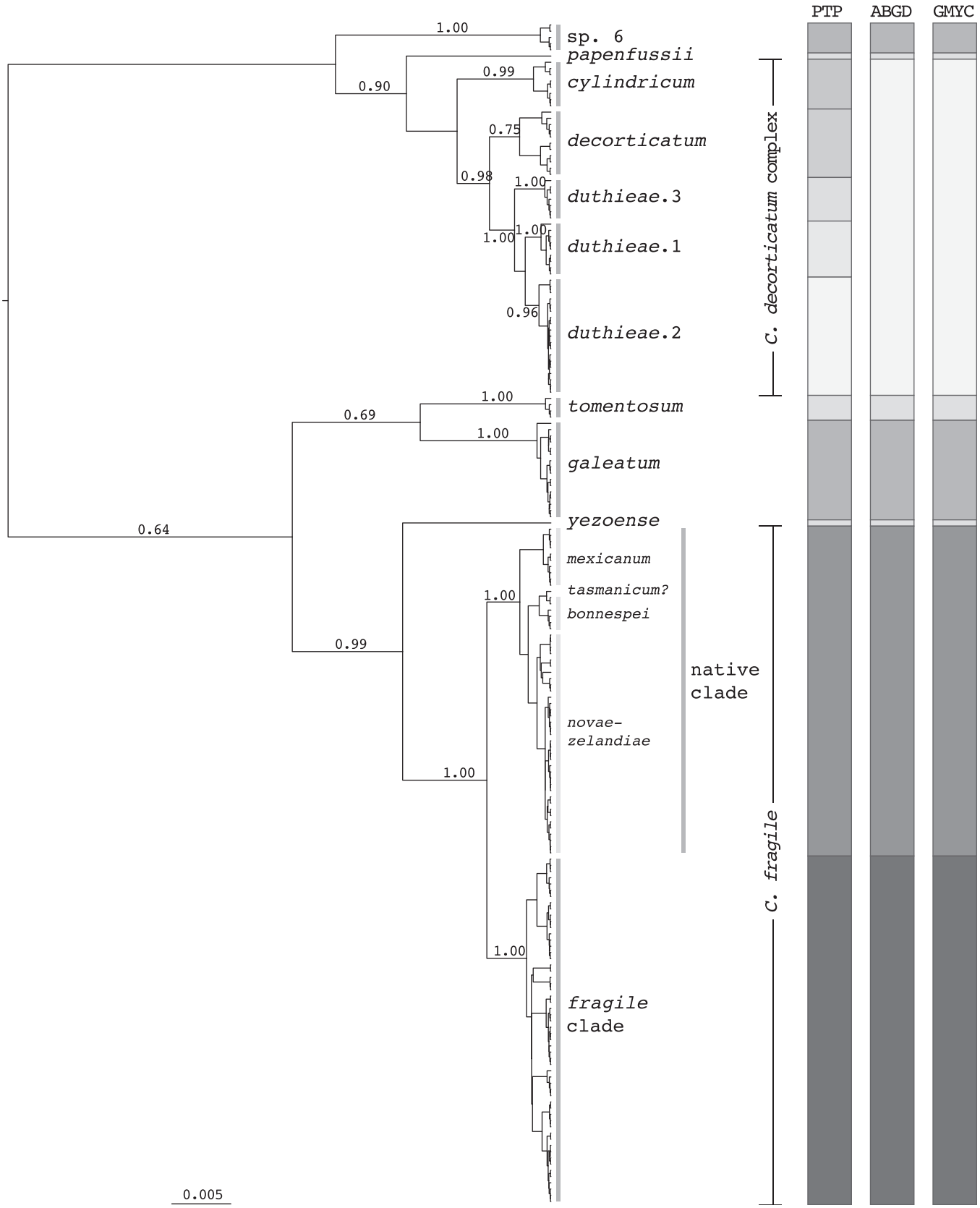
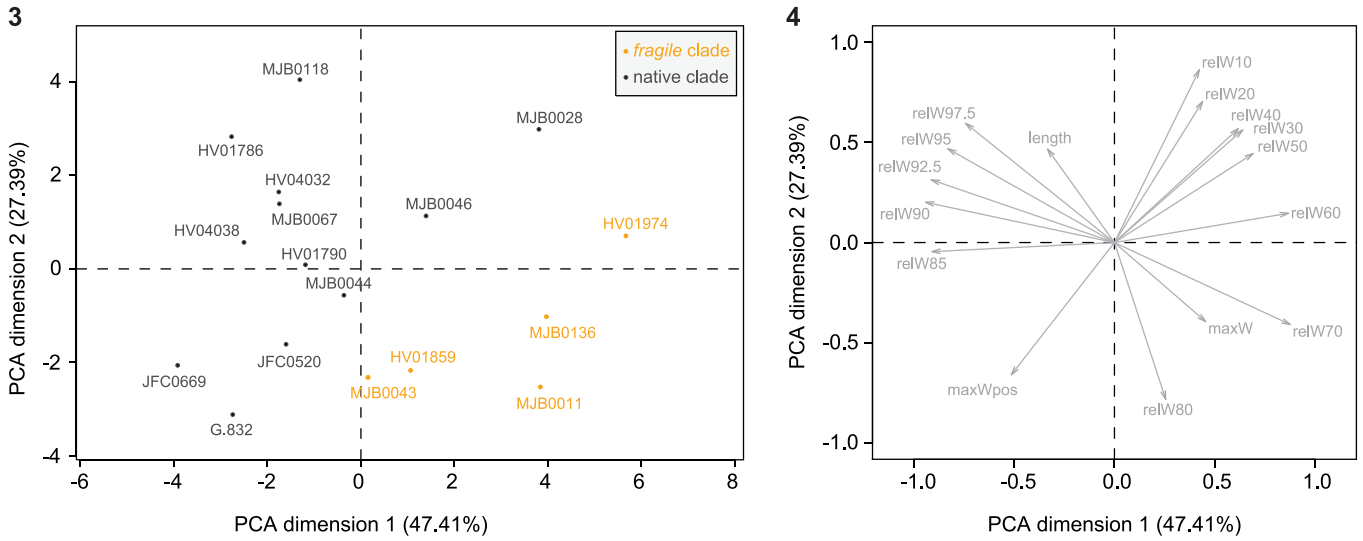


Fig. 2. Haplotype tree of the *Codium tuf*A data set inferred with BEAST. The three vertical bars on the right-hand side show the species boundaries inferred using three algorithmic species delimitation methods (PTP = Poisson tree process, ABGD = automatic barcode gap detection, GMYC = generalized mixed Yule coalescent). Values at nodes are Bayesian posterior probabilities. Note that this is a tree inferred with a coalescent prior, so nonzero branch lengths result even when sequences are identical. Sample numbers and specimen information could not be shown here because of the dense nature of the tree, but a full version of the tree with specimen information and Genbank numbers is available in Supplement 1.



Figs 3, 4. Principal components analysis (PCA) of morphometric data. The colours refer to the online PDF version of the paper. In print, the *fragile* clade data are shown in light grey.

Fig. 3. Biplot of specimens showing the different positioning of specimens from the *fragile* clade (orange) and the native clade (charcoal).

Fig. 4. Structure plot showing correlation of the variables to the main PCA axes.

utricles are shorter and fatter than native clade utricles (Fig. 6). When the size aspect of the utricles is annulled and only shape is compared, the two clades can be seen to differ in the utricle tip region, with the *fragile* clade narrowing earlier and having a longer mucro (Fig. 7).

DISCUSSION

Because *Codium fragile* subsp. *fragile* is among the most problematic invasive algae on the planet (Nyberg & Wallentinus 2005), it is critical that we gain a solid understanding of the taxonomy of the species and its different subspecies. Our data show that there are two clear-cut clades among *C. fragile tuf*A sequences, which we call the *fragile* and native clades. Algorithmic species delimitation methods agree that these clades likely represent two distinct species, and our morphometric data indicate that utricle morphology also differs statistically between these clades.

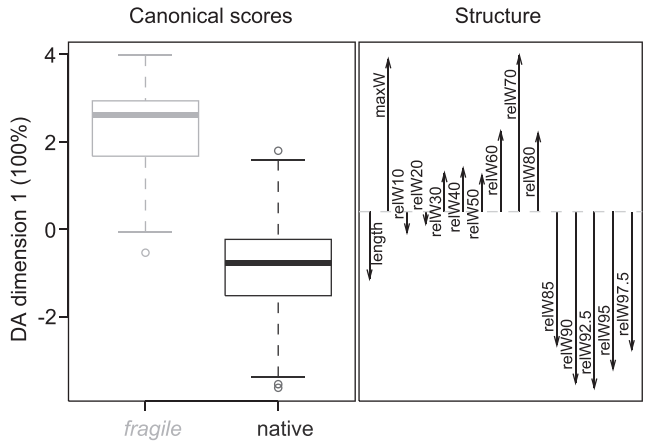
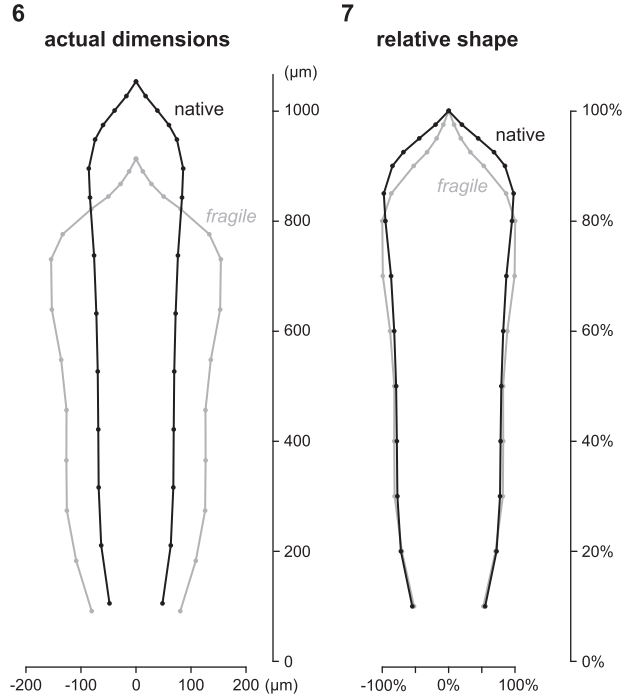


Fig. 5. Discriminant analysis (DA) of morphometric data. The left panel summarizes scores of utricles against the DA axis, showing separation of *fragile* clade utricles (light gray) and native clade utricles (black). The structure plot in the right panel shows correlation of the variables to the DA axis.

Algorithmic species delimitation methods for DNA data are a useful addition to the toolkit for algal systematists (reviewed by Leliaert *et al.* 2014). For this study, three methods were applied to the *tuf*A gene, a chloroplast marker that outperforms most other loci in a comparison of potential DNA barcode markers (Saunders & Kucera 2010), and has since been used widely as a DNA barcode region for green seaweeds including *Codium* (Lee & Kim 2015; Verbruggen & Costa 2015). In addition to the target group (*C. fragile*) we also included samples from eight closely related species to improve the ability of the species delimitation algorithms to converge onto meaningful parameter values. The three species delimitation algorithms differ substantially in their underlying method. ABGD is a simple method based on detecting the so-called barcode gap, a difference between inter- and intraspecific genetic distances observed in the frequency histogram of pairwise distances between specimens (Puillandre *et al.* 2012). GMYC is a more formal test of species boundaries. It is based on the notion that branching processes in a haplotype tree will differ below the species boundary (where interbreeding occurs) and above the species boundary. By optimizing the optimal transition point between a Yule branching process deeper in the tree and a coalescent model nearer the tips, it infers what can be considered the optimal species boundaries (Pons *et al.* 2006). The PTP is similar in nature to the GMYC but uses numbers of substitutions in the DNA data to optimize the model and infer transition points where branching changes from an among-species to a within-species pattern (Zhang *et al.* 2013).

Despite the species delimitation methods being based on different principles, they show completely congruent results



Figs 6, 7. Model representation of the shape of utricles from the *fragile* and native clades. These models are derived from the group means of the morphometric variables given in the discriminant analysis.

Fig 6. Comparison based on actual dimensions shows that *fragile* clade utricles are wider on average than native clade utricles.

Fig. 7. Data rescaled to the same total height and width to illustrate the shape differences better.

for *C. fragile*, providing a strong argument to consider the *fragile* and native clades as different species. However, we cannot conclude this with certainty for two reasons. First, there is conflicting evidence from the UCP6 marker that suggests a much shallower molecular structure of haplotypes within *C. fragile*, and no strong subdivision of *fragile* and native clades. Second, our results are based on a single marker (*tuf*A), and single-gene analyses have limited power to detect species boundaries because the chosen gene may not always be representative of organismal history, and inferred thresholds may not always match actual species boundaries (Hickerson *et al.* 2006; Knowles & Carstens 2007; Fujisawa & Barraclough 2013). For these reasons, we have decided not to make a taxonomic decision at this time and simply refer to the two groupings as the native and *fragile* clades.

Whereas the *fragile* clade matches with a single defined subspecies (*fragile*), the native clade comprises several subspecies including *novae-zelandiae*, *mexicanum* and *bonnespei* and possibly *tasmanicum*. It is interesting that with *tuf*A, we did not recover as many subspecies as previous work based on UCP6 did (Provan *et al.* 2008). This can be attributed to several factors. First, our sampling is not as widely spread geographically as that of the previous study, and we are undoubtedly missing some of the haplotypes for that reason. Second, it is possible that some of the UCP6 haplotypes are joined in a single clade in the *tuf*A tree. For instance, there are several samples from Pacific Canada among the samples labelled as *mexicanum* in our *tuf*A phylogeny. We labelled them as *mexicanum* because they cluster with samples that originated from Pacific Mexico and matched the UCP6 haplotype known as subsp. *mexicanum* (HV01790, HV01791). However, subsp. *mexicanum* is only known from Mexico (Pacific Baja California), whereas farther north subsp. *californicum* is found (Maggs & Kelly 2007). This raises the possibility that the Canadian samples are subsp. *californicum* and that this subspecies is identical in *tuf*A sequence to *mexicanum*. We cannot confirm this because the Canadian sequences were downloaded from Genbank and there are no matching UCP6 sequences for them. Another interesting case is that of HV04038 from Tasmania, which has a UCP6 haplotype that was not known before this study but shows some affinity with South African samples of subsp. *bonnespei* in the *tuf*A tree. The sequence is different from those of *bonnespei*, and the clade in question has no statistical support, so we cannot attach any firm conclusions to its possible identity (is it *tasmanicum*?) or its relationship to *bonnespei*. It is also interesting that despite intensive sampling along the mainland of southern Australia, only subspecies *novae-zelandiae* was found, whereas literature records suggest that both *novae-zelandiae* and *tasmanicum* are present here, calling the molecular distinctness of these subspecies into question.

Our morphometric analysis of utricle size and shape shows clear differences between the native and *fragile* clades. The PCA (Figs 3, 4) showed a separation between the two clades and the DA (Fig. 5) reinforced this finding and showed mostly accurate classification of individual utricles and specimens. The utricle models (Figs 6, 7) show that, on average, *fragile* clade specimens have shorter, fatter utricles with more acute mucros. We have not attempted to design formal morphological identification tools on the basis of these results for three reasons. First, it is important to realize that the utricle models are based on averages, and that there is substantial variation within both clades that needs to be accounted for in identification tools. Second, although we do consider our work to provide proof of concept that morphometrics can serve as an identification tool, building reliable identification tools requires more intensive sampling of utricle morphologies from both clades, including subspecies that we have not been able to include in this study. Third, it must be taken seriously that sample MJB0043 was classified (using morphometrics) as belonging to the native clade, whereas sequence data indicate that it belongs to the *fragile* clade. This sample was also on the edge of the morphological range of the *fragile* clade in the PCA biplot (Fig. 3). We can only guess what has caused this problem (perhaps a hybrid or a labelling problem), but such outliers require careful attention when designing identification tools.

Before the description of a swathe of subspecies based solely on molecular data (Maggs & Kelly 2007), it was widely accepted that utricle morphology was informative in distinguishing between subspecies (Cotton 1912; Silva 1955, 1957, 1959; Silva & Womersley 1956), but the same authors also warned that intergrades existed between the typical morphologies of the subspecies. These authors did not perform quantitative analyses, so it would be useful to interpret their qualitative assessments in the light of our morphometric analyses. Our results clearly confirm that

there are trends in utricle morphology, with utricles of the *fragile* clade being largely distinct from those of the native clade. Within the native clade, we also see some patterning. For example, our South African samples belonging to subsp. *bonnespei* (JFC0669, JFC0520, G.832) are all found in the lower left corner of the PCA biplot, away from the remaining samples in the native clade. They are morphologically similar to samples from the *fragile* clade in the sense that they are both in the lower half of the PCA, confirming Silva's (1959) suggestion that there are similarities between what he called subsp. *capense* and subsp. *fragile* (as *tomentosoides*). Our samples of subspecies *novae-zelandiae* (HV04032, MJB0028, MJB0044, MJB0046, MJB0067, MJB0118), *mexicanum* (HV01786, HV01790) and what may be subsp. *tasmanicum* (HV04038) are not clearly separated in the biplot, suggesting that, at least along these two major axes of morphological variability, they do not differ much. Silva (1955, 1957) also discussed differences between subspecies *atlanticum*, *scandinavicum* and *fragile* (as *tomentosoides*). Regrettably, we did not have samples of subsp. *atlanticum* or *scandinavicum* at our disposal, but a comparison of the drawings and descriptions in Silva (1955, 1957) with our utricle models (Figs 6, 7) would suggest that *atlanticum* is morphologically closer to the native clade than the *fragile* clade, whereas *scandinavicum* is closer to the *fragile* clade. Molecular data have in fact shown synonymy of *scandinavicum* and *fragile* (Maggs & Kelly 2007). Silva (1955) noted that Japanese populations (i.e. subsp. *fragile*) featured a very wide range of utricle morphologies, which appears to be confirmed in our analysis: our two Japanese samples of the *fragile* clade (HV01859, HV01974) were very distant from one another in the biplot.

Quantitative morphological analyses have been carried out on NW Atlantic material. Using a set of measurements of thallus shape and utricles, Hubbard & Garbary (2002) recognized two morphotypes in Canadian waters. Although overall utricle shape was similar, mucro length and certain macromorphological features differed between morphotypes. The authors considered the morphotype with the long mucros to be subsp. *fragile* and the other subsp. *atlanticum*. A later study confirmed that mucro length differed significantly between populations, and that these populations were also genetically distinct on the basis of microsatellite analysis (Kusakina *et al*. 2006). A comparison of the UCP6 sequences of Canadian material confirmed that the morphotypes have slightly different haplotypes but indicates that both should be considered subsp. *fragile* (Benton 2014). There have also been morphometric studies of external appearance of the plant (i.e. not utricle shape) aimed at subspecies identification at a regional scale (Trowbridge 1996) as well as the effects of wave exposure on morphology and ecology (D'Amours & Scheibling 2007).

Some studies have also commented on the variability of utricle shape within individual thalli (e.g. Silva 1955; McDonald *et al*. 2015). In new growth, utricles may not be fully developed yet and mucros could wear down in older parts. On the basis of utricle shape variability within thalli, it has been argued that utricle microstructure does not provide an unambiguous diagnostic character for subspecies identification (Provan *et al*. 2008). Verbruggen *et al*. (2005a) performed a quantitative analysis of such problems in the genus *Halimeda*, showing that the inclusion of 'deviant' segments from apical and basal parts reduced the power of the data to assign samples to the correct species. They also showed that these problems could be eliminated by systematically using data from segments that are neither too young nor too old. A second approach that typically improves the taxonomic power of morphometric data sets is to assemble data sets for a series of replicate structures for each sample, and then averaging these replicates to yield a single data point per sample (Verbruggen *et al*. 2005b). Our study shows that in *Codium*, too, the systematic use of subapical branches for dissection and digitising multiple utricles per sample yields a data set with high taxonomic resolution in which utricle morphology is very informative (yet not completely accurate) about membership to either the *fragile* or native clade. Needless to say, our data set is rather small and we do not address differences between all of the subspecies, so there is more work to be done on morphological differentiation between subspecies. Our results suggest that microscopic identification methods can be designed despite variation of utricle shape within thalli. Identification approaches are needed especially for areas where a native subspecies and the *fragile* clade live in sympatry. Because the addition of accurate morphology-based methods to the identification toolkit would lead to wider accessibility of identification methods, we consider it worth pursuing, particularly at a regional scale.

Our molecular and morphometric results suggest that *Codium fragile* may consist of two species. Because of the importance of the taxon to a wide range of stakeholders, it seemed appropriate to delay making formal taxonomic decisions until the data prove beyond doubt that two species are involved. At the moment, uncertainty remains because (1) the shallow structure of the UCP6 data raises doubt about the two-species hypothesis, (2) the suggestion of two species is based on a single chloroplast marker (*tuf*A) and (3) morphometric data are suggestive of two species but not conclusive. In our opinion, four avenues of research would contribute to a better understanding of the complex. First and foremost, superior species delimitation can be achieved with multigene data sets. We would consider it especially useful to target multiple nuclear genes, and use species delimitation methods based on the multispecies coalescent (see Leliaert *et al*. 2014). Second, it would be useful to have information about whether samples from the *fragile* and native clades can interbreed in a natural setting to guide decisions about species limits. Nuclear gene sequencing of sympatric populations could provide that information. Third, if taxonomic changes need to be made, the nomenclature of species and subspecies should be supported by molecular data from historical specimens including types. This has not yet been done for subsp. *novae-zelandiae* and *tasmanicum*. Considering the results from this study, this work should use more than the small number of polymorphisms from the UCP6 region. High-throughput shotgun sequencing of historical samples may provide the required data. In the context of updating the taxonomy of subspecies, we feel it is also necessary to apply a more evolutionarily meaningful interpretation of subspecies. In recent studies subspecies have been equated with UCP6 haplotypes (Maggs & Kelly 2007; Provan *et al*. 2008), but considering that

populations are naturally polymorphic and that the history of individual loci does not necessarily reflect evolutionary lineages, a multimarker approach should be taken. Finally, we consider it worth expanding the morphometric work, particularly at a regional scale where there are clear practical benefits of being able to distinguish between the native and *fragile* clades for management decisions.

ACKNOWLEDGEMENTS

This research was largely carried out as two undergraduate research projects. Part of the molecular work was done for MB's science research subject in the second semester of 2015. Two thirds of the morphometric work was done by a group of three students as a mini-research project that formed part of the Marine Botany subject at the University of Melbourne in November/December 2015. We thank Chris Ma, Nick Craig and Lillian Haines for this contribution. We thank R.J. Anderson, C. Boedecker, J. Bolton, B. L. Brooks, R. Brown, F. Bunker, E. Coppejans, O. De Clerck, P. Diaz, S. Draisma, C. Fernandez, G. Hansen, J. Huisman, J. Hernandez-Kantun, L. Le Gall, D.S. Littler, M.M. Littler, L. McIvor, F. Mineur, F. Rousseau, J. Tsarahevitra, T. Schils, K. Pauly, E. Demeulenaere, G.W. Saunders, C.W. Schneider, F. Steen, E. De Swaef, M. Taylor, E. Tronchin, S. Wilson, L. Tyberghein, J. West and B. Wysor for collecting samples. Funding for this study was provided by the Australian Research Council (FT110100585 to HV), the Australian Biological Resources Study (RFL213-08 to HV) and the University of Melbourne (MIRS/MIFRS to JFC). Some of the molecular data presented here were generated in the Saunders lab (University of New Brunswick) and we thank Gary Saunders, Daniel McDevit and Tanya Moore for that work and sharing the data. Work in the Saunders Lab was funded through the Canadian Barcode of Life Network from Genome Canada (through the Ontario Genomics Institute), and other sponsors listed at www.BOLNET.ca.

SUPPLEMENTARY DATA

Supplementary data associated with this article can be found online at http://dx.doi.org/10.2216/16-54.1.s1.

REFERENCES

Agardh J.G. 1887. Till algernes systematik. Nya bidrag. (Femte afdelningen.). *Acta Universitatis Lundensis* 23(2): 1–174.

Begin C. & Schiebling R.E. 2003. Growth and survival of the invasive green alga *Codium fragile* ssp. *tomentosoides* in tide pools on a rock shore in Nova Scotia. *Botanica Marina* 46: 404–412.

Benton C.S. 2014. *Evaluating taxonomic variation, ecology and the transcriptome of* Codium fragile *in the NW Atlantic*. PhD thesis. University of New Hampshire.

Campbell S.J. 1999. Occurrence of *Codium fragile* subsp. *tomentosoides* (Chlorophyta: Bryopsidales) in marine embayments of southeastern Australia. *Journal of Phycology* 35: 938–940.

Cotton A.D. 1912. Marine Algae. *Proceedings of the Royal Irish Academy. Section B: Biological, Geological and Chemical Science* 31 (section 1, 15): 1–178.

D'Amours O. & Scheibling R.E. 2007. Effect of wave exposure on morphology, attachment strength and survival of the invasive green alga *Codium fragile* ssp. *tomentosoides*. *Journal of Experimental Marine Biology and Ecology* 351: 129–142.

Dromgoole F.I. 1975. Occurrence of *Codium fragile* subspecies *tomentosoides* in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research* 9: 257–264.

Drummond A.J., Suchard M.A., Xie D. & Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Fujisawa T. & Barraclough T.G. 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* 62: 707–724.

Hickerson M.J., Meyer C.P. & Moritz C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology* 55: 729–739.

Hubbard C.B. & Garbary D.J. 2002. Morphological variation of *Codium fragile* (Chlorophyta) in eastern Canada. *Botanica Marina* 45: 476–485.

Knowles L.L. & Carstens B.C. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* 56: 887–895.

Kusakina J., Snyder M., Kristie D.N. & Dadswell M.J. 2006. Morphological and molecular evidence for multiple invasions of *Codium fragile* in Atlantic Canada. *Botanica Marina* 49: 1–9.

Lee H.W. & Kim M.S. 2015. Species delimitation in the green algal genus *Codium* (Bryopsidales) from Korea using DNA barcoding. *Acta Oceanologica Sinica* 34: 114–124.

Leliaert F., Verbruggen H., Vanormelingen P., Steen F., Zuccarello G.C. & De Clerck O. 2014. DNA-based species delimitation in algae. *European Journal of Phycology* 49: 179–196.

Maggs C.A. & Kelly J. 2007. *Codium*. In: *The green seaweeds of Britain and Ireland* (Ed. by J. Brodie, C.A. Maggs & D.M. John), pp.189–201. British Phycological Society, Dunmurry.

McDonald J.I., Huisman J.M., Hart F.N., Dixon R.R.M. & Lewis J.M. 2015. The first detection of the invasive macroalga *Codium fragile* subsp. *fragile* (Suringar) Hariot in Western Australia. *BioInvasions Records* 4: 75–80.

Neill P.E., Alcalde O., Faugeron S., Navarrete S.A. & Correa J.A. 2006. Invasion of *Codium fragile* ssp. *tomentosoides* in northern Chile: a new threat for *Gracilaria* farming. *Aquaculture* 259: 202–210.

Nyberg C.D. & Wallentinus I. 2005. Can species traits be used to predict marine macroalgal introductions? *Biological Invasions* 7: 265–279.

Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamound S., Sumlin W.D. & Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 55: 595–609.

Provan J., Murphy S. & Maggs C.A. 2004. Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology* 39: 43–50.

Provan J., Murphy S. & Maggs C.A. 2005. Tracking the invasive history of the green alga *Codium fragile* ssp. *tomentosoides*. *Molecular Ecology* 14: 189–194.

Provan J., Booth D., Todd N.P., Beatty G.E. & Maggs C.A. 2008. Tracking biological invasions in space and time: elucidating the invasive history of the green alga *Codium fragile* using old DNA. *Diversity and Distributions* 14: 343–354.

Puillandre N., Lambert A., Brouillet S. & Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877.

R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Saunders G.W. & Kucera H. 2010. An evaluation of *rbc*L, *tuf*A, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Cryptogamie, Algologie* 31: 487–528.

Saunders G.W. & McDevit D.C. 2012. Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. In *DNA barcodes* (Ed. by W.J. Kress & D.L. Erickson), pp.207–222. Humana Press, Totowa, NJ, USA.

Schmidt O.C. 1923. *Beiträge zur Kenntnis der Gattung Codium Stackh*. Bibliotheca Botanica Vol. 91. 67 pp.

Silva P.C. 1955. The dichotomous species of *Codium* in Britain. *Journal of the Marine Biological Association of the United Kingdom* 34: 565–577.
Silva P.C. 1957. *Codium* in Scandinavian waters. *Svensk Botanisk Tidskrift* 51: 117–134.
Silva P.C. 1959. The genus *Codium* (Chlorophyta) in South Africa. *Journal of South African Botany* 25: 101–165.
Silva P.C. & Womersley H.B.S. 1956. The genus *Codium* (Chlorophyta) in southern Australia. *Australian Journal of Botany* 4: 261–289.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
Tamura T. & Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.
Trowbridge C.D. 1996. Introduced versus native subspecies of *Codium fragile*: how distinctive is the invasive subspecies *tomentosoides*? *Marine Biology* 126: 193–204.
Trowbridge C.D. 1998. Ecology of the green macroalga *Codium fragile* (Suringar) Hariot 1889: invasive and non-invasive subspecies. *Oceanography and Marine Biology: An Annual Review* 36: 1–64.
Van Goor A.C.J. 1923. Die Holländischen Meeresalgen (Rhodophyceae, Phaeophyceae und Chlorophyceae) insbesondere der Umgebung von Helder, des Wattenmeeres und der Zuidersee. *Koninklijke Akademie van Wetenschappen, Amsterdam* Sect. 2, 23: 1–232.
Verbruggen H. & Costa J.F. 2015. Molecular survey of *Codium* species diversity in southern Madagascar. *Cryptogamie, Algologie* 36: 171–187.
Verbruggen H., De Clerck O. & Coppejans E. 2005a. Deviant segments hamper a morphometric approach towards *Halimeda* taxonomy. *Cryptogamie, Algologie* 26: 259–274.
Verbruggen H., De Clerck O., Kooistra W. & Coppejans E. 2005b. Molecular and morphometric data pinpoint species boundaries in *Halimeda* section *Rhipsalis* (Bryopsidales, Chlorophyta). *Journal of Phycology* 41: 606–621.
Verbruggen H., Leliaert F., Maggs C.A., Shimada S., Schils T., Provan J., Booth D., Murphy S., De Clerck O., Littler D.S., Littler M.M. & Coppejans E. 2007. Species boundaries and phylogenetic relationships within the green algal genus *Codium* (Bryopsidales) based on plastid DNA sequences. *Molecular Phylogenetics and Evolution* 44: 240–254.
Watanabe S., Metaxas A. & Scheibling R.E. 2009. Dispersal potential of the invasive green alga *Codium fragile* ssp. *fragile*. *Journal of Experimental Marine Biology and Ecology* 381: 114–125.
Zhang J., Kapli P., Pavlidis P. & Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29: 2869–2876.

Received 20 April 2016; accepted 19 July 2016