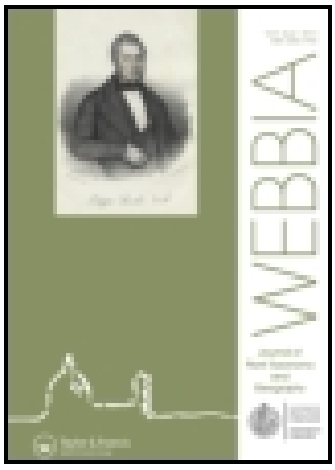

Webbia: Journal of Plant Taxonomy and Geography
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tweb20

Brown barcoded as red but reality is green! How epiphytic green algae confuse phycologists?

Felix Bast[a], Satej Bhushan[a] & Aijaz Ahmad John[a]
[a] Centre for Biosciences, Central University of Punjab, Bathinda, Punjab, India
Published online: 19 Feb 2015.

To cite this article: Felix Bast, Satej Bhushan & Aijaz Ahmad John (2015): Brown barcoded as red but reality is green! How epiphytic green algae confuse phycologists?, Webbia: Journal of Plant Taxonomy and Geography, DOI: 10.1080/00837792.2015.1014217

To link to this article: http://dx.doi.org/10.1080/00837792.2015.1014217

Webbia: Journal of Plant Taxonomy and Geography, 2015
http://dx.doi.org/10.1080/00837792.2015.1014217

Brown barcoded as red but reality is green! How epiphytic green algae confuse phycologists?

Felix Bast*, Satej Bhushan and Aijaz Ahmad John

Centre for Biosciences, Central University of Punjab, Bathinda, Punjab, India

(*Received 1 November 2014; final version received 29 January 2015*)

Promises and perils of DNA barcoding are now well-known, but no studies have revealed the extent of taxonomic misidentification of algal specimens available in primary DNA sequence repositories. Our original objective was to assess the molecular identity of the ubiquitous brown alga *Turbinaria ornata* (Sargassaceae:, Fucales) from the southeast Indian coast. We extracted total genomic DNA from freshly collected algal thalli and sequenced the nuclear ribosomal DNA Internal Transcribed Spacer-1 (nrDNA ITS1) barcode locus. Following a BLASTn DNA sequence similarity search, the identity of our alga was *Laurencia thyrsifera*, a Pacific red alga that has never been reported in India, which came as a big surprise. Further analyses of BLAST hits using a robust phylogenetic framework of Bayesian Inference led to the conclusion that our sequence belonged to an epiphytic Ulvellacean green algal genus *Ulvella*, which might have been extracted and amplified with our universal ITS primers. This is the first report for *Ulvella* from India, and detection of this alga growing on *Turbinaria*. Our Bayesian analyses revealed that a number of GenBank accessions of this epiphyte are misidentified as red algae, which are published in some of the reputed phycological and botanical journals. This finding could have a profound impact on several of the fallacious phylogenetic conclusions arrived at in these publications.

Keywords: DNA barcode; *Laurencia*; misidentification; *Turbinaria*; *Ulvella*

DNA barcoding is now a standard technique for the taxonomic identification of organisms, especially for those taxa exhibiting phenotypic plasticity. The popularity of this technique in taxonomy somehow owes an analogy to its use in forensics – identifying samples that are otherwise impossible to key out. This technique is now routinely employed for algal identification, as a number of macroalgae are known to change their morphology in response to changing ecophysiological conditions, including herbivory (Lewis et al. 1987). Misidentification of marine algae was previously thought to be because of these conditions. Chromatic adaptation of benthic marine algae, which is considered a consequence of changing ecophysiological conditions, has been a topic of debate for more than 50 years (Crossett et al. 1965).

There are reports suggesting that changes in light intensity with depth affect the overall metabolism in marine algae, so causing the change in colour and appearance (Dring 1981). Although this used to be the main reason behind the confusion when identifying and cataloguing marine algal species, our study suggests that this risk can be overcome by DNA barcoding. In the last decade a number of reports have highlighted the risks associated with being overly reliant on DNA barcoding methods for routine taxonomic identification, because the "big three" DNA databases (DDBJ, EMBI and NCBI) are not curated and it is impossible to verify the taxonomic identity of a deposited specimen with certainty (Moritz & Cicero 2004). Extending from its use in taxonomic identification, these barcodes are also routinely used in molecular phylogenetics and phylogeographic studies and pose a significant threat to the conclusions based on mistaken identity. To minimize such errors, inclusion of only those accessions that have been published in literature is gaining popularity among researchers. Although it is believed that a number of algal accessions in GenBank are misidentified, no studies have focused on the extent of this uncertainty. We sought to barcode one of the common brown algae, *Turbinaria ornata* (Sargassaceae:, Fucales), distributed on the southeast Indian coast using nuclear ribosomal DNA Internal Transcribed Spacer-1 (ITS1) locus – a widely used universal DNA barcode for fungi and plants (Baldwin 1992). The objective had been to characterize the genetic heterogeneity of this species distributed along the Indian coast.

Specimens were collected from Pamban Strait, Tamil Nadu, India (9.16 N, 79.11 E) and transported to the laboratory in cold conditions. Collection locations did not include any region designated as protected by the Government of India and therefore no specific permissions were required for the sampling expedition. Further, our field studies did not involve endangered or protected species. The alga was dark brown in colour with thick conical blades bearing spines on the periphery – a typical morphological feature of *Turbinaria ornata* (Figure. 1). Total genomic DNA was extracted from the algae using a HiPurA Algal Genomic DNA extraction kit (HiMedia Laboratories Pvt. Ltd., Mumbai, India) following the manufacturer's protocol. Extracted DNA was polymerase chain reaction (PCR) amplified using ITS-1 (forward) and ITS-2 (reverse) primers for the target ITS1 amplicon

*Corresponding author. Email: felix.bast@cup.ac.in

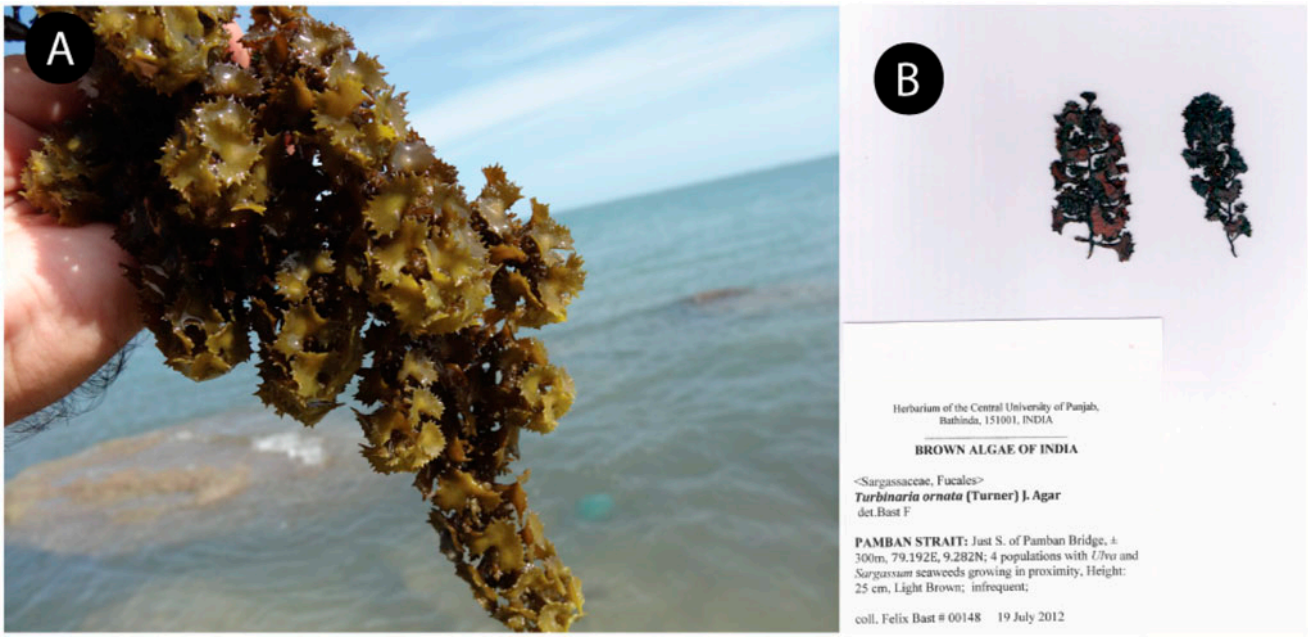

Figure 1. Morphology of the collected *Turbinaria ornata* specimen. (A) Photo of the specimen immediately upon collection and (B) pressed herbarium voucher of the same specimen.

as per standard PCR protocols (White et al. 1990). PCR amplicons were purified using an ExoSAP-IT® PCR clean-up kit following the manufacturer's instructions (USB Corporation, Cleveland, OH, USA) and subjected to direct DNA sequencing using a dideoxy chain termination protocol with an ABI BigDye Terminator Cycle Sequencing Ready® Reaction Kit v3.1 (Applied Biosystems, Foster City, CA, USA) and a programmable thermal cycler (Veriti, Applied Biosystems,). Reactions were then purified using a Centri-Sep® spin column (Applied Biosystems), vacuum dried and subsequently bidirectional DNA sequencing was performed (Applied Biosystems 3730*xl* Genetic Analyzer). Unabridged step-by-step protocol for sequence assembly and phylogenetic analyses are as per Bast, 2013. Contigs were assembled using CodonCodeAligner (CodoneCode Corporation, Centerville, MA, USA) and subjected to BLASTn homology search using GeneiousPro (http://www.genious.com). The assembled contiguous sequence is available in GenBank under the Accession number KF309181. The top seven hits were downloaded and aligned using the same software. Original electropherograms and alignments are available from the first author upon request. Best-fitting nucleotide substitution models were calculated using ML-ModelTest within MEGA (http://www.megasoftware.net). The nucleotide substitution model with the lowest Bayesian Information Criterion score was Jukes–Cantor 69 with a Bayesian Information Criterion score of 649.780 and this model was chosen in all of our phylogenetic analyses. Pairwise distances were calculated using the same software with the complete deletion option for treating gaps and missing data, presented as Table 1.

Phylogenetic analysis using Bayesian Inference with the Jukes Cantor 69 model of nucleic acid substitution was conducted using MrBayes plug-in v3 (Ronquist and Huelsenbeck 2003) inside computer program Geneious v6 (available at http://www.genious.com). Analyses were run with four Markov chains for 10^6 generations with a tree saved every 100th generation. The first 1000 trees were discarded as burn-in. A consensus tree was then constructed using the consensus tree builder within Geneious. Our data sets, including alignments, electropherograms and trees, are available from the corresponding author.

Results were indeed staggering, with barcode identity of this specimen as the Pacific red alga *Laurencia thyrsifera*, with BLASTn E-value of 5e-69 and corrected (Jukes Cantor 69 model) pairwise distance of 0.024 (Table 2). Other top hits included *Laurencia perforata*, *Ulvella leptochaete* (Green), *Gracilaria corticata* (Red) and *Ulva flexuosa* (Green). Both corrected (range 0.024–0.199) and uncorrected (range 0.23–0.174) pairwise distances were within the typical intrageneric range reported for a number of algal species, thereby suggesting that all of these sequences might affiliate to the same genus or very similar genera. Of the first 100 hits, 95 were from Ulvophyceae, which is suggestive of the phylogenetic inclusion of our query in this class. With these confusing homology search results we sought out for careful scrutiny of GenBank flat files. Interestingly, the first four hits were from the same source, who have published their findings in a journal with a moderate Impact Factor (Lewis et al. 2008). The report included detailed phylogenetic analysis of *Laurencia* among other members of Ceramiales using the same set of sequences that hit our BLASTn search, and it was in turn cited by a number of other reports that dealt with phylogenetics of related red algal taxa (e.g., Cassano et al. 2009; Machín-Sánchez et al.

Table 1. Pairwise distance between aligned sequences. Bottom-left part of the matrix is distance calculated using Jukes–Cantor 69 model of nucleotide substitution and top-right part is calculated using p-distance. Top and bottom values are highlighted in both the matrices.

	AF082340	AF082342	AF082343	AF082344	EU933991	EU937777	JN104107	*Turbinaria ornata*
AF082340_*Laurencia* sp. *amarilla*		**0.023**	0.093	0.140	0.105	0.128	**0.174**	0.116
AF082342_*Laurencia perforata*	0.024		0.093	0.140	0.105	0.128	0.174	0.116
AF082343_*Laurencia thyrsifera*	0.099	0.099		0.070	0.081	0.058	0.093	**0.023**
AF082344_*Laurencia* cf. *perforata*	0.154	0.154	0.073		0.128	0.105	0.128	0.093
EU933991_*Ulva flexuosa*	0.113	0.113	0.086	0.140		0.093	0.140	0.081
EU937777_*Gracilaria corticata*	0.140	0.140	0.061	0.113	0.099		0.128	0.035
JN104107_*Ulvella leptochaete*	**0.199**	**0.199**	0.099	0.140	0.154	0.140		0.093
Turbinaria ornata	0.126	0.126	**0.024**	0.099	0.086	0.036	0.099	

Table 2. Detailed blast results of top seven hits.

S. No.	Accession no.	Species	% Overlap	E-value	% Identity	% Gap
1	AF082343	*Laurencia thyrsifera*	100	5e-69	87	7%
2	HG931702	*Ulvella leptochaete*	100	2e-52	82	6%
3	AF082340	*Laurencia* sp.	100	8e-37	79	7%
4	AF082342	*Laurencia perforata*	100	8e-32	78	9%
5	EU937777	*Gracilaria corticata*	35	3e-31	95	0%
6	AF082344	*Laurencia perforata*	99	3e-26	77	9%
7	LM653285	*Ulvella leptochaete*	46	1e-25	85	3%

2012). The original publication of the sequence of *Ulvella* (JN104107) raises a very interesting insight (Deng et al. 2012). A phylogram of *Ulvella* (reported as a currently obsolete synonym *Acrochaete*) among other green algae of Ulvales was presented in which *Laurencia* (AF082343) oddly clustered within, although the authors did not discuss this issue elsewhere in the paper. This is the only ITS1 sequence of *Ulvella* available in GenBank. Because of the very high proportion of Ulvales in our hit, and the straightforward evidence of a maximum likelihood phylogram in Deng's paper, we are of the opinion that the query is indeed *Ulvella*, a microscopic epiphyte that was extracted and subsequently amplified by our universal ITS primers. *Ulvella* grows on a number of conspicuous seaweeds, including *Chetomorpha*, *Gracilaria*, *Laurencia*, *Ulva*, and members of Sargassaceae including *Turbinaria*; as a result many accessions could have been misidentified with the true identity being this epiphyte. This is the first report on *Ulvella* from India, and detection of this alga growing on *Turbinaria*.

Now the question is how much of our BLASTn hits represent this epiphyte instead of the misidentified taxa. Given its distant phylogenetic affinity with rhodophytes, all accessions identified as *Laurencia* and *Gracilaria* are most certainly *Ulvella*. Because this epiphyte belongs to Ulvales, *in silico* detection of misidentifications with *Ulva* (i.e., due to the amplification of this epiphyte growing on *Ulva*) possesses a significant challenge. Our strategy had been to align full-length unique sequences among our first 50 BLASTn hits, along with other accessions from the same study (Lewis et al. 2008) and *Cladophora* as the out-group. Subsequent phylogenetic analysis of this data set using Bayesian Inference performed using MrBayes add-in inside GeneiousPro yielded a phylogenetic cluster (Cluster A) that contained both the query and *Ulvella*, with robust Posterior Probability support (Figure. 2). It is therefore highly likely that sequences in this cluster belong to the genus *Ulvella*. The accession (EU937777) identified as *Gracilaria corticata* in a recent, freqently cited compendium of seaweed diversity in India (Reddy et al. 2009) is therefore *Ulvella*. All accessions of *Ulva* formed another cluster (Cluster B) of robust Posterior Probability support and therefore (fortunately) the identities of most similar *Ulva* accessions were not affected.

The impact of this finding is tremendous, with core conclusions derived in a number of highly cited articles that have used these misidentified sequences being adversely affected. This finding further emphasizes the need for extreme diligence to be observed when

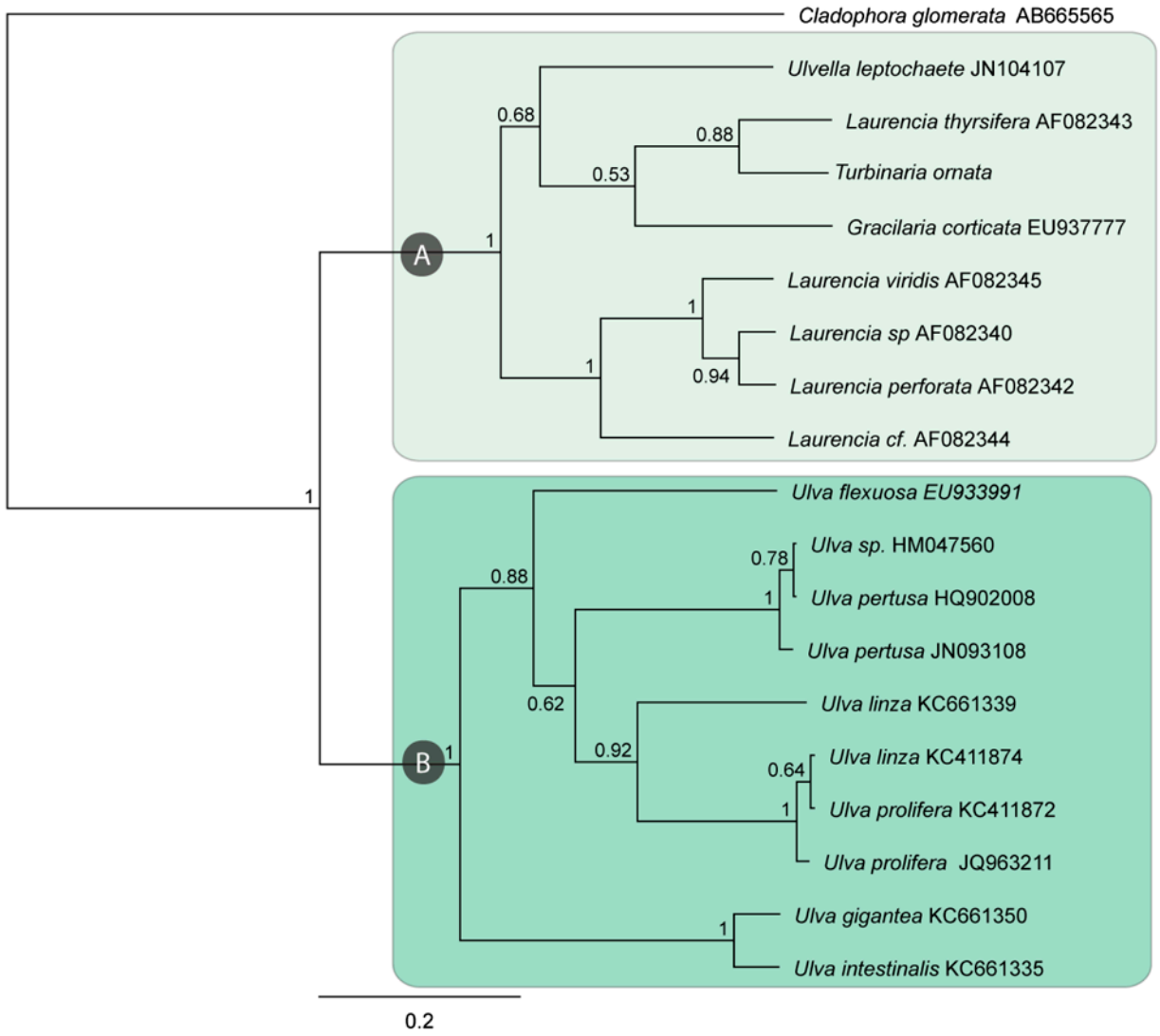

Figure 2. Bayesian Inference phylogram based on genomic DNA data, rooted with *Cladophora glomerata* as outgroup. Bayesian Posterior Probabilities exceeding 0.5 are indicated above the branches. Total chain length = 1,100,000 and mean *LnL* = –2497.475. Scale bar is on the unit of average nucleotide substitutions per site.

depositing generated sequences in the sequence repositories as the same sequence might be used in a number of unrelated studies. For taxa that are well represented in sequence databases such as the Ulvales used in the present study, careful scrutiny of top BLAST hits, especially under the robust statistical framework of Bayesian Inference, might provide some insights for finding the odd one out. However, for samples with poor coverage in repositories, no easy method exists for the efficient *in silico* detection of misidentifications;, at least we are not aware of any. Probably the best approach as far as sequence repositories are concerned is to be a responsible contributor and a cautious habitué.

Acknowledgements

We thank the Vice Chancellor, Central University of Punjab for his administrative help for the execution of this report. We also thank Dr. Vaibhav Mantri, CSMCRI, Mandapam, Dr. Lydiane Mattio, University of Cape Town, South Africa, Dr. Kuldeep Singh, Punjab Agricultural University, Ludhiana and Dr. Pankaj Bhardwaj, Central University of Punjab for their help in improving this manuscript.

References

Baldwin BG. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Mol Phylogenet Evol. 1:3–16.

Bast F. 2013. Sequence similarity search, multiple sequence slignment, model selection, distance matrix and phylogeny reconstruction. Nature Protocol Exchange. Nature Publishing Group. doi:10.1038/protex.2013.065.

Cassano V, Díaz-Larrea J, Sentíes A, Oliveira MC, Gil-Rodríguez MC, Fujii MT. 2009. Evidence for the conspecificity of Palisada papillosa with P. perforata (Ceramiales, Rhodophyta) from the western and eastern Atlantic Ocean on the basis of morphological and molecular analyses. Phycologia. 48:86–100.

Crossett R, Drew E, Larkum A. 1965. Chromatic adaptation in benthic marine algae. Nature. 207:547–548.

Deng Y, Tang X, Huang B, Teng L, Ding L. 2012. Molecular identification and culture observation on *Acrochaete leptochaete* (Chaetophoraceae, Chlorophyta) from China. Chinese J Oceanol Limnol. 30:476.

Dring M. 1981. Chromatic adaptation of photosynthesis in benthic marine algae: an examination of its ecological significance using a theoretical model. Limnol Oceanogr. 26:271–284.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munroe MN, editor. Mammalian protein metabolism, Vol III. Waltham, MA: Academic Press Inc; p. 21–132.

Lewis S, Gacesa P, Gil-Rodríguez M, Valdés F, Frías I. 2008. Molecular systematics of the genera *Laurencia*, *Osmundea* and *Palisada* (Rhodophyta) from the Canary Islands – analysis of rDNA and RUBISCO spacer sequences. Anales-Jardin Botanico De Madrid. 65:97.

Lewis SM, Norris JN, Searles RB. 1987. The regulation of morphological plasticity in tropical reef algae by herbivory. Ecology. 636–641.

Machín-Sánchez M, Díaz-Larrea J, Fujii M, Sentíes A, Cassano V, Gil-Rodríguez M. 2012. Morphological and molecular evidences within Osmundea (Ceramiales, Rhodophyta) from the Canary Islands, eastern Atlantic Ocean. Afr J Marine Sci. 34:27–42.

Moritz C, Cicero C. 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2:e354.

Reddy CRK, Thakur Mukund C. 2009. Seaweeds of India: the diversity and distribution of seaweeds of the Gujarat Coast. New Delhi: Springer.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Waltham, MA: Academic Press.