

Assessment of Four Molecular Markers as Potential DNA Barcodes for Red Algae *Kappaphycus* Doty and *Eucheuma* J. Agardh (Solieriaceae, Rhodophyta)

Ji Tan^{1,2}, Phaik-Eem Lim^{1,2*}, Siew-Moi Phang^{1,2}, Dang Diem Hong³, H. Sunarpi⁴, Anicia Q. Hurtado⁵

1 Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia, **2** Institute of Ocean and Earth Sciences (IOES), University of Malaya, Kuala Lumpur, Malaysia, **3** Institute of Biotechnology, Vietnamese Academy of Science and Technology, Cau Giay, Hanoi, Vietnam, **4** Faculty of Science and Mathematics, Mataram University, Mataram, Lombok, Indonesia, **5** Integrated Services for the Development of Aquaculture and Fisheries, Tabuc Suba, Iliolo City, Philippines

Abstract

DNA barcoding has been a major advancement in the field of taxonomy, seeing much effort put into the barcoding of wide taxa of organisms, macro and microalgae included. The mitochondrial-encoded *cox1* and plastid-encoded *rbcl* has been proposed as potential DNA barcodes for rhodophytes, but are yet to be tested on the commercially important carrageenophytes *Kappaphycus* and *Eucheuma*. This study gauges the effectiveness of four markers, namely the mitochondrial *cox1*, *cox2*, *cox2-3* spacer and the plastid *rbcl* in DNA barcoding on selected *Kappaphycus* and *Eucheuma* from Southeast Asia. Marker assessments were performed using established distance and tree-based identification criteria from earlier studies. Barcoding patterns on a larger scale were simulated by empirically testing on the commonly used *cox2-3* spacer. The phylogeny of these rhodophytes was also briefly described. In this study, the *cox2* marker which satisfies the prerequisites of DNA barcodes was found to exhibit moderately high interspecific divergences with no intraspecific variations, thus a promising marker for the DNA barcoding of *Kappaphycus* and *Eucheuma*. However, the already extensively used *cox2-3* spacer was deemed to be in overall more appropriate as a DNA barcode for these two genera. On a wider scale, *cox1* and *rbcl* were still better DNA barcodes across the rhodophyte taxa when practicality and cost-efficiency were taken into account. The phylogeny of *Kappaphycus* and *Eucheuma* were generally similar to those earlier reported. Still, the application of DNA barcoding has demonstrated our relatively poor taxonomic comprehension of these seaweeds, thus suggesting more in-depth efforts in taxonomic restructuring as well as establishment.

Citation: Tan J, Lim P-E, Phang S-M, Hong DD, Sunarpi H, et al. (2012) Assessment of Four Molecular Markers as Potential DNA Barcodes for Red Algae *Kappaphycus* Doty and *Eucheuma* J. Agardh (Solieriaceae, Rhodophyta). PLoS ONE 7(12): e52905. doi:10.1371/journal.pone.0052905

Editor: Sofia Consuegra, Aberystwyth University, United Kingdom

Received: September 7, 2012; **Accepted:** November 23, 2012; **Published:** December 20, 2012

Copyright: © 2012 Tan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Project funding is by the E-Science Fund, Project No. 14-02-03-4027, Department of Fisheries Malaysia (53-02-03-1062) and University of Malaya Malaya, PPP (Phylogenetic analysis of *Kappaphycus* spp. and *Eucheuma* spp. with the application of various molecular markers; grant number: PV014/2011A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: phaikem@um.edu.my

Introduction

The introduction of DNA barcodes by Herbert and co-workers [1,2,3] has brought about large impacts on the advancement of systematic biology; where short, easily amplified regions of DNA exhibiting large variation among species, yet sufficiently variable within species, are constantly used for species delineation, identification as well as archiving with reference to known, established species [4,5,6]. The Barcode of Life Data System (BOLD) is notably the largest initiative in establishing a worldwide DNA barcode library, signifying its importance and popularity for the scientific community [7,8,9]. The usefulness of DNA barcoding is evident when dealing with taxa displaying phenotypic plasticity throughout diphasic or triphasic life cycles as well as taxa involving cryptic species. These phenomena are generally predominant in marine macroalgae, thereby enticing the application of DNA barcoding, as reported in numerous studies encompassing the order Gelidiales [10], Gigartinales, [11,12,13], Gracilariales [14,15], Laminariales [16], and Fucales [17]. Studies on DNA barcoding over broader taxa of rhodophytes have also been reported with promising results [18,19,20].

The rhodophytes *Kappaphycus* and *Eucheuma*, commercially known as “cottonii” and “spinosum”, respectively are widely established as lucrative sources of carrageenan, with Indonesia and the Philippines being the largest carrageenophyte producers worldwide [21]. Despite being extensively farmed, the morphologically diverse nature of *Kappaphycus* and *Eucheuma* still poses difficulties in species identification [22,23,24,25,26,27,28], even leading to the cultivation of mixed populations that inevitably reduces overall yield [29]. These have resulted in the subsequent employment of molecular phylogenetic studies which all share one main objective – to infer and understand the phylogenetic relationships between *Kappaphycus* and *Eucheuma* congeners. As of now, various molecular markers have been introduced for the molecular taxonomy of these carrageenophytes, namely the mitochondrial-encoded partial *cox1* and *cox2-3* spacer, nuclear-encoded ribosomal Internal Transcribed Spacer (ITS) and 28S large subunit (LSU), plastid-encoded *rbcl*, RuBisCO spacer and the 23S Universal Plastid Amplicon (UPA) [23,28,30,31]. However, the suitability of these genetic markers as DNA barcodes are to date, unassessed.

Table 1. Details of samples used in this study.

No. Sample Name	Operational Taxonomic Unit (OTU)	Sampling Location	Collection Code	GenBank Accession Numbers			
				Cox1	Cox2-3 spacer	Cox2	rbcl
1	<i>Kappaphycus alvarezii</i> 13 "Buaya" *	Sabangkat, Sabah, Malaysia	PSM11996-UMSS0144	-	JN663762	-	-
2	<i>Kappaphycus alvarezii</i> 18 "Tambalang Giant" *	Sabangkat, Sabah, Malaysia	PSM12001-UMSS0154	-	JN663768	-	-
3	<i>Kappaphycus alvarezii</i> 52 "Buaya" *	Omadal, Sabah, Malaysia	PSM12029-UMSS0196	-	JN663763	-	-
4	<i>Kappaphycus alvarezii</i> 53 "Tangan-tangan" *	Omadal, Sabah, Malaysia	PSM12030-UMSS0198	-	JN663773	-	-
5	<i>Kappaphycus alvarezii</i> 58 "Tangan-tangan" *	Omadal, Sabah, Malaysia	PSM12035-UMSS0203	JX624014	JN663774	JX624043	JX623985
6	<i>Kappaphycus alvarezii</i> 63 "Tambalang Giant" *	Sisipan, Sabah, Malaysia	PSM12043-UMSS0214	-	JN663769	-	-
7	<i>Kappaphycus alvarezii</i> 89 "Tambalang Brown" *	Sandakan, Sabah, Malaysia	PSM12059-UMSS0230	JX624015	JN663766	JX624044	JX623986
8	<i>Kappaphycus alvarezii</i> 103*	Sabangkat, Sabah, Malaysia	PSM12072-UMSS0243	JX624016	JN663776	JX624045	JX623987
9	<i>Kappaphycus alvarezii</i> 109 "Tangan-tangan" *	Semporna, Sabah, Malaysia	PSM12078-UMSS0249	-	JN663775	-	-
10	<i>Kappaphycus alvarezii</i> 121 "Tambalang Green" *	Pangkor Island, Perak, Malaysia	PSM12105-UMSS0260	JX624017	JN663772	JX624046	JX623988
11	<i>Kappaphycus alvarezii</i> 123 "Tambalang Brown" *	Pangkor Island, Perak, Malaysia	PSM12107-UMSS0262	-	JN663767	-	-
12	<i>Kappaphycus alvarezii</i> BA*	Semporna, Sabah, Malaysia	-	-	JN234760	-	-
13	<i>Kappaphycus alvarezii</i> BN*	Semporna, Sabah, Malaysia	-	-	JN234759	-	-
14	<i>Kappaphycus alvarezii</i> YF*	Semporna, Sabah, Malaysia	-	-	JN234762	-	-
15	<i>Kappaphycus alvarezii</i> 433	Teluk Ekas, Indonesia	PSM12290-UMSS0433	JX624018	JX624072	JX624047	JX623989
16	<i>Kappaphycus alvarezii</i> ZAM4 "Milo"	Zamboanga City, Mindanao, Philippines	AQHZA004-UMSS0380	JX624019	JX624073	JX624048	JX623990
17	<i>Kappaphycus alvarezii</i> V7 "Dark Green"	Son Hai, Vietnam	PSM12380-UMSS0525	JX624020	JX624074	JX624049	JX623991
18	<i>Kappaphycus alvarezii</i> E3*	Venezuela	-	-	AY687427	-	-
19	<i>Kappaphycus alvarezii</i> 2614*	Hawaii	-	-	FJ554862	-	-
20	<i>Kappaphycus alvarezii</i> UR13*	Tanzania	-	-	JQ713902	-	-
21	<i>Kappaphycus alvarezii</i> E130*	Tanzania	-	-	AY687427	-	-
22	<i>Kappaphycus alvarezii</i> Reef4*	Paje-Jambiani, Tanzania	-	-	JQ713901	-	-
23	<i>Kappaphycus alvarezii</i> E16*	Madagascar	-	-	AY687430	-	-
24	<i>Kappaphycus alvarezii</i> E57*	Hawaii	-	-	AY687432	-	-
25	<i>Kappaphycus alvarezii</i> E71*	Hawaii	-	-	AY687433	-	-
26	<i>Kappaphycus alvarezii</i> 919*	Hawaii	-	-	FJ554860	-	-
27	<i>Kappaphycus alvarezii</i> 3955*	Hawaii	-	-	FJ554861	-	-
28	<i>Kappaphycus striatum</i> 1 "Yellow Flower" *	Sabangkat, Sabah, Malaysia	PSM11984-UMSS0128	JX624021	JN663779	JX624050	JX623992
29	<i>Kappaphycus striatum</i> 31 "Green Flower" *	Sabangkat, Sabah, Malaysia	PSM12011-UMSS0170	JX624022	JN663780	JX624051	JX623993
30	<i>Kappaphycus striatum</i> 59 "Green Flower" *	Bum-Bum Island, Malaysia	PSM12039-UMSS0208	JX624023	JN663777	JX624052	JX623994
31	<i>Kappaphycus striatum</i> 60 "Green Flower" *	Bum-Bum Island, Malaysia	PSM12040-UMSS0209	-	JN663778	-	-
32	<i>Kappaphycus striatum</i> AG*	Semporna, Sabah, Malaysia	-	-	JN234763	-	-
33	<i>Kappaphycus striatum</i> GF*	Semporna, Sabah, Malaysia	-	-	JN234765	-	-

Table 1. Cont.

No. Sample Name	Operational Taxonomic Unit (OTU)	Sampling Location	Collection Code	GenBank Accession Numbers		
				Cox1	Cox2-3 spacer	Cox2
34 <i>Kappaphycus striatum</i> GTF*		Semporna, Sabah, Malaysia	-	JN234764	-	-
35 <i>Kappaphycus striatum</i> 460		Kertasari, Indonesia	PSM12293-UMSS0460	JX624024	JX624075	JX624053
36 <i>Kappaphycus striatum</i> GUI4 "Cottonii"		Guimaras Is. Panay, Philippines	AQHGUI004-UMSS0360	JX624025	JX624076	JX624054
37 <i>Kappaphycus striatum</i> SIT5 "Cottonii light green (sacoli)"		Sitangkai, Tawi Mindanao, Philippines	AQHSIT005-UMSS0394	JX624026	JX624077	JX624055
38 <i>Kappaphycus striatum</i> V6 "Payaka Green"		Cam Ranh, Vietnam	PSM12379-UMSS0524	JX624027	JX624078	JX624056
39 <i>Kappaphycus striatum</i> E89*		Philippines	-	-	AY687434	-
40 <i>Kappaphycus striatum</i> 83*	KS2	Sabangkat, Sabah, Malaysia	PSM12053-UMSS0224	JX624028	JN663781	JX624057
41 <i>Kappaphycus striatum</i> 98*		Sabangkat, Sabah, Malaysia	PSM12067-UMSS0238	JX624029	JN663782	JX624058
42 <i>Kappaphycus striatum</i> 105*		Sabangkat, Sabah, Malaysia	PSM12074-UMSS0245	JX624030	JN663783	JX624059
43 <i>Kappaphycus striatum</i> D13*		Semporna, Sabah, Malaysia	-	-	JN645177	-
44 <i>Kappaphycus striatum</i> D14*		Semporna, Sabah, Malaysia	-	-	JN645178	-
45 <i>Kappaphycus striatum</i> SIT4 "Kab-kab green"		Sitangkai, Tawi Mindanao, Philippines	AQHSIT004-UMSS0393	JX624031	JX624079	JX624060
46 <i>Kappaphycus striatum</i> E117*		Indonesia	-	-	AY687435	-
47 <i>Kappaphycus striatum</i> E48*		Indonesia	-	-	AY687431	-
48 <i>Kappaphycus</i> sp. 14 "Aring-aring" *	KAR	Sabangkat, Sabah, Malaysia	PSM11997-UMSS0146	-	JN663784	-
49 <i>Kappaphycus</i> sp. 49 "Aring-aring" *		Sabangkat, Sabah, Malaysia	PSM12026-UMSS0192	JX624032	JN663785	JX624061
50 <i>Kappaphycus</i> sp. 93 "Aring-aring" *		Sabangkat, Sabah, Malaysia	PSM12063-UMSS0234	JX624033	JN663786	JX624062
51 <i>Kappaphycus</i> sp. 115 "Aring-aring"		Sabangkat, Sabah, Malaysia	PSM12100-UMSS0255	JX624034	JX624080	JX624063
52 <i>Eucheuma denticulatum</i> 44 "Spinusum" *	ED1	Sabangkat, Sabah, Malaysia	PSM12021-UMSS0187	JX624035	JN663787	JX624064
53 <i>Eucheuma denticulatum</i> 45 "Spinusum" *		Sabangkat, Sabah, Malaysia	PSM12022-UMSS0188	JX624036	JN663788	JX624065
54 <i>Eucheuma denticulatum</i> 46 "Spinusum" *		Sabangkat, Sabah, Malaysia	PSM12023-UMSS0189	-	JN663789	-
55 <i>Eucheuma denticulatum</i> 56 "Spinusum" *		Omada, Sabah, Malaysia	PSM12033-UMSS0201	JX624037	JN663790	JX624066
56 <i>Eucheuma denticulatum</i> 57 "Spinusum" *		Omada, Sabah, Malaysia	PSM12034-UMSS0202	-	JN663791	-
57 <i>Eucheuma denticulatum</i> 99 "Spinusum" *		Sabangkat, Sabah, Malaysia	PSM12068-UMSS0239	-	JN663792	-
58 <i>Eucheuma denticulatum</i> DM*		Semporna, Sabah, Malaysia	-	-	JN234756	-
59 <i>Eucheuma denticulatum</i> AD*		Semporna, Sabah, Malaysia	-	-	JN980403	-
60 <i>Eucheuma denticulatum</i> AB*		Semporna, Sabah, Malaysia	-	-	JN234758	-
61 <i>Eucheuma denticulatum</i> E13*		Indonesia	-	-	AY687429	-
62 <i>Eucheuma denticulatum</i> 454	ED2	Kertasari, Indonesia	PSM12292-UMSS0454	JX624038	JX624081	JX624067
63 <i>Eucheuma denticulatum</i> BOHS "Spinusum"		Bohol, Central Visayas, Philippines	AQHBOH005-UMSS0371	JX624039	JX624082	JX624068
64 <i>Eucheuma denticulatum</i> 41 "Cacing"		Sabangkat, Sabah, Malaysia	PSM12018-UMSS0181	JX624040	JX624083	JX624069
65 <i>Eucheuma denticulatum</i> 42 "Cacing"		Sabangkat, Sabah, Malaysia	PSM12019-UMSS0183	JX624041	JX624084	JX624070

Table 1. Cont.

No. Sample Name	Operational Taxonomic Unit (OTU)	Sampling Location	Collection Code	GenBank Accession Numbers			
				Cox1	Cox2-3 spacer	Cox2	rbcl
66 <i>Euचेuma denticulatum</i> 97 "Cacing"		Sabangkat, Sabah, Malaysia	PSM12066-UMSS0237	JX624042	JX624085	JX624071	JX624013
67 <i>Euचेuma denticulatum</i> CG *		Semporna, Sabah, Malaysia	-	-	JN234757	-	-
68 <i>Euचेuma denticulatum</i> E32*		Indonesia	-	-	AY687437	-	-
69 <i>Euचेuma denticulatum</i> 888*		Hawaii	-	-	FJ554859	-	-
70 <i>Euचेuma denticulatum</i> E8*	ED3	Madagascar	-	-	AY687428	-	-
71 <i>Euचेuma denticulatum</i> PAC5*		Paje-Jambiani	-	-	JQ713903	-	-
72 <i>Euचेuma denticulatum</i> E60*		Mauritius	-	-	AY687439	-	-
73 <i>Euचेuma denticulatum</i> 3953		Hawaii	-	-	FJ561733	-	-
74 <i>Euचेuma platycladum</i> E65*	EP	Tanzania	-	-	AY687423	-	-
75 <i>Euचेuma platycladum</i> E111*		Kenya	-	-	AY687422	-	-
76 <i>Solieria</i> 120		Merambong, Johor, Malaysia	PSM12104-UMSS0259	-	JN663793	-	-

Specimens were grouped into Operational Taxonomic Units (OTU) for selected analyses. Dashes (–) indicate non-available or irrelevant data. Asterisks (*) indicate samples where corresponding *cox2-3* spacer sequences were obtained from the GenBank and used for *Large Dataset Assessment*.
doi:10.1371/journal.pone.0052905.t001

Having high quality DNA barcodes for *Kappaphycus* and *Euचेuma* would hasten the identification and recording of these seaweeds throughout the globe, particularly those within Southeast Asian waters, where it is believed that many species are yet to be described. Accurate identification of undescribed species would be most useful in developing improved commercial varieties, preferably ones with higher carrageenan yield, faster growth as well as better disease resistance. Apart from species diversity, data on the distribution of *Kappaphycus* and *Euचेuma* would also be useful in the detection of invasive strains [23,32]. The potential benefits of DNA barcoding for *Kappaphycus* and *Euचेuma* underline the main purpose of the present study, which is to gauge the efficacy of four molecular markers, including the commonly used mitochondrial *cox1* and plastid *rbcl* barcodes, the mitochondrial *cox2-3* spacer and a newly designed *cox2* molecular marker as potential DNA barcodes for selected *Kappaphycus* and *Euचेuma* specimens found in Southeast Asia. Marker assessment and comparison were carried out using the tree-based DNA identification technique as well as the distance technique- *Best Match* (BM), *Best Close Match* (BCM) and *All Species Barcodes* (ASB) criteria [1,33]. Additionally, the effect of larger datasets on the accuracy and robustness of DNA barcoding was assessed by incorporating sequences of the widely used *cox2-3* spacer marker which includes specimens collected from different parts of the world, to that generated in the present study.

Materials and Methods

Ethics Statement

Kappaphycus and *Euचेuma* specimens were not collected from any national parks or protected areas, thus not requiring any specific permits for sampling. Specimens were collected from open coastal areas as well as from aquaculture sites with consents from the respective owners. Members of the two genera are not endangered or protected species.

Sample Processing

Details of samples used in this study are summarized in Table 1. Specimens were either sampled directly from farms, retrieved via snorkeling or scuba diving. Sample preservation and genomic DNA extraction were based on protocols described by [29]. Details on molecular markers and primers utilized in this study are summarized in Table 2. The mitochondrial encoded cytochrome *c* oxidase subunit 2 (*cox2*) primers were designed based on the *cox2* region of the complete mitochondrial genome of *Chondrus crispus* (GenBank no. NC_001677) [34]. PCR parameters of the *cox2* molecular marker are similar to that of *rbcl*, but with 30 amplification cycles. Amplicons were electrophoresed using a SYBR®Safe (Invitrogen, USA) stained 1.0% agarose gel, prior to purification using LaboPass™ (Cosmo Gentech, Korea) gel and PCR purification kits. Purified products were sent to Lucigen (Taiwan) for ABI SoLiD sequencing.

Resulting electropherograms were viewed, truncated at terminals and contigs generated using ChromasPro V1.5 (Technelysium Pty Ltd). Multiple sequence alignments were computed using ClustalX V2.0 [35] and converted into NEXUS format prior to subsequent analyses.

Considering the poor taxonomic status of *Kappaphycus* and *Euचेuma*, certain molecular marker assessments were conducted using two different criteria, where (i) species were named and processed based on Operational Taxonomic Units (OTU) (Table 1) [36], or (ii) original species names were used with no alterations whatsoever throughout the analyses, herein termed non-OTU. The 2nd criterion is only applicable under the *Large Dataset*

Table 2. Primer details and corresponding annealing temperatures for the *cox1*, *cox2*, *cox2-3* spacer and *rbcl* genetic markers used in this study.

DNA Markers	Primers	Primer Sequences*	Reference	Annealing temperature, T _m
Cox1	COXI43F	5'-TCAACAAATCATAAAGATATTGGWACT-3'	[46,47]	52°C
	C622F	5'-CCTGTNTTAGCAGGWGCTATTACAATGC-3'		
	C880R	5'-ACAGTATACATATGATGNGCTCAAAC-3'		
	COXI1549R	5'-AGGCATTCTTCAAANGTATGATA-3'		
Cox2-3 spacer	Cox2_for	5'-GTACCWTCITTDGRRRKDAATGTGATGC-3'	[55]	50°C
	Cox3_rev	5'-GGATCTACWAGATGRAAWGGATGTC-3'		
Cox2	Kcox2_F71	5'-TTCAAGATCCTGCACTCC-3'	This Study	51°C
	Kcox2_R671	5'-ATTTCACTGCATTGGCCAT-3'		
rbcl	F-7	5'-AACTCTGTAGTAGAACGNACAAG-3'	[58,59]	50°C
	F-577	5'-GTATATGAAGGTCTAAAAGGTGG-3'		
	R-753	5'-GCTCTTTCATACATATCTCC-3'		
	R-rbcS start	5'-GTTCTTTGTGTAATCTCAC-3'		

Ambiguous nucleotide codes are in accordance to IUPAC: K = G/T; R = A/G; W = A/T; N = A/T/C/G.
doi:10.1371/journal.pone.0052905.t002

Assessment. An OTU in this context refers to a cluster of species-specific specimens constituting a monophyletic clade which is (i) sufficiently variable genetically from the sister taxa such that bifurcating patterns are observed or (ii) geographically distinct. Morphological criteria were not applied due to the extensive morphological plasticity of these red algae.

Distance-based Assessment

Distance analyses were performed using TaxonDNA's Species Identifier v1.7.7 [33] for each molecular marker dataset. The pairwise distances for intra- and interspecific frequencies; calculated using both the uncorrected and the Kimura 2-parameter (K2P) corrected pairwise distances, were plotted to observe overlaps in genetic variability, if present. These pairwise distances were computed using the (i) total overlap range and (ii) 90% overlapping range- the largest 5% of the intraspecific and lowest 5% of the interspecific samples excluded. Minimum base pairs in common for distance calculations were set at 300 bp for all molecular markers assessed. The effectiveness of respective molecular markers was tested based on the *Best Match* (BM) and *Best Close Match* (BCM) criteria comprehensively described by Meier and co-workers [33]. In short, the BM criterion assigns a species name to the query sequence based on its best barcode match, regardless of the magnitude of similarity between the query and the barcode sequences. BCM, on the other hand involves the initial identification of the best-matched barcode, determination of whether the barcode is sufficiently similar to the query, followed lastly by the assignment of a species name [33]. The smallest interspecific distance generated using the Pairwise Summary module for each molecular marker was used as threshold value for BM and BCM computations.

Tree-based Assessment

Neighbor-Joining (NJ) trees were generated based on the Kimura 2-parameter model using default PAUP 4.0b10 [37] settings for each DNA marker to provide visual displays of genetic variation within and between species. Tree nodal supports were generated via 1,000 bootstrap replicates. Resulting NJ trees were analyzed and processed using Figtree v1.3.1 [38]. For all trees, the success of species identification was determined based on the criteria proposed in [1] and [33]. Under Hebert and co-workers [1], identification was considered a success when the query clusters along with all conspecific

sequences; and considered a failure when the query matches conspecific sequences occurring in multiple clusters or clades within the tree. Ambiguous were all singletons within the dataset. Meier and co-workers [33] used a revised, and more stringent identification criteria for the tree-based approach. Queries were considered correctly identified when in polytomy with conspecifics, or at least one node into a clade of conspecifics. Misidentification were those queries in polytomy with only allospecific sequences or those at least one node into an allospecific clade. Queries without conspecific sequences or queries sister to conspecifics were considered as ambiguous or unidentified. Additionally, the phylogeny of *Kappaphycus* and *Eucheuma* were interpreted and briefly described based on the resulting NJ, Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian Inference (BI) trees.

Large Dataset Assessment

The *cox2-3* spacer, currently the most widely used genetic marker for *Kappaphycus* and *Eucheuma*, were used to simulate empirically the effectiveness of a molecular marker in relatively larger datasets. *Cox2-3* spacer sequences from recent studies [23,29,32,39] were obtained from the GenBank. Specimens of ambiguous or uncertain identity were excluded from the dataset. Multiple sequence alignments were generated and subjected to similar K2P distance and tree-based analyses. Datasets were analyzed based on the application of the OTU and non-OTU criteria. An additional distance-based identification criterion termed *All Species Barcodes* (ASB) was also employed using Species Identifier v1.7.7 [33]. This relatively more subtle approach compares the query to barcodes generated using the same threshold for BM and BCM. Queries were considered as successfully identified when matched with at least two conspecific barcodes of the species in question. Ambiguous were all queries followed by only one conspecific barcode or only a portion of the conspecific sequences. Queries were designated misidentified when matched with allospecific barcodes.

Phylogenetic trees were also inferred for the *cox2-3* spacer based on Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (BI) algorithms. *Solieria* was used as the outgroup for each analyses [29]. Parsimony analyses were conducted using PAUP 4.0b10 [37] as heuristic searches using 1,000 bootstrapping replications; with 100 stepwise random sequence addition and

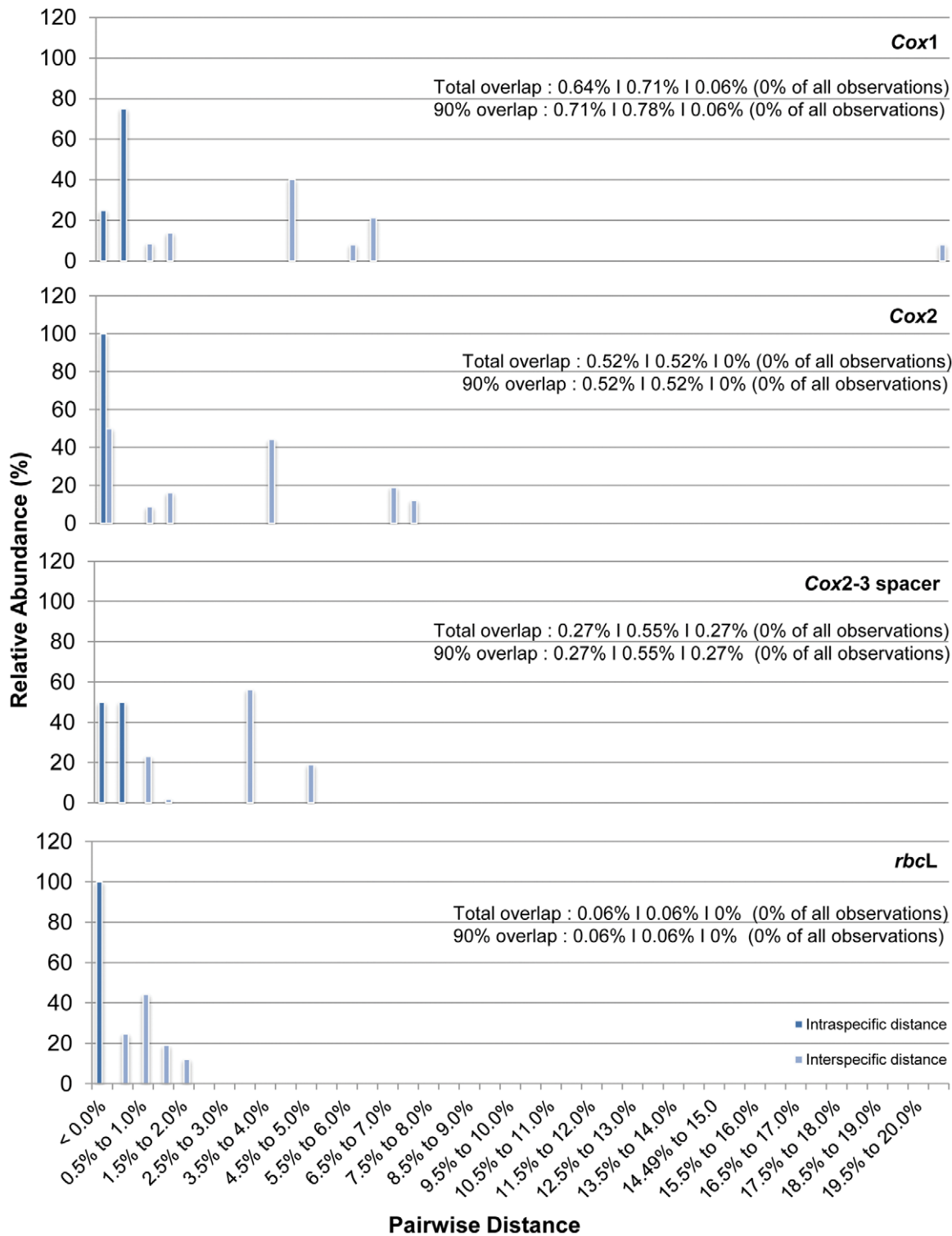


Figure 1. Plot of intra- and interspecific genetic distances for the *cox1*, *cox2*, *cox2-3* spacer and *rbcL* DNA markers. Numeric values are arranged according to: the difference between the smallest interspecific but intrageneric distance and the largest intraspecific distance|smallest pairwise distance between interspecific but intrageneric sequences|largest pairwise distance between intraspecific sequences; followed by the number of observations affected (in brackets).
doi:10.1371/journal.pone.0052905.g001

tree bisection reconnection (TBR) branch swapping. All characters were assigned unordered and unweighted. Retention indices (RI) and Consistency indices (CI) were also generated.

ML analysis was performed using raxmlGUI [40] and BI via Mr. Bayes v3.2.1 [41,42]. Best fit nucleotide substitution models were determined for each DNA region using Kakusan 4 [43],

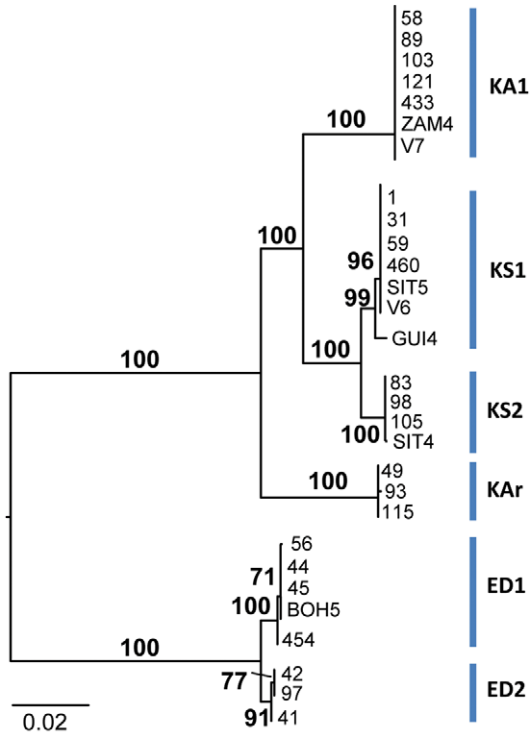
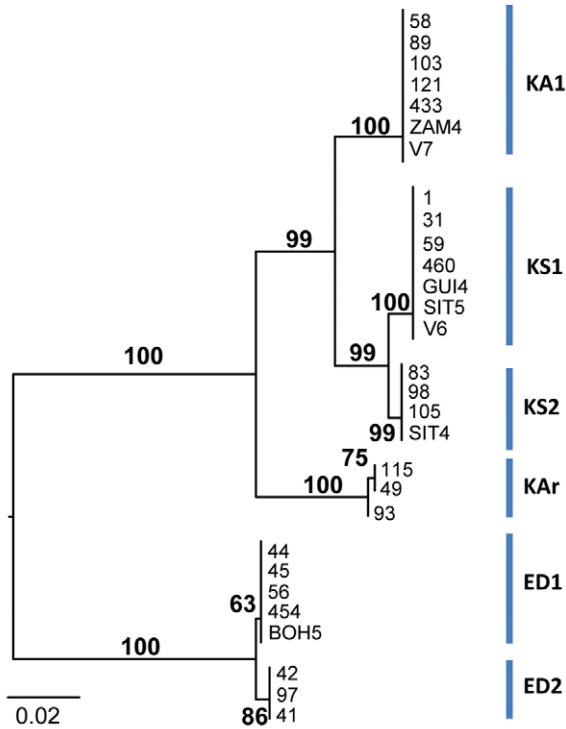
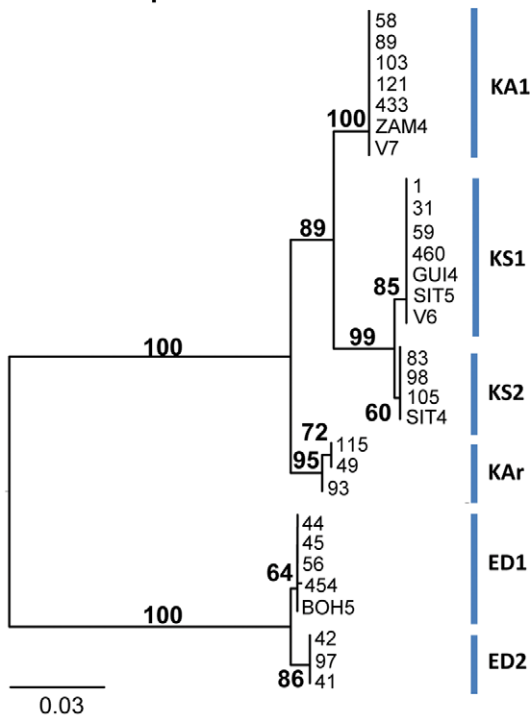
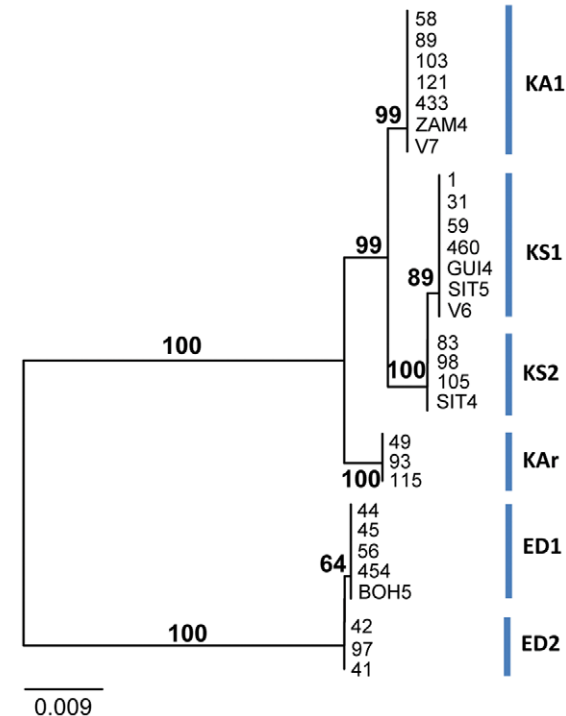
A *Cox1*B *Cox2*C *Cox2-3 spacer*D *rbcL*

Figure 2. Neighbour Joining (NJ) trees of selected *Kappaphycus* and *Eucheuma* from Southeast Asia based on (a) *cox1*; (b) *cox2*; (c) *cox2-3 spacer*; (d) *rbcL* molecular markers. Numbers at node indicate corresponding bootstrap values over 1,000 replicates. Clade annotations represent Operational Taxonomic Units (OTU), where KA = *Kappaphycus alvarezii*; KS = *Kappaphycus striatum*; KAr = *Kappaphycus* sp. "Aring-aring"; ED = *Eucheuma denticulatum*.

doi:10.1371/journal.pone.0052905.g002

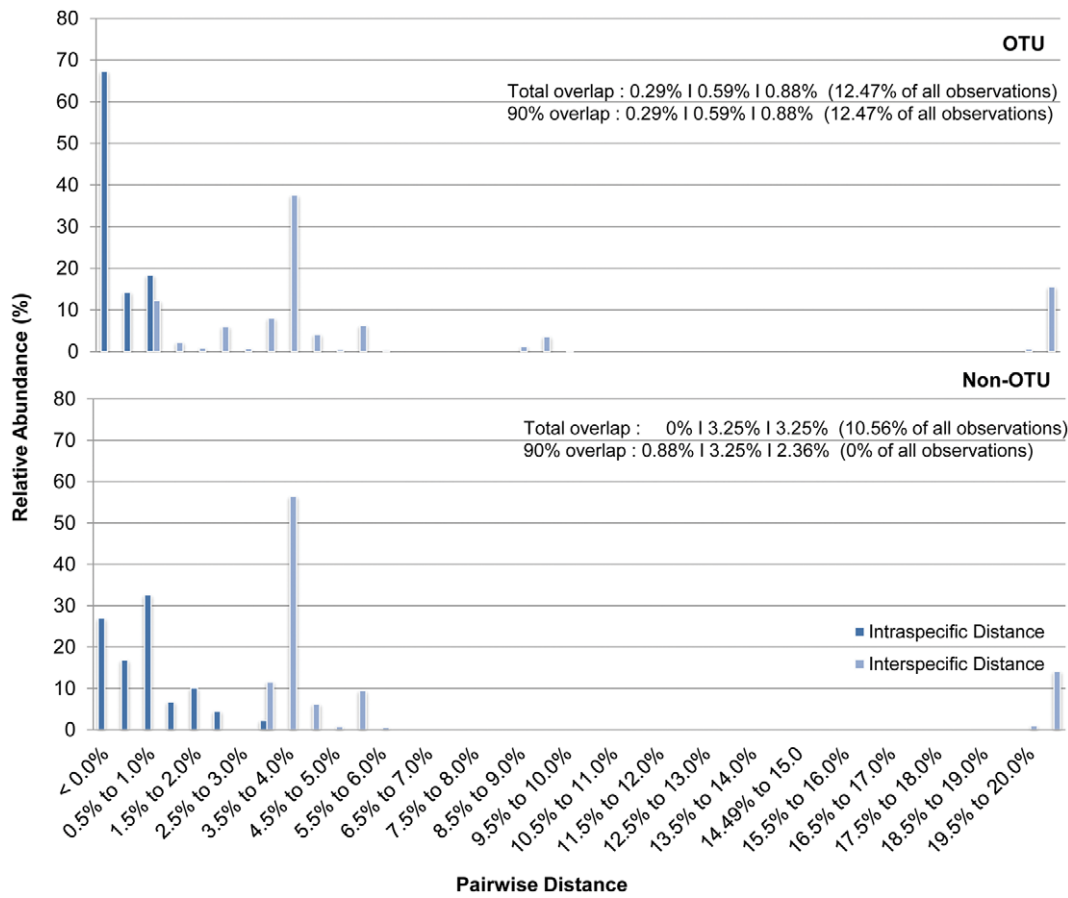


Figure 3. Plot of intra- and interspecific genetic distances of the *cox2-3* spacer with the application of OTU and non-OTU criteria. Numeric values are arranged according to: the difference between the smallest interspecific but intragenetic distance and the largest intraspecific distance|smallest pairwise distance between interspecific but intragenetic sequences|largest pairwise distance between intraspecific sequences; followed by the number of observations affected (in brackets). doi:10.1371/journal.pone.0052905.g003

generating command files for RAxML and input files for MrBayes. ML trees were inferred using the ML + thorough bootstrap algorithm, with the GTR+GAMMA model over 20 independent searches and 1,000 non-parametric bootstrap replicates. Identical sequences were not omitted for the analysis. For the BI analyses, two sets of four Monte Carlo Markov Chains (MCMC) were performed in parallel over 2,000,000 generations, with trees sampled every 500th generation. Tracer v1.5 (<http://196.bio.ed.ac.uk/software/tracer/>) was used to assess convergence of log likelihood values, where 100,000 generations were discarded as burn-ins, well after stationarity was achieved. Results were used to generate a 50% majority-rule consensus tree. All trees were again analysed and processed using Figtree v1.3.1 [38].

Results

All 29 sequences were continuous and easily aligned for all four molecular markers, generating datasets of different final lengths: *cox1* (1,411 bp), *cox2* (575 bp), *cox2-3* spacer (365 bp) and for *rbcL* (1,464bp). *Cox1* exhibited the most phylogenetically informative sites of 277 (19.6%) characters, followed by *rbcL* with 126 (8.6%) characters, *cox2* with 115 (20%) characters and the *cox2-3* spacer with 76 (20.8%) characters. All sequences were deposited into GenBank (see Table 1). The *cox1* sequences amplified in this study encompasses the ones by Saunders et al. [19] and Robba et al. [18].

Distance-based Assessment

Pairwise distances (based on the corrected Kimura 2-parameter) showing the intra- and interspecific genetic variability for *cox1*, *cox2*, *cox2-3* spacer and *rbcL* were plotted in Figure 1. Both the corrected and uncorrected pairwise distances generated similar results for *cox2*, *cox2-3* spacer and *rbcL*. No overlapping in terms of intra- and interspecific genetic divergence was observed for all four molecular markers. However, the *rbcL* marker showed the least distance (0.06% differences for total and 90% overlaps) between the smallest pairwise distance among interspecific but intragenetic samples and the largest pairwise distance among intraspecific sequences- smaller “barcoding gap”. This was followed by *cox2-3* spacer (0.06% for total and 90% overlaps), *cox2* (0.52% for total and 90% overlaps) and *cox1* (K2P: 0.64% for total overlaps and 0.71% for 90% overlaps; uncorrected pairwise distance: 0.63% for total overlaps and 0.70% for 90% overlaps).

Results based on the *Best Match* (BM) and *Best Close Match* (BCM) criteria are as follows: All four genetic markers were able to correctly identify all 29 species based on the BM criteria. For the BCM criteria, *rbcL* showed the highest success in identification with 100%, followed by *cox2* (96.6%), *cox2-3* spacer (93.1%) and *cox1* (79.3%). No matches were recorded at values 3.44%, 6.89% and 20.68% for *cox2*, *cox2-3* spacer and *cox1* respectively.

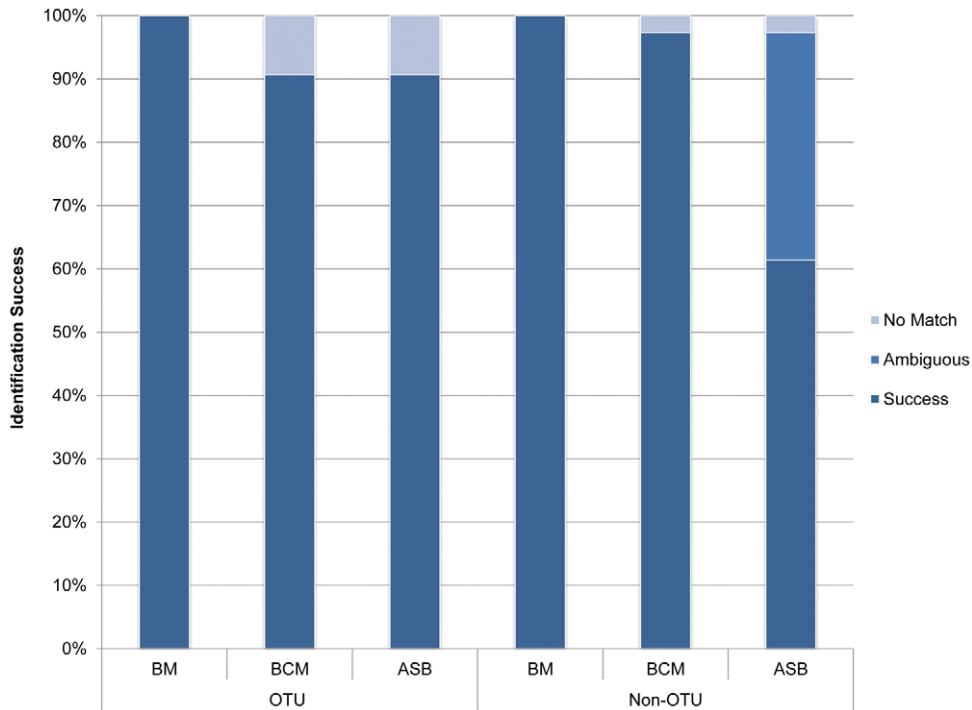


Figure 4. Identification success of the *cox2-3* spacer. *Kappaphycus* and *Eucheuma* samples were categorized under either Operational Taxonomic Units (OTU) or non-OTU conditions. Results were generated based on the Best Match (BM), Best Close Match (BCM) and All Species Barcodes (ASB) criteria.

doi:10.1371/journal.pone.0052905.g004

Tree-based Assessment

Resulting NJ phylogenetic trees are shown in Figure 2. Similar tree topologies were observed based on the *cox1*, *cox2*, *cox2-3* spacer and the *rbcL* molecular markers, all showing an apparent phylogenetic delineation between *Kappaphycus* and *Eucheuma* i.e. Neighbor Joining bootstrap supports (NJ-BS) of 100% respectively. Better resolution was observed for the mitochondrial-encoded genetic markers compared to the plastid-encoded *rbcL* counterpart on the species level, with relatively higher discrimination in the clustering of specimens.

Tree-based DNA identification using Hebert et al.'s [1] criteria showed 100% identification success across all species for all four genetic markers. Apart from the plastid *rbcL* DNA marker, identification success saw a decrease with the application of Meier et al.'s [33] revised identification criteria. For the *cox1* marker, identification success was recorded at 89.7%, with the remaining 10.3% sequences assigned as ambiguous. Both the *cox2* and *cox2-3* spacer showed identification success of 96.6%, with 3.4% ambiguity.

Tree topologies based on ML, MP and BI analyses were identical to that of the NJ trees and were not shown (to be distributed upon request). For the monophyletic *Kappaphycus* clade, four subclades were generated, denoted as OTUs KA1, KS1, KS2 and KA; corresponding to *K. alvarezii*, *K. striatum* 1, *K. striatum* 2 and *Kappaphycus* sp. “*Aring-aring*” respectively. All *K. alvarezii* specimens were inferred to be the same species with no significant genetic variation by all four markers (NJ-BS = 99–100%). Two genotypic strains of *K. striatum*-KS1 (1, 31, 59, 460, GUI4, SIT5 and V6) and KS2 (83, 98, 105 and SIT4) were observed with high support (NJ-BS = 99–100%) for each marker, although the clustering was less obvious for *rbcL*. *Cox1* demonstrated more specific delineation within species, where the Philippine GUI4

specimen was inferred to be genetically different from the rest of the KS1 specimens (NJ-BS = 99%). Based on all four markers, *Kappaphycus* sp. “*Aring-aring*”, first reported by Tan and co-workers (2012), formed a highly supported monophyletic clade (NJ-BS = 95–100%) sister to that of *K. alvarezii* and *K. striatum* (NJ-BS = 100%). Specimens 49 and 115, which were both from Sabah, were shown with moderate support (NJ-BS = 72% and 75%) to be more closely related to one another compared to that of sample 93 by the *cox2-3* spacer and *cox2* genetic marker respectively.

Eucheuma denticulatum specimens formed a monophyletic clade composed of subclades termed as ED1 (44, 45, 56, 454 and BOH5) and ED2 (41, 42 and 97) for all four molecular markers (NJ-BS = 100%). Similar to that observed for *K. striatum*, the assortment of ED1 and ED2 was not that discrete for the *rbcL* marker. Conversely, the *cox1* marker exhibited relatively better intraspecific delineation, where the Indonesian 454 specimen was shown to be different from the remaining ED1 samples (NJ-BS = 100%). Although moderately supported (NJ-BS = 77%), samples 42 and 97 were inferred by *cox1* to be distinct from that of sample 41.

Despite the equal ability of all four markers to cluster species consistently and somewhat accurately, the *cox1*, *cox2* and *rbcL* NJ trees showed increased robustness in terms of nodal supports, of which the *cox1* marker exhibited more specific resolution in terms of intraspecific genetic variations.

Large Dataset Assessment

Multiple sequence alignments with a length of 341bp were generated from the dataset incorporating all selected 76 *cox2-3* spacer sequences (including the *Solieria* outgroup), representing ten OTUs or five non-OTUs. *Cox2-3* spacer sequences generated from this study were truncated to allow better comparisons with shorter

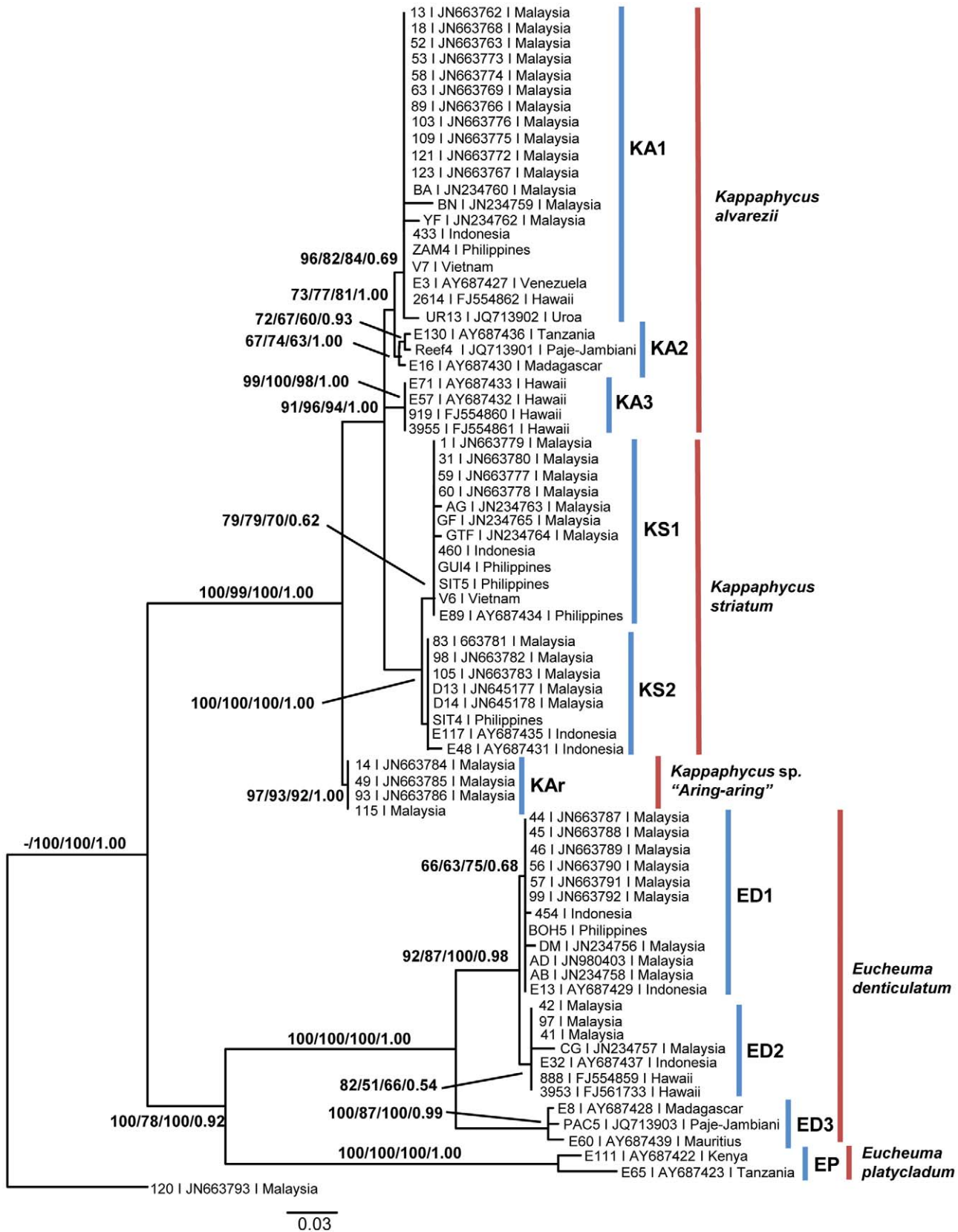


Figure 5. Maximum likelihood 50% majority-rule consensus tree based on the *cox2-3* spacer. Number at nodes is arranged according to NJ bootstrap support/ML bootstrap support/MP bootstrap support/Bayesian posterior probabilities. *Large dataset assessment*: blue lines indicate Operational Taxonomic Units (OTU), whereas red lines represent non-OTU clusters. The *Solieria* outgroup was excluded in the NJ analysis. doi:10.1371/journal.pone.0052905.g005

GenBank counterparts. 118 phylogenetically informative and 192 constant characters were recorded.

Figure 3 illustrates the intraspecific and interspecific genetic divergence based on the K2P model between OTUs and non-OTUs. For the former, a pairwise distance overlap of 0.29% was observed between the smallest interspecific and the largest intraspecific sequences; whereas no overlaps were observed for the latter.

Identification successes based on the BM and BCM criteria are shown in Figure 4. All species were correctly identified under BM for OTU and non-OTU analyses. Under BCM selection, 90.66% of samples were correctly identified when species were considered as OTUs, and 97.33% when species were not considered as OTUs. No matches were found for the remaining queries. Under the *All Species Barcodes* (ASB) identification criteria, an identification success of 90.66% and 61.33% were observed for OTU and non-OTU species categorization respectively. 36% of sequences were considered as ambiguous under the non-OTU criteria whereas the remaining queries were designated matchless.

Due to the similarity in tree topologies based on the NJ, ML, MP and BI algorithms, and for simplicity, resulting data are compiled and depicted as Figure 5. The NJ, MP and BI phylogenetic trees are provided as supporting data Figure S1, S2 and S3 respectively. Tree-based DNA identification is assessed using the NJ tree without taking into account the *Solieria* outgroup not used in the NJ dataset. In accordance to identification criteria by Hebert and co-workers [1], 100% identification success is reported when tree-based identification is based on the OTU concept. When taxa are queried using conventional taxonomic naming (non-OTU), percentage of successful identification is reduced to 94.6%, where the remaining 5.4% indicated misidentifications. Application of Meier et al.'s [33] identification criteria returned comparatively lower successful identification rates, where 95.9% success and 4.1% ambiguity was recorded for OTUs; and 67.6% successful identification, 27% ambiguity and 5.4% misidentification for non-OTUs.

Phylogenetic analyses of the entire *cox2-3* spacer dataset is inclusive of the *Solieria* outgroup. A total of 32 most parsimonious trees were produced by Parsimony analysis, with a Retention Index (RI) and Consistency Index (CI) of 0.9748 and 0.7934 respectively. The tree topology is generally similar to the NJ trees generated for molecular marker assessment, with most specimens clustering into existing clades. Remaining GenBank samples, mostly of different species or locality, were inferred as additional, discrete monophyletic clades. Nevertheless, the overall tree topology is still congruent to that of phylogenetic trees reported in earlier studies [23,28,29] and will not be discussed in detail. Nodal supports of the phylogenetic tree (Figure 5) were arranged in the order of Neighbor Joining bootstrap supports (NJ-BS), Maximum Likelihood bootstrap supports (ML-BS), Maximum Parsimony bootstrap supports (MP-BS) and Bayesian Inference posterior probabilities (BI-PP). Major nodes were mostly moderate to highly supported despite the addition in taxa. Fairly high support were observed for the monophyly of *Kappaphycus* (NJ-BS = 100%; ML-BS = 99%; MP-BS = 100%; BI-PP = 1.00) and *Euचेuma* (NJ-BS = 100%; ML-BS = 78%; MP-BS = 100%; BI-PP = 0.92) used in this study. Although strongly supported (NJ-BS = 91%; ML-BS = 96%; MP-BS = 94%; BI-PP = 1.00), polytomy is observed between *K. alvarezii* (KA1, KA2 and KA3) and *K. striatum*; where KA2 (NJ-BS = 67%; ML-BS = 74%; MP-BS = 63%; BI-PP = 1.00) and KA3 (NJ-BS = 99%; ML-BS = 100%; MP-BS = 98%; BI-PP = 1.00) represent *K. alvarezii* specimens from the African and Hawaiian regions respectively. *Kappaphycus* sp. “*Aring-arang*” (NJ-BS = 97%; ML-BS = 93%; MP-

BS = 92%; BI-PP = 1.00) and *Euचेuma denticulatum* (NJ-BS = 100%; ML-BS = 100%; MP-BS = 100%; BI-PP = 1.00) were inferred to be monophyletic as well. Clade EP (NJ-BS = 100%; ML-BS = 100%; MP-BS = 100%; BI-PP = 1.00) comprises of *Euचेuma platycladum* collected from around Southeast Africa.

Discussion

The dataset in the present study represents a small conglomerate of selected and commonly available *Kappaphycus* and *Euचेuma* samples from Southeast Asia. Sampling size is restricted to an amount supposedly cost-effective for molecular marker assessments as potential DNA barcodes. Owing to the relatively scarce records of *Kappaphycus* and *Euचेuma* in the GenBank, this study will also serve as a preliminary work to increase the amount of reference data using the potential DNA barcode, which can eventually be established as a barcode library.

Molecular Marker Assessment

The accuracy of a DNA barcode is largely determined by the magnitude of overlap between intraspecific variation and interspecific divergence. Ideally, the absence of an overlap would render species identification straightforward but this condition is virtually non-existent in very large datasets, in which the less the overlap, the more accurate it is for species identification [3,6,44]. Results indicated that no overlaps were observed for the *cox1*, *cox2*, *cox2-3* spacer and the *rbcL* genetic markers and have minimal impact on the accuracy of species identification for *Kappaphycus* and *Euचेuma* in this study (Figure 1). However, the absence of overlaps may be attributed to the small taxa limited to Southeast Asia (particularly so the genus *Euचेuma*) at this time, and is expected to change as more samples of different species or from different geographical locations are included in the future [6,45]. The presence of a “barcoding gap” (the absence of overlaps between intra- and interspecific genetic variations) may also be explained by the relatively lower genetic diversity amongst red algae as compared to those observed in arthropods.

All four molecular markers collectively showed an identification success of 100% for the BM criteria. Under the BCM criteria, the plastid encoded *rbcL* gene exhibited the highest identification success of 100% whereas the mitochondrial encoded *cox2*, *cox2-3* spacer and *cox1* spacer recorded slightly lower success of 96.6%, 93.1% and 79.3% respectively. Misidentification of species was not reported for all four molecular markers. The eventual increase in conspecific DNA sequences for *Kappaphycus* and *Euचेuma* is expected to reduce the probability of queries not meeting any matching sequences. All these results reflect the relatively less variable *rbcL* region as compared to mitochondrial counterparts, which was also reported by Geraldino et al. [46] and Yang et al. [47]. Although widely championed as a good potential DNA barcode, the relatively higher intraspecific variation of the *cox1* marker (<0.43% for *Kappaphycus*; <0.07% for *Euचेuma*) requires caution in avoiding misidentifications. Similar or higher intraspecific patterns were pointed out in previous studies associated with rhodophytes [18,19,46,47,48,49].

Tree-based DNA identification approach returned 100% success in species identification for all four molecular markers using the criteria by Hebert and co-workers [1]. Although the application of the relatively stricter identification criteria by Meier and co-workers [33] generally lowered the successful identification scores, they were still higher compared to the results derived from the distance-based approach. For instance, *cox1* and *cox2* showed a higher identification success of 89.7% and 96.6% respectively. Ambiguous identification in this study was mostly caused by

queries that formed a sister group to a cluster of conspecific sequences. This is also expected to decrease as more reference sequences are deposited in the GenBank for *Kappaphycus* and *Euचेuma*. Contrary to the better results obtained by the tree-based method in this study, empirical studies involving much larger taxa coverage have reported the preference of distance-based assessment over tree-based ones in terms of accuracy and robustness [33,50].

Based on both the distance and tree-based DNA identification approaches, the relatively conserved nature and the absence of overlaps between inter- and intraspecific genetic variability of the *rbcL* gene serves as a better potential DNA barcode for *Kappaphycus* and *Euचेuma*. However, the reduced genetic variation would also imply the incapability of the *rbcL* marker to detect incipient speciation or genetic diversity within species [5,51]. This will not only result in an underestimate of the actual genetic richness of these seaweeds, but will most likely overestimate interspecific variation as well due to the unavailability of closely related species [6]. The drawbacks of the plastid encoded *rbcL* marker can be accounted for using supplementary molecular data generated from the relatively more variable, mitochondrial derived *cox1*, *cox2* or *cox2-3* spacer. The concept of combined molecular data, in this case of molecular markers, is not new considering the occasional failures of the DNA barcodes in correctly and consistently identifying species, as observed in Karner blue butterflies [52], seagrasses [53] and flowering plants [54]. Utilization of genomic DNA from different origin i.e. mitochondrial, plastid or nuclear with different evolutionary rates would offer a better picture with respect to gene genealogy and evolutionary lineage.

The *cox1* and *cox2* DNA markers hold better potential as supplementary DNA barcodes compared to the *cox2-3* spacer because of their protein coding properties. These markers are more conserved across taxa with less indel mutations and are much simpler to check for errors through amino-acid translation. Lack of recombination and uniparental inheritance of mitochondrial markers are added advantages [55]. The relatively high mitochondrial copy numbers also enable ease of amplification. The *cox2* genetic marker, with its moderately high interspecific divergences within generic level and non-existent intraspecific variation, has demonstrated relatively successful results in terms of species identification in this study and is thus regarded as a potential mitochondrial DNA barcode for *Kappaphycus* and *Euचेuma*. Large scale application of this molecular marker is expected to bring about advancements mainly in the field of agriculture, taxonomy and biomonitoring. However, considering the already widely utilized *cox2-3* spacer exhibited almost similar barcoding traits as that of the mitochondrial *cox2* marker, continual implementation is recommended, in spite of its non-coding nature. The implementation of the *cox2* molecular marker across taxa may pose a problem as well since the full potential of the *cox2* marker has yet to be tested in other rhodophytes. Additionally, the massively abundant and readily available *cox1* sequences within the GenBank (although not extensively tested on *Kappaphycus* and *Euचेuma* at that time, *cox1* has been proposed as the potential universal DNA barcode for red algae) is apparently more practical and economical to work with, despite the relatively higher intraspecific variations which may reduce the accuracy for species identification [10,18,19,46,47,48,49]. Based on the results of this study, the authors have come to a consensus that the mitochondrial *cox2-3* spacer was in overall the best potential DNA barcode for *Kappaphycus* and *Euचेuma*; whereas the combined mitochondrial-encoded *cox1* and the plastid-encoded *rbcL* markers serve better as DNA barcodes encompassing the entire rhodophyte taxa.

The robustness and efficiency of DNA barcoding tends to increase with increased reference sequences and taxonomic scrutiny [1,6,44,50]. This includes genetically distinct individuals within a species' range to account for molecular markers with high variability such as the *cox1* [33]. Although this would greatly increase GenBank data and hence lead to an inevitable increment in terms of computational demand, larger and properly annotated datasets would expedite phylogeography, evolutionary biology, biodiversity or population genetic studies in the future.

Large Dataset Assessment

An additional 46 GenBank *cox2-3* spacer sequences of *Kappaphycus* and *Euचेuma* were compiled to simulate the effectiveness of a molecular marker in large, empirical datasets. Certain specimens of unknown or uncertain identity i.e. *Kappaphycus cottonii*, *Euचेuma isiforme* etc. were not used to avoid data confounding. Morphological plasticity of these seaweeds has rendered species identification and description challenging, even to seasoned taxonomists. As of now, distinctive morphological characters are still undiscovered despite ample of DNA evidence supporting the possible existence of new, or perhaps cryptic species. This has led to the amplification of Operational Taxonomic Units (OTU) in this context; representing genotypic diversity possibly overlooked via conventional morphological traits.

Distance-based results on the *cox2-3* spacer dataset have shown that the incorporation of more sequences decreases the "barcoding gap" (when not overlapping), and to the extent of forming overlaps between inter- and intraspecific divergence. When *Kappaphycus* and *Euचेuma* species were regarded as non-OTUs, the "barcoding gap" for the entire length of *cox2-3* spacer genetic marker decreased from 0.27% to 0%. When under OTU assortment, the intra- and interspecific genetic divergences formed an overlap of 0.29%. These observations were not surprising considering the larger sample size would eventually lead to higher occurrence of specimens with varying genetic composition [6,45]. This would undoubtedly affect the identification accuracy of DNA barcoding. BCM results saw a slight dip from 93.1% to 90.7% when samples were regarded as OTUs. Similar patterns were observed for tree-based DNA identification under Meier's (2006) [33] criteria, where identification success reduced from 96.6% to 95.9% for non-OTUs, caused mainly by ambiguous sequences. These "singleton" sequences or query sequences sister to known species can be avoided with increased data and taxonomic rectification or reformation.

The distance-based *All Species Barcodes* (ASB) assessment, being relatively stricter compared to BM and BCM; has reflected our poor taxonomic comprehension on *Kappaphycus* and *Euचेuma*, in which at our current state (represented by non-OTU species assortment); a mere 61.3% of queries could be identified correctly using DNA barcoding. This figure is probably overestimated considering the large amount of *Kappaphycus* and *Euचेuma* yet to be discovered and described. Misidentifications under non-OTU conditions were caused by the African (KA2) and Hawaiian (KA3) *K. alvarezii* specimens, where the latter remained unresolved using the *cox2-3* spacer and *cox1* (results not shown) molecular markers. Still, combined *cox2-3* spacer and RuBisCO spacer data has shown with moderate support that the KA3 specimens were more closely related to *K. alvarezii* than *K. striatum* [28]. This implies the possible limitation in resolving power of individual markers at intraspecific levels although the *rbcL* and *cox2* molecular markers are yet to be tested. On the other hand, ASB has returned relatively higher identification successes (90.7%) when *Kappaphycus* and *Euचेuma* species were regarded as OTUs, thus providing invaluable insights

into the possible taxonomic structure of these rhodophytes in the future.

Phylogenetic Interpretation

The inclusion of GenBank sequences did not cause significant alterations to the phylogenetic tree topology (Figure 2 and Figure 5), apart from the additional clade KA3 composed of Hawaiian *Kappaphycus alvarezii* and clade EP consisting of *Euclidean platycladum*. The monophyly of both *Kappaphycus* (NJ-BS = 100%; ML-BS = 99%; MP-BS = 100%; BI-PP = 1.00) and *Euclidean* (NJ-BS = 100%; ML-BS = 78%; MP-BS = 100%; BI-PP = 0.92) was moderate to highly supported. Taxonomic statuses of specimens excluded in this study i.e. *Kappaphycus cottonii*, *Kappaphycus procrusteanus*, *Betaphycus philippinensis* and *Euclidean isiforme* remained unaddressed at this time. Solid inferences can only be made when type specimens are sequenced, or when more reference sequences are available.

Kappaphycus

Similar patterns of biodiversity as to that earlier reported were observed [23,28,29]. *K. alvarezii* KA1, being the most widely distributed genotype, was reported throughout Southeast Asia, Africa, Columbia, Panama as well as Hawaii. *Cox2-3* spacer genealogy suggested that the African and Hawaiian strains of *K. alvarezii* KA1 may possibly be introduced strains, presumably traceable back to the Philippines. This is not surprising as there were efforts to introduce foreign, good strains of *Kappaphycus* for cultivation in the past [23,28,56]. Phylogenetic trees generated using the *cox1*, *cox2* and *rbcL* DNA markers (Figure 2) have collectively displayed similar topologies constituted of similar specimens for clade KA1. This supports the report by Tan and co-workers (2012) [29] that local varieties of cultivated *K. alvarezii* do not differ genetically despite the distinctive morphologies, at least for Malaysian specimens. *K. alvarezii* KA2 and KA3 represent genotypes unique to Africa and Hawaii respectively. The taxonomic position of the Hawaiian *K. "alvarezii"* specimens (KA3) remained unresolved using the *cox1* (results not shown) and *cox2-3* spacer marker, where the Hawaiian samples occur as a polytomy to both *K. alvarezii* and *K. striatum*. Although GenBank sequences are limited, combined analysis using both the *cox1* and *cox2-3* spacer suggested that clade KA3 is sister to that of *K. alvarezii* and *K. striatum* (Results not shown). This contradicts the phylogenetic results based on the combined *cox2-3* spacer and RuBiSCO spacer dataset, where the Hawaiian *Kappaphycus* specimens were inferred to be sister to *K. alvarezii* (KA1 and KA2 in this context) with moderate support (MP-BS = 70%) [28]. Application of the mitochondrial *cox2* and plastid *rbcL* marker may better elucidate the current confusion associated with clade KA3. As earlier reported, two genotypes were observed for *Kappaphycus striatum* [28,29], denoted here as clades KS1 and KS2. No distinctive differences in terms of gross morphology were identified between these two genotypes as of now and it is unsure whether cryptic species may apply in this situation. However, it appears that cultivated *K. striatum* mostly reside within clade KS1 and are largely common in Southeast Asia, although a recent study has also reported its occurrence in Uroa, Tanzania as a result of strain introduction [32]. *Kappaphycus* sp. "*Aring-aring*" from Malaysia was earlier shown to be phenotypically and genotypically different from *K. alvarezii* and *K. striatum* [29]; comparative studies against type specimens, particularly that of *Kappaphycus cottonii* (Weber-van Bosse) Doty ex P.C.Silva are currently underway in order to determine its validity as a new species.

Euclidean

*Euclidean*atoids are relatively poorly studied compared to *Kappaphycus*, possibly due to identification difficulties as well as lower economic value. *Euclidean denticulatum*, being the more popularly cultivated species, were clustered into three genotypically distinct subclades. Subclade ED1 represents specimens from Southeast Asia and Hawaii, subclade ED2 from Southeast Asia, Hawaii and Tanzania [28] whereas samples within ED3 were exclusively from Africa. Despite coexisting in the South China and Celebes seas, Southeast Asian *Euclidean denticulatum* ED1 and ED2 does not share similar morphological characteristics. This was shown by Ganzon-Fortes and co-workers [57], demonstrating the differences between the "*Endong/Spaghetti*" variety (ED2) from that of the usual *Spinosum* variety (ED1) of *Euclidean denticulatum*. The "*Endong*" variety, thence named *E. denticulatum* (Burman) Collins & Hervey var. *endong* Trono & Ganzon-Fortes var. nov. exhibited smooth, slender terete axes with whorls of determinate branchlets at predictable intervals [57]. Recent samplings in Malaysia have also revealed that the local *Euclidean denticulatum* "*Cacing*" variety fits the morphological and biochemical descriptions of the "*Endong*" variety, and its genotypic affinity also supported by *cox1* and *rbcL* molecular data (data not shown). The apparent and distinctive morphological characters of the "*Endong*" variety does not fit the original descriptions for *E. denticulatum*, thus suggesting that it may be a new species instead of a rare variety. With reference to Figure 2, this was shown to be potentially true using the full length *cox1* genetic marker, where the monophyly of clades ED1 and ED2 were highly supported. Relatively lower nodal supports were displayed by the *cox2-3* spacer and *cox2* DNA markers, followed lastly by the *rbcL* marker which fails to clearly indicate monophyly of ED1 and ED2. These patterns are reflective of the genetic variability of each molecular marker and suggest that clades ED1 and ED2 are probably undergoing divergence or have recently diverged. *E. denticulatum* ED3 was inferred to share a common ancestry with *E. denticulatum* ED1 and ED2, and are to date only reported in Africa. Considering the significantly different morphologies reported for ED2, it would be interesting to relook into the detailed anatomy of the African *E. denticulatum*.

Supporting Information

Figure S1 Neighbor-Joining (NJ) tree based on the *cox2-3* spacer marker. Numeric values at nodes indicate NJ bootstrap supports. *Large dataset assessment*: blue lines indicate Operational Taxonomic Units (OTU), whereas red lines represent non-OTU clusters. The *Solieria* outgroup was omitted from the analysis to enable implementation of tree-based identification criteria. (TIF)

Figure S2 Maximum Parsimony (MP) phylogenetic tree based on the *cox2-3* spacer. Number at nodes indicates MP bootstrap supports. *Large dataset assessment*: blue lines indicate Operational Taxonomic Units (OTU), whereas red lines represent non-OTU clusters. (TIF)

Figure S3 Bayesian (BI) phylogenetic tree based on the *cox2-3* spacer DNA marker. Number at nodes indicates BI posterior probabilities. *Large dataset assessment*: blue lines indicate Operational Taxonomic Units (OTU), whereas red lines represent non-OTU clusters. (TIF)

Acknowledgments

We would like to thank the Department of Fisheries Sabah, Prof. Dr. Suhaimi Md. Yasir, En. Adibi Rahiman bin Md. Nor, Dr. Yeong Hui Yin, Fiona Keng Seh Lin and Japson Wong for helping in seaweed collection. We thank Gaurav Vaidya for his support on using TaxonDNA. We would also like to thank Song Sze Looi, Ng Poh Kheng and Poong Sze Wan for their advice and guidance on molecular taxonomy. We are grateful to the

University of Malaya and the Institute of Earth and Ocean Sciences (IOES) for research facilities.

Author Contributions

Conceived and designed the experiments: PL JT. Performed the experiments: PL JT. Analyzed the data: PL JT. Contributed reagents/materials/analysis tools: PL JT SP DH HS AH. Wrote the paper: PL JT SP DH AH.

References

- Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270: 313–321.
- Hebert PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* 270: S96–S99.
- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol* 2: e312.
- Ellegren H, Hubert N, Hanner R, Holm E, Mandrak NE, et al. (2008) Identifying Canadian Freshwater Fishes through DNA Barcodes. *PLoS ONE* 3: e2490.
- Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. *PLoS ONE* 6: e19254.
- Jimbo U, Kato T, Ito M (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science* 14: 107–124.
- Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364.
- Sarkar IN, Trizna M (2011) The Barcode of Life Data Portal: Bridging the Biodiversity Informatics Divide for DNA Barcoding. *PLoS ONE* 6: e14689.
- Wong LL, Peatman E, Lu J, Kucuktas H, He S, et al. (2011) DNA Barcoding of Catfish: Species Authentication and Phylogenetic Assessment. *PLoS ONE* 6: e17812.
- Freshwater DW, Tudor K, O'Shaughnessy K, Wysor B (2010) DNA barcoding in the red algal order Gelidiales: comparison of COI with *rbcL* and verification of the “barcoding gap”. *Cryptogamie, Algologie* 31: 435–449.
- Clarkston BE, Saunders GW (2010) A comparison of two DNA barcode markers for species discrimination in the red algal family Kallymeniaceae (Gigartinales, Florideophyceae), with a description of *Euthora timburtonii* sp. nov. *Botany* 88: 119–131.
- Le Gall L, Saunders GW (2010) DNA Barcoding Is a Powerful Tool to Uncover Algal Diversity: A Case Study of the Phyllophoraceae (Gigartinales, Rhodophyta) in the Canadian Flora. *Journal of Phycology* 46: 374–389.
- Saunders GW (2008) A DNA barcode examination of the red algae family Dumontiaceae (Gigartinales, Florideophyceae) in Canadian waters reveals substantial cryptic species diversity. 1. The foliose *Dilsea*-*Neodilsea* complex and *Weeksia*. *Botany* 86: 773–789.
- Kim MS, Yang MY, Cho GY (2010) Applying DNA barcoding to Korean Gracilariaceae (Rhodophyta). *Cryptogamie, Algologie* 31: 387–401.
- Saunders GW (2009) Routine DNA barcoding of Canadian Gracilariaceae (Rhodophyta) reveals the invasive species *Gracilaria vermiculophylla* in British Columbia. *Molecular Ecology Resources* 9: 140–150.
- McDevit DC, Saunders GW (2010) A DNA barcode examination of the Laminariaceae (Phaeophyceae) in Canada reveals novel biogeographical and evolutionary insights. *Phycologia* 49: 235–248.
- Kucera H, Saunders GW (2008) Assigning morphological variants of *Fucus* (Fucales, Phaeophyceae) in Canadian waters to recognized species using DNA barcoding. *Botany* 86: 1065–1079.
- Robba L, Russell SJ, Barker GL, Brodie J (2006) Assessing the use of the mitochondrial *cox1* marker for use in DNA barcoding of red algae (Rhodophyta). *American Journal of Botany* 93: 1101–1108.
- Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360: 1879–1888.
- Sherwood AR, Kurihara A, Conklin KY, Sauvage T, Presting GG (2010) The Hawaiian Rhodophyta Biodiversity Survey (2006–2010): a summary of principal findings. *BMC Plant Biology* 10: 258.
- Bixler HJ, Porse H (2010) A decade of change in the seaweed hydrocolloids industry. *Journal of Applied Phycology* 23: 321–335.
- Bindu MS, Levine IA (2010) The commercial red seaweed *Kappaphycus alvarezii*—an overview on farming and environment. *Journal of Applied Phycology*.
- Conklin KY, Kurihara A, Sherwood AR (2009) A molecular method for identification of the morphologically plastic invasive algal genera *Euचेuma* and *Kappaphycus* (Rhodophyta, Gigartinales) in Hawaii. *Journal of Applied Phycology* 21: 691–699.
- Doty MS (1985a) *Euचेuma alvarezii*, sp. Nov. (Gigartinales, Rhodophyta) from Malaysia. In: Abbott IA, Norris JN, editors. *Taxonomy of economic seaweeds: with reference to some Pacific and Caribbean species*. La Jolla, California: California Sea Grant College Program. 37–45.
- Doty MS, Norris JN (1985b) *Euचेuma* species (Solieriaceae, Rhodophyta) that are major sources of carrageenan. In: Abbott IA, Norris JN, editors. *Taxonomy of economic seaweeds: with reference to some Pacific and Caribbean species*. La Jolla, California: California Sea Grant College Program. 47–61.
- Paula EJ, Pereira R, Ohno M (1999) Strain selection in *Kappaphycus alvarezii* var. *alvarezii* (Solieriaceae, Rhodophyta) using tetraspore progeny. *Journal of Applied Phycology* 11: 111–121.
- Phang SM, Yeong HY, Lim PE, Adibi Rahiman Md Nor, Gan KT (2010) Commercial varieties of *Kappaphycus* and *Euचेuma* in Malaysia. *Malaysian Journal of Science* 29: 214–223.
- Zuccarello GC, Critchley AT, Smith J, Sieber V, Lhonneur GB, et al. (2006) Systematics and genetic variation in commercial *Kappaphycus* and *Euचेuma* (Solieriaceae, Rhodophyta). *Journal of Applied Phycology* 18: 643–651.
- Tan J, Lim PE, Phang SM (2012) Phylogenetic relationship of *Kappaphycus* Doty and *Euचेuma* J. Agardh (Solieriaceae, Rhodophyta) in Malaysia. *Journal of Applied Phycology* DOI: 10.1007/s10811-012-9833-1.
- Fredericq S, Freshwater DW, Hommersand MH (1999) Observations on the phylogenetic systematics and biogeography of the Solieriaceae (Gigartinales, Rhodophyta) inferred from *rbcL* sequences and morphological evidence. *Hydrobiologia* 398/399: 25–38.
- Zhao S, He P (2011) Molecular identification based on ITS sequences for *Kappaphycus* and *Euचेuma* cultivated in China. *Chinese Journal of Oceanology and Limnology* 29: 1287–1296.
- Halling C, Wikström SA, Lilliesköld-Sjöög G, Mörk E, Lundsör E, et al. (2012) Introduction of Asian strains and low genetic variation in farmed seaweeds: indications for new management practices. *Journal of Applied Phycology* In press.
- Meier R, Shiyang K, Vaidya G, Ng P (2006) DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic Biology* 55: 715–728.
- Leblanc C, Boyen C, Richard O, Bonnard G, Grienberger JM, et al. (1995) Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization. *Journal of Molecular Biology* 250: 484–495.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) ClustalW and ClustalX version 2.0. *Bioinformatics* 23: 2947–2948.
- Meyer CP (2004) Toward comprehensiveness: Increased molecular sampling within Cypraeidae and its phylogenetic implications. *Malacologia* 46: 127–156.
- Swofford DL (2003) PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Rambaut A, Drummond A. FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom 2010. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Zuccarello GC, West JA (2006) Molecular phylogeny of the subfamily *Bostrychioideae* (Ceramiaceae, Rhodophyta): subsuming *Stictosiphonia* and highlighting polyphyly in species of *Bostrychia*. *Phycologia* 45: 24–36.
- Silvestro D, Michalak I (2011) raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*. 10.1007/s13127-011-0056-0.
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Tanabe AS (2011) Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Molecular Ecology Resources* 11: 914–921.
- Meyer CP, Paulay G (2005) DNA Barcoding: Error rates based on comprehensive sampling. *PLoS Biol* 3: e422.
- Meier R, Zhang G, Ali F (2008) The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the “Barcoding Gap” and Leads to Misidentification. *Systematic Biology* 57: 809–813.
- Geraldino PJJ, Yang EC, Boo SM (2006) Morphology and Molecular Phylogeny of *Hypnea flexicaulis* (Gigartinales, Rhodophyta) from Korea. *Algae* 21: 417–423.
- Yang EC, Kim MS, Geraldino PJJ, Sahoo D, Shin J-A, et al. (2007) Mitochondrial *cox1* and plastid *rbcL* genes of *Gracilaria vermiculophylla* (Gracilariaceae, Rhodophyta). *Journal of Applied Phycology* 20: 161–168.
- Geraldino PJJ, Yang EC, Kim MS, Boo SM (2009) Systematics of *Hypnea asiatica* sp. nov. (Hypneaaceae, Rhodophyta) based on morphology and nrDNA SSU, plastid *rbcL*, and mitochondrial *cox1*. *Taxon* 58: 606–616.

49. Wiryadamrikul J, Park JK, Lewmanomont K, Boo SM (2010) Additional records of *Gelidiella fani* (Gelidiales, Rhodophyta) from the western Pacific based on morphology, *rbcL* and *cox1* analyses. *Botanica marina* 53: 343–350.
50. Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* 11: 206.
51. van Velzen R, Weitschek E, Felici G, Bakker FT (2012) DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods. *PLoS ONE* 7: e30490.
52. Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology* 15: 1759–1768.
53. DeSalle R, Lucas C, Thangaradjou T, Papenbrock J (2012) Development of a DNA Barcoding System for Seagrasses: Successful but Not Simple. *PLoS ONE* 7: e29987.
54. Kress WJ (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences* 102: 8369–8374.
55. Zuccarello GC, Burger G, West JA, King RJ (1999a) A mitochondrial marker for red algal intraspecific relationships. *Molecular Ecology* 8: 1443–1447.
56. Pickering TD, Skelton P, Sulu RJ (2007) Intentional introductions of commercially harvested alien seaweeds. *Botanica Marina* 50: 338–350.
57. Ganzon-Forbes ET, Trono GC, Villanueva RD, Romero JB, Montaño MNE (2011) 'Endong', a rare variety of the farmed carrageenophyte *Eucheuma denticulatum* (Burman) Collins & Hervey from the Philippines. *Journal of Applied Phycology* DOI: 10.1007/s10811-011-9740-x.
58. Freshwater DW, Rueness J (1994) Phylogenetic relationships of some European *Gelidium* (Gelidiales, Rhodophyta) species, based on *rbcL* nucleotide sequence analysis. *Phycologia* 33: 187–194.
59. Gavio B, Fredericq S (2002) *Grateloupia turuturu* (Halymeniaceae, Rhodophyta) is the correct name of the non-native species in the Atlantic known as *Grateloupia doryphora*. *European Journal of Phycology* 37: 349–359.