Plant Syst Evol (2014) 300:1503–1514
DOI 10.1007/s00606-013-0977-5

ORIGINAL ARTICLE

Utility of low-copy nuclear markers in phylogenetic reconstruction of *Hypericum* L. (Hypericaceae)

Andrea Sánchez Meseguer · Isabel Sanmartín · Thomas Marcussen · Bernard E. Pfeil

Received: 8 August 2013 / Accepted: 20 December 2013 / Published online: 19 January 2014
© Springer-Verlag Wien 2014

Abstract Primers and sequence variation for two low-copy nuclear genes (LCG) not previously used for phylogenetic inference in the genus *Hypericum*, PHYC and EMB2765, are presented here in comparison with the fast-evolving nuclear intergenic spacer ITS. Substitution rates in the LCG markers were half those reported in ITS for *Hypericum*, which might help avoid the problems caused by substitution saturation and difficulties to establish homologies that afflict the latter marker. We included representatives of all major clades within *Hypericum* and found that levels of phylogenetic resolution, clade support values and internal character consistency were similar to, or even higher than, those of ITS-based phylogenies. The presence of at least two copies in EMB2765 in *Hypericum* imposed a methodological challenge that was circumvented by the design of an effective clade-specific primer. Both EMB2765 and, especially, PHYC appear to be good alternatives to the ITS marker, confirming the main phylogenetic relationships found in previous studies, but with improved resolution and support values for some basal relationships.

Keywords *Hypericum* · Low-copy nuclear genes · Network · Paralogues · Phylogeny

A. S. Meseguer (✉) · I. Sanmartín
Department of Conservation and Biodiversity, Real Jardín Botanico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain
e-mail: asanchezmeseguer@gmail.com

T. Marcussen · B. E. Pfeil
Department of Biological and Environmental Sciences, University of Gothenburg, Box 100, 405 30 Göteborg, Sweden

Introduction

In biosystematics, species phylogenies are generally estimated from gene phylogenies. As the gene phylogenies are contained within the species phylogeny, they represent different levels of organisation and may differ in both topology and relative branch lengths. Processes acting at the gene level, such as gene duplication or incomplete lineage sorting (ILS), or at the species level, such as introgression or hybrid speciation, can produce gene-to-gene inconsistences or gene-tree/species tree conflicts that may hinder the reconstruction of the species phylogeny (Doyle 1992). This recognition has led to advice in favour of using multiple unlinked data sets in phylogenetic analyses (Small et al. 2004), and to the development of new methods to deal with gene-to-gene inconsistencies, discriminate between different types of gene incongruence, and obtain a more accurate estimation of the species phylogeny, in what is nowadays an active field of research (Meng and Kubatko 2009; Heled and Drummond 2010; Maureira-Butler et al. 2008; Bloomquist and Suchard 2010; Blanco-Pastor et al. 2012). All these methods have in common that they require the use of several independent loci or markers for disentangling the role of alternative biological processes and recovering the species phylogeny.

The nuclear ribosomal intertranscribed spacer ITS is by far the most widely used marker in plant (as well as fungal) systematics because of the facility with which it can be amplified using near-universal primers (White et al. 1990; Baldwin et al. 1995). However, its particular structure, with large tandem arrays of hundreds or thousands of more or less similar copies of which the amplified sequence is a weighted average, can lead to problems of incomplete concerted evolution, associated paralogy issues, and the presence of non-functional pseudogenes (Wendel et al.

Table 1 Species included in this study and GenBank accession numbers

Species	ID	Morphological section	References	GenBank accessions		
				PHYC	EMB	ITS
Hypericeae						
H. aegypticum L.	C136	*Adenotrias*	Meseguer et al. (2013)	–	KJ123884	KC709380*
H. aethiopicum Thunb.	C110	*Adenosepalum*	Meseguer et al. (2013)	KJ123856/ KJ123855	KJ123878/ KJ123877	KC709367*
H. balearicum L.	C61	*Psorophytum*	Meseguer et al. (2013)	KJ123864	KJ123876	KC709338*
H. balfouri Robson	C171	*Campylosporus*	Meseguer et al. (2013)	KJ123871/ KJ123870	–	KC709397*
H. canariense L.	C151	*Webbia*	Meseguer et al. (2013)	KJ123866	KJ123892	KC709389*
H. cerastoides (Spach) Robson	C72	*Campylopus*	Meseguer et al. (2013)	KJ123862/ KJ123861	–	KC709341*
H. cistifolium Lam.	C177	*Myriandra*	Meseguer et al. (2013)	KJ123865	KJ123886	KC709402*
H. coris L.	C23	*Coridium*	Meseguer et al. (2013)	KJ123867	–	KC709429*
H. elodes L.	C166	*Elodes*	Meseguer et al. (2013)	–	KJ123885	KC709393*
H. empetrifolium Willd.	C200	*Coridium*	Meseguer et al. (2013)	–	–	KC709416*
H. empetrifolium Willd.	–	*Coridium*	Davis and Chase (2004)	AY425113*	–	–
H. formosum Kunth	C175	*Hypericum*	Meseguer et al. (2013)	KJ123859/ KJ123860	–	KC709400*
H. grandiflorum Choisy	C146	*Adenosepalum*	Meseguer et al. (2013)	KJ123869/ KJ123868	KJ123881	KC709385*
H. hypericoides (L.) Crantz	C185	*Myriandra*	Meseguer et al. (2013)	–	–	KC709407*
H. hypericoides (L.) Crantz	–	*Myriandra*	Wurdack and Davis (2009)	–	FJ669779*	–
H. nummularioides Trautv.	C243	*Taeniocarpum*	Sanchez 164 (MA)	KJ123854	KJ123880	–
H. peplidifolium Rich	C28	*Humisfusoid.*	Aldasoro 10431 (MA)	KJ123857	KJ123883	KJ123872
H. perforatum L.	C56	*Hypericum*	Meseguer et al. (2013)	KJ123858	–	KC709333*
H. reflexum L.	C143	*Adenosepalum*	Meseguer et al. (2013)	KJ123863	KJ123879	KC709382*
H. revolutum Vahl (Schweinf)	C42	*Campylosporus*	Castroviejo 17247 (MA)	KJ123853	KJ123882	KJ123873
H. tortuosum Balf.	C170	*Triadenioides*	Aldasoro 14645 (MA)	–	KJ123891/ KJ123890	–
Triadenum walterii (Gmel.) Gleason	–	–	Wurdack and Davis (2009)	FJ669909*	FJ669780*	–
Triadenum petiolatum (Pursh) Britton	C16	–	Meseguer et al. (2013)	–	–	KC709312*
Vismieae						
Vismia glaziovii Ruhland	C192	–	Meseguer et al. (2013)	–	KJ123852	KC709411*
Vismia sp.	C190	–	Meseguer et al. (2013)	–	–	KC709410*
Vismia sp.	–	–	Wurdack and Davis (2009)	FJ669910*	FJ669781*	–
Cratoxyleae						
Cratoxylum formosum Benth. & Hook. f. ex Dyer	–	–	Wurdack and Davis (2009)	FJ669907*	FJ669777*	–
Cratoxylum formosum Benth. & Hook. f. ex Dyer	–	–	Nürk et al. (2013)	–	–	HE653674*
Eliea articulata Cambess.	C189	–	Wurdack and Davis (2009), Meseguer et al. (2013)	FJ669908*	FJ669778*	KC709409*
Outgroups						
Byrsonima crassifolia (L.) Kunth	–	–	Davis et al. (2002)	AF500526*	–	–
Byrsonima sp.	C152	–	Aldasoro 9931 (MA)	–	KJ123889/ KJ123888	KJ123874
Clusia gundlachii Stahl	–	–	Davis and Chase (2004)	AY425095*	–	–
Croton polyandrus Spreng.	–	–	van Ee et al. (2011)	–	HM564312*	–
Euphorbia polychroma L.	–	–	Wurdack and Davis (2009)	–	FJ669757*	–
Garcinia sp.	C153	–	Aldasoro 9930 (MA)	–	KJ123887	–
Garcinia latissima Miq.	–	–	Wurdack and Davis (2009)	–	FJ669743*	–

Table 1 continued

Species	ID	Morphological section	References	GenBank accessions		
				PHYC	EMB	ITS
Pentaphalangium	–	–	Wurdack and Davis (2009)	FJ669891*	–	–
Phyllanthus liukiuensis Hayata	–	–	Kawakita and Kato (2009)	FJ235364*	–	–
Podostemum ceratophyllum Michx.	–	–	Tippery et al. (2011)	–	–	HM470367*
Podostemum ceratophyllum Michx.	–	–	Davis and Chase (2004)	AY425129*	–	–
Rheedia macrophylla (Mart.) Planch. & Triana	–	–	Davis and Chase (2004)	AY425095*	–	–
Viola pubescens Ait.	–	–	Wurdack and Davis (2009)	–	FJ669844*	–
Clusia sp.	–	–	MG-2010	–	–	HM045517*

Morphological classification of *Hypericum* is based on Robson (1977–2010). Herbaria acronyms follow the abbreviations published in the Index Herbariorum

* Sequences obtained from GenBank. GenBank accession numbers for paralogues copies are presented: "paralogue1/paralogue2"

Table 2 Low-copy nuclear regions and sequences of primers screened in this study

Region	Primer name	Sequence 5′-3′	References
Phytochrome C	PHYC_Hyp_1F	CCAGCCACCGACATACCTCAAG	Own
	PHYC_Hyp_1R	GTAAGCTCCGCCACTTGAC	Own
	PHYC-INT1F	CCAGCTACTGATATACCWCARGCTTC	Wurdack and Davis (2009)
	PHYC-INTR	CCAGCTTCCATAAGGCTATCAGTRCT	Wurdack and Davis (2009)
Chalcone synthetase	CHS_Hyp_1F	GGAAGAAGTCAGGAAGGCGCAG	Own
	CHS_Hyp_1R	GGTCTCAACGGTAAGCCCAG	Own
	CHS_Hyp_2F	ACCGTGATGGCCATCGGAAC	Own
	CHS_Hyp_2R	CCAAAAAGCACTCCCCACTCGA	Own
Chloroplast-expressed glutamine synthetase	GScp687f	GATGCTCACTACAAGGCTTG	Emshwiller and Doyle (1999)
	GScp994r	AATGTGCTCTTTGTGGCGAAG	Emshwiller and Doyle (1999)
	GScp853f	TTACYGAACAAGCTGGYGTTGT	Emshwiller and Doyle (1999)
	GScp856r	AGSACAACRCCAGCTTGTTC	Emshwiller and Doyle (1999)
GBSSI	Wax1f	CTGGTGGACTTGGTGATG	Own
Granule-bound starch synthase	Wax1r	GGCYCCCATDTGRAATCCTGTG	Own
	Wax2f	CCTGKCTGCTCTKGARGCAC	Own
	Wax2f	CCTTGGCAAGWGGAGCRATCTCS	Own
Beta-carotene hydroxylase mRNA	Chyb_1F	TTG GCARATGGAGGGTGGAGA	Own
	Chyb_1R	GGCSTAYATGTTTGTMCAYGAYGG	Own
Embryo-defective 2765	EMB2765ex9F2	TATCCAAATGAGCAGATTATGTGGGA	Wurdack and Davis (2009)
	EMB2765ex9R	TTGGTCCAYTGTGCWGCAGAAGGRT	Wurdack and Davis (2009)
	EMB_Hyp_3F	TGATTCCAAAAT TGCCTTGAAG	Own
	EMB_Hyp_4F	TGTCCAAGGCRATAGTTACAGTTCTC	Own
	EMB_Hyp_3R	CCAGGAAGCTGTCCCACA	Own
TAFII15, Salt tolerance during germination 1	STG_Hyp_1F	CATCCCTGTTGATGGGCTRT	Duarte et al. (2010)
	STG_Hyp_1R	GAAATTTGTTGCAGADGTTGC	Duarte et al. (2010)
	STG_Hyp_2F	CTTGGACAGATCATCCATNGTCA	Duarte et al. (2010)
Glucose-6-phosphate isomerase	GPI_Hyp_12_1F	CGTGGTGCCACTGTCTCT	Own
	GPI_Hyp_16_1R	AGTTGRTAAAAGCTRTGCTG	Own
	GPI_Hyp_12_2F	CAATATGGTTTTCCAGTTGTTGA	Own
	GPI_Hyp_16_2R	GTTCCAGGTTCACCGAARTC	Own
	GPI_Hyp_13_3F	AGGTGCTGCAAGCATTGAT	Own

(a)

Clade E
H. aethiopicum_C110_2
H. cerastoides_C72_1
H. aethiopicum_C110_1
H. reflexum_C143
H. perforatum_C56
H. peplidifolium_C28
H. nummularioides_C243
H_formosum_C175_2
H. formosum_C175_1
H. cerastoides_C72_2
H. empetrifolium_AY425113
H. coris_C23
H. balearicum_C61
H. revolutum_C42
H. balfouri_C171_2
H. balfouri_C171_1
Clade D
Clade B
Triadenum walteri_FJ669909
H. canariense_C151
H. grandifolium_C146_1
H. grandifolium_C146_2
Clade C
Eliea. articulata_FJ669908
Cratoxylum formosum_FJ669907
H. cistifolium_C177
Vismia sp_FJ669910
Clade B
0.01

(b)

Phyllanthus liukiuensisFJ235364
Byrsonima crassifoliaAF500526
Podostemum ceratophyllumAY425129
Clusia gundlachiiAY425095
Rheedia macrophyllaFJ669891
Pentaphalangium latissimumFJ669890
Cratoxylum formosumFJ669907
Vismia sp_FJ669910
Eliea articulataFJ669908
Triadenum walteriFJ669909
H. cistifolium_C177
H. canariense_C151
H. grandifolium_C146_2
H. grandifolium_C146_1
H. balearicum_C61
H. revolutum_C42
H. balfouri_C171_2
H. balfouri_C171_1
H. empetrifoliumAY425113
H. coris_C23
H. nummularioides_C243
H. formosum_C175_1
H. formosum_C175_2
H. reflexum_C143
H. aethiopicum_C110_2
H. aethiopicum_C110_1
H. peplidifolium_C28
H. perforatum_C56
H. cerastoides_C72_2
H. cerastoides_C72_1
0.98
1
1
0.99
0.85
1
0.95
0.99
1
0.92
1
0.56
1
0.89
1
1
0.98
1
1
0.93
1
1
0.0
Hypericaceae
Hypericum
Vismieae
Cratoxyleae
B *Triadenum* sect. *Myriandra*
New World group
C *Androsaemum* group
D *Ascyreia* and Afromontane group
E1 *Hirtella* group
E2 *Hypericum* group
Old World group

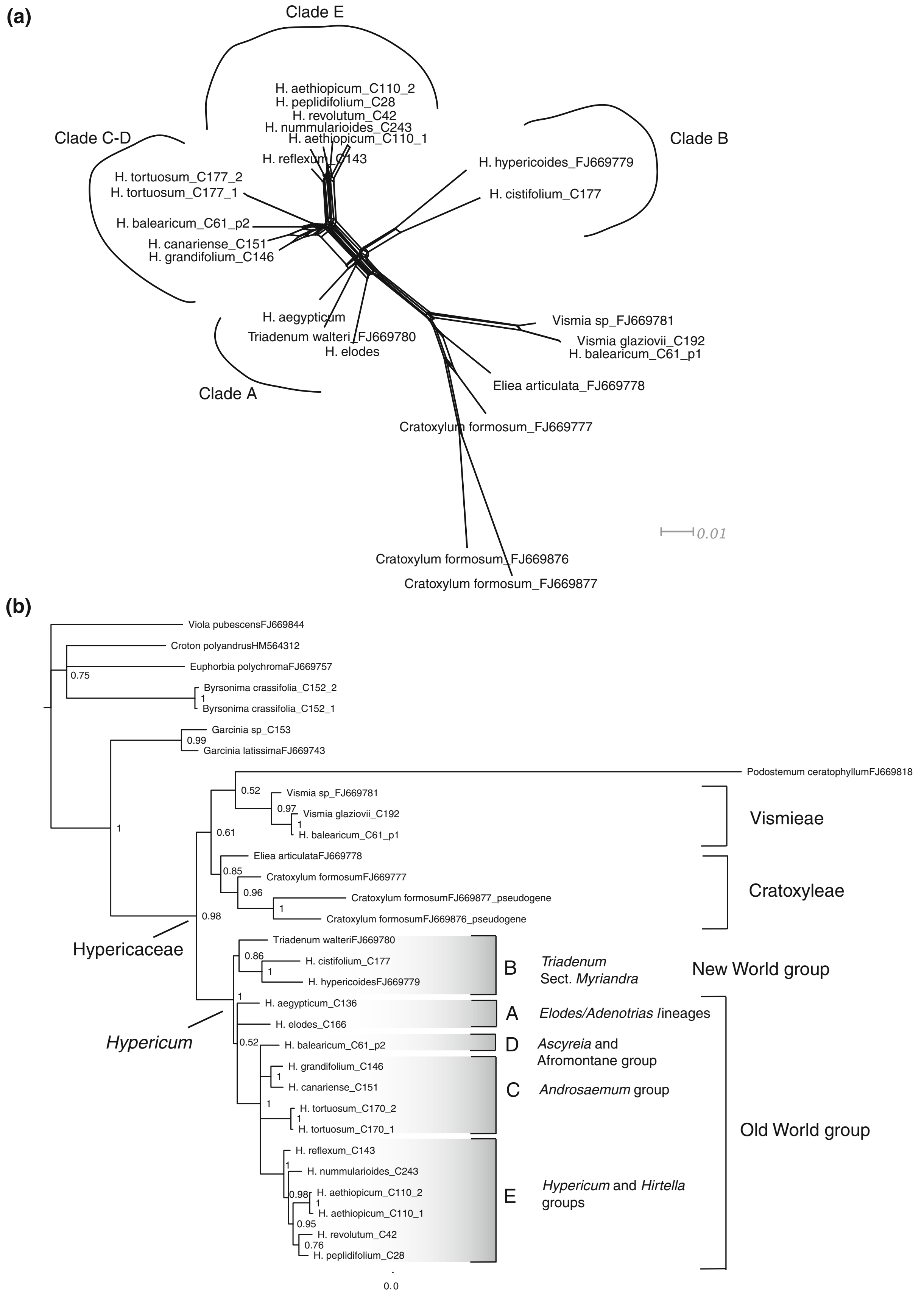
◀**Fig. 1** **a** PHYC phylogenetic network in the genus *Hypericum*. Weight threshold = 0.001. **b** Phylogenetic relationships in *Hypericum* and related taxa inferred from the nuclear PHYC marker. 50 % Bayesian Majority-Rule consensus tree showing posterior probabilities. *A to E* letters indicate major clades as defined in Meseguer et al. (2013). *Podostemum* (Podostemaceae), *Phyllanthus* (Phyllanthacea), *Byrsonima* (Malpighiaceae), *Rheedia, Pentaphalangium* and *Clusia* (Clusiaceae) were used as outgroups

1995; Álvarez and Wendel 2003; Nieto-Feliner and Rosselló 2007, but see Razafimandimbison et al. 2004). Lower thermodynamical stability and a higher rate of mutation than other ribosomal markers can also cause problems in phylogenetic inference (Mayol and Rosselló 2001) or estimation of lineage divergence times (Kay et al. 2006).

Low-copy or single-copy nuclear genes (LCGs) represent a vast but generally unexplored number of unlinked genetic markers rich in phylogenetic information (Sang 2002). LCGs have proven useful to resolve relationships at low taxonomic levels and, unlike the fast-evolving nuclear ITS, are expected to exist in one copy per chromosome set (Crawford and Mort 2004; Small et al. 2004), an assumption that is readily testable as more genomes are sequenced. Nevertheless, nuclear gene families may have complex evolutionary dynamics. For example, duplicate gene copies are frequently found in plant genomes as a consequence of both local (paralogues) and genome-wide (homoeologues) duplication processes (Clegg et al. 1997; Innes et al. 2008). Next to alleles, the occurrence of paralogues or homoeologues creates phylogenetic methodological difficulties, which has made more challenging the extensive use of LCGs in plant systematics (Sang 2002). In the ideal case, a set of primers would amplify a gene at a single locus, making downstream analysis fairly straightforward (allelic variation notwithstanding). However, primers for a gene that is single copy in one clade or pilot study (e.g. Denton et al. 1998) may instead amplify more than one copy in another clade (e.g. Oxelman et al. 2004; Pfeil et al. 2004; Ekenäs et al. 2012), thus making the transfer of existing primers to new groups challenging.

Hypericum L. is the largest genus within the family Hypericaceae. It comprises nearly 500 species of shrubs, small trees or perennial, sometimes annual, herbs, with yellow flowers and frequently glandulous petals, sepals, or leaves (Robson 2012). The present diversity of the genus has been classified in 36 morphological sections distributed worldwide and covering different environments—*Hypericum* is only absent in the poles, deserts and low-altitude tropical areas (Robson 1981). The most recent systematic revisions included the family Hypericaceae within the informal "clusioid clade" of order Malpighiales, which also includes tropical families such as Clusiaceae, Bonnetiaceae, and Podostemaceae (Davis et al. 2005; Ruhfel et al. 2011). Hypericaceae comprises three tribes: the mainly tropical tribes Vismieae (*Vismia*, *Harungana*, and *Psorospermum*) and Cratoxyleae (*Cratoxylum*, *Eliea*), and the cosmopolitan tribe Hypericeae which includes the genera *Triadenum, Thornea, Santomasia, Lianthus*, and *Hypericum* (Stevens 2007). Recent molecular results do not support current morphology-based classifications (Stevens 2007) but instead show that *Hypericum* is paraphyletic to *Triadenum, Thornea*, and *Santomasia* (Ruhfel et al. 2011; Nürk et al. 2013; Meseguer et al. 2013). These studies also reject the traditional infrageneric classification and recover many of the large taxonomic sections as non-monophyletic (e.g. *Ascyreia*, *Hirtella*, *Hypericum*, and *Brathys*). To date, all molecular studies have relied on either chloroplast markers or the nuclear ribosomal ITS marker. Meseguer et al. (2013) compared the phylogenetic signal of these two genomes and found overall congruence, supporting a geographical dichotomy between a New World lineage group and an Old World lineage. The New World lineage comprised the sections *Myriandra*, *Brathys*, and *Trigynobrathys* sister to genus *Triadenum*, which has been recently synonymised to *Hypericum* (Ruhfel et al. 2011). The Old World lineage comprised the remaining species and sections of *Hypericum*, e.g. *Ascyreia*, *Hypericum, Campylosporus*, and *Hirtella*. The Western Palearctic, species-poor sections *Elodes* and *Adenotrias* form the sister group to the New World–Old World clade, although this relation needs to be clarified as it received little support. Meseguer et al. (2013) also reported some cases of incongruence between nuclear and plastid markers, mainly affecting species or species groups, and a general lack of support for both basal and distal relationships in the ITS phylogeny.

Numerous studies suggest that LCGs have the potential to compensate for the lack of resolution and support values of phylogenies based on cpDNA and nrDNA, as well as the ability to recover reticulate phylogenetic relationships (Sang 2002; but see Rauscher et al. 2002). Previous efforts to use LCGs in *Hypericum* have been limited to two species of pharmacological importance, *H. perforatum* and *H. androsaemum*, to study protein expression in relation to hypericine biosynthesis and the genes encoding it (Liu et al. 2003; Bais et al. 2003; Karppinen and Hohtola 2008). Wurdack and Davis (2009) explored the use of LCGs in resolving phylogenetic relationships within order Malpighiales, and concluded that some of these genes, in particular the rapidly evolving exon 1 of PHYC and exon 9 of EMB2765 in the *Populus* genome (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html), could be useful in the systematics of this order. Here, we evaluate the utility of LCGs in resolving relationships in genus *Hypericum*, and between *Hypericum* and its closest relatives, and explore their potential to improve branch support values and resolution in comparison with the nuclear ribosomal ITS marker (Meseguer et al. 2013). We assess levels of



(a)
Clade E
H. aethiopicum_C110_2
H. peplidifolium_C28
H. revolutum_C42
H. nummularioides_C243
H. aethiopicum_C110_1
H. reflexum_C143
Clade C-D
H. tortuosum_C177_2
H. tortuosum_C177_1
H. balearicum_C61_p2
H. canariense_C151
H. grandifolium_C146
Clade B
H. hypericoides_FJ669779
H. cistifolium_C177
H. aegypticum
Triadenum walteri_FJ669780
H. elodes
Clade A
Vismia sp_FJ669781
Vismia glaziovii_C192
H. balearicum_C61_p1
Eliea articulata_FJ669778
Cratoxylum formosum_FJ669777
0.01
Cratoxylum formosum_FJ669876
Cratoxylum formosum_FJ669877
(b)
Viola pubescensFJ669844
Croton polyandrusHM564312
Euphorbia polychromaFJ669757
0.75
Byrsonima crassifolia_C152_2
1
Byrsonima crassifolia_C152_1
Garcinia sp_C153
0.99
Garcinia latissimaFJ669743
Podostemum ceratophyllumFJ669818
0.52
Vismia sp_FJ669781
0.97
Vismia glaziovii_C192
1
H. balearicum_C61_p1
Vismieae
0.61
1
Eliea articulataFJ669778
0.85
Cratoxylum formosumFJ669777
0.96
Cratoxylum formosumFJ669877_pseudogene
1
Cratoxylum formosumFJ669876_pseudogene
Cratoxyleae
0.98
Hypericaceae
Triadenum walteriFJ669780
0.86
H. cistifolium_C177
1
H. hypericoidesFJ669779
B
Triadenum
Sect. Myriandra
New World group
1
H. aegypticum_C136
H. elodes_C166
A
Elodes/Adenotrias lineages
Hypericum
0.52
H. balearicum_C61_p2
D
Ascyreia and
Afromontane group
H. grandifolium_C146
1
H. canariense_C151
C
Androsaemum group
1
H. tortuosum_C170_2
1
H. tortuosum_C170_1
Old World group
H. reflexum_C143
1
H. nummularioides_C243
0.98
H. aethiopicum_C110_2
1
H. aethiopicum_C110_1
E
Hypericum and Hirtella
groups
0.95
H. revolutum_C42
0.76
H. peplidifolium_C28
0.0

◀ **Fig. 2** **a** EMB2765 phylogenetic network in the genus *Hypericum*. Weight threshold = 0.001. **b** Phylogenetic relationships in *Hypericum* and related taxa inferred from the nuclear EMB2765 marker. 50 % Bayesian Majority-Rule consensus tree showing posterior probabilities. *A to E* letters indicate major clades as defined in Meseguer et al. (2013). *Podostemum* (Podostemaceae), *Garcinia* (Clusiaceae), *Byrsonima* (Malpighiaceae), *Viola* (Violaceae), *Croton* and *Euphorbia* (Euphorbiaceae) were used as outgroups

variation for the two low-copy nuclear regions PHYC and EMB2765, and present newly developed PHYC primers specific to *Hypericum* and clade-specific primers to isolate paralogous copies in EMB2765.

Materials and methods

Taxonomic and gene sampling

Species sampling included representatives of 13 out of 36 morphological sections of *Hypericum* and focused on representing all major clades within the group (*Elodes*/*Adenotrias*, *Myriandra*, *Brathys* s.l., *Triadenum*, *Androsaemum* s.l., Afromontane group, *Ascyreia* s.l., *Hirtella* s.l., and *Hypericum* s.l.), as found in a comprehensive phylogenetic analysis of the genus based on ITS and three different plastid markers (trnL-trnF, trnS-trnG, psbA-trnH; Meseguer et al. 2013). The sample also included representatives of closely related genera and families. DNA was extracted from fresh material collected in the field and preserved in silica gel, and from dry material preserved at several herbaria. GenBank accessions from previous studies were also included (Table 1).

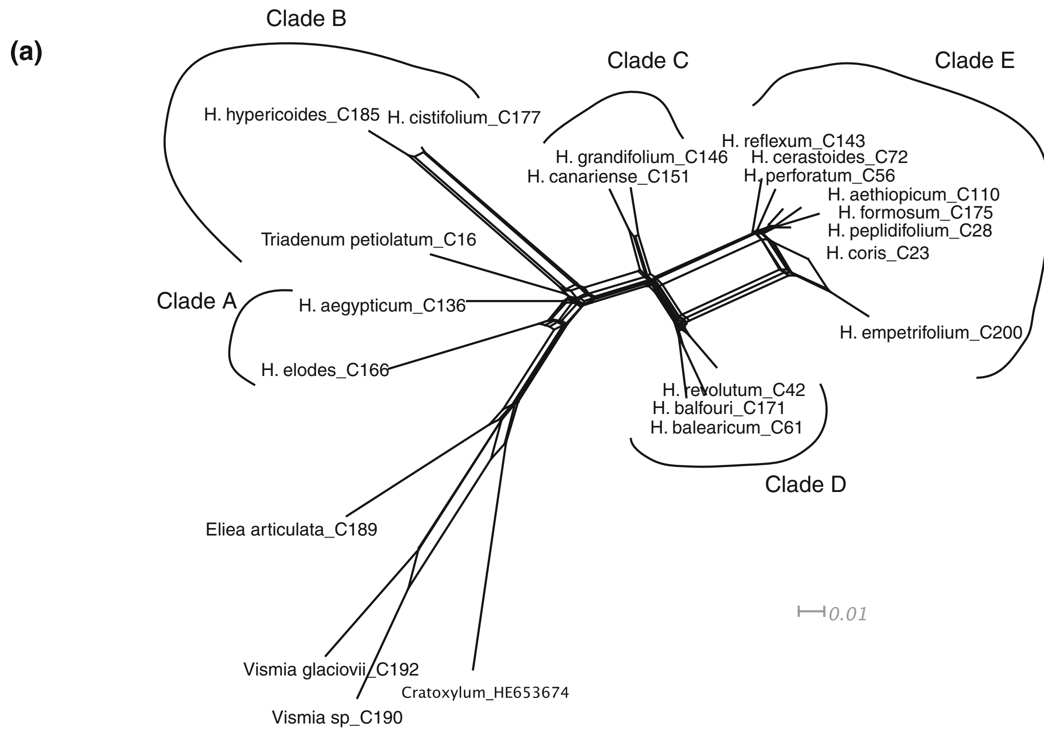
Amplification and sequencing

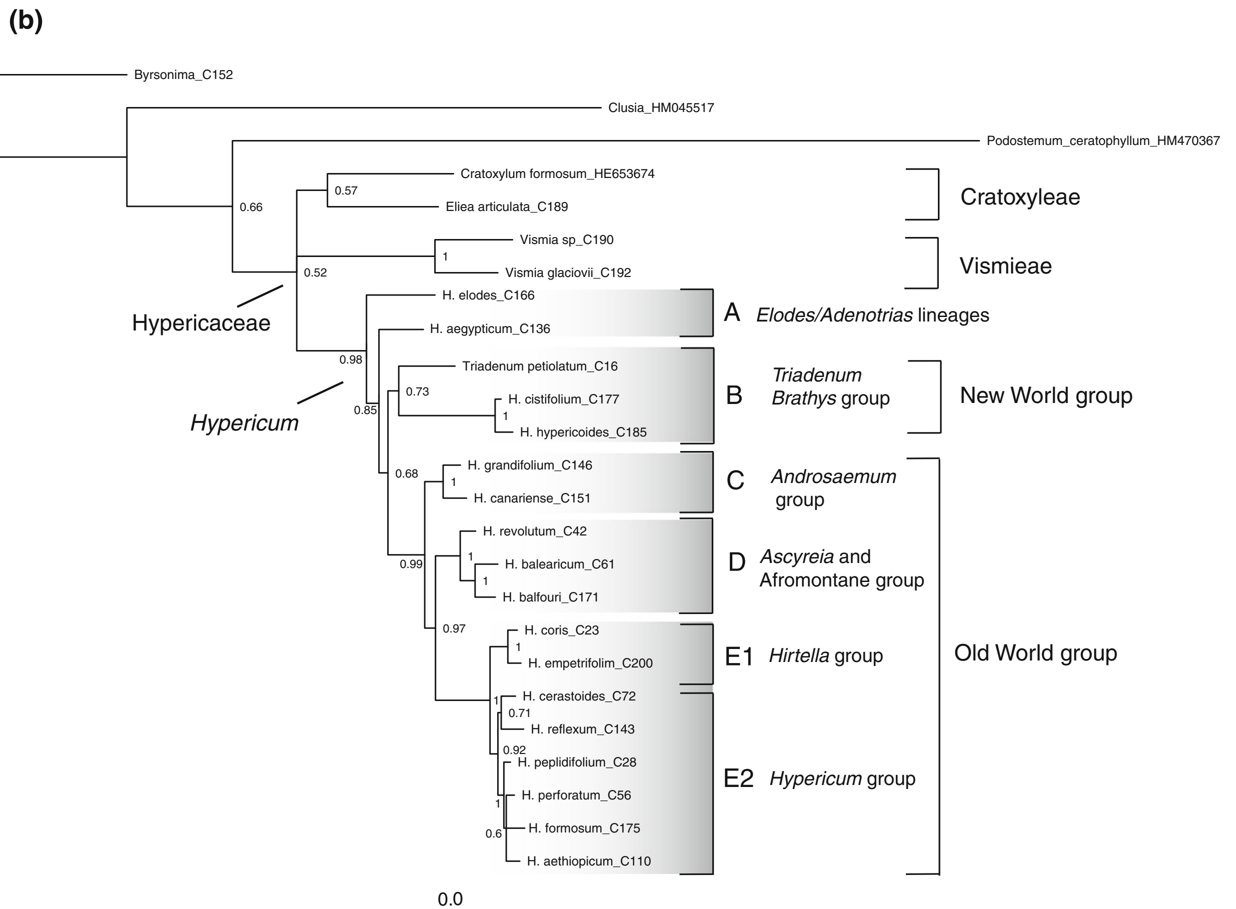
We initially screened eight low-copy nuclear regions, using primers published in the literature and others newly designed in this study for *Hypericum* (Table 2). PCR products range from 800 to 1,100 base pairs. The regions were: phytochromeC (PHYC), embryo-defective 2765 (EMB2765, "At2g38770"), chalcone synthase (CHS), waxy (GSSBI), chloroplast-expressed glutamine synthase (ncpGS, GS2, glnII or gln), glucose-6-phosphate isomerase (GPI, "PGIc"), salt tolerance during germination 1 (STG1, TAFII15, "At4 g31720"), and beta-carotene hydroxylase (Chyb). Internal primers designed by Wurdack and Davis (2009) in Malpighiales were initially used to amplify PHYC and EMB2765 markers, but later we designed a new set of PHYC primers specific to *Hypericum* to increase the length of the amplified region. For EMB2765, we also designed clade-specific primers to isolate paralogous copies. For comparative purposes, we also sequenced the ITS region using primers ITS1 and ITS4 (Aguilar et al. 1999; White et al. 1990; Table 1). DNA was extracted from leaf tissue samples using the QIAGEN DNeasy plant kit (Qiagen, Hilden, Germany) at the laboratories of the Real Jardín Botánico-CSIC (Madrid, Spain), and following the manufacturer's protocol. The PCR cycling conditions were as follows: 95 °C for 5 min, 35 cycles of [94 °C for 30 s, 52 °C for 30 s, 72 °C for 1.5 min] and a final extension step of 5 min at 72 °C. PCR products were checked on 1 % agarose gels and sequencing was performed at Macrogen, Inc. (Seoul, Korea) using the PCR primers. In all, we generated 44 sequences for 21 species. Several low-copy genes (GBSS, GS2, and STG) did not amplify or showed multiple unspecific bands, indicating low-primer specificity (Table 1). Others, such as GP1, were successfully amplified in a pilot study of a few individuals, but later sequencing indicated high intragenomic polymorphism that would require extensive subcloning to resolve. Only two regions, PHYC and EMB2765, were successfully amplified and sequenced in a majority of taxa. Hence, phylogenetic analysis and discussion of results were based on these regions.

Phylogenetic analysis

DNA sequences were edited using Sequencher 4.7 (Gene Codes, Ann Arbor, MI). The alignment was done with the online version of MAFFT (Stamatakis et al. 2008, using the L-INS algorithm) and manually adjusted in the editor Se-Al v. 2.0a11 (Rambaut 2002). Although gaps are a potential source of information in phylogenetic analysis, they often exhibit high levels of homoplasy and can be difficult to align, especially when taxon sampling is sparse; therefore, gaps were treated as missing data in the analysis. The alignment is available from the corresponding author. Resulting matrices were analysed under Bayesian inference using MrBayes 3.2cvs (Ronquist et al. 2012), with two parallel runs of four chains each for 2 million generations and sampling every 1,000 generations. Nucleotide substitution models were chosen based on the Akaike Information Criterion (Akaike 1973) as implemented in MrModeltest 2.3 (Nylander 2004). The GTR model was selected for EMB2765, and the HKY model for PHYC, with rate variation among sites in both. We used the program Tracer v.1.5 (Rambaut and Drummond 2003–2009) to verify that all the parameters had reached the stationary phase in log-likelihood values, and the split frequency criterion in MrBayes to assess convergence among chains. The initial 35,000 generations were discarded as burn-in samples and the remaining trees pooled to estimate the posterior probability distribution of the phylogeny and Bayesian clade posterior probabilities (pp).

Evolutionary data are most often presented as a phylogenetic tree with the underlying assumption that evolution


(a)
Clade B
Clade C
Clade E
Clade A
Clade D
H. hypericoides_C185
H. cistifolium_C177
H. grandifolium_C146
H. canariense_C151
H. reflexum_C143
H. cerastoides_C72
H. perforatum_C56
H. aethiopicum_C110
H. formosum_C175
H. peplidifolium_C28
H. coris_C23
H. empetrifolium_C200
Triadenum petiolatum_C16
H. aegypticum_C136
H. elodes_C166
H. revolutum_C42
H. balfouri_C171
H. balearicum_C61
Eliea articulata_C189
0.01
Vismia glaciovii_C192
Cratoxylum_HE653674
Vismia sp_C190


(b)
Byrsonima_C152
Clusia_HM045517
Podostemum_ceratophyllum_HM470367
Cratoxylum formosum_HE653674
Eliea articulata_C189
Vismia sp_C190
Vismia glaciovii_C192
H. elodes_C166
H. aegypticum_C136
Triadenum petiolatum_C16
H. cistifolium_C177
H. hypericoides_C185
H. grandifolium_C146
H. canariense_C151
H. revolutum_C42
H. balearicum_C61
H. balfouri_C171
H. coris_C23
H. empetrifolim_C200
H. cerastoides_C72
H. reflexum_C143
H. peplidifolium_C28
H. perforatum_C56
H. formosum_C175
H. aethiopicum_C110
Cratoxyleae
Vismieae
Hypericaceae
Hypericum
A Elodes/Adenotrias lineages
B Triadenum Brathys group
New World group
C Androsaemum group
D Ascyreia and Afromontane group
E1 Hirtella group
E2 Hypericum group
Old World group
0.0

◀ **Fig. 3** **a** ITS phylogenetic network in *Hypericum*. Weight threshold = 0.001. **b** Phylogenetic relationships in the genus *Hypericum* and related taxa inferred from the nuclear ITS marker. 50 % Bayesian Majority-Rule consensus tree showing posterior probabilities. *A to E letters* indicate major clades as defined in Meseguer et al. (2013). *Podostemum* (Podostemaceae), *Clusia* (Clusiaceae) and *Byrsonima* (Malpighiaceae) were used as outgroups

is a branching process. However, nuclear markers are subject to different types of recombination—in vivo by meiotic crossing over, i.e. interallele or interlocus, and in vitro during PCR. The probability of observing recombination increases with gene copy number and their similarity. Recombination results in conflicting phylogenetic signals within the sequence and cannot be expressed as a branching topology (Martin et al. 2011). We applied splits networks to visualise phylogenetically ambiguous signals in the alignments that might result from recombination, using the neighbour-net method implemented in SplitsTree v 4.0 (Huson and Bryant 2006) and difference distance-based algorithms to analyse each marker separately. We also used RDP v.3.34 (Martin et al. 2010) to test for possible recombination events in the LCG alignments. This software applies a number of recombination detection and analysis algorithms for detecting putative recombination breakpoints. We used all eight methods available, with a p value of 0.1 with Bonferroni correction to initially provide a low stringency examination of putative breakpoints. We did, however, require phylogenetic evidence for recombination, with internal reference sequences for RDP, with default options for the other methods.

Results and discussion

Phylogenetic utility of DNA sequence loci

Not every taxon analysed could be sequenced for all markers. Most technical difficulties were apparently related to low primer specificity owing to base mismatches in the primer site. The amplification sometimes yielded multiple bands and after sequencing we occasionally found polymorphic sites. We excluded specimens with multiple signals excepting those with single polymorphic sites; in such cases, we resolved the polymorphy by creating two sequences. Split networks showed a small number of contradictory characters (visualised as a box in the figure) in all three markers analysed (Figs. 1, 2, 3a), but RDP did not detect significant ($p \leq 0.05$) evidence of recombination in any of the markers.

Levels of sequence variation varied between loci (Table 3). ITS had the highest number of parsimony informative characters and number of variable characters. The two nuclear exons EMB2765 and PHYC exhibited similar levels of variation to one another, exceeding those reported in *Hypericum* for some commonly used fast-evolving chloroplast spacers (Meseguer et al. 2013). On the other hand, substitution rates for the LCGs were half those of ITS (Table 3). Meseguer et al. (2013) reported high rates of nucleotide substitution in the ITS ribosomal spacer, which made alignment, especially with outgroups, exceptionally difficult probably owing to saturation (Meseguer et al. 2013). Alignment was considerably more straightforward in EMB2765 and PHYC, which, together with good phylogenetic support and levels of resolution (see below), make these LCGs a good alternative to ITS in *Hypericum* phylogenetic inference. This agrees with Wurdack and Davis (2009) who found PHYC and EMB2765 to be useful markers for sequencing across a range of Malphigiales and more distant outgroups. Exploring a far deeper phylogeny than ours, the authors had excluded the third codon position in their analyses; analysing our data with or without this position resulted in the same phylogenetic topologies.

Character consistency within each marker was calculated using the consistency and retention indices (CI and RI, respectively; Table 3) based on the Bayesian topologies in PAUP*v4.0b10 (Swofford 2002). This showed that the RI within EMB2765 is comparable to that found in ITS, whereas the CI is more favourable in EMB2765 than in ITS. Within PHYC, the two indices have scores that are considerably better than those found in ITS (Table 3).

Phylogenetic relationships

The Bayesian 50 % majority-rule consensus trees for PHYC, EMB2765 and ITS are given in Figs. 1b, 2b and 3b, respectively. Comparison among the three phylogenies is not straightforward because of amplification failure in some of the taxa. However, the LCG phylogenies (Figs. 1b, 2b) are largely congruent with the ITS tree (Fig. 3b) and with those reported in previous molecular studies of *Hypericum* (Ruhfel et al. 2011; Nürk et al. 2013; Meseguer et al. 2013; Park and Kim 2004; Crockett et al. 2004). Splits networks showed similar groupings (Figs. 1a, 2, 3a). LCGs presented higher support values for some basal nodes than those of ITS (Fig. 3; Meseguer et al. 2013; Nürk et al. 2013; Park and Kim 2004; Crockett et al. 2004) or plastid phylogenies (Meseguer et al. 2013). Among the studied LCGs, resolution and clade support values were highest in the PHYC phylogeny (Fig. 1). The PHYC phylogeny (Fig. 1) shows a geographic dichotomy between the New World lineage (*H. cistifolium*) and an Old World lineage, the latter divided into several clades that generally correspond to those found by Meseguer et al. (2013) and Nürk et al. (2013): clade C (*Androsaemum* group), clade D (*Ascyreia* and Afromontane groups) and clade E. Within the latter, two subclades can be

Table 3 Character status summary for the nuclear low-copy markers PHYC and EMB2765, and for the nuclear intergenic spacer ITS

Region	# Of characters	# Parsimonious uninformative	# Parsimonious informative	% Informative characters	TL mean	CI	RI
EMB2765	819	105	89	10.86	1.822462	0.782	0.749
PHYC	822	103	94	11.43	2.047945	0.853	0.879
ITS	740	105	164	22.16	3.929994	0.729	0.759

TL mean mean of the total tree length estimated over the two independent Bayesian runs, *CI* consistency index, *RI* retention index

distinguished: the *Hirtella* group (clade E1) and the larger *Hypericum* group (clade E2, Meseguer et al. 2013), although the last showed little internal resolution. The EMB phylogeny showed generally lower support values for basal relationships but slightly better resolution at species level (clade E2) than PHYC. This marker also recovers *Triadenum walteri* (=*Hypericum walteri*) as part of the New World lineage, confirming other phylogenetic studies that found *Hypericum* to be non-monophyletic (Meseguer et al. 2013; Nürk et al. 2013; Ruhfel et al. 2011). In contrast, the PHYC phylogeny shows *Hypericum walteri* as the sister group of *Hypericum*, although this relationship is not well supported. Other incongruences between the two LCG markers affect phylogenetic relationships at the tribal level. In PHYC, the family Hypericaceae and the tribe Hypericeae are both recovered as well-supported monophyletic groups, whereas tribe Cratoxyleae (*Cratoxylum formosum* and *Eliea articulata*) appears as non-monophyletic (Fig. 1). In the EMB2765 tree, tribal relationships are better resolved, with tribe Hypericeae sister to Cratoxyleae-Vismieae. However, one species of *Hypericum*, *H. balearicum*, appears within tribe Vismieae, as sister group to *Vismia* (Fig. 2). The anomalous position of this sequence in the EMB2765 tree could be attributed to incomplete lineage sorting or to an ancient duplication event in which a different paralogue has been amplified in *H. balearicum*. The long branch separating the specimen *H. balearicum*_C61_p1 from the rest of *Hypericum*, and the short branch between this species and the outgroup *Vismia*, supports the later explanation as the most probable. To solve this issue, we designed a new internal primer for EMB based on the right *Hypericum* sequences to recover the orthologous copy of this marker in *H. balearicum*. The new sequence *H. balearicum*_C61_p2 was resolved in a congruent position with the other markers in the phylogeny (Fig. 2), lending support to the hypothesis that EMB2765 is present in more than one copy in *Hypericum*. Only diploid specimens have been described in *H. balearicum* (Robson, 1985), suggesting that intragenomic polymorphism probably owed to local duplications (paralogues). Further cloning strategies could help to do a better screening of gene copies. Nevertheless, our study suggests that the design of copy-specific primers can also be a useful strategy to address paralogy. Interestingly, Wurdack and Davis (2009) found non-functional pseudogene copies of EMB in *Cratoxylum* (Hypericaceae, Fig. 2), but these shorter copies grouped with the functional, full-length sequence (and were isolated using the same primers), suggesting a more recent duplication event.

Conclusions

Our study, including representatives from all major clades of *Hypericum*, shows the potential of two LCGs, EMB2765 and PHYC, for reconstructing phylogenetic relationships in genus *Hypericum* and related clades. A lower mutation rate in these markers in comparison with the ITS ribosomal spacer makes it easy to establish homologies in the alignment with outgroups. In addition, pilot phylogenetic studies showed improved resolution and clade support values for basal nodes compared with ITS and other fast-evolving plastid markers. Further, the internal character consistency of the new markers is comparable to, or better than, that found in ITS. In this study, we have also discovered a paralogous copy in EMB2765 for *Hypericum* that was isolated through the design of copy-specific primers.

Acknowledgments This work was funded by the Spanish Ministry of Education and Science (projects CGL2009-13322-C03-01/BOS and CGL2012-40129-C02-01) to I.S. and a PhD research grant AP-2007-01698 to A.S.M. B.E.P. is supported by grants from the Swedish Research Council, the Royal Swedish Academy of Sciences, Lars Hiertas Minne fund, The Royal Physiographic Society in Lund, Helge Ax:son Johnsons fund and the Lundgrenska fund. The authors are grateful to B. Oxelman, Y. Bertrand, A. Hilpold and F. De Sousa for help at different stages of this research.

References

Aguilar JF, Rosselló JA, Feliner GN (1999) Molecular evidence for the compilospecies model of reticulate evolution in *Armeria* (Plumbaginaceae). Syst Biol 48:735–754

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Kiado A (ed.) Second international symposium on information Theory. Budapest, pp 267–281

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenetics Evol 29:417–434

Bais HP, Vepachedu R, Lawrence CB, Stermitz FR, Vivanco JM (2003) Molecular and biochemical characterization of an enzyme responsible for the formation of hypericin in St. John's wort (*Hypericum perforatum* L.). J Biol Chem 278:32413–32422

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA—A valuable source of evidence on angiosperm phylogeny. Ann Missouri Bot Gard 82:247–277

Blanco-Pastor JL, Vargas P, Pfeil BE (2012) Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. Plos One 7:e39089

Bloomquist EW, Suchard MA (2010) Unifying vertical and nonvertical evolution: a stochastic arg-based framework. Syst Biol 59:27–41

Crawford DJ, Mort ME (2004) Single-locus molecular markers for inferring relationships at lower taxonomic levels: observations and comments. Taxon 53:631–635

Clegg MT, Cummings MP, Durbin ML (1997) The evolution of plant nuclear genes. Proc Natl Acad Sci USA 92:7791–7798

Crockett SL, Douglas AW, Scheffler BE, Khan IA (2004) Genetic profiling of *Hypericum* (St. John‘s wort) species by nuclear ribosomal ITS sequence analysis. Pl Med 70:1–7

Davis CC, Chase MW (2004) Elatinaceae are sister to Malpighiaceae; Peridiscaceae belong to Saxifragales. Am J Botany 91:262–273

Davis CC, Bell CD, Mathews S, Donoghue MJ (2002) Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. Proc Natl Acad Sci USA 99:6833–6837

Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ (2005) Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. American Naturalist 165:E36–E65

Denton AL, McConaughy BL, Hall BD (1998) Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. Mol Biol Evol 15:1082–1085

Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. Syst Bot 17:144–163

Duarte J, Wall PK, Edger P, Landherr L, Ma H, Pires JC, Leebens-Mack J (2010) Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 10:61

Ekenäs C, Heidari N, Andreasen K (2012) *Arnica* (Asteraceae) phylogeny revisited using RPB2: Complex patterns and multiple d-paralogues. Mol Phylogenet Evol 64:261–270

Emshwiller E, Doyle JJ (1999) Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). Mol Phylogenet Evol 12:310–319

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570–580

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NWG, Couloux A, Dalwani A, Denny R et al (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol 148:1740–1759

Karppinen K, Hohtola A (2008) Molecular cloning and tissue-specific expression of two cDNAs encoding polyketide synthases from *Hypericum perforatum*. J Plant Physiol 165:1079–1086

Kawakita A, Kato M (2009) Repeated independent evolution of obligate pollination mutualism in the Phyllantheae-Epicephala association. Proc Biol Sci 276:417–426

Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6:36. doi:10.1186/1471-2148-6-36

Liu B, Falkenstein-Paul H, Schmidt W, Beerhues L (2003) Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. Plant J 34:847–855

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462–2463

Martin DP, Lemey P, Posada D (2011) Analysing recombination in nucleotide sequences. Mol Ecol Resour 11:943–955

Mayol M, Rosselló JA (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Mol Phylogenet Evol 19:167–176

Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ (2008) The reticulate history of *Medicago* (Fabaceae). Syst Biol 57:466–482

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor Popul Biol 75:35–45

Meseguer AS, Aldasoro JJ, Sanmartín I (2013) Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St John's wort (*Hypericum*). Mol Phylogenet Evol 67:379–403

Nieto-Feliner G, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. Mol Phylogenet Evol 44:911–919

Nürk NM, Madriñán S, Carine MA, Chase MW, Blattner FR (2013) Molecular phylogenetics and morphological evolution of St. John's wort (*Hypericum*; Hypericaceae). Mol Phylogenet Evol 66:1–16

Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University

Oxelman B, Yoshikawa N, McConaughy BL, Luo J, Denton AL, Hall BD (2004) RPB2 gene phylogeny in flowering plants, with particular emphasis on asterids. Mol Phylogenet Evol 32:462–479

Park SJ, Kim KJ (2004) Molecular phylogeny of the genus *Hypericum* (Hypericaceae) from Korea and Japan: Evidence from nuclear rDNA ITS sequence data. J Plant Biol 47:366–374

Pfeil BE, Brubaker CL, Craven LA, Crisp MD (2004) Paralogy and orthology in the Malvaceae rpb2 gene family: investigation of gene duplication in *Hibiscus*. Mol Biol Evol 21:1428–1437

Rambaut A, Drummond AJ (2003–2009) Tracer, version 1.5, MCMC trace analysis package. http://tree.bio.ed.ac.uk/software/

Rambaut A (2002) Se-Al: Sequence Alignment Editor. Available from http://tree.bio.ed.ac.uk/software/seal/

Rauscher JT, Doyle JJ, Brown AH (2002) Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol Ecol 11:2691–2702

Razafimandimbison SG, Kellogg EA, Bremer B (2004) Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from *Naucleeae* (Rubiaceae). Syst Biol 53:177–192

Robson NKB (1981) Studies in the genus *Hypericum* L. (Guttiferae). 2. Characters of the genus. Bull Brit Mus (Nat Hist) Bot 8:55–226

Robson NKB (1985) Studies in the genus *Hypericum* L. (Guttiferae). 3. Sections 1. *Campylosporus* to 6a. *Umbraculoides*. Bull Brit Mus (Nat Hist) Bot 12:163–211

Robson NKB (2012) Studies in the genus *Hypericum* L. (Hypericaceae) 9. Addenda, corrigenda, keys, lists and general discussion. Phytotaxa 72:1–111

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Hpohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542

Ruhfel BR, Bittrich V, Bove CP, Gustafsson MHG, Philbrick CT, Rutishauser R, Xi Z, Davis CC (2011) Phylogeny of the clusioid clade (Malpighiales): Evidence from the plastid and mitochondrial genomes. Am J Bot 98:306–325

Sang T (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit Rev Biochem Mol Biol 37:121–147

Small RL, Cronn RC, Wendel JF (2004) L. A. S. Johnson review no. 2. Use of nuclear genes for phylogeny reconstruction in plants. Austral Syst Bot 17:145–170

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758–771

Stevens PF (2007) Hypericaceae. In: Kubitzki K (ed) The families and genera of vascular plants. Springer, Berlin, pp 194–201

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods)

Tippery NP, Philbrick CT, Bove CP, Les DH (2011) Systematics and phylogeny of neotropical Podostemaceae. Syst Bot 36:105–118

van Ee BW, Riina R, Berry PE (2011) A revised infrageneric classification and molecular phylogeny of New World Croton (Euphorbiaceae). Taxon 60:791–823

Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92:280–284

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications, 315–322

Wurdack KJ, Davis CC (2009) Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. Am J Bot 96:1551–1570