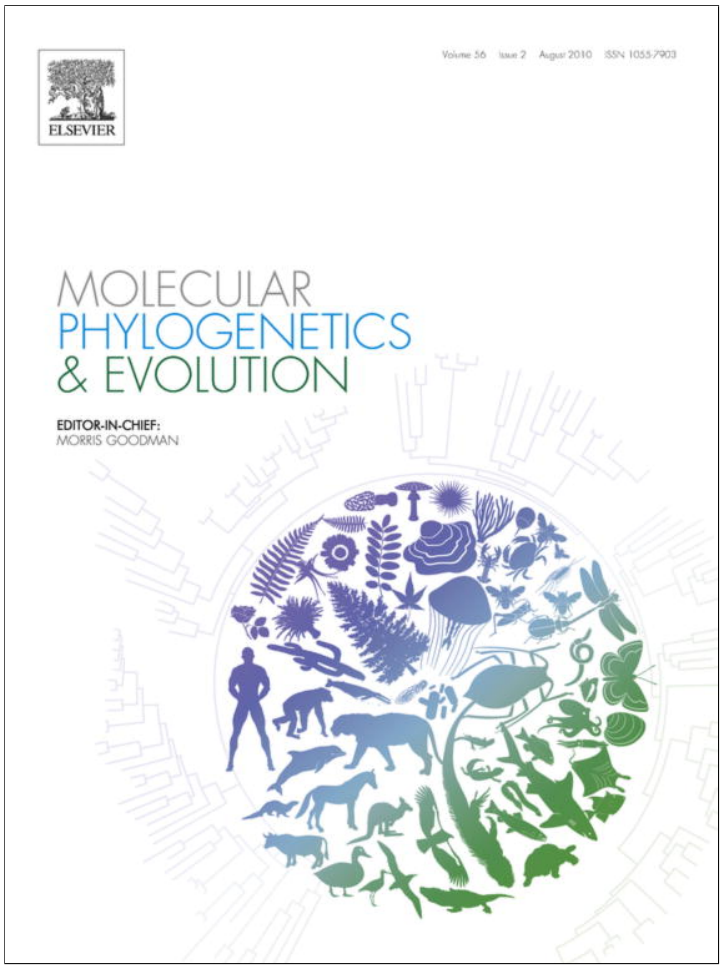

Molecular Phylogenetics and Evolution 56 (2010) 718–726

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Polymorphism and divergence within the ascidian genus *Ciona*

Marie L. Nydam *, Richard G. Harrison

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 5 December 2009
Revised 16 March 2010
Accepted 31 March 2010
Available online 18 April 2010

Keywords:
Polymorphism
Heterozygosity
Divergence
Gene genealogies
Ciona
Ascidian

ABSTRACT

The genus *Ciona*, a widely distributed group of solitary ascidians, has long been an important model in embryology and developmental biology. *Ciona* has also recently attracted the attention of evolutionary biologists because of the remarkably high levels of heterozygosity found within single individuals. Surprisingly, genealogical relationships in *Ciona* have received little attention. Here, we expand our knowledge of relationships among the members of the *Ciona* genus and estimate levels of polymorphism in natural populations.

Previous studies have documented the outgroup status of *Ciona savignyi* among the shallow-water *Ciona* and revealed the existence of two distinct forms (Types A and B) of the widespread *Ciona intestinalis*. Here, using gene genealogies of six nuclear gene loci, we show Types A and B to be well-supported monophyletic groups. In spite of their morphological similarity, Type A vs. Type B divergences range from 0.035 to 0.124. In contrast, the morphologically distinct *Ciona roulei* is embedded within Type B in all genealogies, and a new species, *Ciona* sp., appears to be associated with Type B/*C. roulei* to the exclusion of Type A. Levels of polymorphism in natural populations are similar to levels reported in other organisms that are considered to be highly polymorphic.

1. Introduction

The genus *Ciona*, a widely distributed group of solitary ascidians, has been of interest to biologists for over a century, largely because two of its members, *Ciona intestinalis* (Linne, 1767) and *Ciona savignyi* (Herdman, 1882) are model organisms in developmental biology and embryology. Recently published genome sequences of both species (Dehal et al., 2002; Vinson et al., 2005) have increased the utility of these species as models in developmental biology and have also attracted the attention of population geneticists because of the remarkably high levels of heterozygosity found within the single individuals for which whole genome sequences have been produced. Allelic polymorphism across the entire genome in *C. intestinalis* is 1.2% (including both Single Nucleotide Polymorphisms (SNPs) and insertion/deletions) (Dehal et al., 2002). This level of polymorphism is 15 times that of *Homo sapiens* (Dehal et al., 2002) and three times that of the Japanese puffer fish, *Fugu rubripes,* which has been described as having a high level of polymorphism (Aparicio et al., 2002; Dehal et al., 2002). The average genome-wide SNP heterozygosity in *C. savignyi*, again based on a single individual, is 4.5% (Small et al., 2007).

Surprisingly, given the prominence and utility of these ascidians as models, genealogical and phylogenetic relationships within and among *Ciona* species have received little attention. As we uncover the evolutionary history of these species, our understanding of the complexity of this group continues to increase.

For example, it is now evident that *C. intestinalis* comprises two distinct, highly divergent entities (Suzuki et al., 2005; Caputi et al., 2007; Nydam and Harrison, 2007). Type A *C. intestinalis* (the type for which the genome sequence is available), is thought to be native to the Northwest Pacific Ocean and is now found throughout the Pacific Ocean, the Mediterranean Sea, the Atlantic coast of South Africa, and the Black Sea (Van Name, 1945; Kott, 1952). Type B *C. intestinalis,* originally described by Linnaeus from the North Atlantic Ocean (Linne, 1767), is found in the Western Atlantic Ocean (Nydam and Harrison, 2007). The two types are ~12% divergent at the mitochondrial COI gene (cytochrome oxidase I, uncorrected *p*-distance (Nydam and Harrison, 2007)).

A third taxon, *Ciona roulei* (Lahille, 1887), is endemic to the Northwestern Mediterranean Sea, having been found at five locations along the French coast, near the Spanish border (Harant and Vernieres, 1933; Fiala-Medioni, 1974). This species is morphologically distinct from all other *Ciona* species (Harant and Vernieres, 1933). In a mtCOI tree, *C. roulei* is embedded within Type B, i.e. Type B is paraphyletic with respect to *C. roulei* (Nydam and Harrison, 2007).

In a previous study (Nydam and Harrison, 2007), three morphologically distinct individuals (*Ciona* sp.) were collected from a depth of <5 m in Banyuls Harbor, southern France. These individuals were determined to be genetically distinct from all other *Ciona*

* Corresponding author. Fax: +1 607 255 8088.
E-mail addresses: mln32@cornell.edu (M.L. Nydam), rgh4@cornell.edu (R.G. Harrison).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.03.042

species based on mtCOI sequence data (10.9% and 12.6% uncorrected *p*-distance from Type A and Type B, respectively) (Nydam and Harrison, 2007). However, the phylogenetic placement of *Ciona* sp. within the *Ciona* genus remained unresolved using this single marker.

A fourth species among the shallow-water *Ciona* species is *C. savignyi*, which was first described from the Western Pacific Ocean (Herdman, 1882) and subsequently invaded the Eastern Pacific Ocean (Hoshino and Nishikawa, 1985). Based on mtCOI sequencing, *C. savignyi* is a well-supported monophyletic group that occupies a basal position in the phylogeny as a sister group to the *C. intestinalis/C. roulei/C.* sp. clade (Nydam and Harrison, 2007).

While the relationships in the mtCOI tree are well-supported, inferences of phylogenetic relationships from a single marker should not be considered definitive. A genealogy from a single locus reflects the particular evolutionary history of that locus, which may not be the evolutionary history of the genome as a whole (Maddison, 1997; Nichols, 2001). Therefore, multiple markers need to be examined to construct a species phylogeny with more confidence.

In the present study, we examine patterns of genetic variation within and among the shallow-water members of the genus *Ciona*. Gene genealogies based on six nuclear loci allow us to determine whether patterns revealed by mtCOI are consistent across the nuclear genome; we focus on the distinctness of the two types of *C. intestinalis*, the paraphyly of Type B with respect to *C. roulei*, and the uncertain relationship between *Ciona* sp. and its congeners. We also estimate polymorphism in population samples of Type A and B *C. intestinalis* and in *C. roulei* for each of the six nuclear loci and compare these with estimates for *Ciona* from whole genome sequences and with levels of heterozygosity in natural populations of other taxa.

Table 1
Collection sites for *Ciona* species and number of individuals from each site included in genealogies.

Species	Region/state or country	Site	# of individuals
Ciona intestinalis			
Type A	Northwest Pacific/Japan	Morotsu	1
		Nishiura	1
		Onagawa	1
		Shikoku Island	1
		Yokohama	1
	Northeast Pacific/California	Newport Harbor	1
		Alamitos Bay	1
		Santa Barbara Harbor	1
		Half Moon Bay	1
		Sausalito	1
	English Channel/United Kingdom	Falmouth	1
		Plymouth	1
	Mediterranean/France	Banyuls-sur-Mer	2
Type B	Northwest Atlantic/NH to CT	Newcastle	1
		Gloucester	1
		Winthrop	1
		Mystic	1
	Northeast Atlantic/Sweden	Fiskebäckskil	3
	Northeast Atlantic/The Netherlands	Breskens	3
	English Channel/United Kingdom	Falmouth	1
		Plymouth	1
	English Channel/France	Granville	1
Ciona roulei	Mediterranean/France	Banyuls-sur-Mer	3
Ciona sp.	Mediterranean/France	Banyuls-sur-Mer	2

2. Materials and methods

2.1. DNA extraction, amplification and sequencing

Ovaries were dissected from freshly-collected individuals, cut into several pieces, immediately preserved in DMSO (dimethyl sulfoxide), and ultimately (within 12 d) stored at −80 °C until needed. We use ovaries as they are accessible by simple dissection and ovary tissue is easily lysed for DNA extraction. Samples from Sweden and The Netherlands were preserved and shipped in ethanol soon after collection. Upon arrival, the ovaries from these individuals were dissected and preserved in DMSO. Total DNA was extracted from the ovaries using the Qiagen DNeasy® Tissue Kit (Qiagen Corporation, Santa Clarita, CA).

To provide a mitochondrial COI tree that reflects a better sample of the geographic range of the *Ciona* species, we sequenced thirteen *C. intestinalis* (from Japan and the English Channel) and 7 *C. savignyi* individuals (from Japan) and added these sequences to the mitochondrial COI data matrix used previously (Nydam and Harrison, 2007). The nuclear gene genealogies each included a 22-individual subset of the individuals present in the mtCOI tree plus 10 additional *C. intestinalis* samples from the southern UK, northern France and Japan (Table 1). A *C. savignyi* individual from the Eastern Pacific Ocean served as the outgroup individual for each nuclear genealogy.

Approximately 1 kilobase from each of six nuclear genes was amplified using primers developed from the published Type A genome sequence. The genes are: vesicular acetylcholine transporter (*vAChTP*), Cellulose synthase (*CiCesA*), Fibroblast Growth Factor orthologous to vertebrate Fibroblast Growth Factor 4/5/6 (*Ci-Fgf4/5/6*), Forkhead (*Ci-fkh*) (5′ regulatory region), Jade (*jade*) and Patched (*Ci-Patched*). Loci were selected to represent a range of polymorphism, based on aligning whole genome shotgun reads from the Type A genome. This was done so that the six loci could be used for genealogical analyses, as well as for analyses of levels of polymorphism. Four of the loci were considered to have "low" levels of polymorphism: (*vAChTP*, *CiCesA*, *Ci-fkh*, and *Ci-Patched*), and two were considered to have "high" levels of polymorphism (*Ci-Fgf4/5/6* and *jade*). For *Ci-Fgf4/5/6* and *jade*, primers developed from Type A individuals did not consistently amplify Type B individuals; Type B individuals were amplified and sequenced using specific primers developed from preliminary sequences of several Type B individuals. *Ciona* sp. and *C. savignyi*-specific primers were also developed for several loci. Sequences from three of the loci (*vAChTP*, *CiCesA*, *jade*) include both coding and noncoding regions, whereas sequences from *Ci-Fgf4/5/6*, *Ci-fkh*, *Ci-Patched* are entirely noncoding (Supplementary Table 1). Primer sequences and thermocycling conditions for the six nuclear loci are available from the authors. Primers and PCR amplification conditions for the mtCOI sequences have been published previously (Nydam and Harrison, 2007).

PCR amplification of nuclear genes was performed in a 10-μl total reaction volume with 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 μl of 10× buffer (50 mM KCl, 20 mM Tris (pH 8.4)), 0.2 μM of each primer, 0.08 U of *Taq* Polymerase (Gibco-BRL) and 1 μl of template DNA. To obtain clean sequence data, cloning of PCR products was necessary for the *Ci-Fgf4/5/6* and *jade* loci and for certain individuals at the other nuclear loci. Cloning was performed using a pGEM®-T kit (Promega Corporation, Madison, WI). PCR products (obtained directly from DNA or from clones) were incubated with 0.25 μl each of Exonuclease I and Shrimp Antarctic Phosphatase at 37 °C for 30 min, followed by 90 °C for 10 min. The products were purified on Sephadex® columns (Sigma–Aldrich). The purified product was sequenced with a Big Dye Terminator Cycle sequencing kit and an Automated 3730 DNA Analyzer (Applied Biosystems) using the

primers listed above. All unique haplotypes have been submitted to GenBank (Accession Numbers HM151021–HM151268). Sequences were edited, trimmed and aligned with Aligner version 3.0 (CodonCode Corporation, Dedham, MA).

2.2. Analyses

We examined each locus for evidence of intragenic recombination, calculating the minimum number of recombination events (R_m) in DnaSP 5.0 (Rozas et al., 2003). For each locus, we used 1000 replicate coalescent simulations in DnaSP 5.0 to obtain a confidence interval for the R_m value and a probability that the true R_m is less than or equal to the observed R_m.

For each locus which showed evidence of intragenic recombination we created recombination networks in the program SplitsTree4 (Huson and Bryant, 2006). Phylogenetic trees were constructed using Maximum Likelihood (ML) and Bayesian inference, for each locus alone, and for all loci concatenated. Analyses for each locus were performed using a data set composed only of the unique haplotypes. While we are aware that ML and Bayesian methods may not be entirely appropriate for loci showing evidence of intragenic recombination, it is only by using these methods that we can assess confidence in the topologies.

The best-scoring ML tree for each locus/concatenated loci and bootstrap support for each node on this tree were obtained using the program RAxML v. 7.0.0 on the CIPRES web portal (Stamatakis et al., 2008). The GTR + G likelihood model of nucleotide substitution was used in all analyses (General Time Reversible + Gamma Rate Distribution, RAxML only supports GTR-based models of nucleotide substitution). The concatenated data set was partitioned by locus, so α-shape parameters, GTR-rates, and base frequencies were estimated separately for each locus. All nodes with less than <50% support were then collapsed using TreeView 1.6.6 (Page, 1996).

Bayesian analyses were performed with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The GTR + G model of nucleotide substitution was applied to all data sets (Nset = 6) so that Bayesian trees could be compared to ML trees. In the concatenated data set, data were partitioned and the parameters for each locus (α-shape, GTR-rates, and base frequencies) were determined independently. Each analysis was run for 10 million generations, with sampling every 1000 generations. The first 2000 trees were eliminated as burn-in and a 50% majority-rule consensus tree was created from the remaining 8000 trees using PAUP*4.0 (Swofford, 2003). The MrBayes runs were carried out using the resources of the Computational Biology Service Unit at Cornell University which is partially funded by the Microsoft Corporation.

p-Distances (the proportion of nucleotide sites that differ between two sequences) were calculated for every Type A vs. Type B combination; average *p*-distance between the two types was then calculated in MEGA 4.0 (Tamura et al., 2007). Both uncorrected *p*-distances and *p*-distances corrected for the locus-specific best likelihood model of nucleotide substitution (as determined by AIC (Akaike Information Criterion) in Modeltest 3.7 (Posada and Crandall, 1998) were calculated. Additionally, the number of net nucleotide substitutions per site between Types A and B (D_a) was calculated for each locus in DnaSP 5.0 (Rozas et al., 2003) in order to correct for within-type variation (Nei, 1987).

We tested the position of *C. roulei* within the Type B clade by obtaining two best-scoring ML trees in RAxML using the GTR + G model for each gene (Stamatakis et al., 2008). One tree was constrained by Type B monophyly, the other by *C. roulei* monophyly. The likelihood of each constraint tree was compared to the unconstrained ML tree using the Shimodaira–Hasegawa (S–H) test in PAUP*4.0 (Swofford, 2003) with 10,000 resampling estimated log-likelihood (RELL) bootstrap replicates under a GTR + G model. These tests could not be performed for *vAChTP* because Type B and *C. roulei* individuals shared haplotypes, and for *Ci-Fgf4/5/6* in the case of *C. roulei* monophyly because only 1 *C. roulei* individual was sequenced for *Ci-Fgf4/5/6*.

Similar methodology was used to investigate the placement of *Ciona* sp. within each genealogy, with a best-scoring ML tree produced under the constraint that Type B/*C. roulei* and *Ciona* sp. formed a monophyletic group, to the exclusion of Type A.

We calculated two measures of nucleotide polymorphism within Type A, Type B and *C. roulei*: nucleotide diversity (π) and theta ($4N_e\mu$) estimated from segregating sites (θ_w) using DnaSP 5.0 (Rozas et al., 2003).

3. Results

Significant levels of intragenic recombination were detected at five of the six loci (Table 2), with Type B showing more recombination than Type A for all loci. The average minimum number of recombination events (R_m) across all six loci was 3 for Type A and 13.67 for Type B. Recombination networks created in SplitsTree4 (Huson and Bryant, 2006) revealed Types A and B within *C. intestinalis* to be very distinct, with *C. roulei* embedded within Type B (Fig. 1). Because SplitsTree4 (Huson and Bryant, 2006) was unable to create recombination networks when all variable sites were included for *Ci-fkh* and *Ci-Patched* and when no recombination events were detected in *vAChTP*, only recombination networks for *CiCesA*, *Ci-Fgf4/5/6* and *jade* are shown in Fig.1.

The published mtCOI genealogy for the genus *Ciona* revealed Type A and Type B individuals as members of two distinct and well-supported clades (100% bootstrap support for each) (Nydam and Harrison, 2007). This result, including the 100% bootstrap support for these clades, also holds in the updated mtCOI tree, which includes 13 new *C. intestinalis* individuals from Japan and the English Channel (tree not shown). The reciprocal monophyly of Types A and B is also evident when we use ML or Bayesian approaches to construct nuclear gene trees: either the Type A clade, the Type B clade, or both the Type A and B clades are well-supported although we acknowledge Bayesian posterior probabilities are often high (Fig. 2). There are no instances in which a Type A individual is found in a Type B clade, or vice versa (Figs. 1 and 2).

Among the nuclear gene loci, divergence estimates between the Type A and Type B groups varied by locus, from 0.035 for *vAChTP* to 0.116 for *Ci-Fgf4/5/6* (Table 3, uncorrected *p*-distances). Corrected *p*-distances ranged from 0.039 for *vAChTP* to 0.151 for *Ci-Fgf4/5/6*; D_a varied from 0.029 for *Ci-fkh* to 0.09 for *Ci-Fgf4/5/6* (Table 3).

In all trees and networks, Type B is consistently paraphyletic with respect to *C. roulei*, a pattern which was evident in the original mtCOI tree (Nydam and Harrison, 2007). Where multiple *C. roulei* individuals are included (all trees or networks except those for *Ci-Fgf4/5/6*), these individuals never form a monophyletic group within the Type B clade (Fig. 1). The S–H tests found no statistical differences for any of the genes between the unconstrained tree and the tree constrained by Type B monophyly ($p > 0.05$), (*CiCesA*: $p = 0.1828$, *Ci-Fgf4/5/6*: $p = 0.4595$, *Ci-fkh*: $p = 0.4393$, *jade:* $p = 0.4989$, mtCOI: $p = 0.4636$, *Ci-Patched*: $p = 0.4675$). The S–H tests also found no statistical differences between the unconstrained tree and the tree constrained by *C. roulei* monophyly for three genes ($p > 0.05$), (*Ci-fkh*: $p = 0.1333$, mtCOI: $p = 0.4944$, *Ci-*

Table 2
Intragenic recombination, measured as R_m (minimum # of recombination events).

	vAChTP	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*
Type A	0	0	4	2	7	5
Type B	0	6	19	18	21	18

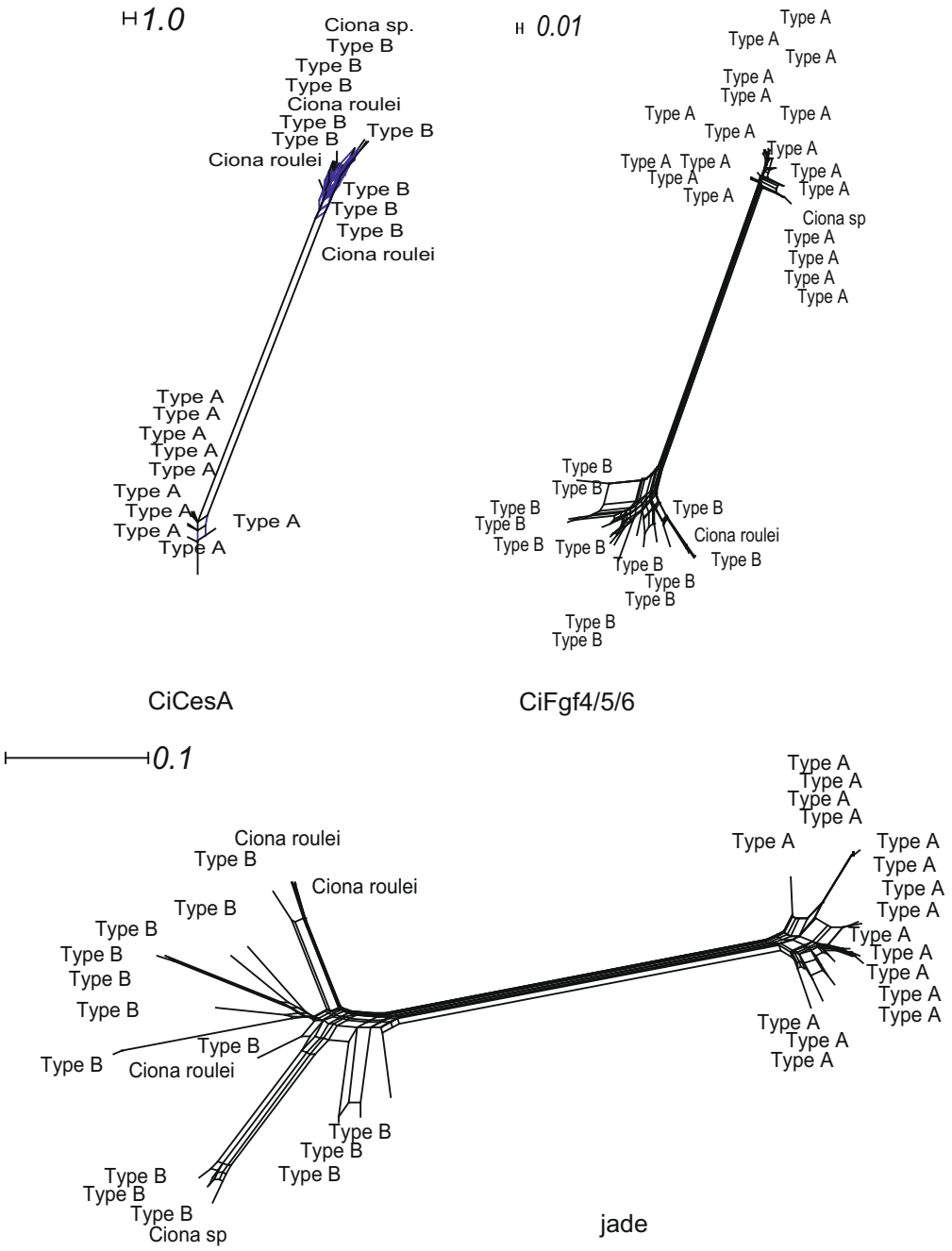

Fig. 1. Recombination networks for *CiCesA*, *Ci-Fgf4/5/6* and *jade*. The scale bar in upper left hand corner of each network is proportional to the amount of molecular difference between the groups.

Patched: $p = 0.3529$). However, two of the genes showed a significantly better likelihood score for the unconstrained tree than the tree constrained by *C. roulei* monophyly ($p > 0.05$), (*CiCesA*: $p = 0.0309$, *jade*: $p < 0.001$).

Using only the mtCOI marker, the relationship of *Ciona* sp. to its congeners is not clearly defined; MP and ML methods are unable to resolve the relationships among *Ciona* sp., Type A and Type B/*C. roulei*, whereas NJ methods placed *Ciona* sp. as sister to the *C. intestinalis* clade (Nydam and Harrison, 2007). In the nuclear gene trees, where *Ciona* sp. individuals can be associated with one type or the other, they are always associated with Type B, except the *Ci-Fgf4/5/6* tree where *Ciona* sp. is in the Type A clade (Fig. 1; ML trees are not shown). Divergence estimates are smaller for Type B vs. *Ciona* sp. than for Type A vs. *Ciona* sp. for all loci except *Ci-Fgf4/5/6* and mtCOI (Table 3).

In order to assess support for defined topologies with respect to placement of *Ciona* sp., we compared the unconstrained best-scoring ML tree for each locus to the best-scoring tree constrained by Type B/*C. roulei/Ciona* sp. monophyly. For four of the loci, the unconstrained tree was not significantly different from the constrained tree ($p > 0.05$), (*vAChTP*: $p = 0.3167$, *CiCesA*: $p = 0.1323$, *Ci-Fgf4/5/6*: $p = 0.2684$, *Ci-fkh*: $p = 0.4369$). For the remaining three loci (*jade*, mtCOI, *Ci-Patched*), the unconstrained tree was significantly better supported than the constrained tree ($p < 0.05$), (*jade*: $p = 0.0304$, mtCOI: $p = 0.000$, *Ci-Patched*: $p = 0.0411$).

The concatenated data set gave a different result (with respect to the placement of *Ciona* sp.) using ML and Bayesian algorithms. In ML analysis *Ciona* sp. is a clade distinct from the well-supported Type A and Type B/*C. roulei* clades; but relationships among these clades cannot be resolved (Fig. 3a). However, in the Bayesian anal-

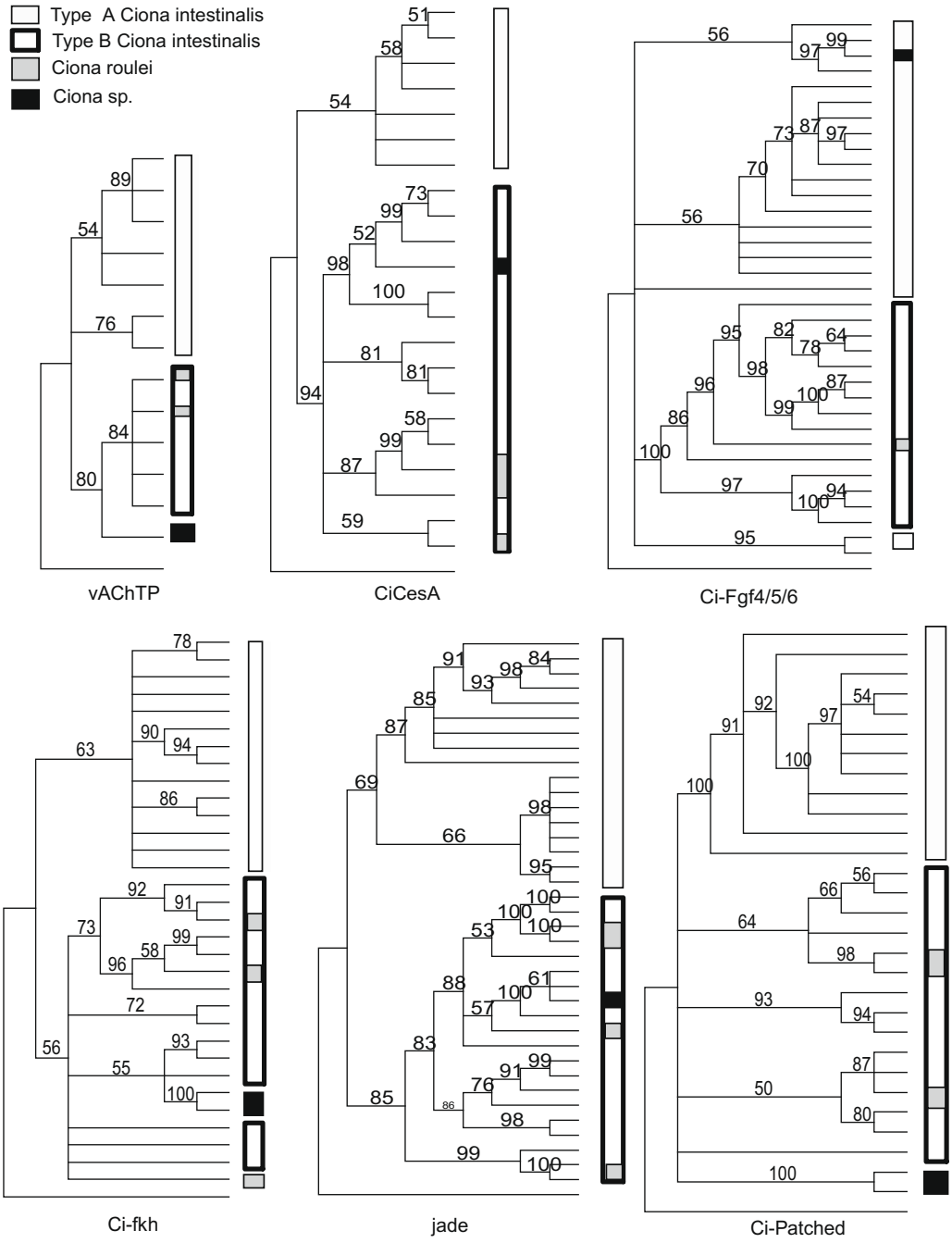

Fig. 2. 50% majority-rule consensus trees created in PAUP* 4.0 from trees obtained by MrBayes 3.1.2. Values on the branches are Bayesian posterior probabilities. Nodes with a posterior probability value less than 50 were collapsed. The outgroup is *Ciona savignyi*.

ysis, Type B/*C. roulei*/*Ciona* sp. form a well-supported monophyletic group, as does Type A (Fig. 3b). Each of the loci used in the concatenated data set have a substantial number of phylogenetically informative sites, so we view these concatenated trees as the product of information assembled from across the genome, rather than from one or two dominant loci.

The π values averaged across all six nuclear loci are 0.0094 (Type A), 0.0361 (Type B), and 0.0324 (*C. roulei*). The θ_w values are 0.0104 (Type A), 0.0388 (Type B), and 0.03 (*C. roulei*). Levels of polymorphism were highly variable across the six nuclear loci (Table 4). The two loci determined to have high level of heterozygosity in the Type A genome (*Ci-Fgf/4/5/6* and *jade*) had the highest π and θ_w values for nearly all loci and species.

4. Discussion

4.1. Relationships within Ciona: molecules and morphology

Our data suggest that levels of morphological and molecular differentiation are not correlated within the shallow-water *Ciona* species. Morphologically similar entities (Type A and Type B) are highly differentiated at all loci examined, whereas morphologically distinct "species" (*C. roulei* and Type B) cannot be distinguished based on the seven loci we surveyed. Indeed, *C. roulei* is always embedded within Type B and never forms a monophyletic group within the Type B clade for any of the loci used.

Table 3
p-Distances and D_a between each *C. intestinalis* type and between each *C. intestinalis* type and *Ciona* sp. Uncorrected *p*-distance is the proportion of nucleotide sites that differ between two sequences, and corrected *p*-distance is obtained by incorporating the locus-specific best likelihood model of nucleotide substitution. D_a is the number of net nucleotide substitutions per site between species.

		vAChTP	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*	*mtCOI*
Type A vs. Type B *C. intestinalis*/*C. roulei*	Uncorrected *p*-distance	0.035	0.077	0.116	0.051	0.113	0.077	0.124
	Corrected *p*-distance	0.039	0.092	0.151	0.055	0.139	0.081	0.156
	D_a	0.035	0.072	0.09	0.029	0.08	0.051	0.102
Type A *C. intestinalis* vs. *Ciona* sp.	Uncorrected *p*-distance	0.191	0.079	0.019	0.184	0.125	0.241	0.109
	Corrected *p*-distance	0.413	0.104	0.024	0.307	0.161	0.388	0.187
	D_a	0.193	0.085	0.015	0.193	0.129	0.229	0.099
Type B *C. intestinalis*/*C. roulei* vs. *Ciona* sp.	Uncorrected *p*-distance	0.177	0.014	0.115	0.176	0.069	0.229	0.125
	Corrected *p*-distance	0.35	0.015	0.094	0.286	0.081	0.358	0.232
	D_a	0.184	0.007	0.097	0.176	0.03	0.207	0.121

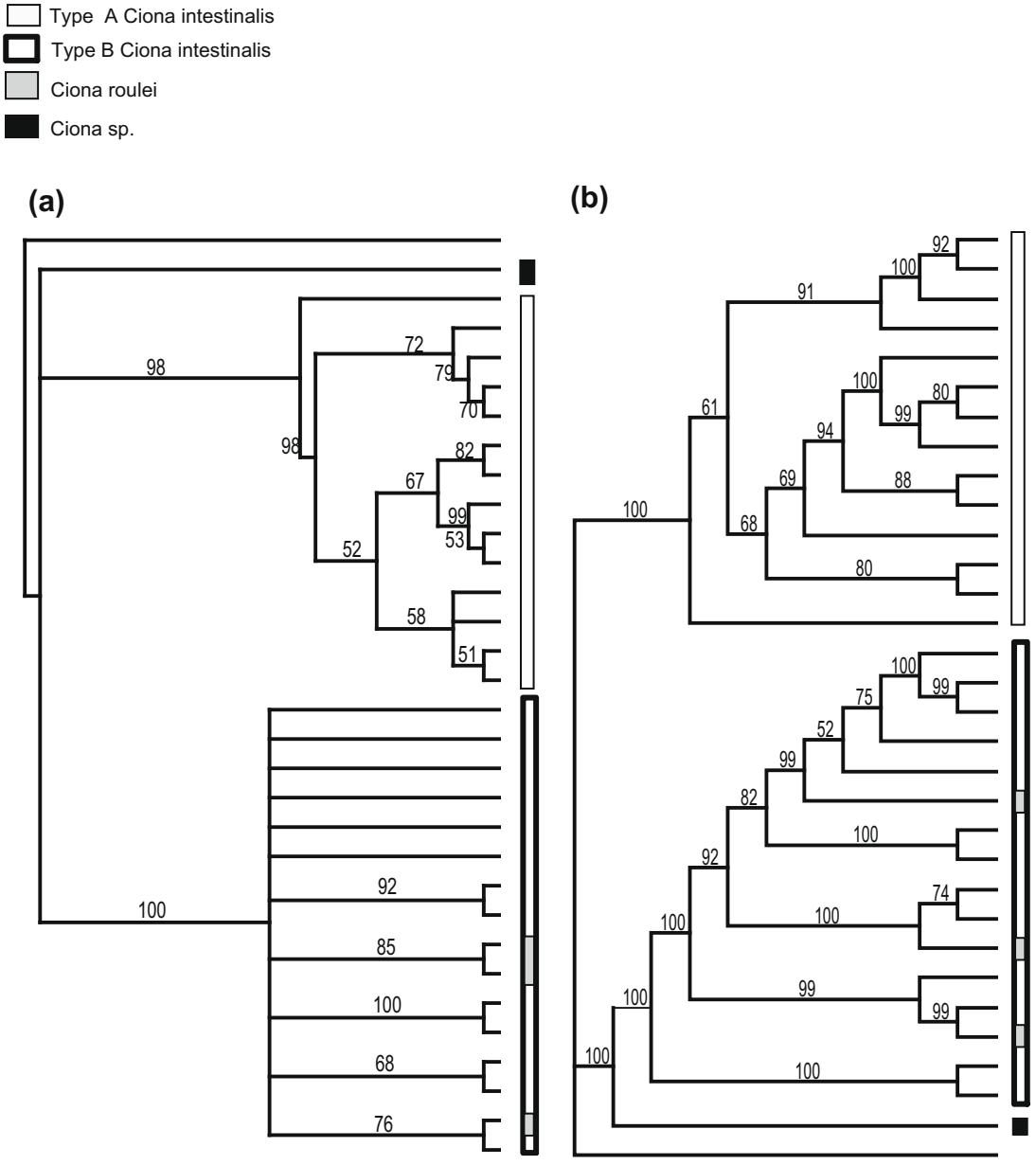

Fig. 3. Trees obtained from concatenating the seven loci into one data set. The outgroup is *Ciona savignyi*. (a) Maximum Likelihood: the best-scoring ML tree was obtained using the program RAxML v. 7.0.0. Values on the branches are bootstrap values. All nodes with less than 50% support collapsed. (b) Bayesian: 50% majority-rule consensus trees created in PAUP* 4.0 from trees obtained by MrBayes 3.1.2. Values on the branches are Bayesian posterior probabilities. Those nodes with a posterior probability value less than 50 were collapsed.

Table 4
Polymorphism within each species. Polymorphism is measured by π, nucleotide diversity, calculated as the weighted average of the proportion of nucleotide differences among all sequences in the population, and Θ_w (4Neμ) calculated from the number of segregating sites.

Species	Measure	*vAChTP*	*CiCesA*	*Ci-Fgf4/5/6*	*Ci-fkh*	*jade*	*Ci-Patched*
Type A	π	0.0027	0.0024	0.0097	0.0070	0.0225	0.0130
	Θ_w	0.0023	0.0029	0.0117	0.0094	0.0220	0.0164
Type B	π	0.0006	0.0340	0.0460	0.0382	0.0616	0.0363
	Θ_w	0.0009	0.0488	0.0446	0.0473	0.0529	0.0382
C. roulei	π	0.0009	0.0190	NA*	0.0343	0.0811	0.0266
	Θ_w	0.0009	0.0185	NA*	0.0343	0.0699	0.0266

* Only one *C. roulei* individual was sequenced for *Ci-Fgf4/5/6*.

Previous workers have always considered *C. roulei* morphologically distinct from *C. intestinalis* (Lahille, 1890; Harant and Vernieres, 1933; Fiala-Medioni, 1974; Lambert et al., 1990). Harant and Vernieres (1933) and Lahille (1890) compared *C. roulei* to sympatric *C. intestinalis* (Type A) using multiple individuals from several populations of each species; these studies remain the only detailed morphological descriptions of *C. roulei* (Lahille, 1890; Harant and Vernieres, 1933). Although several of these distinguishing traits are variable within *C. intestinalis,* these variations do not overlap with *C. roulei*'s phenotype. For instance, *Ciona pulchella* (subsequently synonymized with *C. intestinalis* (Hoshino and Nishikawa, 1985)), from Devon, Cornwall and Guernsey (English Channel coast) was described as a new species distinct from *C. intestinalis* owing in part to its coloration (reddish, pale yellow or hyaline white) (Alder and Hancock, 1907). The coloration of this morph is distinct from *C. roulei*, however (personal observation; Lahille, 1890; Harant and Vernieres, 1933). A orange morph of Type B was also described by Millar (1953), but this morph has little or no pigment in the transverse bars of the branchial sac (Millar, 1953), whereas the reddish coloration of *C. roulei* is due to pigment in the peribranchial and transverse vessels of the branchial sac (Lahille, 1890). Likewise, the tunic of *C. intestinalis* has been described as both "smooth and soft" (Alder and Hancock, 1907) and "wrinkled" (Harant and Vernieres, 1933); the tunic morphology of *C. intestinalis* varies with location, habitat and depth (personal observation). But *C. roulei*'s tunic is consistently smooth (personal observation; Lahille, 1890; Harant and Vernieres, 1933).

Although these morphological differences were based on comparisons with Type A, the two types of *C. intestinalis* are extremely similar morphologically. In fact, they were not distinguished at all before molecular data became available, despite a thorough anatomical study of two populations of each type (Hoshino and Nishikawa, 1985). Furthermore, the few traits by which they may differ are not the same traits that differentiate *C. roulei* and Type A (Caputi et al., 2007). Therefore, the morphological traits that separate Type A and *C. roulei* must also separate Type B and *C. roulei*.

Because *Ciona roulei* and Type B are genetically inseparable yet morphologically distinct, genetic and morphological data conflict with regard to whether *C. roulei* should be considered a species separate from Type B. According to the Biological Species Concept, the inability of two taxa to interbreed confers species status on both taxa, regardless of genetic similarity (Mayr, 1995). Therefore, the presence of reproductive incompatibilities would enable us to label *C. roulei* as a species distinct from Type B. While we know that sympatric *C. roulei* and Type A are incompatible in one direction (Lambert et al., 1990), compatibility between the allopatric *C. roulei* and Type B has not been determined as *C. roulei* is now difficult to obtain. Without data on reproductive isolation between these two types, we have no clear evidence supporting species status for *C. roulei.*

The only evidence for the distinctness of *C. roulei* and Type B lies in their morphological differences, as several genetic observations fail to discriminate these two taxa. First, for all the loci for which multiple *C. roulei* individuals were sequenced, haplotypes found in *C. roulei* are interspersed in the Type B clade (i.e. some haplotypes of *C. roulei* are more similar to Type B haplotypes than to other *C. roulei* haplotypes). At one locus, *vaChTP,* Type B and *C. roulei* even share haplotypes. Second, at two of five loci tested, an unconstrained tree had a statistically better likelihood than a tree constrained by *C. roulei* monophyly. Therefore, it is likely that *C. roulei* is recently derived from Type B individuals that invaded the Mediterranean Sea. It is possible that the observed morphological differences might reflect the new environmental conditions encountered in the Mediterranean Sea. Without additional data, we cannot distinguish between scenarios in which these differences represent genetic changes in response to a new selection regime and scenarios that involve direct responses to the environment (phenotypic plasticity).

Ciona sp., which is found in the Mediterranean Sea, appears to be more closely related to Type B than Type A. However, while six loci result in trees in which *Ciona* sp. is associated with Type B rather than Type A, the S–H tests show that only four loci provide significant support for an association of *Ciona* sp. with Type B/*C. roulei* rather than as sister to a *C. intestinalis/C.roulei* clade. Trees based on the concatenated data give different results depending on which tree-building algorithm is used. Therefore, our results lead us to support an association of *Ciona* sp. with Type B/*C. roulei* to the exclusion of Type A. However, analyses that include additional *Ciona* sp. individuals and additional loci are necessary to resolve these relationships.

The morphology of *Ciona* sp. is distinct from that of *Ciona* located in this region (for a description see Nydam and Harrison (2007), including *Ciona edwardsi* (Roule, 1886), a species which is extremely rare (Copello, 1981), and for which specimens could not be obtained. In the case of *Ciona* sp., genetic and morphological data agree that this species is distinct from other *Ciona*, although no data are available on reproductive compatibilities between this species and other *Ciona* species.

4.2. Levels of polymorphism

Based on whole genome annotation and analyses from single individuals, both *C. intestinalis* and *C. savignyi* have previously been shown to have very high levels of Single Nucleotide Polymorphisms, but levels of heterozygosity vary widely across the genome (Dehal et al., 2002; Small et al., 2007). Indeed, it was argued (Small et al., 2007) that *C. savignyi* harbors the highest levels of SNP variation in a multi-cellular organism. In *Ciona*, high levels of variation have been attributed to large effective population size. This explanation has been proposed not only for sea squirts but also for the Pacific oyster, *Crassostrea gigas* (Sauvage et al., 2007) and the nematode *Caenorhabditis remanei* (Cutter et al., 2006).

For the loci *Ci-Fgf4/5/6, Ci-fkh and Ci-Patched*, the sequences we obtained are entirely noncoding, so we can compare π for these three loci to π for noncoding sites in the oyster and nematode loci. Average π values across loci are 0.038 in oysters (Sauvage et al., 2007) and 0.051 in nematodes (Cutter et al., 2006), whereas averages for Type A, Type B, and *C. roulei* are 0.01, 0.04 and 0.03.

θ_w values are also similar between nematodes and *Ciona* for non-coding sites (not calculated for oysters).

The variability in amounts of population-level polymorphism for all three *Ciona* "species" among the six nuclear loci examined in this paper confirms the pattern seen in the single Type A individual from which genome sequence was obtained. Not only is this variability also present in populations of Type B and *C. roulei*, but levels of diversity in these taxa are substantially higher than in Type A at all six loci.

We cannot currently explain the difference in polymorphism between Types A and B *C. intestinalis*. Available evidence does not support the notion that differences in current effective population size can explain the remarkable amount of variation found in *C. intestinalis*. Type A, as a cosmopolitan species, likely has a larger effective population size than the geographically restricted Type B. And although a larger number of Type A than Type B populations are invasive and may have experienced a reduction of genetic diversity, the analysis of Boffelli et al. shows both native and invasive Type B populations with much higher heterozygosity than either invasive or native Type A populations (2004).

Levels of polymorphism in *C. roulei* are similar to those in Type B, even though the *C. roulei* estimates were obtained from only 2–3 individuals, whereas Type B estimates were obtained from 11 to 13 individuals. Given the paraphyly of Type B with respect to *C. roulei*, we previously proposed a scenario in which Type B individuals invaded the Mediterranean Sea from the North Atlantic Ocean (Nydam and Harrison, 2007). This invasion must have occurred relatively recently, given that *C. roulei* haplotypes are embedded within Type B haplotypes. Although the signature of a bottleneck may be obscured by population expansion, the current diversity of *C. roulei* does not support the idea that the ancestors of this species suffered a substantial reduction in genetic diversity due to a bottleneck effect when entering the Mediterranean Sea.

The genus *Ciona* provides an excellent system for investigating the evolution of molecules and morphology. *Ciona* species pairs represent the entire range of possible combinations of morphological and genetic divergence. Types A and B are cryptic species: divergent at the DNA sequence level but morphologically nearly identical. Discovery of cryptic species in the marine environment is accelerating with the widespread use of molecular tools; examining the biology of existing cryptic species will allow us to understand how and why substantial molecular divergence has evolved without concomitant morphological change. In contrast, Type B and *C. roulei* are genetically indistinguishable but morphologically divergent. This species pair provides an opportunity to address questions in evolutionary ecology related to local adaptation and phenotypic plasticity. Finally, *Ciona* sp. is both morphologically and molecularly distinct from all other *Ciona* spp. The discovery of a previously unknown and morphologically distinct *Ciona* species in a harbor immediately adjacent to a marine laboratory highlights how little is known about a genus that includes important model organisms. Surely a continued effort to understand diversity within the genus *Ciona* will increase the utility of *C. intestinalis* and *C. savignyi* as models for diverse biological processes. Levels of polymorphism in natural populations are similar to levels reported in other organisms that are considered to be highly polymorphic (nematodes and oysters). This study provides the first estimate of population-level heterozygosity in *Ciona* and provides a critical first step in elucidating the origin and maintenance of these extraordinary high levels of polymorphism.

Acknowledgments

We would like to thank members of the Harrison Lab, C. Aquadro, and A. Clark for giving advice and commenting on an earlier version of the manuscript, D. Chong for help with data collection, J. Bishop for providing logistical assistance and G. Miller-Messner for collecting the Japanese samples. This work was funded by an AMNH Lerner-Gray Grant for Marine Research, a Cornell University Research Travel Grant to M.L.N., and by NSF Grant DEB-0639904 to R.G.H.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.03.042.

References

Alder, J., Hancock, A., 1907. The British Tunicata, vol. 2. Ray Society, London..

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christofells, A., Rash, S., Hoon, S., Smit, A., Gelpke, M.D., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., Brenner, S., 2002. Whole genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301–1310.

Boffelli, D., Weer, C.V., Weng, L., Lewis, K., Shoukry, M.I., Pachter, L., Keys, D.N., Rubin, E.M., 2004. Intraspecies sequence comparisons for annotating genomes. Genome Res. 14, 2406–2411.

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M., Sordino, P., 2007. Cryptic speciation in a model invertebrate chordate. Proc. Natl. Acad. Sci. USA 104, 9364–9369.

Copello, M., 1981. *Ciona edwardsi* (Roule, 1886) Espece littorale de Mediterranee distincte de *Ciona intestinalis* (Linne, 1767). Vie Milieu A Mar. 31, 243–253.

Cutter, A., Baird, S.E., Charlesworth, D.A., 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics 174, 901–913.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., Harafuji, N., Hastings, K.E.M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I.A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D.N., Haga, S., Hayashi, H., Hino, K., Imai, K.S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K.W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-I, T., Spagnuolo, A., Stainier, D., Suzuki, M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P.D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., Rokhsar, D.S., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298, 2157–2167.

Fiala-Medioni, A., 1974. Ascidians of the rocky benthos of Banyuls-sur-Mer. Faunistic inventory and ecology notes. Vie Milieu B Oceanog. 24, 193–207.

Harant, H., Vernieres, P., 1933. Tuniciers. Faune de France 27, 1–93.

Herdman, W.A., 1882. Report on the Tunicata Collected During the Voyage of the HMS Challenger During the Years 1873–1876. Part I: Ascidiae Simplices. Report on the Scientific Results of the Voyage of HMS Challenger 1882. 6, 236–237.

Hoshino, Z.I., Nishikawa, T., 1985. Taxonomic studies of *Ciona intestinalis* (L.) and its allies. Publ. Seto. Mar. Biol. Lab. 30, 61–79.

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

Kott, P., 1952. The ascidians of Australia. Aust. J. Mar. Fresh. Res. 3, 206–233.

Lahille, M.F., 1890. I. Recherches sur les Tuniciers II. Propositions donnees par la Faculte. PhD. Thesis. Faculte des Sciences Paris.

Lambert, C.C., LaFargue, F., Lambert, G., 1990. Preliminary note on the genetic isolation of *Ciona* species (Ascidiacea, Urochordata). Vie Milieu A Mar. 40, 293–295.

Linne, C., 1767. Vindobonae: Typis Ioannis Thomae, Systema Naturae, pp. 1767–1770.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Mayr, E., 1995. Species, classification, and evolution. In: Arai, R., Kato, M., Doi, Y. (Eds.), Biodiversity and Evolution. National Science Museum Foundation, Tokyo, pp. 3–12.

Millar, R.H., 1953. *Ciona*. The University Press of Liverpool, Liverpool.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 37–45.

Nydam, M.L., Harrison, R.G., 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). Mar. Biol. 151, 1839–1847.

Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: bayesian inference under mixed models. Bioinformatics 19, 1572–1574.

Rozas, J., Sánchez-Del Barrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Sauvage, C., Bierne, N., Lapegue, S., Boudry, P., 2007. Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. Gene 406, 13–22.

Small, K.S., Brudno, M., Hill, M.M., Sidow, A., 2007. Extreme genomic variation in a natural population. Proc. Natl. Acad. Sci. USA 104, 5698–5703.

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

Suzuki, M., Nishikawa, T., Bird, A., 2005. Genomic approaches reveal unexpected difference within *Ciona intestinalis*. J. Mol. Evol. 61, 627–635.

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0b 10. Sinauer Associates, Sunderland.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Van Name, W., 1945. The north and south American Ascidians. Bull. Am. Mus. Nat. Hist. 84, 1–476.

Vinson, J., Jaffe, D., O'Neill, K., Karlsson, E., Stange-Thomann, N., Anderson, S., Mesirov, J., Satoh, N., Nusbaum, C., Birren, B., Galagan, J., Lander, E., 2005. Assembly of polymorphic genomes: algorithms and application to *Ciona savignyi*. Genome Res. 15, 1127–1135.