MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 225–238 doi: 10.1111/mec.12584

Evaluating summary statistics used to test for incomplete lineage sorting: mito-nuclear discordance in the reef sponge *Callyspongia vaginalis*

MELISSA B. DEBIASSE, BRADLEY J. NELSON and MICHAEL E. HELLBERG
Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

Abstract

Conflicting patterns of population differentiation between the mitochondrial and nuclear genomes (mito-nuclear discordance) have become increasingly evident as multilocus data sets have become easier to generate. Incomplete lineage sorting (ILS) of nucDNA is often implicated as the cause of such discordance, stemming from the large effective population size of nucDNA relative to mtDNA. However, selection, sex-biased dispersal and historical demography can also lead to mito-nuclear discordance. Here, we compare patterns of genetic diversity and subdivision for six nuclear protein-coding gene regions to those for mtDNA in a common Caribbean coral reef sponge, *Callyspongia vaginalis*, along the Florida reef tract. We also evaluated a suite of summary statistics to determine which are effective metrics for comparing empirical and simulated data when testing drivers of mito-nuclear discordance in a statistical framework. While earlier work revealed three divergent and geographically subdivided mtDNA *COI* haplotypes separated by 2.4% sequence divergence, nuclear alleles were admixed with respect to mitochondrial clade and geography. Bayesian analysis showed that substitution rates for the nuclear loci were up to 7 times faster than for mitochondrial *COI*. Coalescent simulations and neutrality tests suggested that mito-nuclear discordance in *C. vaginalis* is not the result of ILS in the nucDNA or selection on the mtDNA but is more likely caused by changes in population size. Sperm-mediated gene flow may also influence patterns of population subdivision in the nucDNA.

Keywords: coalescent, Florida keys, Porifera, simulation

Received 17 January 2013; revision received 25 October 2013; accepted 28 October 2013

Introduction

The number of studies finding conflicting phylogeographical patterns between mitochondrial and nuclear genomes (mito-nuclear discordance) has increased in the last three decades as it has become easier to generate multilocus data sets (Fig. 3 in Toews & Brelsford 2012). Mito-nuclear discordance may be expected due to the intrinsic characteristics of each genome. The mitochondrial DNA (mtDNA) of most animals is generally free of recombination (Birky 2001), is haploid and maternally inherited, while nuclear DNA (nucDNA) recombines, is diploid and is transmitted by both parents. Mitochondrial DNA thus has a fourfold smaller effective population size than the nuclear genome in taxa with separate sexes. Mitochondrial genes are generally expected to undergo lineage sorting faster than nuclear genes, as the rate of sorting is inversely proportional to effective size (Funk & Omland 2003; Zink & Barrowclough 2008) and may provide signals of population differentiation when nucDNA does not. Mito-nuclear discordance may also arise as a function of the difference in rates of nucleotide substitution between the mtDNA and nucDNA. Typically, nucDNA has lower rates of nucleotide substitution compared with mtDNA (Brown *et al.* 1979), and in some cases, variation in the nucDNA may be insufficient to reveal phylogeographical patterns observed in the mtDNA (Hare 2001; Hickerson & Cunningham 2005).

Correspondence: Melissa B. DeBiasse, Fax: 1 225 578 2597;
E-mail: melissa.debiasse@gmail.com

Most studies finding structure in the mtDNA, but not in nucDNA, have corresponding historical or biogeographical evidence that suggests mtDNA haplotypes diverged in allopatry (Toews & Brelsford 2012). In these cases, the absence of complimentary genetic divergence in the nucDNA is often explained by incomplete lineage sorting (ILS) of the nucDNA (Zink & Barrowclough 2008; McKay & Zink 2010). However, mito-nuclear discordance without strong barriers to gene flow can arise due to selection (Cheviron & Brumfield 2009), nonrandom mating (Pardini *et al.* 2001) or differences in substitution rates (Orozco-Terwengel *et al.* 2008; Eytan & Hellberg 2010), among other mechanisms (Irwin 2002).

Coalescent simulations have been used to test whether mito-nuclear discordance results from ILS. In this approach, nucDNA are simulated under the demographic history inferred from the mtDNA, and the distribution of a summary statistic calculated for the simulated data is constructed. This null distribution represents the range of values that would be expected in the nucDNA given the differences in effective population size (and therefore sorting time) of the nuclear and mitochondrial genomes. The hypothesis of ILS is rejected if the test statistic of the empirical nuclear data falls outside of the expected distribution. For example, Peters *et al.* (2012) rejected ILS in favour of male-biased dispersal in a sea duck based on observed and simulated Φ_{ST} values. Wilson & Eigenmann Veraguth (2010) failed to reject ILS in a pipefish when F_{ST} and D_{EST} values for the empirical nucDNA sequence data fell within their respective null distributions and concluded that mito-nuclear discordance was due to the stochasticity of the coalescent. Wilson & Eigenmann Veraguth (2010) and Peters *et al.* (2012) used summary statistics that describe how genetic variation is partitioned among populations. However, it is possible that some summary statistics might not fully capture the signal of ILS in coalescent genealogies. To our knowledge, no one has evaluated the performance of different statistics to compare empirical and simulated data when testing drivers of mito-nuclear discordance.

Sponges offer a novel perspective on the mechanisms underlying mito-nuclear discordance due to their reproductive biology and attributes of their mtDNA that differ from those of most animals. Mitochondrial nucleotide substitution rates among some basal metazoans (including the Porifera and Anthozoa) are up to 100 times slower than in most bilateral animals (Shearer *et al.* 2002; Hellberg 2006), and studies surveying *COI* sequence variation in the Porifera have found little to no intraspecific variation over distances up to 20 000 km (Duran *et al.* 2004c; Wörheide 2006; Whalan *et al.* 2008; López-Legentil & Pawlik 2009; Duminil *et al.* 2011). In corals, slow mtDNA rates are not matched by corresponding slow rates for nucDNA (Eytan *et al.* 2009), and the pattern of fast mtDNA evolution compared with nucDNA seen in bilateral animals (Brown *et al.* 1979) is reversed in corals (Hellberg 2006) as it is in plants (Wolfe *et al.* 1987). Rates for mtDNA and nucDNA genes are lacking for sponges, but Duran *et al.* (2004a,c) found higher sequence variation for multicopy nuclear *ITS* than for *COI* in the Mediterranean sponge *Crambe crambe*. More equitable rates of nucleotide substitution in the sponge nucDNA should reduce mito-nuclear discordance caused by difference in the resolving power of the mtDNA and nucDNA.

Although factors such as mating system and variation in reproductive success can alter effective population size (Nunney 1993), the hermaphroditic reproductive strategy in most sponges (Bergquist 1978) allows every individual to pass on a mitochondrial genome and should theoretically make the ratio of the effective size of the mtDNA to the nucDNA 1:2 as opposed to 1:4 as in gonochoristic taxa. The reduction in the difference between effective sizes of the two genomes may decrease the likelihood of discordance due to ILS, thus making sponges a useful model to reveal other drivers of mito-nuclear discordance.

Callyspongia vaginalis is a common sponge that occurs on coral reefs throughout Florida and the Caribbean (Humann & DeLoach 2003) and broods larvae that are competent to settle almost immediately after release (Maldonado 2006). Sequence data from the mitochondrial *COI* gene revealed significant genetic subdivision among populations of *C. vaginalis* collected from 465 km of the Florida reef tract ($\Phi_{ST} = 0.33$, $P < 0.0001$, DeBiasse *et al.* 2010). Most sampled sponges possessed one of three divergent mitochondrial haplotypes, each separated by an average of 12 mutational steps (2.4% sequence divergence). Preliminary *COI* data from other locations show that the three major Florida haplotypes are more divergent from each other than they are from any other haplotype sampled in *C. vaginalis* across the Caribbean.

Given the low rates of mtDNA evolution in sponges and problems associated with drawing phylogeographical conclusions based on one locus, we collected sequence data from six novel, nuclear protein-coding genes to test for concordant patterns of genetic diversity and subdivision between the mtDNA and nucDNA in *C. vaginalis* in Florida. In contrast to the patterns observed in the mtDNA, nucDNA sequences showed no geographical subdivision nor differentiation corresponding to the three mitochondrial clades. Motivated by these findings, we tested for selection on the mtDNA and used coalescent simulations to test the hypotheses that the mito-nuclear discordance in *C. vaginalis* is due to ILS of the nucDNA and that changes in population size might

differentially affect the mtDNA and nucDNA. We also evaluated five commonly used summary statistics for their ability to reveal discordance.

Materials and methods

Geographical sampling and development of novel nuclear markers

Our aims were to determine how variation in the nucDNA was partitioned within and among populations of *Callyspongia vaginalis* and how well any patterns of structure and/or sequence divergence in nucDNA corresponded to those observed the mtDNA. Therefore, we randomly subsampled 61 individuals from the study by DeBiasse *et al.* (2010) from the three mtDNA lineages in each of locations where the lineages occurred sympatrically (Fig. 1) and generated nucDNA sequences for each individual. Genetic resources for sponges are lacking compared with other taxa, particularly at the species level. Therefore, nuclear markers for this study were generated *de novo* from a *C. vaginalis* cDNA library following the protocol of Hale *et al.* (2009) using 454 sequencing technology. Details about the development of nuclear markers, PCR amplification and sequencing are in the Supporting Information.

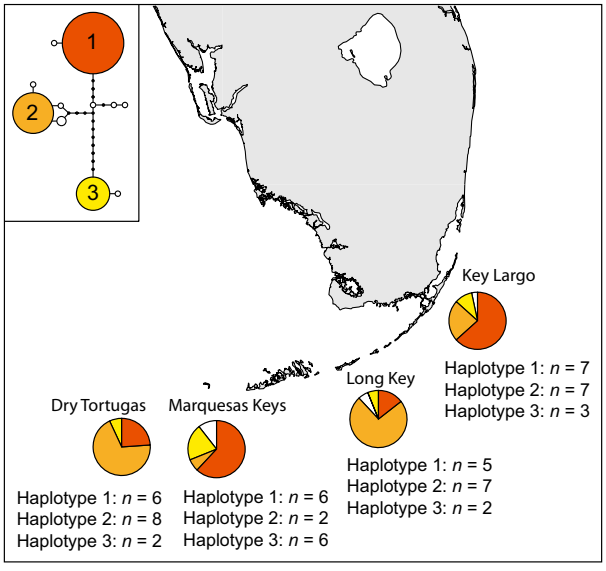

Fig. 1 Distribution of *C. vaginalis COI* haplotypes along the Florida reef tract in four locations sampled by DeBiasse *et al.* (2010). The three most common haplotypes are represented in yellow, light orange and dark orange. Less frequent haplotypes are in white. The area of each sector is proportional to haplotype frequency at each site. Sample sizes listed below each pie chart show the number of individuals of each haplotype for which nucDNA sequences were collected. Inset shows haplotype network with common haplotypes 1, 2 and 3 represented by the colours dark orange, light orange and yellow, respectively.

Primers for six nuclear genes showed consistent PCR and sequencing results and had sequence variation between individuals (Table S1, Supporting information). Some genes failed to amplify or sequence for some individuals (eight for *cata*; three for *cps*; five for *ef*). As appropriate, analyses were performed on data sets including all individuals for three sequenced genes (*cir*, *fil*, *mep*), only the individuals sequenced for all six genes and/or individuals with missing data. Sequences generated here are available from the European Nucleotide Archive (Accession nos. HG000670 – HG001246 and HG737726 – 737847) and the Dryad Digital Repository: doi:10.5061/dryad.hn104.

Phasing of alleles and tests for recombination

We resolved alleles in heterozygous individuals using PHASE v2.1 (Stephens *et al.* 2001). Individuals heterozygous for one insertion/deletion were resolved using CHAMPURU v1.0 (Flot 2007). Individuals with alleles that could not be phased probabilistically to a probability >90% were cloned using the Invitrogen TOPO TA kit. At least eight clones per reaction were sequenced. Individuals that could not be resolved after several rounds of cloning and phasing (one for *cata*; three for *cps*; four for *ef*) were removed from the data set. Each gene region was tested for intralocus recombination using GARD and SBP implemented in Hy-Phy (Pond & Frost 2005; Pond *et al.* 2006) and with the DSS method implemented in TOPALi v2.0 (Milne *et al.* 2004). Recombination was not detected in any gene region.

Analysis of nuclear sequence data

We used three approaches to determine the distribution of nucDNA genetic variation in *C. vaginalis* according to the geographical location and mitochondrial lineage from which each individual was sampled. To visualize the relationships between nucDNA sequences, we constructed allelic networks for each nuclear locus using TCS v1.21 (Clement *et al.* 2000). These analyses were conducted using the default settings and resulted in the most parsimonious connections between alleles at the 95% confidence level. STRUCTURE v2.3.2 (Pritchard *et al.* 2000) was used to infer genetic subdivision for the nucDNA. First, nucDNA sequences were recoded into frequency data using DnaSP v5.0 (Rozas *et al.* 2003; Librado & Rozas 2009), and individuals were genotyped. No individuals shared a multilocus genotype, meaning subdivision measures cannot be skewed by clonal reproduction. We used the admixed

ancestry model with correlated allele frequencies and set the number of clusters, '*K*', to 1–8, eight being twice the number of geographical locations. We ran the program with 1 million MCMC steps and discarded the first 10 000 steps as burn-in. We performed 20 iterations for each *K*. We used the Evanno method (Evanno *et al.* 2005) implemented in STRUCTURE HARVESTOR (Earl & vonHoldt 2011) to determine the most likely number of clusters in the data. Finally, we used a hierarchical AMOVA (Excoffier *et al.* 1992) implemented in Arlequin (Excoffier *et al.* 2005) (sequence data) and Genodive (Meirmans & Van Tienderen 2004) (frequency data) to determine the partitioning of genetic variation when individuals were grouped according to geographical location or mitochondrial lineage. The AMOVA s were performed on data sets consisting of only individuals sequenced for all six genes and on data sets consisting of all individuals for three genes (*cir*, *fil*, *mep*).

To compare genetic diversity between nuclear genes and between mitochondrial and nuclear genes, we calculated summary statistics in DnaSP. To test for selection in *COI*, we performed the McDonald-Kreitman (MK) (McDonald & Kreitman 1991) and Hudson-Kreitman-Aguadé (HKA) (Hudson *et al.* 1987) tests using *COI* sequences from *Niphates digitalis* (GenBank Accession nos. EF519655-8), in the sister genus to *Callyspongia* (Erpenbeck *et al.* 2007), as the out-group. Tests for selection that do not rely on allelic distributions, such as the MK and HKA tests, are robust to potentially confounding nonequilibrium processes, such as demographic change and recombination (Hudson *et al.* 1987; Sawyer & Hartl 1992; Nielsen 2001; Eyre-Walker 2002). We also calculated Tajima's *D* (Tajima 1989b), although cognizant that changes in population size can affect the outcome of this test (Tajima 1989a).

Hypothesis testing of population processes in a coalescent framework

In contrast to our expectation, nucDNA alleles did not cluster according to mtDNA lineage or sampling site: nucDNA was admixed with respect to both and had a different Tajima's *D* than mtDNA (see Results). To test (nonmutually exclusive) hypotheses for explaining mito-nuclear discordance in *C. vaginalis*, we used two approaches employing coalescent simulations in a statistical framework. First, we tested whether ILS could explain why mtDNA haplotypes cluster in divergent clades, while nucDNA remains admixed using a parametric bootstrapping approach. Second, we tested whether changes in population size could simultaneously produce significant and nonsignificant Tajima's *D* values in the mtDNA and nucDNA, respectively.

Coalescent simulations of lineage sorting in the nucDNA

We used the following procedure to test for ILS of nucDNA in *C. vaginalis* (see Fig. S3, Supporting information for an example). We simulated 1000 nuclear coalescent gene trees matching the demographic history of *COI* in *C. vaginalis* using ms (Hudson 2002). We then simulated sequence data with the same model of nucleotide evolution [estimated in jModelTest v2.0.2 (Posada 2008)] for each empirical nuclear gene on the 1000 coalescent trees using seq-gen (Rambaut & Grassly 1997). Gene trees were estimated for each simulated nuclear data set in PAUP* (Swofford 2003). We calculated the summary statistics theta, nucleotide diversity and Φ_{ST} for the simulated nuclear sequence data using the R package *pegas* (Paradis 2010) and calculated the number of deep coalescent events (DCEs) and Slatkin and Maddison's *s* for simulated nuclear gene trees compared with the mtDNA tree in Mesquite v2.74 (Maddison & Maddison 2010). We calculated the same suite of summary statistics for the empirical nuclear data and compared the empirical values to the distribution of each summary statistic calculated for the simulated data (i.e. the null distributions).

Because different statistics summarize data in different ways, we may expect variation in their performance when comparing null and simulated data in tests for ILS. For example, Φ_{ST} describes how variation is structured within and among populations. Slatkin and Maddison's *s* and the number of DCEs document the discordance between an individual's position in a gene tree and another feature of the data – in the case of *s*, the geographical location, and in the case of the DCE, the species/population tree. In contrast, theta and nucleotide diversity are determined by variation in the sequences themselves, not by how variation is partitioned in geographical or genealogical space. We expect that Φ_{ST}, DCEs and *s* will be better able to reveal ILS than nucleotide diversity and theta because their values should better trace the accumulation of sequence differences among populations (Φ_{ST}) and lineages (DCEs and *s*). We test this prediction here.

The estimation of parameters from the mtDNA used in ms (effective population sizes, divergence times) to simulate coalescent trees requires mitochondrial and nuclear substitution rates. Although slow substitution rates are well documented for sponge mtDNA, less is known about specific rates of nucleotide substitution for either mitochondrial or nuclear genes. Therefore, we identified two rates encompassing the upper and lower range of mitochondrial evolution for the best-studied and most closely related animal clade with slow mtDNA, the anthozoans [lower rate: 5.0×10^{-10} substitutions/site/year

(Hellberg 2006); upper rate: 1.0×10^{-9} substitutions/site/year (Romano & Palumbi 1996)], that likely contain the similarly slow mtDNA rate in sponges (Schröder *et al.* 2003; Duran *et al.* 2004a; Erpenbeck *et al.* 2006; Wörheide 2006). We used the lower rate as reported by Hellberg (2006) and increased the fast rate to 5.0×10^{-9} to make the difference between rates one order of magnitude. Although uncertainty in substitution rates leads to variance around parameter estimates, here it works to our advantage because we are less interested in knowing specific population parameters in *C. vaginalis* and more interested in creating a broad distribution of values that will encompass the true parameters. Using a wide range of values makes the test of ILS more conservative and our ability to reject the null hypothesis more robust.

To obtain substitution rates for each nuclear gene, we estimated their rates relative to the *C. vaginalis COI* gene in BEAST (Heled & Drummond 2010) and multiplied the relative rate by the fast and slow mitochondrial substitution rates. We estimated parameters for the demographic history of the *COI* gene (lineage divergence times and effective sizes: Fig. S3.1, Table S2, Supporting information), which are required for the simulations in the parametric bootstrap, from phylogenies constructed in BEAST and *BEAST v1.7.1 (Heled & Drummond 2010). For each run, two independent MCMC analyses were conducted for 10 million (BEAST) or 50 million (*BEAST) steps, sampling every 1000 steps. Convergence was determined by viewing the log files in Tracer v1.5. All parameters had effective sample sizes (ESS) >300. Treefiles were combined in LogCombiner v1.7.1 with a 10% burn-in, and the maximum clade credibility (MCC) tree for the combined file was calculated in TreeAnnotator v1.7.1. Specific details about the BEAST settings are available in the Supporting Information.

To determine which summary statistics were useful for detecting ILS, we calculated the statistics mentioned above for data sets simulated with various population divergence times, thus producing trees with varying levels of lineage sorting (Rosenberg 2003). We simulated five sets of 1000 coalescent trees with divergences of 0.1N, 0.5N, 1.0N, 5N and 10N generations in ms and simulated sequence data onto those trees in seqgen. We calculated nucleotide diversity, theta and Φ_{ST} in R and DCEs and *s* in Mesquite as described above.

Coalescent simulations of demographic change

Fay & Wu (1999) showed with coalescent simulations that the discordance in Tajima's *D* for humans (negative for the mtDNA and positive for the nucDNA) could be explained by a recent population expansion after a bottleneck. Because mtDNA has a smaller effective population size than nucDNA, it responds more quickly to population expansion after a contraction and will

Table 1 Genetic diversity indices for *Callyspongia vaginalis* for each gene by sampling location. *n*, sample size (alleles); *bp*, base pairs; *S*, segregating sites; *h*, haplotypes; *Hd*, haplotype diversity; π, nucleotide diversity. Bold Tajima's *D* values are significant

Location	*n*	*bp*	*S*	*h*	*Hd*	π	Tajima's *D*
mt*COI*							
Key Largo	17	511	18	3	0.669	0.0143	**2.1103**
Long Key	14	511	18	3	0.641	0.0137	1.6144
Marquesas Keys	14	511	18	3	0.659	0.0150	**2.1341**
Dry Tortugas	16	511	18	3	0.633	0.0133	1.6176
All Locations	61	511	18	3	0.656	0.0142	**3.3527**
cata							
Key Largo	30	346	13	10	0.770	0.0099	−0.0765
Long Key	26	346	13	8	0.776	0.0077	−0.7390
Marquesas Keys	20	346	17	10	0.758	0.0131	−0.1949
Dry Tortugas	30	346	13	9	0.851	0.0095	0.0175
All Locations	106	346	20	22	0.802	0.0103	−0.1982
cir							
Key Largo	34	159	7	6	0.729	0.0090	−0.0748
Long Key	28	159	7	6	0.635	0.0064	−0.9755
Marquesas Keys	28	159	7	6	0.730	0.0094	−0.5035
Dry Tortugas	32	159	10	9	0.833	0.0121	−0.4323
All Locations	122	159	11	10	0.742	0.0095	−0.4758
cps							
Key Largo	34	113	7	6	0.754	0.0261	**2.0813**
Long Key	26	113	7	7	0.763	0.0277	**2.1556**
Marquesas Keys	28	113	7	6	0.733	0.0231	−0.6803
Dry Tortugas	28	113	7	7	0.741	0.0255	1.8119
All Locations	116	113	8	10	0.758	0.0240	1.9038
ef							
Key Largo	28	161	3	6	0.726	0.0064	0.8728
Long Key	26	161	6	8	0.843	0.0110	0.3764
Marquesas Keys	26	161	7	6	0.794	0.0145	0.8309
Dry Tortugas	28	161	6	6	0.698	0.0071	−0.7663
All Locations	112	161	10	13	0.808	0.0108	−0.1935
fil							
Key Largo	34	127	12	14	0.884	0.0240	0.7571
Long Key	28	127	7	9	0.839	0.0256	**2.4217**
Marquesas Keys	28	127	10	11	0.749	0.0199	−0.0509
Dry Tortugas	32	127	10	8	0.792	0.0234	0.6120
All Locations	122	127	16	22	0.840	0.0251	0.5938
mep							
Key Largo	34	91	10	8	0.816	0.0351	0.9391
Long Key	28	91	10	8	0.812	0.0297	0.1677
Marquesas Keys	28	91	9	6	0.759	0.0274	0.2448
Dry Tortugas	32	91	9	7	0.786	0.0201	−0.5622
All Locations	122	91	10	11	0.835	0.0204	1.0192

achieve a negative Tajima's *D*, while the value for the nucDNA is still positive. In *C. vaginalis,* we found Tajima's *D* was positive for the mtDNA and not different from zero for the nucDNA (Table 1), consistent with the early stages of a bottleneck (Fig. 2 in Fay & Wu 1999). However, the difference in the effective population size of mtDNA and nucDNA is smaller in *C. vaginalis* than in humans (1:2 vs. 1:4), and the results of Fay & Wu (1999) may not apply to our system. We used coalescent simulations to test whether we could recover higher Tajima's *D* values in the mtDNA than nucDNA in the early stages of a bottleneck given the ratio of population effective sizes in *C. vaginalis.* We used ms to simulate population contractions of various reductions and timescales (Table 3). Initial population sizes were scaled to reflect the ratio of the effective population size of the mtDNA and nucDNA in *C. vaginalis* and in a typical gonochoric species such as humans. Each simulation was run for 5000 replicates. Sample_stats (Hudson 2002) was used to calculate Tajima's *D* for each replicate.

Results

Genetic diversity and distribution of genetic variation in nuclear loci

Nuclear loci were more diverse than *COI* in terms of haplotype number and nucleotide diversity (Table 1). *Cata* and *fil* had the highest number of alleles (22) and *fil* had the highest nucleotide diversity (0.025) Results for the MK ($G = 0.348$, $P = 0.555$) and HKA ($P = 0.944$) tests for selection were nonsignificant. Tajima's *D* values were significantly positive for *COI* for Key Largo and the Marquesas Keys and for all locations combined (Table 1). Tajima's *D* was nonsignificant for each sampling location individually and for all locations combined for all nuclear genes, with the exception of *cps* in Key Largo and Long Key and *fil* in Long Key (Table 1). Substitution rates for nuclear genes were faster relative to *COI* (*cata*, 2.3 times faster; *cir*, 1.6; *cps*, 2.7; *ef*, 3.0; *fil*, 7.1; *mep*, 3.6).

All methods used to determine how nucDNA variation was partitioned showed nuclear alleles corresponded to neither mitochondrial lineage nor geography. TCS networks for each nuclear gene joined all but one allele (*mep*) at the 95% confidence level, and most networks had equally parsimonious connections between alleles. The network for each nucDNA locus also showed that alleles did not cluster according to mitochondrial lineage or geography (Fig. 2). The Evanno *et al*. (2005) method concluded that the most likely number of clusters in the nucDNA was 2, but the clusters did not correspond to either mitochondrial lineage or geographical location (Fig. 3). Because the Evanno *et al*. (2005) method determines the most likely *K* based on the second-order rate of change in the likelihood function with respect to *K*, it is unable to find the best *K* if *K* is 1. In the plot for the mean likelihood of *K*, $K = 1$ had the best likelihood score and the likelihood for $K = 2$ through 8 declined from there (Fig. S1, Supporting information). This suggests that the true *K* for the nucDNA in Florida is 1. All hierarchical AMOVA s for the nucDNA indicated that most genetic variation (>90%) occurred within, not among groups,

Table 2 Results of hierarchical analyses of molecular variance (AMOVA) from Arlequin (sequence data). Groups were defined as mitochondrial lineages or geographical locations. Two data sets were analysed: A) only individuals sequenced for all six nuclear genes and B) all individuals sequenced for *cir*, *mep* and *fil*. Bolded values are significant

Source of variation	% variance	Φ statistic	*P* value
Data set: Arlequin_A	Groups: mtDNA lineages		
Among groups	−0.77	$\Phi_{CT} = -0.008$	0.736
Among populations within groups	7.11	$\mathbf{\Phi_{SC} = 0.071}$	**0.002**
Within populations	93.66	$\mathbf{\Phi_{ST} = 0.063}$	**< 0.001**
Data set: Arlequin_A	Groups: Geographical locations		
Among groups	5.47	$\mathbf{\Phi_{CT} = 0.055}$	**< 0.001**
Among populations within groups	1.94	$\Phi_{SC} = 0.021$	0.173
Within populations	92.59	$\mathbf{\Phi_{ST} = 0.074}$	**0.002**
Data set: Arlequin_B	Groups: mtDNA lineages		
Among groups	0.23	$\Phi_{CT} = 0.002$	0.468
Among populations within groups	8.29	$\mathbf{\Phi_{SC} = 0.083}$	**< 0.001**
Within populations	91.48	$\mathbf{\Phi_{ST} = 0.085}$	**< 0.001**
Data set: Arlequin_B	Groups: Geographical locations		
Among groups	3.64	$\Phi_{CT} = 0.036$	0.068
Among populations within groups	5.39	$\mathbf{\Phi_{SC} = 0.056}$	**0.004**
Within populations	90.97	$\mathbf{\Phi_{ST} = 0.090}$	**< 0.001**

Table 3 Parameters used in the coalescent simulations of population contraction and Tajima's *D* values for each set of parameters. Population sizes reflect the ratio of the mitochondrial to nuclear genome for *C. vaginalis* (1:2) and humans (1:4). N_1, initial population size; N_0, final population size; θ, theta [4*N_0*μ*locus length (bp)]; T, duration of bottleneck in generations; *D*, mean of Tajima's *D* (and range of *D*)

Simulation 1a–c:		
(a) mtDNA	(b) 1:2 nucDNA	(c) 1:4 nucDNA
N_1 = 10 000 000	N_1 = 20 000 000	N_1 = 40 000 000
N_0 = 1000	N_0 = 2000	N_0 = 4000
θ = 0.04	θ = 0.8	θ = 1.6
T = 1000	*T* = 1000	*T* = 1000
D = 2.02	*D* = 1.63	*D* = 1.13
(−3.02–4.91)	(−2.84–4.81)	(−2.26–4.42)
Simulation 2a–c:		
(a) mtDNA	(b) 1:2 nucDNA	(c) 1:4 nucDNA
N_1 =1 000 000	N_1 = 2 000 000	N_1 = 4 000 000
N_0 = 10 000	N_0 = 20 000	N_0 = 40 000
θ = 0.4	θ = 8	θ = 16
T = 1000	*T* = 1000	*T* = 1000
D = 0.527	*D* = 0.280	*D* = 0.103
(−2.31–3.98)	(−2.46–3.48)	(−2.40–3.40)
Simulation 3a–c:		
(a) mtDNA	(b) 1:2 nucDNA	(c) 1:4 nucDNA
N_1 = 1 000 000	N_1 = 2 000 000	N_1 = 4 000 000
N_0 = 200 000	N_0 = 400 000	N_0 = 800 000
θ = 8	θ = 160	θ = 320
T = 500	*T* = 500	*T* = 500
D = −0.088	*D* = −0.111	*D* = −0.109
(−2.48–3.06)	(−2.40–3.11)	(−2.43–3.12)
Simulation 4a–c:		
(a) mtDNA	(b) 1:2 nucDNA	(c) 1:4 nucDNA
N_1 = 1 000 000	N_1 = 2 000 000	N_1 = 4 000 000
N_0 = 200 000	N_0 = 400 000	N_0 = 800 000
θ = 8	θ = 160	θ = 320
T = 1000	*T* = 1000	*T* = 1000
D = 0.079	*D* = −0.041	*D* = −0.109
(−2.40–3.04)	(−2.38–3.49)	(−2.43–3.12)

regardless of whether groups were based on mitochondrial lineage or geography (Tables 2 and S4, Supporting information). The trend in the results was the same regardless of which data set (only individuals sequenced for all 6 nuclear genes or all individuals sequenced for *cir, fil and mep*) was tested or which analysis (Arlequin or Genodive) was conducted.

Distribution of simulated and empirical summary statistics

The distributions of summary statistics calculated from data sets simulated under the fast and slow substitution rates for all nuclear genes were in general agreement, although values for Φ_{ST} were smaller and more variable under the slow rate (Figs 4 and S4–8, Supporting information). The empirical value fell within the null distribution for theta and nucleotide diversity for each gene and fell outside the null distribution for Φ_{ST}, DCEs and *s*. Based on these results, we fail to reject the hypothesis of ILS based on nucleotide diversity and theta and reject ILS based on Φ_{ST}, DCEs and *s*.

The results of the simulations to evaluate summary statistics are shown in Fig. 5. The distribution of theta values for all population divergence times overlapped. There was also overlap in nucleotide diversity for data sets simulated with population divergences of 0.1N, 0.5N and 1N generations (Fig. 5). Nucleotide diversity for data sets with older divergences (5N and 10N generations) had less overlap and generally increased as divergence times increased. As expected, Φ_{ST} was lowest in the 0.1N divergence data set and highest in data sets with 5N and 10N population divergences. There was a wide range of Φ_{ST} values (−0.030–0.877) for 0.5N and 1N diverged data sets. DCEs and *s* decreased as divergence increased. There was overlap in DCEs and *s* values in 5N and 10N data sets, and the 0.1N divergence data set had a wide variation in DCEs and *s* values.

Coalescent simulations of demographic change

Tajima's *D* values for each replicate within a simulation varied widely, but the average Tajima's *D* for those replicates matched our expectations when comparing across population effective size categories (Table 3). For example, in simulations 1a-c, the mtDNA had the fastest response to the population contraction and therefore the highest Tajima's *D* (2.02; Table 3, simulation 1a), while the simulated nucDNA data sets had slower responses and lower positive Tajima's *D* values. The simulated nucDNA with four times the effective size of the mtDNA had the lowest positive Tajima's *D* (1.13; Table 3, simulation 1c), while the simulated nucDNA with double the mtDNA effective size had an intermediate Tajima's *D* (1.63; Table 3, simulation 1b).

Discussion

Our data show that patterns of genetic diversity and population subdivision in the mitochondrial and nuclear genomes in *C. vaginalis* are discordant, in contrast to previously published sponge studies where genetic data are available from both genomes (Duran *et al.* 2004a,b,c; Whalan *et al.* 2008; Dailianis *et al.* 2011; Voigt *et al.* 2012). Our analyses show that discordance in *C. vaginalis* is not due to ILS of nucDNA or selection on mtDNA but may be explained by population size change or sperm-mediated dispersal.

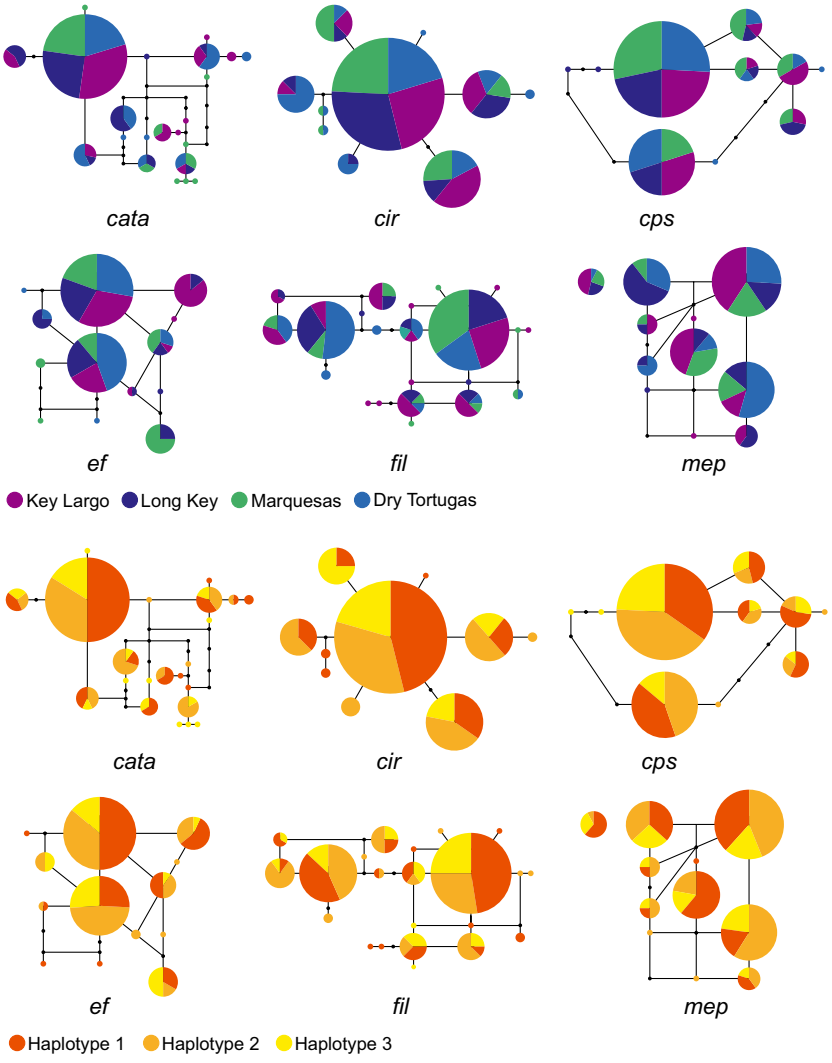

Fig. 2 Unrooted statistical parsimony networks for nuclear genes, with geographical locations (top) and mitochondrial lineage (bottom) indicated. Circles represent individual alleles with circle size proportional to frequency of occurrence. Lines connecting alleles represent one mutational step, and small black circles represent possible, but not sampled, alleles. The area of each sector or circle is proportional to haplotype frequency at that site.

Mito-nuclear discordance in Callyspongia vaginalis cannot be explained by incomplete lineage sorting or selection

We simulated 1000 data sets under each of 5 different population divergence times (0.1N, 0.5N, 1N, 5N and 10N generations). The range of nucleotide diversity values calculated for each group of 1000 data sets was the same regardless of the divergence time under which the data were simulated. The same pattern was also true for theta. Given this insensitivity, neither of these summary statistics provides resolution for testing the null hypothesis of ILS as the cause of mito-nuclear discordance.

In contrast, distributions of Φ_{ST}, DCEs and s across data sets with different divergence times were more distinct, suggesting they are superior in comparing simulated and empirical data sets in a hypothesis-testing framework. Based on these findings, we reject the hypothesis that ILS explains the failure of the nucDNA to correspond to the mtDNA lineages in *C. vaginalis*. Given the population size and split times estimated for the mtDNA, nucDNA loci had sufficient time to sort and their failure to cluster according to mtDNA lineage is not due to differences in effective size of the two genomes. The parametric bootstrapping approach we employed here generates the expected distribution of summary statistics for the nucDNA if discordance in the mtDNA and nucDNA was due differences in effective population sizes (and therefore sorting time). Had we employed only nucleotide diversity and/or theta, two widely used summary statistics, we would have erroneously failed to reject the hypothesis of ILS for our data. Our findings show that summary statistics that describe a genealogical or geographical pattern in the data (Φ_{ST}, DCE, s) are superior to those that describe sequence-level variation (nucleotide diversity, theta) when testing for ILS by comparing simulated and empirical data. We suggest researchers test a suite of

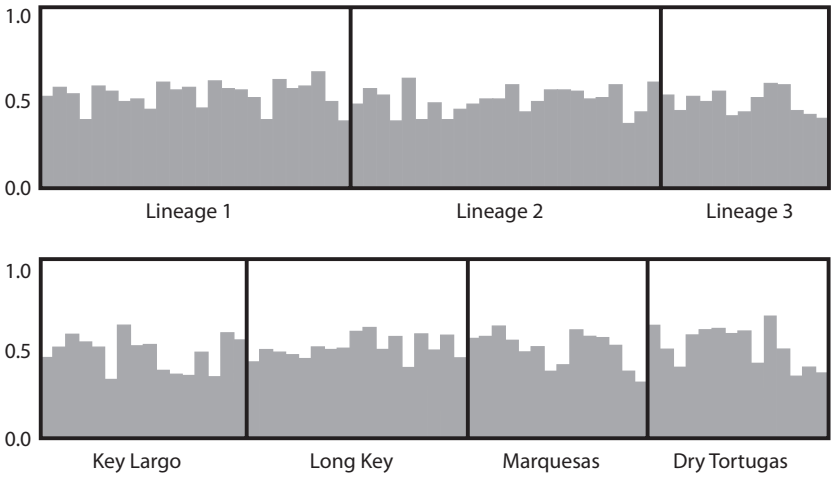

Fig. 3 Bayesian clustering indicated by STRUCTURE, with individuals grouped by mitochondrial haplotype (top) and geography (bottom). Each individual is represented by a vertical line, with the coloured portion of the line representing the proportion of that individual's genome originating from each inferred clusters.

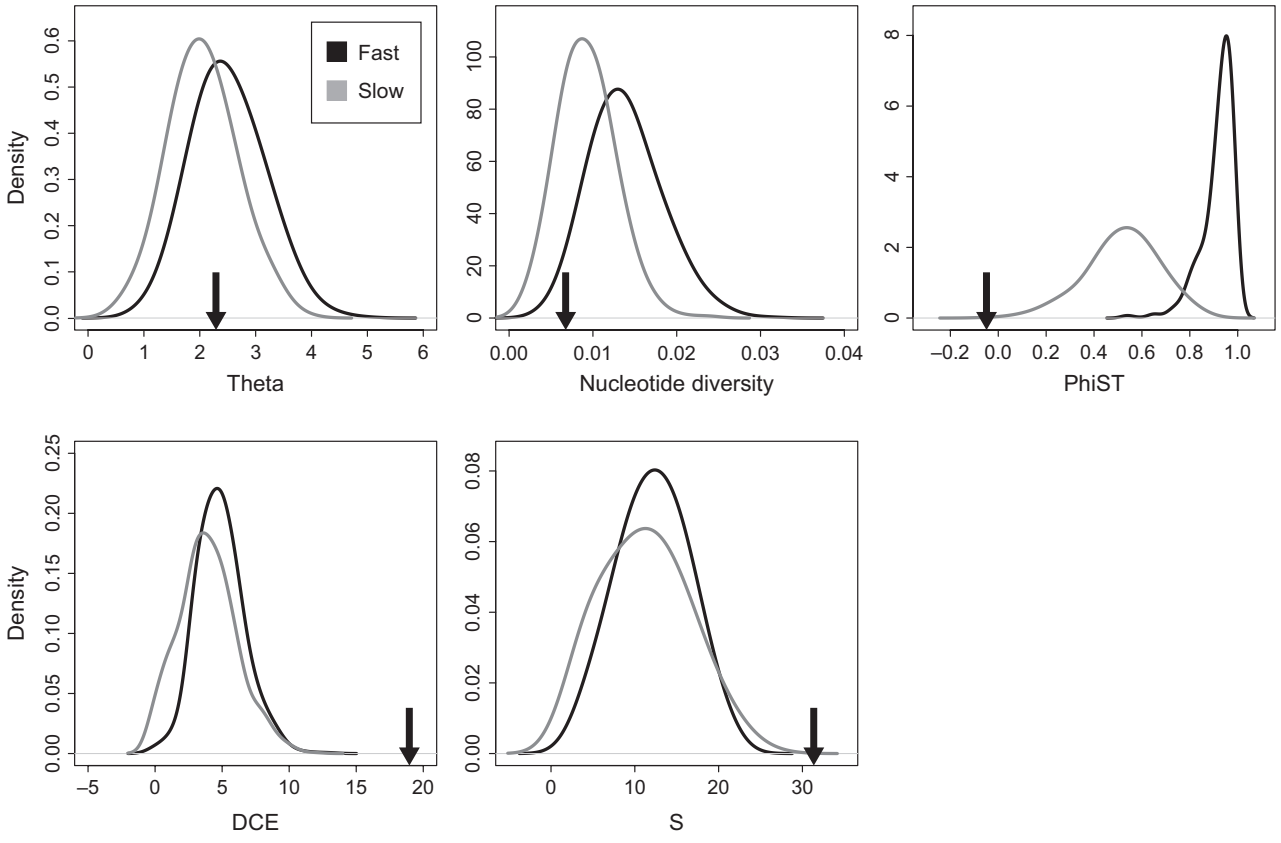

Fig. 4 Density plots showing distributions of summary statistics calculated for data sets simulated with population parameters matching the mtDNA demographic scenario and model of nucleotide substitution for *cata*. Black and grey curves represent values for data sets simulated assuming a fast and slow substitution rate, respectively. Arrows indicate the value calculated for the empirical data. Density plots for the other nuclear genes are in Figures S4–8.

metrics to be sure that they are using the summary statistics most informative for their data when hypothesis testing.

Selection can create discordant patterns of genetic diversity between the mtDNA and nucDNA. However, the results of the MK and HKA tests indicated that mtDNA in *C. vaginalis* is evolving according to neutral expectations. Selective sweeps on maternally inherited symbionts (i.e. *Wolbachia*, Hurst & Jiggins 2005) can also eliminate variation in the mtDNA with respect to the

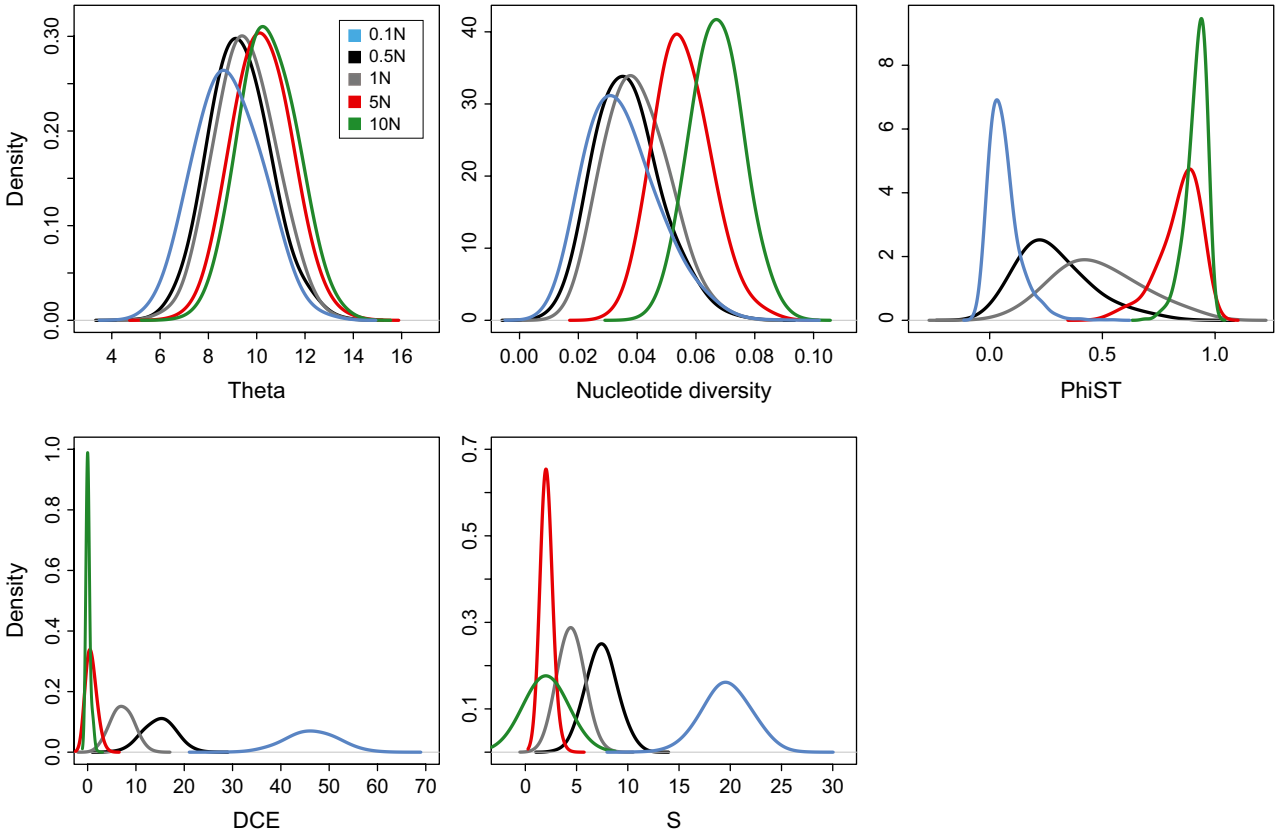

Fig. 5 Density plots showing the distribution of summary statistics calculated for data sets simulated under various population divergences: 0.1N generation population divergence, blue; 0.5N, black, 1N, grey; 5N, red; 10N, green.

nucDNA (Meiklejohn *et al.* 2007). Many sponges have complex (Schmitt *et al.* 2011), often vertically transmitted (Webster *et al.* 2009) microbiomes, but *C. vaginalis* harbours small amounts of bacterial cells and may acquire symbionts from the environment (Giles *et al.* 2012), reducing the likelihood that the low variation in the mtDNA is due to selection on a symbiont.

Discordant mito-nuclear patterns of geographical subdivision are consistent with sperm-mediated dispersal

Male-biased dispersal is often invoked to explain mito-nuclear discordance when mtDNA is more structured than nucDNA. The pattern of geographically structured mtDNA and unstructured nucDNA for the hermaphroditic *C. vaginalis* could result if dispersal by sperm (which transmit only nucDNA) is greater than that for eggs and larvae (which also transmit mtDNA).

Many coral reef taxa are broadcasters that release enormous quantities of gametes during synchronous mass spawning events (Harrison *et al.* 1984; Colin 1996). In contrast, many sponges, including *C. vaginalis*, employ spermcast mating where sperm released over an extended period disperse to find eggs retained for internal fertilization (Bishop & Pemberton 2006). The combined effects of finite gamete viability and dilution can reduce the fertilization success of broadcast spawners (Oliver & Babcock 1992; Levitan & Petersen 1995). However, gametes that remain viable for up to 72 h (Bishop 1998) and fertilization at low sperm concentrations have been documented in spermcasting marine invertebrates (Pemberton *et al.* 2003). Sponges, which continuously pump water at high rates (3.5 l/s/kg dry mass for *C. vaginalis*, Weisz *et al.* 2008), may avoid the effects of gamete dilution by filtering and concentrating sperm (Bishop 1998).

Sponge larvae are competent to settle almost immediately after release from the parent and dispersal is brief (Maldonado 2006) so sperm (i.e. nucDNA) dispersal potential in spermcasting sponges may be substantial relative to larval (i.e. mtDNA and nucDNA) dispersal. Although, to our knowledge, sperm dispersal has not been measured directly in sponges, and most fertilizations in spermcasters come from individuals close to the broodparent (Grosberg 1991), fertilization success at

distances of 207 metres was recorded in the field for the ascidian *Botryllus schlosseri* (Yund *et al.* 2007). Two studies examining multiple paternity in spermcasting marine invertebrates found that some offspring were fathered by individuals outside the study area (Yund 1998, 300 m^2; Lasker *et al.* 2008, 400 m^2). In two spermcasters, a gorgonian and a sponge, reproductive success was not correlated with proximity to mates, suggesting distance may not limit fertilization (Lasker *et al.* 2008, 400 m^2; Maritz *et al.* 2010, 50 m^2). In the light of our results, these examples indicate that sex-biased dispersal could contribute to mito-nuclear discordance in *C. vaginalis*.

Discordant mito-nuclear patterns of genetic diversity are consistent with a population bottleneck

Changes in population size can create conflicting patterns in mtDNA and nucDNA due to differences in the effective sizes of the two genomes; mtDNA responds faster to contractions and expansions than nucDNA (Fay & Wu 1999). Our simulations showed that Tajima's *D* can be higher for mtDNA than for nucDNA under a model of population contraction despite the 1:2 ratio of effective population size for mtDNA and nucDNA in *C. vaginalis* (Table 3). When we simulated a bottleneck of similar size and duration as one produced by a cyanobacterial bloom that killed 80% of surveyed sponges (including *C. vaginalis*) over 700 km^2 in the Florida Keys (Butler *et al.* 1995), the average Tajima's *D* was positive for the mtDNA and negative for the nucDNA (Table 3, simulations 4a–c). Field studies in Florida and the Caribbean show sponge populations can experience rapid, severe declines (Butler *et al.* 1995; Wulff 1995; Laboy-Nieves *et al.* 2001; Cowart *et al.* 2006), supporting the idea that demography plays a role in determining genetic diversity in the Porifera.

Genetic population structure in the Porifera

Of the few population-level studies in sponges employing nuclear sequences, most report minimal subdivision among populations, although on much larger spatial scales (along the Great Barrier Reef, Bentlage & Wörheide 2007; within the Mediterranean Sea, Dailianis *et al.* 2011) than the approximately 260 km considered in this study. However, other studies report significant subdivision at large spatial scales (from the Red Sea to central South Pacific Ocean, Wörheide *et al.* 2008) and at spatial scales similar to those here for *C. vaginalis* (tens to hundreds of kilometres, Blanquer & Uriz 2010). Phylogeographical studies in the Porifera have been hampered by the conservation of the sponge mitochondrial genome, but the high levels of sequence variation and substitution rates uncovered here suggest nucDNA sequence data hold promise for future population-level work.

Acknowledgements

We thank Vince Richards, Mahmood Shivji, and Ted and Christine Testerman for assistance with sample collection. Bryan Carstens, Joseph Neigel, Ron Eytan, Carlos Prada, Tara Pelletier, Sarah Hird and Reid Brennan provided helpful discussion about data analyses and/or comments on drafts of this manuscript. We also appreciate the constructive comments given by the anonymous reviewers. This research was supported by a Louisiana Board of Regents fellowship and Sigma Xi and LSU Biology Graduate Student Association grants to MBD, NSF grant OCE-0550270 to MEH and Iliana Baums, and funds from the LSU Department of Biological Sciences.

References

Bentlage B, Wörheide G (2007) Low genetic structuring among *Pericharax heteroraphis* (*Porifera: Calcarea*) populations from the Great Barrier Reef (Australia), revealed by analysis of nrDNA and nuclear intron sequences. *Coral Reefs*, **26**, 807–816.

Bergquist P(1978) *Sponges* Hutchinson and Company, London.

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics*, **35**, 125–148.

Bishop J (1998) Fertilization in the sea: are the hazards of broadcast spawning avoided when free-spawned sperm fertilize retained eggs? *Proceedings of the Royal Society of London Series B: Biological Sciences*, **265**, 725–731.

Bishop J, Pemberton A (2006) The third way: spermcast mating in sessile marine invertebrates. *Integrative and Comparative Biology*, **46**, 398–406.

Blanquer A, Uriz MJ (2010) Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evolutionary Biology*, **10**, 13.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, **76**, 1967–1971.

Butler I, Hunt J, Herrnkind W *et al.* (1995) Cascading disturbances in Florida Bay, USA: cyanobacteria blooms, sponge mortality, and implications for juvenile spiny lobsters *Panulirus argus*. *Marine Ecology Progress Series*, **129**, 119–125.

Cheviron ZA, Brumfield RT (2009) Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*, **63**, 1593–1605.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colin PL (1996) Longevity of some coral reef fish spawning aggregations. *Copeia*, **1996**, 189–192.

Cowart J, Henkel T, McMurray S, Pawlik J (2006) Sponge orange band (SOB): a pathogenic-like condition of the giant barrel sponge, *Xestospongia muta*. *Coral Reefs*, **25**, 513–513.

Dailianis T, Tsigenopoulos C, Dounas C, Voultsiadou E (2011) Genetic diversity of the imperilled bath sponge *Spongia officinalis* Linnaeus, 1759 across the Mediterranean Sea: patterns

of population differentiation and implications for taxonomy and conservation. *Molecular Ecology*, **20**, 3757–3772.

DeBiasse M, Richards V, Shivji M (2010) Genetic assessment of connectivity in the common reef sponge, *Callyspongia vaginalis* (*Demospongiae: Haplosclerida*) reveals high population structure along the Florida reef tract. *Coral Reefs*, **29**, 47–55.

Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ (2011) Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp. (*Meliaceae*). *Molecular Phylogenetics and Evolution*, **62**, 275–285.

Duran S, Giribet G, Turon X (2004a) Phylogeographical history of the sponge *Crambe crambe* (*Porifera, Poecilosclerida*): range expansion and recent invasion of the Macaronesian islands from the Mediterranean Sea. *Molecular Ecology*, **13**, 109–122.

Duran S, Pascual M, Estoup A, Turon X (2004b) Strong population structure in the marine sponge *Crambe crambe* (*Poecilosclerida)* as revealed by microsatellite markers. *Molecular Ecology*, **13**, 511–522.

Duran S, Pascual M, Turon X (2004c) Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (*Poecilosclerida*). *Marine Biology*, **144**, 31–35.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 1–3.

Erpenbeck D, Hooper J, Wörheide G (2006) CO1 phylogenies in diploblasts and the 'Barcoding of Life'- are we sequencing a suboptimal partition? *Molecular Ecology Notes*, **6**, 550–553.

Erpenbeck D, Duran S, Rützler K *et al.* (2007) Towards a DNA taxonomy of Caribbean Demosponges: a gene tree reconstructed from partial mitochondrial CO1 gene sequences supports previous rDNA phylogenies and provides a new perspective on the systematics of Demospongiae. *Journal of the Marine Biological Association of the United Kingdom*, **87**, 1563–1570.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47.

Eyre-Walker A (2002) Changing effective population size and the McDonald-Kreitman test. *Genetics*, **162**, 2017–2024.

Eytan RI, Hellberg ME (2010) Nuclear and mitochondrial sequence data reveal and conceal different demographic histories and population genetic processes in Caribbean reef fishes. *Evolution*, **64**, 3380–3397.

Eytan RI, Hayes M, Arbour-Reily P, Miller M, Hellberg ME (2009) Nuclear sequences reveal mid-range isolation of an imperilled deep-water coral population. *Molecular Ecology*, **18**, 2375–2389.

Fay JC, Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution*, **16**, 1003–1005.

Flot JF (2007) Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Notes*, **7**, 974–977.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.

Giles EC, Kamke J, Moitinho-Silva L *et al.* (2012) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*, **83**, 232–241.

Grosberg RK (1991) Sperm-mediated gene flow and the genetic structure of a population of the colonial ascidian *Botryllus schlosseri*. *Evolution*, **45**, 130–142.

Hale MC, McCormick CR, Jackson JR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics*, **10**, 203.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **16**, 700–706.

Harrison PL, Babcock RC, Bull GD *et al.* (1984) Mass spawning in tropical reef corals. *Science*, **223**, 1186–1189.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Hellberg M (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology*, **6**, 24.

Hickerson MJ, Cunningham CW (2005) Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution*, **59**, 344–360.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Humann P, DeLoach N(2003) *Reef Creature Identification*. New World Publications Inc., Jacksonville, Florida.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1525–1534.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, **56**, 2383–2394.

Laboy-Nieves EN, Klein E, Conde JE *et al.* (2001) Mass mortality of tropical marine communities in Morrocoy, Venezuela. *Bulletin of Marine Science*, **68**, 163–179.

Lasker HR, Gutierrez-Rodriguez C, Bala K, Hannes A, Bilewitch JP (2008) Male reproductive success during spawning events of the octocoral *Pseudopterogorgia elisabethae*. *Marine Ecology Progress Series*, **367**, 153–161.

Levitan DR, Petersen C (1995) Sperm limitation in the sea. *Trends in Ecology & Evolution*, **10**, 228–231.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

López-Legentil S, Pawlik J (2009) Genetic structure of the Caribbean giant barrel sponge *Xestospongia muta* using the I3-M11 partition of COI. *Coral Reefs*, **28**, 157–165.

Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis. Version 2.74. http://mesquite project.org.

Maldonado M (2006) The ecology of the sponge larva. *Canadian Journal of Zoology*, **84**, 175–194.

Maritz K, Calcino A, Fahey B, Degnan B, Degnan SM (2010) Remarkable consistency of larval release in the spermcast-mating demosponge *Amphimedon queenslandica* (Hooper and van Soest). *The Open Marine Biology Journal*, **4**, 57–64.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.

McKay BD, Zink RM (2010) The causes of mitochondrial DNA gene tree paraphyly in birds. *Molecular Phylogenetics and Evolution*, **54**, 647–650.

Meiklejohn CD, Montooth KL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genetics*, **23**, 259–263.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Milne I, Wright F, Rowe G *et al.* (2004) TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics*, **20**, 1806–1807.

Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641–647.

Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution*, **47**, 1329–1341.

Oliver J, Babcock R (1992) Aspects of the fertilization ecology of broadcast spawning corals: sperm dilution effects and *in situ* measurements of fertilization. *The Biological Bulletin*, **183**, 409–417.

Orozco-Terwengel P, Nagy ZT, Vieites DR, Vences M, Louis E Jr (2008) Phylogeography and phylogenetic relationships of Malagasy tree and ground boas. *Biological Journal of the Linnean Society*, **95**, 640–652.

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Pardini AT, Jones CS, Noble LR *et al.* (2001) Sex-biased dispersal of great white sharks. *Nature*, **412**, 139–140.

Peters JL, Bolender KA, Pearce JM (2012) Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. *Molecular Ecology*, **21**, 3562–3575.

Pemberton AJ, Hughes RN, Manríquez PH, Bishop JDD (2003) Efficient utilization of very dilute aquatic sperm: sperm competition may be more likely than sperm limitation when eggs are retained. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S223–S226.

Pond SLK, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, **21**, 2531–2533.

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics*, **22**, 3096–3098.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945.

Rambaut A, Grassly NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer applications in the biosciences: CABIOS*, **13**, 235–238.

Romano SL, Palumbi SR (1996) Evolution of scleractinian corals inferred from molecular systematics. *Science*, **271**, 640–641.

Rosenberg NA (2003) The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465–1477.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. *Genetics*, **132**, 1161–1176.

Schmitt S, Tsai P, Bell J *et al.* (2011) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal*, **6**, 564–576.

Schröder H, Efremova S, Itskovich V *et al.* (2003) Molecular phylogeny of the freshwater sponges in Lake Baikal. *Journal of Zoological Systematics and Evolutionary Research*, **41**, 80–86.

Shearer T, Van Oppen M, Romano S, Wörheide G (2002) Slow mitochondrial DNA sequence evolution in the *Anthozoa* (*Cnidaria*). *Molecular Ecology*, **11**, 2475–2487.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, **68**, 978–989.

Swofford D(2003) PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0 b10. Sinauer, Sunderland, MA.

Tajima F (1989a) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tajima F (1989b) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Voigt O, Eichmann V, Wörheide G (2012) First evaluation of mitochondrial DNA as a marker for phylogeographic studies of *Calcarea*: a case study from *Leucetta chagosensis*. *Hydrobiologia*, **687**, 101–106.

Webster NS, Taylor MW, Behnam F *et al.* (2009) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*, **12**, 2070–2082.

Weisz JB, Lindquist N, Martens CS (2008) Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia*, **155**, 367–376.

Whalan S, De Nys R, Smith-Keune C *et al.* (2008) Low genetic variability within and among populations of the brooding sponge *Rhopaloeides odorabile* on the central Great Barrier Reef. *Aquatic Biology*, **3**, 111–119.

Wilson AB, Eigenmann Veraguth I (2010) The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Molecular Ecology*, **19**, 4535–4553.

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloro-

plast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, **84**, 9054–9058.
Wörheide G (2006) Low variation in partial cytochrome oxidase subunit I (COI) mitochondrial sequences in the coralline demosponge *Astrosclera willeyana* across the Indo-Pacific. *Marine Biology*, **148**, 907–912.
Wörheide G, Epp LS, Macis L (2008) Deep genetic divergences among Indo-Pacific populations of the coral reef sponge *Leucetta chagosensis* (*Leucettidae*): founder effects, vicariance, or both? *BMC Evolutionary Biology*, **8**, 24.
Wulff J (1995) Effects of a hurricane on survival and orientation of large erect coral reef sponges. *Coral Reefs*, **14**, 55–61.
Yund PO (1998) The effect of sperm competition on male gain curves in a colonial marine invertebrate. *Ecology*, **79**, 328–339.
Yund PO, Murdock K, Johnson SL (2007) Spatial distribution of ascidian sperm: two-dimensional patterns and short vs. time-integrated assays. *Marine Ecology Progress Series*, **341**, 103–109.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

M.B.D. and M.E.H. designed research, B.J.N. provided custom R scripts, M.B.D. collected and analysed the data, and M.B.D. and M.E.H. wrote the article.

Data accessibility

Sequences for all genes are available in the European Nucleotide Archive under the Accession nos. HG000670 – HG001246 and HG737726 – 737847. R, ms, and seqgen scripts, simulated sequence data, empirical sequence alignments, and a table with ENA Accession nos. for all individuals is available from the Dryad Digital Repository: doi:10.5061/dryad.hn104.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Primer sequences and expected amplicon size for nuclear gene markers used in this study

Table S2 Demographic parameters estimated in BEAST and *BEAST from slow and fast mitochondrial substitution rates. N_e, effective size; HPD, highest posterior density

Table S3 Results of hierarchical analyses of molecular variance (AMOVA) from Genodive (genotype data). Groups were defined as mitochondrial lineages or geographic locations. Two datasets were analyzed: A) only individuals sequenced for all six nuclear genes and B) all individuals sequenced for *cir*, *mep*, and *fil*. Bolded values are significant

Fig. S1 Graph of the mean log likelihood score from 20 iterations for each value of K run in STRUCTURE.

Fig. S2 *COI* gene tree estimated in BEAST to confirm the relationships among the *C. vaginalis* mtDNA clades. Posterior probabilities are listed at the nodes of the tree.

Fig. S3 Representation of the steps in the parametric bootstrap analysis.

Fig. S4 Density plots showing the distribution of summary statistics calculated for datasets simulated with population parameters matching the mtDNA demographic scenario and a model of nucleotide substitution matching the *cirhin* gene.

Fig. S5 Density plots showing the distribution of summary statistics calculated for datasets simulated with population parameters matching the mtDNA demographic scenario and a model of nucleotide substitution matching the *cathepsin* gene.

Fig. S6 Density plots showing the distribution of summary statistics calculated for datasets simulated with population parameters matching the mtDNA demographic scenario and a model of nucleotide substitution matching the *elongation factor 1 alpha* gene.

Fig. S7 Density plots showing the distribution of summary statistics calculated for datasets simulated with population parameters matching the mtDNA demographic scenario and a model of nucleotide substitution matching the *filamin* gene.

Fig. S8 Density plots showing the distribution of summary statistics calculated for datasets simulated with population parameters matching the mtDNA demographic scenario and a model of nucleotide substitution matching the *macrophage expressed protein* gene.