ORIGINAL PAPER

Protistan Biogeography: A Snapshot Across a Major Shipping Corridor Spanning Two Oceans

Katrina M Pagenkopp Lohan^{a,b,1}, Robert C Fleischer^b, Mark E Torchin^c, and Gregory M Ruiz^a

Deciphering patterns of protistan taxa is a crucial step for understanding anthropogenic and environmental impacts on biogeography. We characterized and compared protistan communities from environmental samples collected along a major shipping corridor, the Panama Canal, and the Bocas del Toro archipelago. We used metabarcoding with high throughput sequencing (HTS) with the V4 hypervariable region of the ribosomal gene complex (rDNA). We detected many protistan taxa, including a variety of parasitic and toxic taxa. There were 1,296 OTUs shared across all three regions, with an additional 342–1,526 OTUs occurring across two or more regions, suggesting some mixing within the Caribbean and across the Isthmus. In general, this mixing did not impact community similarity, which was primarily distinct across regions. When OTUs identified as gregarines were analyzed separately, most samples still grouped by region and there was no overlap of communities on either side of the Canal. Shipping traffic through the Panama Canal could move some taxa across regions; however, different environmental conditions in the two oceans may limit their establishment. Overall our results suggest that contemporary protistan biogeographic patterns are likely caused by a complex combination of factors, including anthropogenic dispersal and environmental tolerance.

Key words: Panama Canal; protistan ecology; protistan communities; dispersal; metabarcode; protistan diversity.

49 **Introduction**

50

51 Protists are a major component of marine ecosystems. In addition to being extremely
52 abundant and occurring nearly everywhere in the world, they also perform important
53 functions as primary producers, decomposers, and parasites (Caron et al. 2012). Though
54 they often comprise large standing stocks of biomass, their small size and general
55 absence of distinguishing morphological features make them difficult to observe.
56 Consequently, we lack basic knowledge regarding taxonomy, systematics, and
57 biogeographic patterns for most protistan species (Caron et al. 2012). One major question
58 is whether microbes, including protists, exhibit biogeographic patterns based on regional
59 endemism or if these small organisms, which have huge population sizes and potentially
60 high dispersal, are globally cosmopolitan (Martiny et al. 2006). Current evidence
61 suggests that macroecological patterns in distribution can vary among protistan taxa, as
62 some occur over wide geographic regions while others appear to be geographically
63 restricted (Bass et al. 2007, Martiny et al. 2006). We compared marine protistan diversity
64 across the Isthmus of Panama, where the potential for mixing across the Panama Canal
65 may be high.

66 The Panama Canal has had a major impact on global shipping patterns by
67 reducing transit times, changing source and recipient ports, and altering global trade
68 routes (Ruiz et al. 2006, 2009). Ships that arrive at the Canal without cargo can take up
69 ballast water to facilitate passage through the narrower parts of the Canal (Chesher 1968;
70 Cohen 2006; Sheffey 1968). While some ships add ballast water in transit, others take up
71 ballast water at the Canal entrance and then discharge the extra ballast immediately upon

1
2
3
4 72 exiting the Canal (Cohen 2006; Sheffey 1968). Additionally, some ships refuel after
5
6
7 73 transiting the Canal, resulting in ballast water discharge when fuel is added (Cohen
8
9 74 2006). Historically, oil tankers (generally carrying oil from Alaska) too large to pass
10
11 75 through the Canal would meet smaller vessels on the Pacific side to transport the oil.
12
13
14 76 These smaller vessels discharged their ballast from the Caribbean entrance upon exiting
15
16 77 on the Pacific side (Carlton 1985; Glynn 1982). Thus, since at least the 1950s, a
17
18 78 substantial amount of water, known to contain living organisms, has been transported
19
20
21 79 from one entrance of the Canal to the other (Cohen 2006), creating the potential to
22
23
24 80 homogenize microbial communities across the two sides of the Isthmus.
25
26 81 In addition to their regional movements, ships from around the world transit the
27
28 82 Panama Canal and transfer live organisms in ballast water, on the hulls, or in sea chests
29
30
31 83 (Carlton 1985; Coutts and Dodgshun 2007; Davidson et al. 2009). Given the vast number
32
33
34 84 of ships that have transited the Canal over the past century, the cumulative number and
35
36 85 diversity of organisms delivered by vessel traffic to Panamanian waters is thought to be
37
38 86 high relative to other global regions (Ruiz et al. 2006). While some organisms can cross
39
40
41 87 the Canal unaided (e.g., passive drifting, aerial dispersal), the system of locks and
42
43 88 exposure to freshwater in the Canal is considered to prevent significant natural dispersal
44
45
46 89 of marine organisms (Cohen 2006; Rubinoff 1970; Schlöder et al. 2013). Thus, shipping
47
48 90 is probably a primary mechanism for movement across the Canal. Multiple previous
49
50
51 91 studies have reported finding protists in the sediments of ballast tanks (Drake et al. 2007;
52
53 92 Galil and Hülsmann 1997; Villac and Kaczmarek 2011) suggesting that this is a viable
54
55
56 93 transport mechanism. Currently, there is limited information on microbes transported and
57
58
59
60
61
62
63
64
65

introduced by shipping, indicating a fundamental gap in our knowledge of the invasion dynamics and potential of these taxa (Pagenkopp Lohan et al. 2016; Ruiz et al. 2000).

To begin evaluating patterns of microbial distributions and potential for anthropogenic homogenization, we compared protistan taxa and communities from multiple sites along both sides of the Panama Canal (Pacific-Canal and Caribbean-Canal) and the Bocas del Toro archipelago. Protists were collected in environmental samples (i.e., water and sediment) and were identified using a metabarcode approach with high throughput sequencing (HTS) to obtain a large number of sequences per sample, thus identifying as many protists as possible for comparison across regions. Given the potentially large volume of water that has subsequently been exchanged among study regions, we predicted that a large proportion of protists would be shared across the Isthmus of Panama and that this would be similar to or greater than the number of protists shared across the Caribbean regions. We also focused particular attention on examining the diversity and distribution of gregarines, since they were abundant in the samples and were expected to be geographically restricted, given their host specificity and limited dispersal.

Results

We collected 35 environmental samples from 19 locations across three regions within Panamanian waters (Supplementary Material Table A1). After all quality filtering, we obtained a total of 1,588,807 sequences (1,742–123,920 sequences per sample) that generated 9,048 OTUs (285–1,793 per sample) (Supplementary Material Table A1).

1
2
3
4 117 Within the OTUs generated, 6.2% (n = 563) could not be assigned to any taxonomic
5
6 118 group and 20.4% (n = 1,847) could only be assigned to Eukaryota (Table 1). In the water
7
8
9 119 and sediment datasets, 19% and 21% of the OTUs could only be assigned to Eukaryota
10
11 120 (Table 1). Overall a large fraction of OTUs (21.1%, n = 1,906) was assigned to the
12
13 121 Alveolata, with 45% and 35% of those OTUs identified as members of the Dinophyta and
14
15 122 Apicomplexa, respectively (Table 1). Another large fraction of OTUs (20.8%, n = 1,879)
16
17 123 was assigned to the Stramenopiles (Table 1). These trends were also visible in the water
18
19 124 and sediment datasets, regardless of whether we used rarified data or not (Table 1). A
20
21 125 number of parasitic taxa were identified in the environmental samples. These parasitic
22
23 126 taxa include Apicomplexa (*Rhytidocystis* spp., *Coccidia*, *Cryptosporidium* spp.,
24
25 127 Gregarines, novel Apicomplexa Class 1 and 2), Syndiniales (Dino Group 1 and 2,
26
27 128 *Syndinium* sp.), Perkinsea (*Perkinsus* spp., *Parvilucifera* spp.), Fungi (*Candida* spp.,
28
29 129 *Chytridium* spp.), Ichthyosporea, and Labyrinthulea.

30
31
32
33
34
35
36 130 Examining relative abundance within individual OTUs revealed that 40 and 12
37
38 131 OTUs in the water and sediment datasets, respectively, had high relative abundance,
39
40 132 containing >5,000 sequences (Table 2). In the water dataset, these OTUs ranged from
41
42 133 5,015 (OTU 256, Ostreopsidaceae) to 50,680 (OTU 56; Dinophyceae) sequences per
43
44 134 OTU. In the sediment dataset, these OTUs ranged from 5,180 (OTU 189,
45
46 135 Cephaloidophoroidea) to 39,231 (OTU33, Navicula) sequences per OTU (Table 2). Four
47
48 136 OTUs (OTU 31, 33, 54, 37) were found in high abundance in both datasets. Eleven OTUs
49
50 137 are identified as parasitic taxa (Gregarines and Syndiniales). More specifically, across all
51
52 138 35 samples in both datasets the gregarines accounted for 495 OTUs (37 - 120 per sample)
53
54 139 and 303,755 sequences (191-40,550 per sample), while the syndinids accounted for 315
55
56
57
58
59
60
61
62
63
64
65

OTUs (9 – 144 per sample) and 72,877 sequences (26 – 19,669 per sample). Another six OTUs were identified as organisms that can be toxic or cause harmful algal blooms (i.e., *Heterocapsa* sp., *Skeletonema* spp., *Thalassiosira* spp.; Table 2).

The distribution of high relatively abundant OTUs exhibited considerable variation across sampling locations from the two Canal regions. Some of these OTUs were found at the same relative abundance across locations within the same region (e.g., OTU31 in the water dataset), but most varied greatly in their relative abundance across locations (Table 3). Three OTUs (i.e., OTU56, OTU13, OTU54) were found in high relative abundance in the same habitat across the regions (Table 3). Two OTUs (i.e., OTU31, OTU37) were found in high relative abundance in both the water and sediment datasets, with similar relative abundances in the same locations (Table 3).

Of the 9,048 OTUs determined in this study, 14.3% (n=1,296) were found in all three regions (Fig 1A). In the Canal-Caribbean, 2,661 OTUs were unique and 3,286 OTUs were shared among two or more regions (Fig. 1A). In the Canal-Pacific, 1,948 OTUs were unique and 3,164 were shared among two or more regions (Fig. 1A). In Bocas del Toro, 811 OTUs were unique and 2,102 were shared among two or more regions (Fig. 1A). Of the 7,689 OTUs in the water dataset, 14.6% (n = 1,122) were found in all three regions (Fig. 1B). In this dataset, the Canal-Caribbean and Pacific regions shared the most OTUs (n=2,111), while the Canal-Caribbean and Bocas del Toro shared the least OTUs (n=1,520). Of the 4,848 OTUs in the sediment dataset, 20.8% (n=1,010) were shared across the Canal regions (Fig. 1C). These trends were similar in the rarified datasets as well. In the rarified water dataset, 12.1% (n=837) of the 6,876 OTUs were found in all three regions (Supplementary Material Fig. A1A). Again, the Canal-

1
2
3
4 163 Caribbean and Canal-Pacific regions shared the greatest number of OTUs (n=1,590). In
5
6 164 the rarified sediment dataset, 14.2% (n=483) were found in both regions (Supplementary
7
8
9 165 Material Fig. A1B).

10
11
12 166 We determined the core OTUs by first determining what OTUs were present in
13
14 167 100% of samples within a single region, then determining whether those OTUs were
15
16 168 unique (i.e., only found in 100% of samples within a single region) or shared (i.e., found
17
18 169 in 100% of samples from multiple regions). Overall, 14 core OTUs occurred in 100% of
19
20
21 170 samples across all three regions, which increases to 28 core OTUs when examining
22
23 171 across two or more regions (Fig. 2). Within regions, Bocas del Toro had the greatest
24
25 172 number of unique core OTUs (n= 91), while the number of core OTUs unique to the
26
27 173 Pacific-Canal and Caribbean-Canal regions was much lower (Fig. 2A). For water
28
29 174 samples, 29 core OTUs were shared across all three regions (Fig. 2B), while an additional
30
31 175 9–18 were shared across only two regions. For sediment samples, 17 core OTUs were
32
33 176 shared across the Canal (Fig. 2C). A number of taxa that are potentially toxic, HAB
34
35 177 forming, or parasitic were identified as core OTUs including *Heterocapsa* spp.,
36
37 178 *Thalassiosira* spp., *Trichotokara* spp., Syndiniales, *Skeletonema* spp., *Pseudo-nitzschia*
38
39 179 sp., Labyrinthulea, and Gregarine. Also, many of the core OTUs were also found in
40
41 180 relatively high abundance in the water and sediment datasets (Table 2).
42
43
44
45
46
47
48 181 The OTU richness in the entire dataset was high as even the extrapolated
49
50 182 rarefaction curve did not show an asymptote, estimating >10,000 OTUs would be found
51
52 183 in 65 samples (Fig. 3). For the water dataset, the extrapolated rarefaction curve closely
53
54 184 mirrored that of the entire dataset, estimating that >9,000 OTUs would be recovered from
55
56 185 >50 water samples, while the extrapolated rarefaction curve for the sediment dataset
57
58
59
60
61
62
63
64
65

1
2
3
4 186 estimated that >6,000 OTUs would be recovered from >40 samples, which appeared to be
5
6 187 the point at which this curve began to asymptote (Fig. 3). On the other hand, the
7
8
9 188 gregarine dataset did asymptote, showing that all estimated species (n=470–515) are
10
11 189 likely included in ~30-40 samples (Fig. 3). As most of these extrapolated rarefaction
12
13
14 190 curves do not reach an asymptote, it is clear that the sampling effort for this study was
15
16 191 insufficient to characterize the total OTU richness in these regions and provide only an
17
18
19 192 estimate of OTU overlap among regions.

21 193 **Beta Diversity**

23
24 194 The beta diversity analyses for the water and sediment datasets showed the protist
25
26 195 communities within each region were generally distinct, but with interesting exceptions.
27
28 196 In the UPGMA generated from the water dataset, there are three major clades (Fig. 4A)
29
30
31 197 with the first clade containing three samples from the Caribbean and all the samples from
32
33 198 the Pacific were in the second clade. This second clade also contained one sample from
34
35
36 199 the Caribbean (ACPW), which was taken from inside the Caribbean entrance of the
37
38 200 Canal. The third clade contained two samples from near the Caribbean entrance to the
39
40
41 201 Canal and all five samples from Bocas del Toro. The UPGMA for the sediment dataset
42
43 202 had two major clades (Fig. 4B), one containing only samples from the Pacific and the
44
45
46 203 other containing only samples from the Caribbean. The results from the NMDS plots
47
48 204 show similar distributions (Supplementary Material Fig. A2). The Mantel test showed
49
50
51 205 that the communities in the water samples were significantly correlated with distance
52
53 206 between sampling locations (Mantel r statistic = 0.33, p = 0.002, two-tailed), as were the
54
55
56 207 communities in the sediment samples (Mantel r statistic = 0.83, p = 0.002, two-tailed).

For the gregarine dataset, the UPGMA showed three major clades (Fig. 5). The first clade consisted primarily of samples from the Caribbean, with a single sample from Bocas del Toro. The second clade contained the remaining samples from Bocas del Toro with one sample from the Caribbean. The third clade contained only samples from the Pacific. The NMDS plot showed similar results (Supplementary Material Fig. A2).

Discussion

In both habitat types, we found multiple OTUs that occurred in all regions, including across both sides of the Panama Canal, including several parasitic, toxic, and HAB forming taxa (primarily in sediments). Although many OTUs were shared across regions, the community composition between regions appeared to be generally distinct, with an interesting exception (APCW), possibly indicating the influence of mixing across the Canal. For the gregarines, we found higher than anticipated levels of dispersal across the Caribbean regions and no major overlap in communities across the isthmus. Our results indicate that the high shipping volume across the Canal may have some influence in dispersing taxa across the Isthmus, but this dispersal has a limited impact on the measured similarity of the protistan communities in both oceans. Thus, other factors, particularly the environmental differences (discussed further below), are likely responsible for the observed differences in communities in the two oceans.

Given the breadth of protistan diversity and potential spatio-temporal variation in abundances, our sampling was insufficient to capture the total diversity in the sampled regions. Rather, we provide a snapshot of protistan diversity across these regions.

1
2
3
4 231 Rarefaction analyses indicated that additional sampling would be needed to fully assess
5
6 232 total richness. Further, a comparison of analyses of relatively abundant OTUs and core
7
8
9 233 OTUs (Tables 2 and 3) indicate that our sampling probably excluded rare taxa.
10
11 234 Nonetheless, with our sampling was targeted within a few regions, focusing on the
12
13
14 235 entrances of the Panama Canal, we were able to determine that many OTUs are shared
15
16 236 across regions, including across the Canal. These data represent a minimum estimate of
17
18
19 237 OTU overlap and suggest mixing across the oceans. The differences in the protistan
20
21 238 community similarity across the Canal indicate that other factors, such as environmental
22
23
24 239 influences, also play a strong role in shaping these communities.

25
26 240 While there is still some debate in the literature regarding the ubiquity of
27
28 241 microbial species, including protists, many studies have found evidence for regional
29
30
31 242 restrictions in some protistan taxa on both broad and narrow spatial scales. For example,
32
33 243 Bass et al. (2007) used the fast-evolving ITS1 region to show that the distributions of the
34
35
36 244 observed ITS-types of cercozoans were mixed, with some being widely spread across
37
38 245 continents and habitats while others were patchy or regionally restricted. Additional
39
40
41 246 studies have shown similar patterns in microbial communities between latitudes (Baldwin
42
43 247 et al. 2005) and depths (Bik et al. 2011). In a comprehensive statistical analysis of
44
45
46 248 published data, Hillebrand et al. (2001) showed that for both diatoms and ciliates,
47
48 249 similarity in species composition decreased with geographic distance, suggesting
49
50 250 dispersal limitation at larger spatial scales. This pattern was much stronger for metazoans
51
52
53 251 than for unicellular organisms, indicating that unicellular organisms may disperse farther
54
55 252 than multicellular ones and that macroecological patterns likely differ between the two
56
57
58 253 groups. The results of our study are consistent with the results of these previous studies

(Baldwin et al. 2005, Bik et al. 2011, Hillebrand et al. 2001) showing that some OTUs are widespread across locations and habitats, but many are restricted at these spatial scales.

Protistan Diversity in Panamanian Waters

While the role of shipping in transporting macroorganisms around the globe is well-known, comparatively less research has focused on the potential for shipping to spread microorganisms (Pagenkopp Lohan et al. 2016; Ruiz et al. 2000). Even in areas with high volumes of shipping traffic (e.g., Panama), the information about species introductions of macroorganisms from other countries are often circumstantial (Cohen 2006). There is considerably less information about the diversity and distribution of protists (or other unicellular organisms) associated with the Panama Canal. We recovered many OTUs identified as parasitic taxa, including alveolates, with the majority of those identified as dinoflagellates and apicomplexans. Many of the OTUs had high relative abundances (based on number of obtained sequences), which likely corresponds to high abundance or biomass in the environment (Lindeque et al. 2013; Pawlowski et al. 2014). However, multiple of these abundant OTUs were identified as either alveolates or Stramenopiles and relative abundance estimates could be overestimated for some these taxa due to their massive rDNA copy numbers (Godhe et al. 2008; Gong et al. 2013). These results demonstrate many protistan taxa, including free-living and parasitic taxa, exist in the waters surrounding the Panama Canal.

Protistan Distribution in Panamanian Waters

We predicted that the high shipping volume would effectively disperse microbial species across the Panama Canal. We found multiple OTUs were widely shared across regions of

1
2
3
4 277 Panama. Of the 9,048 OTUs that were identified, 14% were shared across all three
5
6 278 regions, with 14 OTUs occurring in every sample from all three regions. Additionally,
7
8
9 279 40% of all OTUs were shared across two or more regions, with 28 OTUs occurring in
10
11 280 every sample from two or more regions. Within regions, multiple OTUs were found
12
13
14 281 across all samples in the Pacific and Caribbean (Fig. 1).
15

16 282 When examining community similarity, the protistan communities were generally
17
18
19 283 distinct between regions, with a single exception in the water dataset. A sample obtained
20
21 284 from just within Cristobal Bay (APCW) on the Caribbean side of the Canal was the only
22
23
24 285 sample from the Caribbean that grouped with samples from the Pacific, indicating that
25
26 286 the protistan community in this sample was more similar to communities in the Pacific
27
28
29 287 than to the other communities in the Caribbean. This could be due to high traffic of
30
31 288 empty ships (those most likely to take on ballast), which travel from the Pacific to the
32
33
34 289 Atlantic and generally de-ballast in Cristobal Bay (Chesher 1968). Despite the frequency
35
36 290 of transiting vessels, the protistan communities in the water and sediment samples show
37
38
39 291 distinct separation by region, particularly across the Canal. In fact, protistan communities
40
41 292 appear more similar between the two regions along the Caribbean coast, as demonstrated
42
43 293 by two Caribbean-Canal samples (RAPA2 and SBPA3) grouping with samples from
44
45
46 294 Bocas del Toro, than across the two oceans. These results show that though many
47
48
49 295 microbial taxa occur across regions, there are still major differences in community
50
51 296 composition between regions. Thus, either the Panama Canal constrains dispersal for
52
53 297 many microorganisms, or environmental differences between the Caribbean and Pacific
54
55 298 coasts of Panama limit their establishment or abundance.
56
57
58
59
60
61
62
63
64
65

1
2
3
4 299 Most gregarines parasitize sessile invertebrates and are host-specialists (Leander
5
6 300 2008). Our results showed that these parasites showed little similarity in community
7
8
9 301 composition across the two oceans, suggesting that either dispersal is limited or suitable
10
11 302 hosts or conditions do not exist across regions. However, mixing appears to occur along
12
13
14 303 the Caribbean coast for this group. Further research is needed to identify the parasite
15
16 304 species represented by the OTUs recovered along with other important factors of the
17
18
19 305 ecology of these species, such as their host preferences and specificity, dispersal
20
21 306 potential, and longevity and environmental tolerances of the transmissive stage.
22
23
24 307 The environmental differences between the Caribbean and Pacific coasts of
25
26 308 Panama are considerable, resulting in differences in the biotic compositions across the
27
28
29 309 Isthmus (Leigh et al. 2013). On the Pacific side, the tidal, salinity, and temperature ranges
30
31 310 are greater than in the Caribbean, primarily due to the seasonal upwelling that occurs
32
33
34 311 only on the Pacific coast (Cohen 2006). This upwelling also results in seasonal variation
35
36 312 in key nutrients, such as nitrogen and phosphate, which can reach substantially higher
37
38
39 313 levels in the Pacific than the Caribbean (Cohen 2006). The weather on the Caribbean side
40
41 314 is generally windier and stormier, causing decreased sunlight and increased rainfall
42
43
44 315 seasonally (Cohen 2006). These environmental differences have led to major differences
45
46 316 in the dominant communities across the Isthmus, with the Pacific dominated by higher
47
48
49 317 surface primary production from planktonic organisms and the sea bottom dominated by
50
51 318 suspension and detritus feeders, while the Caribbean is dominated by seagrass beds and
52
53 319 corals with most of the primary production occurring in the sediments (Leigh et al. 2013).

320 **Conclusions**

1
2
3
4 321 We found a substantial number of OTUs that were shared across multiple regions of
5
6 322 Panama, including many that were shared across the Isthmus of Panama. A number of
7
8
9 323 bloom-forming, toxic, and parasitic protists were identified, with some being shared
10
11 324 across regions including near the entrances of the Panama Canal. However, while we
12
13
14 325 found many of the same taxa occurred across all regions, including across the Canal,
15
16 326 there were substantial differences in community similarity across regions in both water
17
18 327 and sediment. Further, gregarines showed little similarity across the two oceans. While
19
20
21 328 our data show that there are some shared taxa, further research with broader spatial and
22
23 329 temporal sampling is needed to understand the endemism of microscopic organisms and
24
25
26 330 for tracking potential microbial invasions.
27
28
29 331

30 31 332 **Methods** 32 33 333

34
35
36 334 **Sample collection:** Surface water, intertidal sediment, and shallow subtidal sediment
37
38 335 were collected from locations around both sides of the Panama Canal and Bocas del Toro
39
40 336 archipelago (Supplementary Material Table A1). To attempt to control for differences
41
42 337 between habitat types, samples were primarily collected from similar habitats (i.e.,
43
44 338 shallow, protected, low-energy waters) on both coasts. Surface water was collected using
45
46 339 sterile 1L bottles, which were submerged until full. Sediment samples were collected
47
48 340 using sterile or bleach-washed 10 or 20cc syringes, which were modified via removal of
49
50 341 the tip of the syringe. For water samples, 500mL of water was filtered through a 3µM
51
52 342 nucleopore Whatman filter (VWR International, Radnor, PA) placed on a single-use
53
54 343 sterile filter apparatus (Fisher Scientific, Inc., Pittsburgh, PA) using a vacuum pump.
55
56
57
58
59
60
61
62
63
64
65

344 Sediment was transferred from the syringe to a 1.5mL microcentrifuge tube using sterile
 345 or bleach-washed forceps. Filters and sediment were stored at or below -20 °C prior to
 346 DNA extraction.

DNA extraction, PCR amplification, and library preparation: Total genomic
 DNA was extracted from filters and sediment using a PowerWater DNA Isolation Kit
 (MoBio Laboratories, Inc., Carlsbad, CA) and a PowerSoil DNA Isolation Kit (MoBio
 Laboratories, Inc.), respectively. A negative control containing no sample was included
 with every set of extractions. These negative controls were treated exactly the same as the
 samples through the entire process, including amplification and sequencing, in order to
 account for the sequencing of aerosolized contaminants in the laboratory. In an effort to
 reduce the impact of primer bias, three primers were utilized in a multiplex PCR and all
 amplifications were carried out in triplicate. We used Nexterra Illumina adaptors with a
 two-step PCR protocol for tailed Illumina amplicon sequencing, which includes adding a
 20bp addition to the forward and reverse primers, which are then used as the priming
 sites in a second indexing PCR (Illumina Support 2013). The primers used in this study
 were 3NDf, V4_euk_R1, and V4_euk_R2 (Bråte et al. 2010), which amplify ~400bp of
 the hypervariable region of the V4 domain of the small subunit (SSU) of the ribosomal
 RNA (rRNA) gene complex and is a universal eukaryotic primer set known to amplify a
 variety of protists (Bråte et al. 2010; Pagenkopp Lohan et al. 2016). The PCR reactions
 consisted of the following final reagent concentrations: 1X PCR Buffer (150mM Tris-
 HCl, pH 8.0, 500mM KCl; Applied Biosystems), 1.5mM MgCl₂, 0.2mM each dNTP,
 0.4mg/mL bovine serum albumin (BSA; New England Biolabs), 0.5µM of forward and
 reverse primers, 0.03units/µL AmpliTaq Gold polymerase (Applied Biosystems), and

367 5μL of DNA with water added to a final reaction volume of 30μL. Thermocycling
 368 conditions were as follows: initial denaturation at 94 °C for 10 minutes, 30 cycles of 94
 369 °C for 30 seconds, 52 °C for 45 seconds, 72 °C for 45 seconds, and a final extension of
 370 five minutes at 72 °C. An aliquot of the PCR product (5μl) was electrophoresed on an
 371 agarose gel (2% w/v), stained with GelRed (Phenix Research, Candler, NC), and
 372 visualized under UV light.

Library construction and sequencing: PCR triplicates for each sample and
 374 negative controls were pooled based on the intensity of bands observed following gel
 375 electrophoresis. No product was visible via electrophoresis for any of the negative control
 376 replicates. To remove excess primers, 10μL of each pooled amplicon was exo-sapped
 377 using 1μL of 1:10 diluted EXO-SAP IT (Affymetrix, Santa Clara, CA) under the
 378 following thermocycling conditions: 40 minutes at 37 °C then 15 minutes at 80 °C.

379 Pooled, cleaned amplicons were dual-indexed with Illumina Nexterra adaptors using the
 380 following final reagent concentrations: 1X of the KAPA HiFi HotStart Ready Mix (Kapa
 381 Biosystems, Wilmington, MA), 0.4 μM of each adaptor, 1μL of pooled, cleaned
 382 amplicons, and water to a final volume of 25μL. Thermocycling conditions were as
 383 follows: initial denaturation at 95 °C for 5 minutes, 4 cycles of 98 °C for 20 seconds, 60
 384 °C for 45 seconds, 72 °C for 45 seconds, and a final extension of five minutes at 72 °C.

385 To confirm that the indexing was successful, aliquots of the both indexed and unindexed
 386 PCR product (5μl) were electrophoresed on an agarose gel (2% w/v), stained with
 387 GelRed (Phenix Research), and visualized under UV light. Excess primer dimer and
 388 nucleotides were removed from pooled amplifications using Sera-Mag SpeedBeads

(ThermoScientific, Inc.), followed by a wash in 70% ethanol, and then eluted in 20µL of TE (10mM Tris-Cl, 0.5mM EDTA; pH 9.0) buffer.

To determine the size range and confirm the removal of small fragments, a subset of samples was run on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) following the manufacturer's instructions for the High Sensitivity DNA Chip. To quantify the libraries for pooling, quantitative PCR was performed on a Stratagene Mx3000P (Agilent Technologies) using the KAPA Library Quantification Kit for Illumina (KAPA Biosystems) following the manufacturer's instructions for 10µL reactions. The samples were run in duplicates at four dilutions (1:1000, 1:2000, 1:4000, 1:8000). The purified reactions were pooled in equal copy numbers. Sequencing was performed on an Illumina MiSeq using the Illumina 2x300bp sequencing kit at the Molecular Biology Core Facilities at the Dana Farber Cancer Institute. Raw sequence files can be found in GenBank under NCBI BioProject PRJNA355887.

Bioinformatics: Sequence quality was initially assessed using FastQC v0.11.2 (www.bioinformatics.babraham.ac.uk/projects/). Using UPARSE v8.0 (Edgar 2013), paired end reads were merged and quality filtered using a maximum expected error rate of 0.5. To facilitate downstream processing, the program Quantitative Insights Into Microbial Ecology (QIIME) v1.8 (Caporaso et al. 2010) was used to add labels to the sequences. All merged reads were then concatenated into a single file. Primer sequences were removed using the online version of PRINSEQ (<http://prinseq.sourceforge.net>, Schmieder and Edwards 2011).

Using UPARSE v8.0, sequences were globally trimmed to 350bp, dereplicated, sorted by abundance, and all singletons were removed. Clustering of the operational

1
2
3
4 412 taxonomic units (OTUs) was conducted at the default setting of 97% similarity. Though
5
6 413 this step includes the removal of chimeric sequences, we did an additional reference-
7
8
9 414 based chimera detection using UCHIME (Edgar et al. 2011) with sequences from the
10
11 415 Protist Ribosomal Reference Database (PR2) (Guillou et al. 2013) as references to ensure
12
13
14 416 that all chimeric sequences were detected and removed from the dataset. Taxonomy was
15
16 417 assigned to OTUs at 80% confidence using the RDP Classifier 2.2 (Wang et al. 2007)
17
18
19 418 implemented in QIIME using the PR2 database (Guillou et al. 2013).
20

21 419 Our approach for removing OTUs from negative controls is detailed in
22
23
24 420 Pagenkopp Lohan et al. (2016). Briefly, we removed OTUs when they were present only
25
26 421 in negative controls, were >10X more abundant in the negative controls compared to
27
28
29 422 samples, or were equally abundant in both negative controls and samples (n=53 OTUs).
30
31 423 All OTUs identified as organelle or mitochondrial sequences were also removed (n=96).
32
33
34 424 Finally, OTUs that were >10X more abundant in samples compared to the negative
35
36 425 controls were considered not likely due to external contamination, but were considered
37
38 426 ecologically valid and kept in the dataset. All OTUs identified as Metazoa or
39
40
41 427 Embryophyta were also removed from the dataset. Due to the lack of taxonomic
42
43 428 resolution in the dataset, as many OTUs were not identified to the species or even genus
44
45
46 429 level, we did not remove OTUs classified in groups that are known to contain macroalgae
47
48
49 430 (i.e., Chlorophyta, Rhodophyta). Thus, though the majority of taxa are generally referred
50
51 431 to as protists and these are the main taxa highlighted in this study, we acknowledge that
52
53 432 our dataset probably contains at least some OTUs identified as macroorganisms
54
55
56 433 (primarily macroalgae), which are probably due to cells or extracellular DNA from these
57
58 434 organisms being captured during filtration or from the sediments. Additionally, given the
59
60
61
62
63
64
65

1
2
3
4 435 sensitivity of this technique, it is possible that these OTUs represent a combination of
5
6 436 active, resting, and dead organisms.
7
8
9 437 We calculated a variety of summary statistics in QIIME. First, we calculated the
10
11 438 total number of sequences and OTUs for all samples as well as the most abundant OTUs
12
13
14 439 across the entire dataset. We then calculated the number of core OTUs
15
16 440 (`compute_core_microbiome.py`), meaning those present in 100% of samples in each
17
18 441 region, habitat, and from samples collected from both entrances to the Canal (Table 1).
19
20
21 442 We used *mothur* (Schloss et al. 2009) to calculate OTUs shared across any samples from
22
23 443 the three regions using the calculation for shared observed richness. With the exception
24
25
26 444 of the analyses of core OTUs, these summary statistics were generated for the entire and
27
28
29 445 rarified datasets.

30
31 446 **Beta diversity comparisons: Water and sediment:** The data was split into two
32
33 447 separate datasets based on habitat (water and sediment), which were independently
34
35
36 448 analyzed using beta diversity indices. In EstimateS v9.1 for MacOS (Colwell 2013), we
37
38 449 calculated extrapolated rarefaction curves, which were extrapolated by an additional 30
39
40
41 450 samples. All standard deviation calculations were performed with 100 randomizations
42
43 451 estimated for every sample with individuals randomized without replacement.

44
45 452 We ran standard rarefaction analyses in QIIME for both observed species and the
46
47
48 453 Chao1 diversity index per sample. Based on the output, we rarified all of the water
49
50
51 454 samples to 30,000 sequences and all of the sediment samples to 15,000 sequences as, for
52
53 455 each graph, these numbers were close to the point where the majority of samples appear
54
55 456 to begin to asymptote. The gregarine dataset was rarified to 2,000. At this step, all
56
57
58 457 samples that did not contain the minimum number of sequences were removed from the
59
60
61
62
63
64
65

datasets (Supplementary Material Table A1). After rarifying all water samples to 30,000 sequences, the water dataset contained 19 samples including five from Bocas del Toro, six from Caribbean-Canal, and eight from Pacific-Canal (Supplementary Material Table A1). In this dataset, the total number of sequences was 570,000 and the total number of OTUs was 6,861. After rarifying all sediment samples to 15,000 sequences, the sediment dataset contained eight samples including six from Caribbean-Canal and two from Pacific-Canal. In this dataset, the total number of sequences was 120,000 and the total number of OTUs was 3,373. After rarifying across all samples with only OTUs identified as gregarines included to 2,000 sequences, the gregarine dataset contained 22 samples, 380 OTUs (24 – 86 per sample), and 44,000 sequences.

Beta diversity was calculated using the binary Sorenson Dice index (beta_diversity.py). Unweighted pair group with arithmetic mean (UPGMA) plots were constructed in QIIME (upgma_cluster.py), then viewed using FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). Using the same beta diversity data, we created non-metric multidimensional scaling (NMDS) plots (nmds.py) for each dataset. Finally, for the sediment and water datasets, we assessed the similarity by distance using the GPS coordinates (Supplementary Material Table A1) for each location (distance_matrix_from_mapping.py) compared to the results from the beta diversity analyses to conduct a Mantel test with 1,000 permutations (compare_distance_matrices.py). Significance was based on $\alpha < 0.05$.

Acknowledgements

1
2
3
4 481 We thank Drs. Christina Simkanin, Ian Davidson, and three anonymous reviewers for
5
6 482 providing feedback on an earlier version of this manuscript. Kristina Hill-Spanik and Dr.
7
8
9 483 Carmen Schlöder were instrumental in assisting with field collections and obtaining
10
11 484 permits. We thank Gabriel Jacome and Plinio Gondola at the Smithsonian Bocas del Toro
12
13
14 485 Marine Station for logistical support. Zach Herbert at Dana Farber Cancer Institute
15
16 486 assisted with protocols for preparation of Illumina libraries. Nancy Rotzel McInerney
17
18
19 487 provided lab logistical support at CCEG. This work was supported with a Smithsonian
20
21 488 Institution MarineGEO Postdoctoral Fellowship to K. Pagenkopp Lohan and Smithsonian
22
23
24 489 Institution funds from a Grand Challenges Award from the Consortium for
25
26 490 Understanding and Sustaining a Biodiverse Planet to all authors. This is contribution
27
28
29 491 number 11 from the Smithsonian's MarineGEO and Tennenbaum Marine Observatories
30
31 492 Network.
32
33
34 493

35 36 494 **Appendix A. Supplementary Data**

37
38 495 Supplementary tables and figures with this article can be found, in the online version, at
39
40
41 496 XXX.
42

43 497 44 45 498 **References** 46 47 48 499

49
50 500 **Baldwin AJ, Moss JA, Pakulski JD, Catala P, Joux F, Jeffrey WH** (2005) Microbial
51 501 diversity in a Pacific Ocean transect from the Arctic to Antarctic circles. *Aquat Microb*
52
53 502 *Ecol* **41**:91–102
54
55
56
57 503 **Bass D, Richards TA, Matthai L, Marsh V, Cavalier-Smith T** (2007) DNA evidence
58
59 504 for global dispersal and probable endemism of protozoa. *BMC Evol Biol* **7**:162
60
61
62
63
64
65

- 505 **Bik HM, Sung W, De Ley P, Baldwin JG, Sharma J, Rocha-Olivares A, Thomas**
506 **WK** (2012) Metagenetic community analysis of microbial eukaryotes illuminates
507 biogeographic patterns in deep-sea and shallow water sediments: metagenetic community
508 analysis. *Mol Ecol* **21**:1048–1059
- 509 **Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS, Shalchian-**
510 **Tabrizi K** (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by
511 pyrosequencing and phylogeny of environmental rDNA. *ISME J* **4**:1144–1153
- 512 **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,**
513 **Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,**
514 **Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J,**
515 **Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J,**
516 **Knight R** (2010) QIIME allows analysis of high-throughput community sequencing data.
517 *Nat Methods* **7**:335–336
- 518 **Carlton JT** (1985) Transoceanic and interoceanic dispersal of coastal marine organisms:
519 the biology of ballast water. *Oceanogr Mar Biol Annu Rev* **23**:313–317
- 520 **Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A** (2012) Marine protistan
521 diversity. *Annu Rev Mar Sci* **4**:467–493
- 522 **Chesher R** (1968) Transport of marine plankton through the Panama Canal. *Limnol*
523 *Oceanogr* **13**:387–388
- 524 **Coats DW** (1999) Parasitic life styles of marine dinoflagellates. *J Eukaryot Microbiol*
525 **46**:402–409
- 526 **Cohen AN** (2006) Species Introductions and the Panama Canal. In Gollash S, Galil BS,
527 Cohen AN (eds) *Bridging Divides: Maritime Canals as Invasion Corridors*. Springer,
528 Dordrecht, pp 127–206
- 529 **Colwell RK** (2013) EstimateS: Statistical estimation of species richness and shared
530 species from samples. Version 9. User's Guide and application published at:
531 <http://purl.oclc.org/estimates>.

- 532 **Coutts ADM, Dodgshun TJ** (2007) The nature and extent of organisms in vessel sea-
533 chests: A protected mechanism for marine bioinvasions. *Mar Pollut Bull* **54**:875–886
- 534 **Davidson IC, Brown CW, Sytsma MD, Ruiz GM** (2009) The role of containerhips as
535 transfer mechanisms of marine biofouling species. *Biofouling* **25**:645–655
- 536 **Drake LA, Doblin MA, Dobbs FC** (2007) Potential microbial bioinvasions via ships’
537 ballast water, sediment, and biofilm. *Mar Pollut Bull* **55**:333–341
- 538 **Edgar RC** (2013) UPARSE: highly accurate OTU sequences from microbial amplicon
539 reads. *Nat Methods* **10**:996–998
- 540 **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R** (2011) UCHIME improves
541 sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194–2200
- 542 **Finlay BJ** (2002) Global dispersal of free-living microbial eukaryote species. *Science*
543 **296**:1061–1063
- 544 **Galil BS, Hülsmann N** (1997) Protist transport via ballast water - biological
545 classification of ballast tanks by food web interactions. *Europ J Protistol* **33**:244–253
- 546 **Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I** (2008)
547 Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples
548 by real-time PCR. *Appl Environ Microbiol* **74**:7174–7182
- 549 **Gong J, Dong J, Liu X, Massana R** (2013) Extremely high copy numbers and
550 polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and
551 peritrich ciliates. *Protist* **164**:369–379
- 552 **Glynn PW** (1972) Observations on the Ecology of the Caribbean and Pacific Coasts of
553 Panama. In Jones ML (ed) *The Panamic Biota: Some Observations Prior to a Sea-Level*
554 *Canal*. Bull Biol Soc Wash No 2, Washington, DC, pp 13–30
- 555 **Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G,**
556 **de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B,**

- 557 **Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana**
558 **R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A, Siano R,**
559 **Siano R, Stoeck T, Vaultot D, Zimmermann P, Christen R** (2012) The Protist
560 Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit
561 rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**:D597–D604
- 562 **Hillebrand H, Watermann F, Karez R, Berninger UG** (2001) Differences in species
563 richness patterns between unicellular and multicellular organisms. *Oecologia* **126**:114–
564 124
- 565 **Illumina Support** (2013) 16S Metagenomic Sequencing Library Preparation.
566 [http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s](http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)
567 [-metagenomic-library-prep-guide-15044223-b.pdf](http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)
- 568 **Leander BS** (2008) Marine gregarines: evolutionary prelude to the apicomplexan
569 radiation? *Trends Parasitol* **24**:60–67
- 570 **Leigh EG, O’Dea A, Vermeij GJ** (2013) Historical biogeography of the Isthmus of
571 Panama. *Biol Rev* **89**:148–172
- 572 **Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A** (2013) Next
573 generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS*
574 *ONE* **8**:e81327
- 575 **Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL,**
576 **Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvreås**
577 **L, Reysebach A, Smith VH, Staley JT** (2006) Microbial biogeography: putting
578 microorganisms on the map. *Nat Rev Microbiol* **4**:102–112
- 579 **Pagenkopp Lohan KM, Fleischer RC, Carney KJ, Holzer KK, Ruiz GM** (2016)
580 Amplicon-based pyrosequencing reveals high diversity of protistan parasites in ships’
581 ballast water: Implications for biogeography and infectious diseases. *Microbial Ecol*
582 **71**:530–542
- 583 **Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA** (2014)

- 584 Environmental monitoring through protist next-generation sequencing metabarcoding:
585 assessing the impact of fish farming on benthic foraminifera communities. *Mol Ecol*
586 *Resour* **14**:1129–1140
- 587 **Rubinoff I** (1970) The sea-level canal controversy. *Biol Conserv* **3**:33–36
- 588 **Ruiz GM, Torchin ME, Grant K** (2009) Using the Panama Canal to Test Predictions
589 About Tropical Marine Invasions. In Lang MA, Macintyre IG, Rutzler K (eds)
590 Proceedings Smithsonian Marine Science Symposium, *Smith Contrib Mar Sci*, vol 38,
591 Smithsonian Institution Scholarly Press, Washington, DC, pp 291–299
- 592 **Ruiz GM, Lorda J, Arnwine A, Lion K** (2006) Chapter II Shipping Patterns Associated
593 with the Panama Canal: Effects on Biotic Exchange? In Gollash S, Galil BS, Cohen AN
594 (eds) *Bridging Divides*. Springer, Dordrecht, pp 127–206
- 595 **Ruiz GM, Fofonoff PW, Carlton JT, Wonham MJ, Hines AH** (2000) Invasion of
596 coastal marine communities in North America: apparent patterns, processes, and biases.
597 *Annu Rev Ecol Syst* **31**:481–531
- 598 **Schlöder C, Canning-Clode J, Saltonstall K, Strong EE, Ruiz GM, Torchin ME**
599 (2013) The Pacific bivalve *Anomia peruviana* in the Atlantic: a recent invasion across the
600 Panama Canal? *Aquat Invas* **8**:443–448
- 601 **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB,**
602 **Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger**
603 **GG, Van Horn DJ, Weber CF** (2009) Introducing mothur: Open-source, platform-
604 independent, community supported software for describing and comparing microbial
605 communities. *Appl Environ Microbiol* **75**:7537–7541
- 606 **Schmieder R, Edwards R** (2011) Quality control and preprocessing of metagenomic
607 datasets. *Bioinformatics* **27**:863–864
- 608 **Sheffey JP** (1968) When Caribbean and Pacific waters mix. *Science* **162**:1329
- 609 **Villac C, Kaczmarska I** (2011) Estimating propagule pressure and viability of diatoms

1
2
3
4 610 detected in ballast tank sediments of ships arriving at Canadian ports. Mar Ecol Prog Ser
5
6 611 **425**:47–61
7
8
9 612 **Wang Q, Garrity GM, Tiedje JM, Cole JR** (2007) Naive Bayesian classifier for rapid
10
11 613 assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol
12
13 614 **73**:5261–526
14
15
16 615
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1. The number and taxonomic ranking of OTUs detected in all environmental samples (both), all water samples (water), the rarified water dataset (waterR), all sediment samples, and the rarified sediment dataset (sedimentR) using the V4 hypervariable region of the small subunit (SSU) gene of the ribosomal gene (rDNA) complex. OTUs are grouped at their lowest resolved ranking or at the third rank. Color coding denotes number of OTUs within each column, ranging from the lowest value (white), to the 50th percentile value (orange), to the highest value (red).

Taxonomic Ranking	Both	Water	WaterR	Sediment	SedimentR
Unassigned	563	453	392	297	203
Eukaryota	1,847	1,432	1,229	1,017	688
Alveolata	139	104	95	83	56
Apicomplexa	676	622	569	377	266
Ciliphora	222	192	167	84	61
Dinophyta	857	817	754	369	278
Perkinsea	12	12	8	5	2
Amoebozoa	19	16	13	13	10
Breviatea	34	32	26	10	7
Conosa	40	33	29	22	17
Lobosa	90	74	70	54	36
Apusozoa	1	1	1	1	1
Apusomonadidae	48	37	34	36	26
Hilomonadea	21	21	17	14	8
Mantamonadidea	6	4	4	6	5
Archaeplastida	6	6	3	1	1
Chlorophyta	238	226	209	124	82
Rhodophyta	121	114	91	69	55
Excavata	0	0	0	0	0
Discoba	4	4	2	0	0
Hacrobia	2	2	2	1	1
Centroheliozoa	25	17	14	22	12
Crytophyta	41	38	35	32	24
Haptophyta	46	45	42	20	12
Katablepharidophyta	14	13	13	10	9
Picobiliphyta	7	7	7	4	3
Telonemia	13	13	12	8	6
Opisthokonta	139	107	93	79	47
Choanoflagellida	83	72	67	46	35
Fungi	623	611	567	168	102
Mesomycetozoa	25	25	23	14	8
Rhizaria	16	15	12	8	6
Cercozoa	1,166	951	851	634	421
Radiolaria	25	24	20	8	5
Stramenopiles	772	661	598	411	283
Ochrophyta	1,107	888	807	801	621
TOTAL	9,048	7,689	6,876	4,848	3,397

623

Table 2. The taxonomic identity of the relatively abundant OTUs (those with >5,000 sequences) recovered from the water dataset (top) and sediment dataset (bottom). Designation of core OTU is listed for 1) each region where that designation was made, 2) “All” or “Both” if it was found in all regions in the respective dataset, or 3) N/A if the OTU was not a core OTU. Bold font denotes those OTUs that were in relatively high abundance in rarified datasets. Asterisks denote those that were found in relatively high abundant in both datasets.

	OTU Number	Taxonomic Identification	Number of Sequences	Core OTU Region(s)
Water	OTU_31*	Apicomplexa: Gregarines: Cephaloidophoroidea	8,751	N/A
	OTU_81#	Apicomplexa: Gregarines: Cephaloidophora	18,724	Canal-Caribbean
	OTU_12#	Apicomplexa: Gregarines: Cephaloidophora	42,242	Canal-Pacific
	OTU_11#	Apicomplexa: Gregarines: Cephaloidophoroidea	41,683	Bocas del Toro
	OTU_110#	Apicomplexa: Gregarines: Cephaloidophoroidea	8,078	Bocas del Toro
	OTU_159	Apicomplexa: Gregarines	7,123	N/A
	OTU_36	Apicomplexa: Gregarines: Leidyana	11,744	N/A
	OTU_56	Dinophyta: Dinophyceae	57,680	All
	OTU_71	Dinophyta: Dinophyceae	11,699	All
	OTU_952	Dinophyta: Dinophyceae	5,327	Canal-Pacific, Bocas del Toro
	OTU_387	Dinophyta: Heterocapsa	6,885	All
	OTU_146	Dinophyta: Warnowia	5,186	Canal-Caribbean, Bocas del Toro
	OTU_256	Dinophyta: Ostreopsidaceae	5,015	N/A
	OTU_123	Dinophyta: Suessiales	5,509	All
	OTU_44	Dinophyta: Syndiniales: Dino-Group-I-Clade-1	8,432	All
	OTU_23	Dinophyta: Syndiniales: Dino-Group-II-Clade-12	20,142	Bocas del Toro
	OTU_68	Dinophyta: Syndiniales: Dino-Group-II-Clade-3	8,947	Canal-Pacific, Bocas del Toro
	OTU_101	Chlorophyta: Chlorodendrales	6,058	Canal-Pacific, Bocas del Toro

OTU_41	Chlorophyta: Chlorodendrales	16,389	All
OTU_135	Rhodophyta: Ceramiaceae	5,471	Canal- Caribbean
OTU_45	Rhodophyta: Neosiphonia	9,511	N/A
OTU_32	Rhodophyta: Gigartinales	14,311	Canal- Caribbean
OTU_27	Fungi: Dothideomycetes	20,976	All
OTU_6368	Fungi: Teratosphaeriaceae	7,132	All
OTU_105	Fungi: Eurotiomycetes	5,894	All
OTU_147	Fungi: Gibberella	10,702	All
OTU_80	Fungi: Neurospora	5,375	N/A
OTU_108	Fungi: Pestalotiopsis	7,004	Canal- Caribbean, Bocas del Toro
OTU_13	Bacillariophyta: Chaetoceros	32,919	All
OTU_28	Bacillariophyta: Minutocellus	22,272	All
OTU_59	Bacillariophyta: Skeletonema	7,935	Caribbean- Canal, Pacific- Canal
OTU_19	Bacillariophyta: Skeletonema	22,010	Canal- Caribbean, Canal-Pacific
OTU_38	Bacillariophyta: Thalassiosira	29,366	All
OTU_1011	Bacillariophyta: Thalassiosira	9,254	Canal-Pacific, Bocas del Toro
OTU_48	Bacillariophyta: Thalassiosira	7,711	Canal-Pacific
OTU_109	Bacillariophyta: Raphid-pennate	5,273	All
OTU_51	Bacillariophyta: Cylindrotheca	10,012	All
OTU_54*	Bacillariophyta: Navicula	6,241	All
OTU_33*	Bacillariophyta: Navicula	5,793	All
OTU_37*	Unclassified	9,809	N/A
Sediment			
OTU_31*	Apicomplexa: Gregarines: Cephaloidophoroidea	9,617	Canal- Caribbean
OTU_189	Apicomplexa: Gregarines: Cephaloidophoroidea	5,180	N/A
OTU_47	Chlorophyta: Chaetomorpha	5,338	N/A
OTU_46	Chlorophyta: Ulva	8,262	Both
OTU_88	Cercozoa: Endomyxa-Phytomyxea	6,598	N/A
OTU_26	Bacillariophyta	25,857	N/A
OTU_92	Bacillariophyta	5,717	N/A
OTU_54*	Bacillariophyta: Navicula	25,788	Both
OTU_33*	Bacillariophyta: Navicula	39,231	Both

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

	OTU_75	Bacillariophyta: Psammodictyon	18,980	Canal- Caribbean
	OTU_55	Ochrophyta: Chrysophyceae- Synurophyceae	8,663	N/A
	OTU_37*	Unclassified	9,598	N/A

Table 3. The taxonomic identity of the most relatively abundant OTUs (those with >5,000 sequences) recovered from both the water and sediment datasets from samples taken from within the Canal regions, (A) Caribbean-Canal and (B) Pacific-Canal. Bold font denotes those OTUs that were in high relative abundance across regions. Asterisks denote those that were high relative abundance in both water and sediment datasets within the same region. Color coding denotes number of sequences within each OTU across locations. For water datasets, the colors range from white (low number of sequences) to red (high number of sequences). For sediment datasets, the colors range from white (low number of sequences) to dark blue (high number of sequences).

A. Caribbean-Canal			GAL	FTS	FTS2	ACP	RAPA	SBPA
Water	OTU_31*	Gregarines: Cephaloidophoroidea	1,865	1,416	1,814	0	22	2
	OTU_81	Gregarines: Cephaloidophora	214	131	118	9	10,612	7,399
	OTU_11	Gregarines: Cephaloidophoroidea	0	3	0	23	20,608	63
	OTU_56	Dinophyceae	631	934	435	11,179	99	175
	OTU_23	Syndiniales: Dino-Group-II-Clade-12	7	49	1	0	59	19,082
	OTU_135	Rhodophyta: Ceramiaceae	27	4,470	712	4	2	1
	OTU_45	Rhodophyta: Neosiphonia	0	43	13	9,336	3	0
	OTU_13	Bacillariophyta: Chaetoceros	110	1,917	1,270	14,360	17	11
	OTU_19	Bacillariophyta: Skeletonema	363	5,158	3,831	10,405	10	2
Sediment	OTU_37*	Unclassified	9,700	0	71	1	0	0
	OTU_31*	Gregarines: Cephaloidophoroidea	1,031	840	7,715	4	N/A	N/A
	OTU_189	Gregarines: Cephaloidophoroidea	0	37	5,135	0	N/A	N/A
	OTU_47	Chlorophyta: Chaetomorpha	7	84	0	5,242	N/A	N/A
	OTU_88	Cercozoa: Endomyxa-Phytomyxea	0	6,513	19	0	N/A	N/A
	OTU_26	Bacillariophyta	3	20,174	53	0	N/A	N/A
	OTU_92	Bacillariophyta	170	4,933	13	123	N/A	N/A

	OTU_54	Bacillariophyta: Navicula	13,356	2,694	882	174	N/A	N/A
	OTU_33	Bacillariophyta: Navicula	7,089	26,125	2,876	678	N/A	N/A
	OTU_75	Bacillariophyta: Psammodictyon	4,508	10,425	2,263	2	N/A	N/A
	OTU_55	Ochrophyta: Chrysophyceae-Synurophyceae	0	8,623	26	0	N/A	N/A
	OTU_37*	Unclassified	9,578	13	0	0	N/A	N/A
B. Pacific-Canal			VEC	BYC	BAT	PCUL	BB	BP
Water	OTU_12	Gregarines: Cephaloidophora	786	71	8,765	9	32,014	312
	OTU_56	Dinophyceae	5,474	3,597	2,963	2,512	466	208
	OTU_41	Chlorodendrales	50	62	48	8	10,631	453
	OTU_27	Fungi: Dothideomycetes	112	325	14,316	542	65	14
	OTU_80	Fungi: Neurospora	84	99	4,303	774	37	3
	OTU_13	Bacillariophyta: Chaetoceros	505	428	526	5,415	333	116
	OTU_28	Bacillariophyta: Minutocellus	1,447	259	715	412	13,732	3,557
	OTU_59	Bacillariophyta: Skeletonema	5,968	120	208	267	816	118
	OTU_38	Bacillariophyta: Thalassiosira	2,692	610	579	779	15,519	1,639
	OTU_1011	Bacillariophyta: Thalassiosira	4,872	122	161	346	1,993	277
	OTU_48	Bacillariophyta: Thalassiosira	1	26	27	4	6,725	577
	OTU_51	Bacillariophyta: Cylindrotheca	253	26	74	44	4,067	658
Sediment	OTU_54	Bacillariophyta: Navicula	7,660	N/A	9	N/A	N/A	N/A

Figure Legends

Figure 1. Shared and unique OTUs across the three regions sampled. For this analysis, all samples in each dataset were included. The numbers are those OTUs shared across all three regions (center), those OTUs shared across two of three regions (surrounding the center), and those unique to each region (outside).

Figure 2. Core OTUs (defined as those OTUs present in 100% of samples) including those were unique or shared across (A) all samples across all regions, (B) only water samples across all regions, and (C) only sediment samples across all regions. S is the number of samples that was included for each analysis. N is the total number of OTUs present in 100% of S samples.

Figure 3. Extrapolated rarefaction plots generated in EstimateS v9.1 for MacOS (Colwell 2013). Plots show an extrapolation of an additional 30 samples for each dataset.

Figure 4. UPGMA plots of beta diversity for the (A) water and (B) sediment rarified datasets. Beta diversity was calculated using the binary Sorenson Dice index. In both plots, the samples from the Pacific-Canal are highlighted in blue, samples from the Caribbean-Canal are highlighted in red, and samples from Bocas del Toro are black. A star denotes the Caribbean sample in a clade with Pacific samples. See text for additional details

Figure 5. UPGMA plot of beta diversity for the gregarine rarified dataset. Beta diversity was calculated using the binary Sorenson Dice index. Samples from the Pacific-Canal are highlighted in blue, samples from the Caribbean-Canal are highlighted in red, and samples from Bocas del Toro are black. See text for additional details

A. All
Figure 1

B. Water

C. Sediment

Figure2

Pacific-Canal

S = 14

N = 26

A. All

Caribbean-Canal

S = 15

N = 27

Pacific-Canal

S = 9

N = 110

B. Water

Caribbean-Canal

S = 7

N = 62

Bocas del Toro

S = 6

N = 128

Bocas del Toro

S = 6

N = 118

C. Sediment

Pacific-Canal

S = 5

N = 29

Caribbean-Canal

S = 8

N = 56

Figure 4