

## ORIGINAL ARTICLE

# Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing

Man Kit Cheung<sup>1</sup>, Chun Hang Au<sup>1</sup>, Ka Hou Chu, Hoi Shan Kwan and Chong Kim Wong  
*Department of Biology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China*

**Information on genetic diversity of picoeukaryotes (<2–3 μm) comes mainly from traditional gene cloning and sequencing, but this method suffers from cloning biases and limited throughput. In this study, we explored the feasibility of using the cloning-independent and massively parallel 454 pyrosequencing technology to study the composition and genetic diversity of picoeukaryotes in the coastal waters of the subtropical western Pacific using the hypervariable V4 region of the 18S rRNA gene. Picoeukaryote assemblages between two sites with different hydrography and trophic status were also compared. The approach gave a high coverage of the community at genetic difference ≥5% but still underestimated the total diversity at a genetic difference ≤2%. Diversity of picoeukaryotes was higher in an oligomesotrophic bay than in a eutrophic bay. Stramenopiles, dinoflagellates, ciliates and prasinophytes were the dominant groups comprising approximately 27, 19, 11 and 11%, respectively, of the picoeukaryotes. Water samples collected from the two bays contained different high-level taxonomic groups and phylotype operational taxonomic units of picoeukaryotes. Our study represents one of the first and most comprehensive examinations of marine picoeukaryotic diversity using the 454 sequencing-by-synthesis technology.**

*The ISME Journal* (2010) 4, 1053–1059; doi:10.1038/ismej.2010.26; published online 25 March 2010

**Subject Category:** Microbial ecology and functional diversity of natural habitats

**Keywords:** 18S; microbial diversity; Pacific Ocean; picophytoplankton; picoplankton; pyrosequencing

## Introduction

Picoeukaryotes are eukaryotes smaller than 2–3 μm in diameter. They occur in aquatic environments worldwide and are considered as fundamental components of marine ecosystems in the oceans (Sherr and Sherr, 2000). They contribute significantly to biomass and primary production and are particularly important in nutrient-rich coastal waters in which they are most abundant (Li, 1994; Worden and Not, 2008). During the last decade, studies based on cloning and sequencing of the 18S rRNA gene have revealed an unexpected diversity among picoeukaryotes in the oceans (Díez *et al.*, 2001; López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001). Unfortunately, this approach suffers from potential cloning biases that may mask the real diversity of the community under study (Forns *et al.*, 1997). In addition, community diversity may

also be greatly underestimated by the limited throughput of the method (Bent and Forney, 2008).

The newly emerged 454 sequencing-by-synthesis technology is cloning independent and massively parallel (Margulies *et al.*, 2005). The diversity of bacteria (Sogin *et al.*, 2006) and archaea (Galand *et al.*, 2009) in open oceans has been investigated with high-throughput amplicon sequencing targeting the hypervariable V6 region of the 16S rRNA gene. Targeting the V9 hypervariable region of the 18S rRNA gene, this technology has also been applied to marine protists recently (Amaral-Zettler *et al.*, 2009). However, to the best of our knowledge, no similar studies have been carried out on marine picoeukaryotes. In this study, we explored the feasibility of using the 454 pyrosequencing technology to investigate the diversity and community composition of picoeukaryote assemblages in subtropical coastal waters of the western Pacific. Picoeukaryote assemblages between two sites with different hydrography and trophic status were also compared.

Correspondence: CK Wong, Department of Biology, Simon FS Li Marine Science Laboratory, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China.  
E-mail: chongkimwong@cuhk.edu.hk

<sup>1</sup>These authors contributed equally to this work.

Received 10 September 2009; revised 28 January 2010; accepted 11 February 2010; published online 25 March 2010

## Materials and methods

### *Sample collection and DNA extraction*

Surface seawater samples (~2.5 l) were collected from two bays in northeastern Hong Kong in April

**Table 1** Sampling conditions and sequence characteristics of the two sites

	Mirs Bay	Tolo Harbor
Coordinates	22°30.735'N, 114°21.546'E	22°25.920'N, 114°13.110'E
Temperature (°C)	21.8	23.0
Dissolved oxygen (mg l <sup>-1</sup> )	7.87	7.89
Salinity (‰)	33.5	32.5
Chlorophyll <i>a</i> (µg l <sup>-1</sup> )	1.92	9.06
Total reads	100 911	87 392
Average read length (bp)	259	260
Trimmed tags	87 789	76 817
Average tag length (bp)	272	272
Tags after BLASTN	61 671	48 208

Trimmed tags are sequence reads remaining after the removal of low-quality and nontarget reads. Tags after BLASTN are tags with ≥95% query coverage and top match identity to GenBank entries.

2007. Tolo Harbor (TH) is a landlocked bay with a long history of eutrophication (Wear *et al.*, 1984; Chau, 2007). Located outside TH, Mirs Bay (MB) is relatively unpolluted and more exposed to ocean currents from the South China Sea (Hong Kong Environmental Protection Department, 2003). Water samples were filtered through a 200 µm mesh sieve immediately to remove most of the mesozooplankton and large particles. Water temperatures, salinities and dissolved oxygen levels were measured on board using a Hydrolab sensor (Austin, TX, USA). In the laboratory, chlorophyll *a* concentrations were determined using a Turner Designs 10-AU fluorometer (Sunnyvale, CA, USA) as described by Wong and Wong (2003). At the time of sample collection, chlorophyll *a* concentration was 9.1 µg l<sup>-1</sup> in TH and 1.9 µg l<sup>-1</sup> in MB (Table 1). On the basis of these values, TH and MB were considered to be eutrophic and oligomesotrophic, respectively (Molvaer *et al.*, 1997).

Two liters of water was prefiltered through 3 µm pore size Nuclepore membranes (Whatman, Piscataway, NJ, USA) and the microbial biomass was then collected onto a GF/F filter (Whatman). A gentle vacuum (<20 cm Hg) created by a hand pump was used to facilitate the filtration processes. The filter was then immersed in DNA lysis buffer (0.75 M sucrose, 40 mM EDTA, 50 mM Tris-HCl (pH 8)), immediately frozen in liquid nitrogen and stored at -80 °C until DNA extraction. DNA was extracted after the cetyltrimethylammonium bromide extraction procedure (Doyle and Doyle, 1990).

#### PCR and pyrosequencing

PCR was performed using 454 sequencing adaptor-linked primers flanking the hypervariable V4 region of the 18S rRNA gene: A-528F (5'-gcctcctcgcgccatc ag-GCGTAATTCAGCTCCAA-3') and B-706R (5'-gc cttgccagcccgtcag-AATCCRAGAATTTACCTCT-3') (adaptor sequences shown in lowercase) (Elwood *et al.*, 1985). PCR mixtures (50 µl) were prepared in duplicate and each contained 2 µl of DNA template,

5 µl of 10 × PCR buffer (50 mM KCl, 10 mM Tris-HCl and 1.5 mM MgCl<sub>2</sub>), 200 µM of dNTP, 0.2 µM of each primer and 2.5 U *Taq* polymerase (Promega, Madison, WI, USA). The PCR thermal regime consisted of an initial denaturation of 3 min at 94 °C, followed by 30 cycles of 30 s at 94 °C, 30 s at 60 °C, 1 min at 72 °C and a final cycle of 5 min at 72 °C. PCR products were pooled and purified with the Qiaquick gel purification kit according to the manufacturer's instructions (Qiagen, Hilden, North Rhine-Westphalia, Germany). DNA concentration and quality were determined with a NanoDrop 1000 spectrophotometer (Wilmington, DE, USA).

Pyrosequencing of PCR products was performed on a Genome Sequencer FLX system at 454 Life Sciences (Branford, CT, USA). Sequences and quality scores from our pyrosequencing run were submitted to the NCBI short read archive (accession number SRA009090). Raw sequence reads were filtered before subsequent analyses to minimize the effects of random sequencing errors (Huse *et al.*, 2007). Briefly, we eliminated sequence reads that (1) did not perfectly match the proximal PCR primer, (2) were too short (<244 bp) or too long (>302 bp), as determined by the sequence length distribution plots or (3) contained one or more ambiguous base(s) (N(s)).

#### Assignment of phylotype OTUs

Phylotype operational taxonomic units (OTUs) were assigned according to the best BLASTN hit (Altschul *et al.*, 1990) against NCBI nucleotide sequence database NT (as of August 2008) with the following parameters: *E* value = 10<sup>-5</sup>, minimum query coverage = 95% and minimum identity = 95%. Taxonomic group assignment of reads resembling GenBank sequences from environmental clones was obtained using the web-based software package KeyDNATools as described by Guillou *et al.* (2008). Nontarget reads belonging to viruses, bacteria or metazoa were removed from further analyses.

#### Assignment of similarity-based OTUs and species richness estimators

Sequence reads from each sample were clustered to give similarity-based OTUs using cd-hit-est (Li and Godzik, 2006) with minimum sequence identity set to 90, 95, 98 or 100%. Parametric rarefaction curves were calculated in steps of 1000 specimens using Analytic Rarefaction (v 1.3, <http://www.uga.edu/strata/software/anRareReadme.html>). Nonparametric species richness ACE and classic Chao1 estimates were calculated using EstimateS (v 8, <http://viceroy.eeb.uconn.edu/estimates/>).

## Results

#### Characteristics of the pyrosequencing run

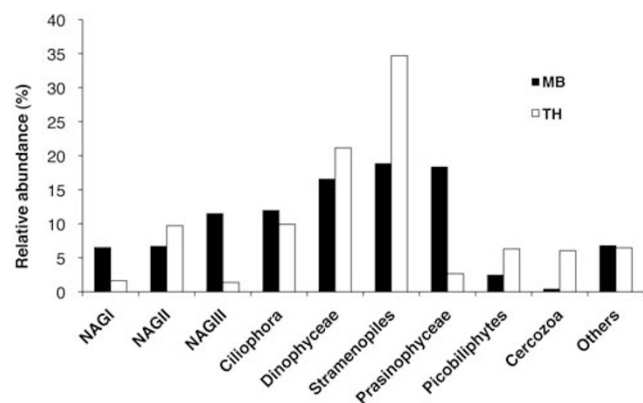
A total of 188 303 sequence reads with an average length of about 260 bp were generated in a single run

of 454 pyrosequencing from the two water samples (Table 1). The filtering process removed about 12% of the raw sequence reads, leaving 87 789 and 76 817 high-quality target tags from the MB and TH samples, respectively. The average tag length was improved to about 272 bp. After BLAST searching, 61 671 and 48 208 sequences with  $\geq 95\%$  query coverage and top match identity remained for samples collected from MB and TH, respectively.

#### Composition and distribution of the picoeukaryotic assemblages

Trimmed tags were grouped under at least 19 high-level taxonomic groups (Supplementary Table 1) and over 93% of them fell into one of the nine major groups (Figure 1). Stramenopiles, dinoflagellates, ciliates and prasinophytes comprising about 27%, 19%, 11% and 11%, respectively, of the picoeukaryotes were the dominant groups. Although stramenopiles and dinoflagellates together contributed more than half (55.9%) of the total community in the sample from TH, five groups (stramenopiles, prasinophytes, dinoflagellates, ciliates and the novel alveolates group III (NAGIII)) shared similar contributions (each between 10% and 25%) for the sample from MB. NAGI, NAGII and prasinophytes were at least threefold more abundant in MB than in TH. Conversely, the abundance of picobiliphytes and cercozoans in TH at least doubled than those in MB. The relative contributions of NAGII, ciliates and dinoflagellates were almost comparable between samples from TH and MB. Dinophyceae and stramenopiles were more abundant in TH, but the relative contributions of their phylotype OTUs were higher in MB (Supplementary Figure). Alternatively, there was more NAGIII in MB than in TH, although both samples shared comparable contributions of phylotype OTUs.

In both samples, the top 20 phylotype OTUs each contributed at least 1% and together represented about 60% of the entire picoeukaryote community



**Figure 1** Relative abundance of the nine dominant high-level taxonomic groups of picoeukaryotes. MB, Mirs Bay; TH, Tolo Harbor; NAG, novel alveolates group.

(61.7% in MB and 56.7% in TH) (Table 2). In addition, more than half of the OTUs represented uncultured environmental clones (MB, 14; TH, 13). Although sequence reads resembling six GenBank entries (*Pseudochattonella verruculosa*, *Verrucophora farcimen*, *Pterosperma cristatum*, uncultured Dinophyceae clone dhot2b10, AMT15\_1B\_26 and uncultured NAGII clone SA1\_4G10) were common in both bays, the abundance of all other top phylotype OTUs differed between TH and MB. More than half of the top phylotype OTUs were at least four times more abundant in terms of percentage contribution in one sample than in the other.

#### Diversity of the picoeukaryotic assemblages

The coverage of libraries at cluster distances  $\geq 0.05$  was high for both samples, with rarefaction curves reaching saturation. However, rarefaction curves at a cluster distance of 0.02 were still in an increasing trend (Figure 2). The numbers of OTUs detected were generally close to those estimated by the nonparametric ACE and Chao1 estimators at cluster distances  $\geq 0.05$ , with percentages of detected OTUs/estimated OTUs over 75% (Table 3), but the difference between the observed and estimated values at cluster distances  $\leq 0.02$  was large, especially when identical tags were grouped into unique OTUs, with percentages of detected OTUs/estimated OTUs less than 20%.

## Discussion

### Methodology

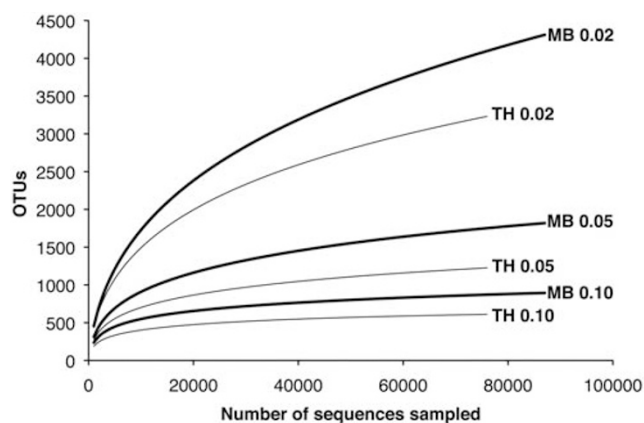
In recent years, the cloning and sequencing approach has been widely used to study the diversity and community composition of picoeukaryotes (Díez *et al.*, 2001; López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001). However, with an examination of typically 100 clones per library (Vaulot *et al.*, 2008), the data generated by this approach only provide general information on the structure of picoeukaryote communities but are not sufficient for meaningful comparisons among libraries. In this study, water samples were collected from two sites of different trophic status and about 100 000 sequence reads were derived from each water sample to facilitate an in-depth investigation into the structure of the picoeukaryote assemblages. However, as only one sample per site was examined, conclusions about picoeukaryotic communities are limited to the respective samples.

In a previous study at the same two sites, picoeukaryotic diversity was studied using the traditional cloning and sequencing method (Cheung *et al.*, 2008). Sequencing in partial PCR products amplified using the widely adopted primer set Euk328f and Euk329r (Moon-van der Staay *et al.*, 2001) with the primer Euk528f (Elwood *et al.*, 1985) allowed us to recover at least 19 high-level taxonomic

**Table 2** Top 20 phylotype OTUs in samples from the two sampling sites

No.	MB					TH				
	%	C%	AN	AF	GenBank ID	%	C%	AN	AF	GenBank ID
1	8.5	0.8	EU780604	NAGIII	UC AMT15_1B_36	4.9	NS	DQ470662	STRA	<i>Heterosigma akashiwo</i>
2	7.9	0.3	DQ120009	CILI	UC NW414.24	4.6	0.2	DQ367045	STRA	<i>Aplanochytrium</i> sp.
3	7.6	NS	EU143384	PRAS	UC PROSOPE.C1-80m.96	4.6	0.6	EU545747	CILI	UC 9_153
4	4.9	3.4	AB217630	STRA	<i>Pseudochattonella verruculosa</i>	4.5	1.3	EF526908	NAGII	UC SA1_4G10
5	4.8	3.2	AM075624	STRA	<i>Verrucophora farcimen</i>	3.8	3.4	EU500021	DINO	UC dhot2b10
6	3.4	3.8	EU500021	DINO	UC dhot2b10	3.8	0.6	EU500204	DINO	UC hotxp3a1
7	3.1	NS	EU143470	PRAS	UC PROSOPE.E1-80m.61	3.4	4.9	AB217630	STRA	<i>Pseudochattonella verruculosa</i>
8	3.1	NS	EU143463	PRAS	UC PROSOPE.EU-30m.113	3.2	4.8	AM075624	STRA	<i>Verrucophora farcimen</i>
9	2.6	0.1	EF172986	NAGI	UC SSRPD84	3.1	0.8	DQ222873	PICO	UC HE000803.72
10	2.1	0.1	EF173005	NAGII	UC H02N5	2.8	NS	AB252743	CERC	UC NAMAKO-3
11	1.9	1.5	EU780595	DINO	UC AMT15_1B_26	2.7	0.6	AY485503	STRA	<i>Eucampia antarctica</i>
12	1.8	1.1	AY664962	DINO	UC SCM37C27	2.7	0.3	EU087273	CILI	UC 9_52
13	1.7	NS	DQ504316	NAGI	UC LC22_5EP_6	2.3	0.1	EU087251	CERC	UC 9_25
14	1.7	0.4	EU780607	NAGIII	UC AMT15_1B_40	1.7	1.0	AJ010407	PRAS	<i>Pterosperma cristatum</i>
15	1.4	0.9	AM075625	STRA	<i>Chattonella verruculosa</i>	1.6	0.2	EF527119	STRA	UC SA2_1C5
16	1.3	4.5	EF526908	NAGII	UC SA1_4G10	1.6	0.2	EU733829	DINO	UC 2C_F07
17	1.2	0.4	AF063239	RADI	<i>Arthracanthida</i> sp.	1.6	0.1	EF526986	STRA	UC MIF_CilE6
18	1.1	0.2	EF492500	DINO	<i>Heterocapsa pygmaea</i>	1.5	1.9	EU780595	DINO	UC AMT15_1B_26
19	1.0	1.7	AJ010407	PRAS	<i>Pterosperma cristatum</i>	1.3	0.2	EF173006	NAGII	UC Q2A03
20	1.0	NS	EU500001	CILI	UC dhot2b7	1.2	0.4	U14387	STRA	<i>Pseudopedinella elastica</i>

Abbreviations: MB, Mirs Bay; TH, Tolo Harbor; C%, percentage of the OTU in the other sample (NS = value <0.05); AN, accession number; AF, affiliation; NAG, novel alveolates group; CILI, Ciliophora; PRAS, Prasinophyceae; STRA, stramenopiles; DINO, Dinophyceae; RADII, radiozoa; PICO, picobiliphytes; CERC, cercozoa; UC, uncultured clone.



**Figure 2** Rarefaction curves of similarity-based operational taxonomic units (OTUs) at cluster distance values of 0.02, 0.05 and 0.10. MB, Mirs Bay; TH, Tolo Harbor.

groups of picoeukaryotes. In this study, at least 19 high-level taxonomic groups were recovered using the primer set, A-528F and B-706R, which targets the same hypervariable V4 region (Supplementary Table 1). Although the dominant groups were retrieved by both methods, sequences belonging to nucleomorphs and ellobiopsids were only detected in the previous study, and sequences belonging to NAGIII, centrohelids and some other groups of chlorophytes apart from prasinophytes were only recovered in this study. However, except for NAGIII, which is a newly defined alveolates group (Guillou *et al.*, 2008), all the other groups were only

**Table 3** Similarity-based OTUs and species richness estimates of the two libraries

Library	Cluster distance	Detected OTU	ACE	Chao1
MB	0.00	33 209	172 662	172 718
	0.02	4326	6664	6670
	0.05	1823	2381	2385
	0.10	896	1075	1079
TH	0.00	30 634	167 610	167 673
	0.02	3243	4657	4662
	0.05	1229	1545	1548
	0.10	612	699	702

Abbreviations: MB, Mirs Bay; TH, Tolo Harbor.

represented in low abundance. The results reveal a comparable coverage of high-level taxonomic groups with the two methods using different primer sets targeting the same DNA region.

It is difficult to assign a correct taxonomic identity to tags that are too divergent from sequences available from the reference database, particularly when the database is not exhaustive enough, similar to the eukaryotic SSU rRNA database under investigation here. Thus, after BLAST searching, sequences that showed a query coverage or top match identity of  $\leq 95\%$  were excluded from further analyses. This stringent process removed about 30% of the MB and 37% of the TH trimmed tags but secured a more accurate taxonomic identification of the remaining reads. Indeed, a significant amount (up to about 30%) of fairly divergent sequence tags, with GAST



distances greater than 0.10, was also recovered in a previous study of protists using the 454 technology (Amaral-Zettler *et al.*, 2009). It is worth noting that as more sequences are accumulated in the database, the assigned taxonomic identity of tags with relatively low similarity to previous entries may change. Therefore, the results presented here were restricted to our current knowledge on picoeukaryotes.

Intrinsic errors of the 454 pyrosequencing technology could lead to overestimation of species richness. We followed the quality filtering strategy proposed by Huse *et al.* (2007), which ensures that the per-base error rates of GS20 pyrosequencing reads will be comparable or lower than the conventional Sanger sequencing method. However, an examination of the 16S rRNA genes of a single *Escherichia coli* strain by Kunin *et al.* (2009) suggested that diversity was still being overestimated in pyrosequencing data sets after the quality filtering process. According to their analysis, our diversity estimates at cluster distances of 0.10, 0.05 and 0.02 may have been overestimated by twofold, threefold and more than ninefold, respectively.

#### *Diversity of the picoeukaryotic assemblages*

Parametric rarefaction curves and nonparametric ACE and Chao1 estimators were used to estimate OTU richness in this study. The nonparametric richness estimators have been used in diversity studies of picoeukaryotes/protists only recently (Massana *et al.*, 2004; Zuendorf *et al.*, 2006; Countway *et al.*, 2007). Using a 95% phylotype cutoff, Countway *et al.* (2007) studied protistan diversity from 18S rRNA clone libraries and reported ACE and Chao1 values of around 400 for the euphotic zone of the Gulf Stream. Using 454 pyrosequencing technology and the same phylotype cutoff, Amaral-Zettler *et al.* (2009) obtained incidence-based ICE and Chao2 estimates of about 4000 for protistan diversity in a polluted estuary site off the coast of Massachusetts. For picoeukaryotic diversity, Massana *et al.* (2004) reported one of the highest Chao1 estimates of nearly 200 in an oligotrophic coastal site at Blanes Bay by grouping clones sharing the same RFLP patterns as unique OTUs. In this study, the calculated ACE and Chao1 estimates for picoeukaryotic diversity in subtropical coastal waters ranged from ~700 in the TH sample at a cluster distance of 0.10 to ~6600 in the MB sample at a cluster distance of 0.02. These values are generally much higher than those published by other investigators for picoeukaryotic/protistan diversity. However, the above comparison is rough as it is difficult to compare estimates based on different OTU defining and/or sequencing strategies. Also, in contrast to prokaryotes, the copy number of SSU rRNA genes varies widely among eukaryotes (Zhu *et al.*, 2005), which in turn may affect the estimates calculated. According to the correlation between cell length and SSU rRNA gene

copy numbers of eukaryotes estimated by QPCR, picoeukaryotes are expected to have less than 10 copies of SSU rRNA genes.

Rarefaction curves predict that additional sampling would lead to significantly higher estimates of total diversity at a cluster distance of 0.02, even by examining nearly 100 000 sequences per sample (Figure 2). As this threshold corresponds roughly to the genus/species level of picoeukaryotes (Romari and Vaultot, 2004), the results suggest that further increasing the sampling effort would reveal more genera/species of picoeukaryotes in the water samples. This finding confirms that the diversity of picoeukaryotes is unexpectedly high and still largely unknown (Mackey *et al.*, 2002).

A unimodal relationship between diversity and productivity for herbaceous vegetation has long been observed in terrestrial systems (Grime, 1973). Recent analysis of a global data set has revealed a similar pattern between marine phytoplankton diversity and phytoplankton biomass (Irigoiien *et al.*, 2004). Lefranc *et al.* (2005) used the cloning and sequencing method to study the genetic diversity of small eukaryotes (<5 µm) in three lakes in France and reported a higher diversity in the oligomesotrophic lake than in the oligotrophic and eutrophic lakes. In this study, both parametric rarefaction curves and nonparametric richness estimators revealed a higher diversity of picoeukaryotes in the oligomesotrophic MB than in the eutrophic TH (Table 3, Figure 2). This finding is in good agreement with that of Lefranc *et al.* (2005) and provides further evidence on the classic unimodal diversity–productivity relationship.

#### *Composition and distribution of the picoeukaryotic assemblages*

Water samples collected from two bays along the subtropical Pacific coast harbored different assemblages of picoeukaryotes, as revealed by the few shared top phylotype OTUs (Table 2) as well as the differential contribution by the most represented groups (Figure 1).

Prasinophytes recovered from this study (Table 2) and from other environmental genetic libraries (Worden, 2006) included mainly uncultured *Bathycoccus*, *Micromonas* and *Ostreococcus* spp. within the order Mamiellales. These organisms are more common in coastal areas than in open oceans (Not *et al.*, 2005). In our study, prasinophytes were more abundant in open waters of MB than in the eutrophic waters of the semi-enclosed TH (Figure 1). A similar result was found in our previous study using the cloning and sequencing strategy (Cheung *et al.*, 2008). Viprey *et al.* (2008) reported a greater dominance by *Micromonas* and *Ostreococcus* environmental sequences in relatively mesotrophic waters than in oligotrophic coastal waters. Taken together, these results suggest that these genera are well adapted to coastal waters of intermediate

productivity, although the role of water temperature should not be neglected (Lovejoy *et al.*, 2007).

Picobiliphytes are a recently defined algal group with a phycobiliprotein-containing plastid (Not *et al.*, 2007). Our previous study showed that picobiliphytes are prevalent in eutrophic waters (Cheung *et al.*, 2008), and this finding is supported by the results obtained in this study (Figure 1). The repeated recovery of picobiliphytes in our samples also provides further evidence on their prevalence in subtropical waters (Cheung *et al.*, 2008, Cuvelier *et al.*, 2008). NAGIII, a newly defined alveolates group, contains environmental sequences from oceanic and coastal waters (Guillou *et al.*, 2008). Since its recent discovery, little is known about its ecology. Our study provides information about its dominance in waters of lower productivity (Figure 1), but more evidence has to be provided for confirmation.

Most of the top phylotype OTU matches in the water sample from MB were from environmental studies performed in oceanic or coastal waters (Supplementary Table 2). In contrast, a noticeable number of OTUs in the water sample from TH were retrieved either from anoxic sediments or from water samples collected during algal blooms. For instance, the third most abundant phylotype OTU in the MB sample (uncultured prasinophyte clone PROSOPE.C1-80m.96) was retrieved from the Mediterranean Sea (Viprey *et al.*, 2008), whereas the third most abundant OTU in the TH sample (uncultured ciliate clone 9\_153) was recovered from the anoxic layer of the East Sea sediment in the northwestern Pacific (Park *et al.*, 2008). In addition, stramenopiles, which have been found abundantly in anoxic and extreme environments (Stoeck and Epstein, 2003; Luo *et al.*, 2005), seemed to dominate the TH sample (Figure 1). Because high levels of phytoplankton biomass could lead to hypoxia/anoxia (Roman *et al.*, 1993), the presence of OTUs with high adaptability to anoxic environments in the eutrophic waters of TH is not unexpected.

## Conclusions

In this study, we successfully applied the massively parallel 454 sequencing-by-synthesis technology to describe the composition and genetic diversity of marine picoeukaryotes off the subtropical western Pacific coast. Our results showed that, even by examining nearly a 100 000 sequences per sample, a greater sampling effort would likely harbor more taxa of picoeukaryotes. Water samples of different trophic status contained different high-level taxonomic groups and phylotype OTUs of picoeukaryotes. With the ultrahigh resolution power provided by 454 pyrosequencing, more in-depth examinations of picoeukaryotic diversity in various aquatic ecosystems are possible. In the future, this technology can be used to answer ecological

questions related to spatial and temporal patterns as well as to identify ecological responses to changing environmental conditions that require an accurate characterization of community composition.

## Acknowledgements

We thank YH Yung for sample collection, and CP Li and KC Cheung for technical assistance. The work was supported by research grants from the Research Committee, The Chinese University of Hong Kong.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4**: e6372.
- Bent SJ, Forney LJ. (2008). The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J* **2**: 689–695.
- Chau KW. (2007). Integrated water quality management in Tolo Harbor, Hong Kong: a case study. *J Cleaner Prod* **15**: 1568–1572.
- Cheung MK, Chu KH, Li CP, Kwan HS, Wong CK. (2008). Genetic diversity of picoeukaryotes in a semi-enclosed harbor in the subtropical western Pacific Ocean. *Aquat Microb Ecol* **53**: 295–305.
- Countway PD, Gast RJ, Dennett MR, Savai P, Rose JM, Caron DA. (2007). Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environ Microbiol* **9**: 1219–1232.
- Cuvelier ML, Ortiz A, Kim E, Moehlig H, Richardson DE, Heidelberg JF *et al.* (2008). Widespread distribution of a unique marine protistan lineage. *Environ Microbiol* **10**: 1621–1634.
- Díez B, Pedrós-Alió C, Massana R. (2001). Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67**: 2932–2941.
- Doyle JJ, Doyle JL. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.
- Elwood HJ, Olsen GJ, Sogin ML. (1985). The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol Biol Evol* **2**: 399–410.
- Forns X, Bukh J, Purcell RH, Emerson SU. (1997). How *Escherichia coli* can bias the results of molecular cloning: preferential selection of defective genomes of hepatitis C virus during the cloning procedure. *Proc Natl Acad Sci USA* **94**: 13909–13914.
- Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. (2009). Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J* **3**: 860–869.
- Grime JP. (1973). Competitive exclusion in herbaceous vegetation. *Nature* **242**: 344–347.

- Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R *et al.* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (Alveolata). *Environ Microbiol* **10**: 3349–3365.
- Hong Kong Environmental Protection Department (2003). *Marine Water Quality in Hong Kong for 2002*. Government Printer, Hong Kong.
- Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**: R143.
- Irgoien X, Huisman J, Harris RP. (2004). Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* **429**: 863–867.
- Kunin V, Engelbrekston A, Ochman H, Hugenholtz P. (2009). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.
- Lefranc M, Thnot A, Lepre C, Debroas D. (2005). Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Appl Environ Microbiol* **71**: 5935–5942.
- Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Li WKW. (1994). Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169–175.
- López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603–606.
- Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R *et al.* (2007). Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic seas. *J Phycol* **43**: 78–89.
- Luo Q, Krumholz LR, Najjar FZ, Peacock AD, Roe BA, White DC *et al.* (2005). Diversity of the microeukaryotic community in sulfide-rich Zoddletone Spring (Oklahoma). *Appl Environ Microbiol* **71**: 6175–6184.
- Mackey D, Blanchot J, Higgins H, Neveux J. (2002). Phytoplankton abundances and community structure in the equatorial Pacific. *Deep-Sea Res II* **49**: 2561–2582.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Massana R, Balague V, Guillou L, Pedrós-Alió C. (2004). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol Ecol* **50**: 231–243.
- Molvaer J, Knutzen J, Magnusson J, Rygg B, Skei J, Sorensen J. (1997). *Environmental Quality Classification in Fjords and Coastal Areas*. Statens Forurensningstilsyn TA-1467, Norway, p 36.
- Moon-van der Staay SY, De Wachter R, Vault D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607–610.
- Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W *et al.* (2005). Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50**: 1677–1686.
- Not F, Valentin K, Romari K, Lovejoy C, Massana R, Töbe K *et al.* (2007). Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**: 253–254.
- Park S-J, Park B-J, Pham VH, Yoon D-N, Kim S-K, Phee S-K. (2008). Microeukaryotic diversity in marine environments, an analysis of surface layer sediments from the East Sea. *J Microbiol* **46**: 244–249.
- Roman M, Gauzens AL, Rhinehart WK, White J. (1993). Effects of low oxygen water on Chesapeake Bay zooplankton. *Limnol Oceanogr* **38**: 1603–1614.
- Romari K, Vault D. (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49**: 784–798.
- Sherr EB, Sherr BF. (2000). Marine microbes. An overview. In: Kirchman, DL (ed). *Microbial Ecology of the Oceans*. Wiley-Liss: New York, pp 13–46.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Stoeck T, Epstein S. (2003). Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Appl Environ Microbiol* **69**: 2657–2663.
- Vault D, Eikrem W, Viprey M, Moreau H. (2008). The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795–820.
- Viprey M, Guillou L, Ferréol M, Vault D. (2008). Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.
- Wear RG, Thompson GB, Stirling HP. (1984). Hydrography, nutrients and plankton in Tolo Harbor, Hong Kong. *Asian Mar Biol* **1**: 59–75.
- Wong CK, Thompson GB. (2003). HPLC pigment analysis of marine phytoplankton during a red tide occurrence in Tolo Harbor, Hong Kong. *Chemosphere* **52**: 1633–1640.
- Worden AZ. (2006). Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43**: 165–175.
- Worden AZ, Not F. (2008). Ecology and diversity of picoeukaryotes. In: Kirchman, DL (ed). *Microbial Ecology of the Oceans* 2nd edn. Wiley-Liss: New York, pp 159–205.
- Zhu F, Massana R, Not F, Marie D, Vault D. (2005). Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52**: 79–92.
- Zuendorf A, Bunge J, Behnke A, Barger KJ-A, Stoeck T. (2006). Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. *FEMS Microbiol Ecol* **58**: 476–491.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)