ECOLOGY AND DIVERSITY OF HETEROTROPHIC PROTISTS IN THE COLUMBIA RIVER COASTAL MARGIN

By

Peter L. Kahn

A DISSERTATION

Presented to the Division of Environmental & Biomolecular Systems and the Oregon Health & Science University School of Medicine
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Environmental Science and Engineering

November 2014

Oregon Health & Science University

Institute of Environmental Health

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Peter L. Kahn

has been approved

Peter Zuber, Ph.D.
Research co-Advisor

Tawnya D. Peterson, Ph.D.
Research co-Advisor
Examination Committee Chair

Byron C. Crump, Ph.D.
Oregon State University
External Committee Member

Lydie Herfort, Ph.D.
Examination Committee Member

Holly M. Simon, Ph.D.
Examination Committee Member

Table of Contents

Acknowledgments

During my time at OHSU many people have had a significant impact on my research and my development as a scientist. First, I want to thank my advisors Dr. Peter Zuber and Dr. Tawnya Peterson, who from day one have offered invaluable support, guidance, editing, and patience. Thanks for always having your doors open and being quick with advice and solutions to problems throughout my time as your student. I am forever grateful for this opportunity. I would also like to thank the other members of my committee, Dr. Byron Crump, Dr. Holly Simon, and Dr. Lydie Herfort, for their scientific advice, flexibility, and support. Lydie also served as my frontline mentor during my REU internship at CMOP, and her enthusiasm and passion got me hooked on the world of protist ecology. I'd also like to thank Byron for providing DNA samples from outside the Columbia River system that assisted greatly in many of our assessments of protist biogeography.

I am grateful to all the members past and present of the Zuber/ Nakano and Needoba/ Peterson labs, who have always provided a welcoming atmosphere to work in. Special thanks go to Michelle Maier and Dr. Suzanne Delorenzo for walking with me down this path and being excellent lab mates, shipmates, and friends. It was always good to have someone, often literally, in the same boat.

I would also like to thank the captains, crews, and fellow scientific personnel of the *R/V Wecoma*, *R/V New Horizon*, *R/V Oceanus*, *R/V Barnes, R/V Magister*, *R/V Sailfin*, and the *M/V Forerunner*. This work included a large amount of ship time and sample collection that would not have been possible without your help. I'd also like to thank all of my former high school and undergraduate interns for all of their

contributions, hard work, and the excitement they showed towards their projects. Thanks also go to Nancy Christie, Vanessa Green, and Nievita Bueno-Watts, for all of their help with the logistics of graduate school.

Finally, I am forever grateful for my family, friends, and girlfriend Nadege for their help and support during my time at OHSU. I'd especially like to thank my parents, Fred and Patti Kahn, for their continual encouragement throughout my academic career.

This research was made possible through National Science Foundation funding of the Center for Coastal Margin Observation & Prediction (CMOP).

Abstract

Heterotrophic protists, a diverse group of microbial eukaryotes characterized by great morphological variability and extensive taxonomic representation, play important ecological roles in aquatic food webs as prey and predators. Difficulty in heterotrophic protist identification has often resulted in lumping them into broad groups, but there is a crucial need to develop methods that increase the spatial and temporal resolution of observations applied to particular organisms in order to discover the drivers of population structure and ecological function. This research characterizes the spatiotemporal distribution of heterotrophic protist assemblages in the Columbia River coastal margin (including the tidal freshwater reaches of the river, the chemical estuary, and the river plume in the adjacent coastal ocean) using DNA sequence and morphological approaches. I analyzed partial small subunit (SSU) rRNA gene sequences of heterotrophic protists from the Columbia River estuary and plume during spring and summer using metagenomic next-generation sequencing (NGS) technology (Illumina HiSeq) and PCR (polymerase chain reaction)-based Sanger sequencing. Ciliates were the dominant heterotrophic protists in the estuary during both spring and summer according to analysis of the Illumina amplicon sequences, with a seasonal transition of abundant species occurring from spring to summer, while the heterotrophic flagellate *Katablepharis* sp. dominated spring protist assemblages estimated by the Sanger method. In the river plume, the assemblage transitioned from one dominated by ciliates in the spring to one dominated by heterotrophic dinoflagellates in the summer. Absolute cell abundances of ciliates and heterotrophic dinoflagellates were determined using light

microscopy, and were strongly positively correlated with the relative proportions of sequences retrieved from the metagenomic dataset for both ciliates and heterotrophic dinoflagellates. However, correlations between the abundances of cells and SSU were weak when the PCR-based approach was used, suggesting that heterotrophic protist diversity is not adequately captured by this method. In contrast, metagenomics provided a reasonable estimate of protist diversity and abundance across a river-to-ocean continuum. Chapter Three details the discovery of a 332 base pair unique sequence element (USE) insertion in the large subunit rRNA gene of the Columbia River Estuary *Katablepharis* (*Katablepharis* CRE) that is not present in other katablepharids or any other organisms. Using USE-specific probes, I determined the spatial and temporal patterns of *Katablepharis* CRE in absolute abundance through quantitative approaches. The presence of USEs in several other protist taxa and the utility of these elements in tracking protist biogeography were determined in Chapter Four. USEs were detected in the parasitic dinoflagellate genus *Euduboscquella* and the flagellate groups Diplonemea and Cercozoa. The distributions of these taxa were tracked with USE-specific probes, and fine-scale genotypic differences were detected amongst closely related strains with putatively restricted biogeography. The USE presented here can be a useful tool for studying protist biogeography due to their highly specific nature, and has wide-reaching potential for a variety of environmental applications. Taken together, this dissertation provides the first detailed characterization of the seasonal distribution of heterotrophic protist assemblages in the Columbia River coastal margin, and presents new tools to track specific protist taxa distribution within a system and globally.

List of Tables

List of Figures

Chapter One: Introduction

"An extensive geographic range, for example, may be an artefact of inadequate taxonomic resolution, which combined with undersampling in the marine environment, could lead to unrecognized cryptic species being amalgamated into a morphospecies with artificially large distributions."

- Watts et al. 2010

The ecological role of heterotrophic protists in aquatic systems

Unicellular eukaryotic microorganisms, referred to hereafter as protists, are the dominant forms of eukaryotes in terms of cell abundance, biomass and species diversity, spanning all five super groups of the eukaryotic domain (Adl et al. 2005). However, the majority of our knowledge of eukaryotic biology is related to the study of animals, land plants, and fungi (Patterson 1999, Katz 2012). Protists, including autotrophs, mixotrophs, and heterotrophs, are an essential component of aquatic ecosystems since they are ubiquitous and abundant in all types of habitats (Sherr & Sherr 2002). The heterotrophic fraction of protist assemblages has often been a poorly characterized and overlooked part of the food web in pelagic systems. Since the seminal paper by Pomeroy (1974) that attributed most of the respiration in marine water to microorganisms, however, it has become increasingly clear that heterotrophic protists play significant roles in the transfer and recycling of carbon and nutrients in pelagic food webs. Numerous field studies have demonstrated that heterotrophic protists, including ciliates, dinoflagellates and

nanoflagellates (2-20 μm) are the most important grazers of phytoplankton biomass in freshwater, brackish, and marine systems (Lessard & Murrell 1998, Lehrter et al. 1999, Calbet & Landry 2004, Hambright et al. 2007, Calbet 2008).

Heterotrophic protists are also the most important bacterivores via the microbial loop (Fig. 1.1) (Azam et al. 1983), and this predation can be a significant source of mortality of planktonic bacteria and a major shaping force for the taxonomic structure of bacterial communities (Strom 2000, Jürgens & Matz 2002, Jürgens et al. 2008). As unicellular predators, heterotrophic protists have growth and metabolic rates comparable to their prey, which allows them to maintain tight trophic coupling with prey populations (Sherr & Sherr 1994, Calbet & Landry 2004). Moreover, heterotrophic protists can also serve as a considerable food source for higher trophic levels, such as mesozooplankton, (Gifford 1991, Calbet & Landry 1999), and act as remineralizers of essential nutrients through the excretion of nitrogen and phosphorus compounds as well as trace metals such as iron (Caron 1991, Dolan 1997).

Coastal margins are an interesting place to investigate heterotrophic protist ecology and population dynamics, since they are highly variable aquatic systems with fluctuating environmental conditions (e.g. salinity, freshwater discharge) that may change considerably over tidal, seasonal, and annual timescales. Heterotrophic protists are often particularly abundant and diverse in coastal and estuarine waters (Rollwagen-Bollens et al. 2006), where they display rapid restructuring of their assemblages in response to changing environmental conditions inherent to coastal margin environments (Vigil & Countway 2009). Changes in protist assemblage patterns may also provide an early warning signal for monitoring the health of coastal margin ecosystems (Odum 1985,

Paerl et al. 2009), and often may precede larger-scale shifts in ecosystem functions, such as changes in nutrient cycles, food webs, and fisheries (Paerl & Peierls 2008). Furthermore, protist biodiversity may be a useful metric to assess environmental health and serve as an indicator for environmental conditions since high diversity has been linked to increased ecosystem stability and resilience against the establishment of invasive species (McGrady-Steed & Morin 2000). Hence, the characterization of protist biodiversity is necessary for the understanding of ecosystem functioning.

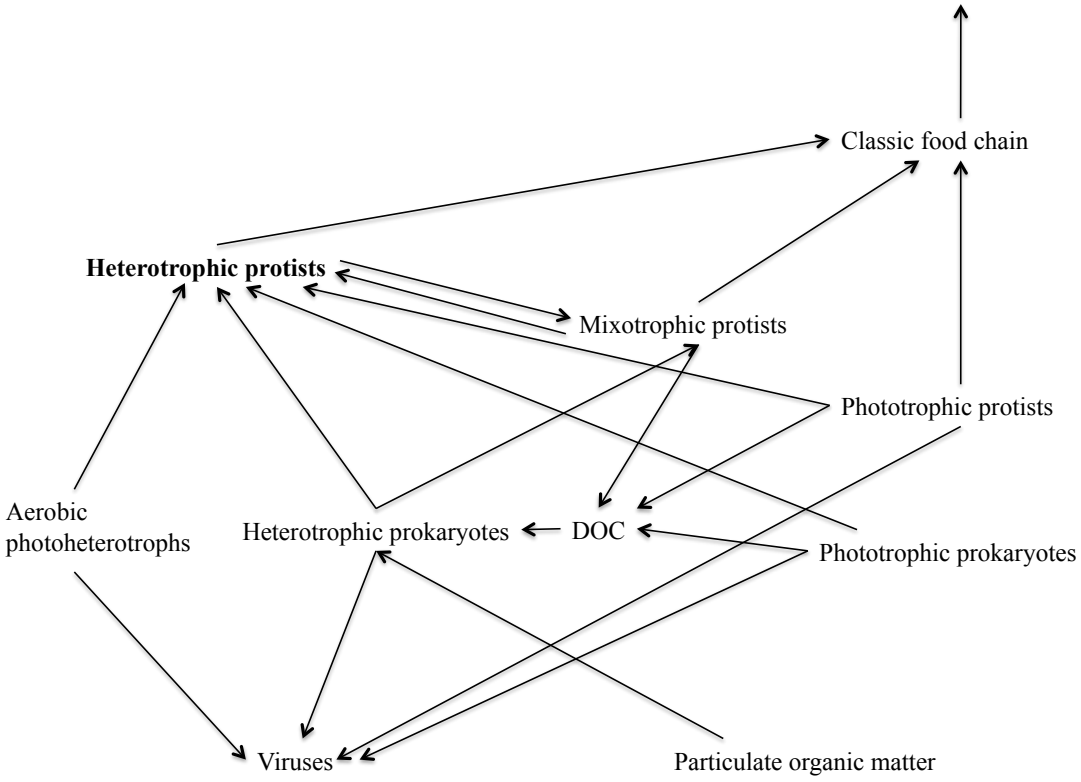


Figure 1.1. The microbial loop with heterotrophic protists indicated in bold. DOC refers to dissolved organic matter. Redrawn from Azam et al. (1983) and Fenchel (2008).

In coastal margin systems, heterotrophic protist assemblages are typically dominated by ciliates, dinoflagellates, and nanoflagellates (Wörner et al. 2000, Strom et al. 2001, Sherr & Sherr 2002, Rollwagen-Bollens et al. 2006, 2011), and often display seasonal variations in assemblage structure, which can be driven by changes in prey availability, such as phytoplankton blooms (Kiss et al. 2009), or top-down grazing pressure (Jurgens & Stolpe 1995). Heterotrophic nanoflagellates, a polyphyletic group spanning several eukaryotic supergroups, have been found to be important primary consumers and bacterivores associated with particles in coastal margin environments (Ploug et al. 2002, Domaizon et al. 2003, Slapeta et al. 2006b). Two ciliate subclasses, Oligotrichia and Choreotrichia (including loricate tintinnid and aloricate species) tend to dominate heterotrophic ciliate assemblages from coastal and estuarine waters (Sherr & Sherr 1987, Dolan 1991, Muylaert et al. 2000, Dolan & Gallegos 2001, Doherty et al. 2007), while members of the order Peniculida (e.g. *Stokesia*, *Frontonia*, *Paramecium*) are found to be dominate in freshwater environments (Urrutxurtu et al. 2003). Over 8000 species of ciliates have been described morphologically, however the number of ciliate species yet to be discovered may be an order of magnitude higher based on DNA sequences from environmental samples related to ciliates (Adl et al. 2007, Foissner et al. 2009). Ciliates are often very fragile organisms to handle, and it has proven difficult to obtain or maintain cultures of many species (Andreoli et al. 2009). Thus, the undersampling of their diversity through culture-based morphological approaches make them an ideal group to explore with culture-independent DNA sequence approaches.

In addition to ciliates and nanoflagellates, heterotrophic dinoflagellates have also been recognized as a significant component of the heterotrophic protist biomass in coastal

margin systems (Sherr & Sherr 2007). Lessard and Swift (1986) revealed that about half of the species found in dinoflagellate assemblages from marine waters did not bear chloroplasts and were consumers of other cells. Subsequent work has shown that athecate dinoflagellates, particularly gymnodinoid forms (e.g. *Gymnodinium*, *Gyrodinium*), dominate heterotrophic dinoflagellate assemblages in marine (Strom 1991) and estuarine (Sherr et al. 1991) waters. Together, ciliates and heterotrophic gymnodinoid dinoflagellates are often the primary grazers of large diatoms in coastal systems (Strom et al. 2001). This is contrary to the earlier assumptions based on the classical food chain, that a transition from microzooplankton to mesozooplankton, such as copepods, accompanies a shift to larger phytoplankton (Moloney & Field 1991). Many heterotrophic dinoflagellates are generalists in terms of their prey selectivity and can consume a wide range of cells, including cells as large or larger than themselves (Lessard 1991, Jeong et al. 2004, 2010, Horner et al. 2005), and can survive longer than ciliates when the ciliates' preferred prey become scarce (Jakobsen & Hansen 1997).

Heterotrophic protist diversity: historical perspective, systematics, and contribution of DNA sequence-based approaches

Despite their importance, the phylogenetic relationship among heterotrophic protists and the scale of protist diversity are poorly resolved (Patterson 1999, Bass & Cavalier-Smith 2004, Caron et al. 2012). A major obstacle to improving protist classification is the lack of a comprehensive and established species concept that includes the morphological, genetic, physiological, and ecological differences of protist taxa (Schlegel & Meisterfeld 2003). The "biological species" concept (Mayr 1942), which

defines species in terms of reproductive isolation of natural populations, does not apply to many protists that reproduce by inbreeding or asexually (Schlegel & Meisterfeld 2003). The "ecological species" concept supported by ecologists to characterize organisms that occupy the same ecological niche has also been troublesome for protists as there has been no consensus on objective criteria to classify "ecospecies" (Fenchel & Finlay 2004). Traditionally, the systematics of heterotrophic protists has been based on morphological characterization through microscopic analysis. Cells that share a characteristic set of morphological features are considered members of the same "morphospecies." However, the presence of cryptic species (Moreira & López-García 2002) that have similar morphologies but differ genetically or physiologically, and the difficulty in cultivation of most heterotrophic protist species (Lim et al. 1999), have made assessments of protist diversity using the morphospecies concept challenging.

While both ciliates and heterotrophic dinoflagellates have a relatively rich history of morphological and taxonomic description, the taxonomic composition of heterotrophic nanoflagellates in aquatic systems is still in its infancy (Cleven & Weisse 2001). Identification based upon morphological traits is difficult for heterotrophic nanoflagellates due to their small cell size and lack of characteristic features. Due to these limitations, heterotrophic protists have often been lumped into broad groups that may not effectively describe their functional and genetic diversity. Conversely, cultivation-independent molecular taxonomy based on DNA sequences has emerged as a powerful tool for broad and relatively rapid assessments of protist assemblage composition and diversity (Caron 2009b). With the advent of DNA sequence-based diversity estimates, numerous studies (López-García & Rodríguez-Valera 2001, Staay et al. 2001, Stoeck &

Epstein 2003, Bass & Cavalier-Smith 2004) have uncovered unexpectedly high levels of protist diversity (including autotrophic, mixotrophic, and heterotrophic taxa) from a wide range of environments. These studies have revealed novel protist lineages (Rodríguez-Martínez et al. 2009), identified cryptic species within morphologically defined species (Škaloud & Rindi, Pfandl & Chatzinotas 2009, Lundholm et al. 2012), and provided a means for distinguishing species lacking distinctive morphologies (Nassonova et al. 2010).

The majority of molecular-based field studies have used traditional DNA sequencing (Sanger 1977) of the small subunit (SSU) or 18S rRNA gene to characterize the diversity of protists (López-García & Rodríguez-Valera 2001, Staay et al. 2001, Stoeck & Epstein 2003, Bass & Cavalier-Smith 2004). The SSU rRNA gene has been widely used as a 'barcode' in protist diversity estimates, since its slow evolutionary rate, presence in all eukaryotic organisms, and strong representation in public database makes it suitable for taxonomic surveys (Bass & Boenigk 2011). However, it has been noted that the limited throughput of Sanger sequencing-based assessments of protist diversity, which generally relies on dozens to hundreds of DNA sequences per sample, may underestimate diversity and fails to detect rare taxa (Bent & Forney 2008, Shokralla et al. 2012). Furthermore, biases inherent to Sanger sequencing methods, such as PCR amplification biases, primer selectivity, and cloning biases have made interpretation of these datasets difficult (Vargas et al. 2009).

In recent years, next-generation sequencing (NGS) technologies, such as 454-pyrosequencing (Margulies et al. 2005) and Illumina Hiseq platforms have begun to replace Sanger sequencing and have made major advances in molecular phylogenetics

due to the massive amounts of sequencing information captured that allows high throughput sequence-based characterization of microbial communities. These new technologies have been applied to protist diversity studies, and have led to very high estimates of species richness and biodiversity in a variety of aquatic environments (Stock et al. 2009, Heywood et al. 2010, Medinger et al. 2010, Eiler et al. 2013, Bachy et al. 2013, Santoferrara et al. 2014). For example, analysis of NGS sequence datasets have uncovered previously undetected members of the protistan 'rare biosphere' and determined they comprise a larger and more diverse component of the protist community than was previously estimated (Sogin & Morrison 2006). When applied to estuarine and coastal system, NGS approaches have shown greatly enhanced richness and diversity estimates when compared to morphological or Sanger sequencing methods (Monchy et al. 2012, Santoferrara et al. 2014).

Despite the fact that DNA sequence-based approaches have greatly improved our knowledge of protist diversity, many significant questions remain (Stoeck & Stock 2009). Quantification of protist abundance based on SSU rRNA sequence analysis has been challenging due to highly variable rRNA gene copy numbers across eukaryotes (Gong et al. 2013). Furthermore, morphological and DNA sequence-based approaches have generally been used separately, and often contain large discrepancies when compared to each other (Savin et al. 2004, Medinger et al. 2010). At this point there is not an established link between the genetic and traditional morphospecies descriptions; however, as DNA sequence-based approaches become the standard for protist diversity and taxonomic assignment, it is important to combine morphological and molecular

methods to validate DNA sequence datasets and determine the best practices in interpretation of diversity and abundance estimates from each approach.

Assessing biogeography through protist taxonomy

The issue of protist biogeography is central to our estimation of global protist diversity (Foissner 2007, Caron 2009a). Biogeography is the study of distribution of biodiversity over space and time to determine where organisms live, at what abundance, and why (Martiny et al. 2006). The notion "Everything is everywhere, but the environment selects," contends that microbial taxa are ubiquitous, and found anywhere that there is suitable habitat. It originated from the ideas of Beijerinck (1913) and was defined in its current form by Baas Becking (1934). It became an early paradigm in microbial ecology, and recently was further contextualized in terms of protist distribution (Fenchel & Finlay 2004). This classical cosmopolitan view, termed the "ubiquity theory", has recently been fiercely debated with an opposing viewpoint, the "moderate endemicity distribution theory", which contends that at least some protists have a restricted geographic distribution (Foissner 2006). A key assumption within the ubiquity model is that protists' large population sizes and short generation times result in high dispersal rates, with an apparent lack of dispersal barriers that prevent speciation as a result of geographic isolation (allopatric speciation) and a capability to disperse over long distances (e.g. through resting cyst formulation), thus resulting in their cosmopolitan distrubution (Finlay 2002). The ensuing growth of protists are then determined by selective pressures of environmental heterogeneity, where environment selection outweighs dispersal limitation and the level of gene flow is high and outweighs any variation caused by adaptation, genetic drift or mutation (Lacap et al. 2011). This

viewpoint has been supported primarily through taxonomic units defined by morphology (i.e. morphospecies) of putatively globally distributed morphospecies (Finlay 2002, Finlay & Fenchel 2004).

Supporters of the moderate endemicity theory, on the other hand, have argued that the spatial scaling of at least some protist species is determined by geographic distance, and dispersal barriers are a greater factor in determining distribution compared to environmental selection (Weisse 2008). For example, geographic distance has been found to be a more accurate predictor of community variability than environmental conditions in fungal communities at a regional-scale (Green et al. 2004). Furthermore, resting cysts from ciliate species, such as the freshwater oligotrich *Pelagostrombidium*, rapidly lose their viability when stored under cold (1-6°C) and dark conditions (Müller 2002), and resting cysts of protists from moist rainforests survived in drought conditions for only a few weeks (Foissner 2006). This suggests that the viability of resting cysts, the most likely means of long distance dispersal (Müller 2000) of protist cells, is likely more restricted than previously thought (Foissner 2007), leading to more limited gene flow between habitats and greater opportunity for allopatric and parapatric speciation (Weisse 2008). For example, high genetic differences in the internal transcribed spacer (ITS) region between populations of the dinoflagellate *Peridinium limbatum* have been observed from adjacent freshwater bodies in Northern Wisconsin, suggesting a faster rate of evolutionary change relative to dispersion and that the neighboring populations are diverging genetically under conditions of limited gene flow (Kim et al. 2004). The moderate endemicity theory has been supported through the use of "flagship species," morphospecies that have distinguishing morphological features and whose presence or

absence can be easily determined during microscopic inspection of water samples (Foissner 2006). Such examples include several species of the testate amoebae (e.g. *Alocodera cockayni*) that are putatively restricted to the south of the Tropic of Cancer desert belt (Smith et al. 2008) as well as many ciliate species (Foissner et al. 2009).

Molecular evidence for ubiquity versus endemism has been split, with DNA sequence analysis studies of protist biogeography supporting both sides of the debate (Bass et al. 2007, Caron 2009a, McManus & Katz 2009, Sharma & Rai 2010, Fontaneto & Brodie 2011, Lara & Heger 2011, Fontaneto & Hortal 2013). However, a key point from these DNA sequence studies, coupled with the morphological assessments, is that the scale of taxonomic resolution is vital when assessing protist biogeography (Mitchell & Meisterfeld 2005, Heger & Mitchell 2009, Bass & Boenigk 2011). Protist distribution patterns are strongly correlated with taxon delimitations, since higher taxon levels (e.g. class and family) have wider distribution patterns, while an individual cell has a distinct occurrence at a single point (Bass & Boenigk 2011). An essential question, then, is: what level of taxonomic resolution is appropriate for the study of protist biogeography?

Biogeographic surveys based on either morphospecies (Fenchel & Finlay 2004) or SSU phylotype suggest that many protists, particularly the most abundant taxa, have cosmopolitan distributions (Fenchel & Finlay 2006, Slapeta et al. 2006a, Darling & Wade 2008). However, more discriminating genetic markers, such as ITS regions of the rRNA gene (Bass et al. 2007, Stoeck et al. 2008) or divergent domains (D1-D12) of the LSU rRNA gene (Wylezich et al. 2010), have shown several lineages with restricted distributions within a single cosmopolitan SSU phylotype. For example, several ITS-derived lineages within the same SSU phylotype differ in biogeographic distribution as

well as phenotype with respect to distinguishing characteristics such as salinity tolerance, morphology, and propensity for cyst formation (Bass et al. 2007). Thus, an appropriate level of taxonomic resolution for protist biogeography is one that most accurately reflects the functional biogeography of protist taxa (Bass & Boenigk 2011).

The Columbia River coastal margin

This dissertation presents a study that focuses on the Columbia River coastal margin, where the population dynamics of heterotrophic protists is virtually unknown. The Pacific Northwest coastal region is strongly influenced by the Columbia River, which is the second largest river in the continental U.S with a mean annual discharge of 7300 m^3s^{-1} (Neal 1972, Hickey et al. 1998). The river drains a 670,000 km^2 watershed encompassing six U.S. states and one Canadian province and culminates in an expansive plume that delivers river-borne dissolved and particulate matter that has a large impact on the chemical, physical, and biological characteristics of the adjacent Washington and Oregon coastal waters (Barnes et al. 1972, Frey et al. 1984, Sullivan et al. 2001). Although the Columbia River flow has been altered dramatically over the past two centuries, due to dam construction, channel diversion, irrigation, and dredging resulting in a decreased overall river discharge and dampened seasonal flow variability (Sherwood & Jay 1990), a seasonality in discharge volume remains. The peak flow is associated with river discharge occurring in the late spring during the spring freshet, and its lowest flow occurs in the late summer to early autumn.

Previous studies (Haertel et al. 1969, Frey et al. 1984, Small et al. 1990, Sullivan et al. 2001) have shown the algal assemblages of the Columbia River estuary to be dominated by freshwater diatoms, while the mixotrophic ciliate *Mesodinium* blooms

annually during the late summer (Herfort et al. 2011). However, the combination of high turbidity and short residence time of the estuary, 2-5 days (Neal 1972), tend to suppress primary productivity and keep phytoplankton standing stocks low relative to freshwaters or to the river plume (Frey et al. 1984). Furthermore, the Columbia River estuary contains large amounts of allochthonous detritus, mainly originating from river inputs, that drive ecosystem processes (Simenstad et al. 1990). Based on microscopic cell counts, the most abundant species of phytoplankton found in the estuary were members of the freshwater genera (such as *Asterionella*, *Melosira*, and *Stephanodiscus*), but also members of marine genera, (e.g. *Thalassiosira*), highlighting the fact that estuarine assemblages are influenced by both marine and freshwater inputs. Pigment analysis by high-performance liquid chromatography (HPLC) indicated that chlorophytes and cryptophytes also constitute a minor portion of the phytoplankton community in the Columbia River estuary, particularly in the freshwater reaches (Sullivan et al. 2001).

Previous studies on heterotrophs in the Columbia River estuary have characterized the mesozooplankton community (0.2-2 mm), which is composed primarily of freshwater, oligohaline, and euryhaline forms (Haertel et al. 1969, Simenstad et al. 1990) and a bacterial assemblage, which is dominated by particle-attached bacteria that have been shown to account for 90% of total bacterial production and to correlate with particulate organic carbon concentration and turbidity (Crump et al. 1998). One study suggests that nanoflagellates and oligotrich ciliates are the most common form of heterotrophic protists in the estuary (Crump & Baross 1996), while the nanoflagellate genus *Katablepharis* has been detected in waters associated with estuarine turbidity maxima through DNA sequence analysis (Herfort et al. 2011). However, heterotrophic

protist diversity and variability, seasonally and across the river-to-ocean continuum, as well as their role as primary consumers, is poorly characterized.

The Columbia River plume is influenced by cycles of upwelling and downwelling favorable winds. Seasonal coastal upwelling of nitrate during the summer months, coupled with nutrient supply from the adjacent Columbia River, can make the Columbia River plume highly productive, although previous studies (Sherr et al. 2005) have clearly demonstrated that the upwelling of nutrient-rich waters in summer results in strong phytoplankton growth as estimated by levels of the pigment, chlorophyll *a*, few studies have characterized the assemblages that are generated through the accelerated productivity. Frame and Lessard (2009) showed that as upwelling relaxes and nutrients are depleted, diatoms are replaced by flagellated phytoplankton forms such as dinoflagellates. Another study showed that heterotrophic dinoflagellates such as members of the gymnodinoids (*Gymnodinium* and *Gyrodinium*), are the most abundant form of primary consumer during upwelling off the Oregon coast, and their abundance covaries with diatom abundance, suggesting predation of bloom-forming diatoms by gymnodinoids (Neuer & Cowles 1994).

Overview of dissertation chapters

Chapter Two of this dissertation presents a study that aimed to characterize the spatiotemporal distribution of heterotrophic protists in the Columbia River coastal margin using DNA sequence and morphological approaches. Specifically, I analyzed SSU rRNA gene sequences using metagenomic (Illumina HiSeq) and amplicon (Sanger sequencing) approaches, and I performed microscopic cell counts to determine the distribution of two major heterotrophic protist groups, the ciliates and dinoflagellates, in the Columbia River

estuary (including tidal freshwater areas) and its plume during the spring and summer. These approaches were employed to answer the following research questions: How well do two molecular methods, Sanger sequencing and Illumina HiSeq, capture the diversity of two major heterotrophic protist groups, the ciliates and heterotrophic dinoflagellates, and are they in agreement with morphological assessment of cell abundance? What changes in heterotrophic protist assemblages occur across a river-to-ocean continuum over the spring and summer months and between years?

Chapter Three of this dissertation further expands on the Sanger sequencing dataset presented in Chapter Two, with particular emphasis on *Katablepharis* CRE, a heterotrophic nanoflagellate that dominated brackish (salinity ~15) protist assemblages during the spring. Further sequence analysis of the large subunit (LSU) rRNA gene sequence of *Katablepharis* CRE was conducted and revealed a 332 base pair unique sequence element (USE) that is not present in any other organism in the National Center for Biotechnology Information (NCBI) database. The presence of this element was further characterized to answer the following research questions: What is the spatial and temporal distribution of organisms bearing this unique element amongst *Katablepharis* CRE and other katablepharids in the Columbia River coastal margin? Is this unique element found in any other organisms in the Columbia River coastal margin and/or elsewhere? Can the unique element be used as a taxonomic marker to facilitate ecological studies of *Katablepharis* CRE? Through utilization of this USE as a taxonomic marker for *Katablepharis* CRE, it was found to be biogeographically restricted to the Columbia River coastal margin.

The unique sequence element (USE) presented in **Chapter Three** is a potentially powerful tool for studying protist biogeography, given its highly discriminating nature and ability to track specific protist taxa distribution both within a system and globally. The utility of the USE as a tool for protist biogeography is further assessed in **Chapter Four**. In this chapter I identified USEs among other protist taxa found primarily in the Columbia River coastal margin, but also in several other coastal margin environments (Amazon River, the Chesapeake Bay, the Susquehanna River and the Beaufort Sea Lagoon), through sequence analysis of LSU D2 region sequences generated through Sanger sequencing and NGS technology (Illumina Hiseq).

Finally, **Chapter Five** presents a summary of the disseration, highlighting the key findings from each chapter as well as suggestions for future studies. The appendix at the end of this dissertation contains supplementary figures as well as a multivariate analysis (**S6**) to link the protist assemblages described in **Chapter Three** to environmental variables.

Chapter Two: Spatiotemporal distribution of heterotrophic protists in the Columbia River coastal margin, USA [1]

2.1. INTRODUCTION

Protists, including autotrophs, mixotrophs, and heterotrophs, play important roles in energy transfer and nutrient transformations in aquatic food webs (Mallin 1994). Since the first introduction of the term "microbial loop" by Azam et al. (Azam, Fenchel, & Field 1983), there has been an increasing appreciation for the important role that heterotrophs play in determining the fate of primary production. For example, in aquatic systems, heterotrophic protists transfer organic carbon produced by major sources of living biomass (e.g. phytoplankton, bacteria, and archaea) to higher trophic levels such as metazoans (Cho & Azam 1990) and facilitate the rapid recycling of nutrients back to primary producers (Sherr & Sherr 2007). Large ciliates, primarily oligotrichs and tintinnids, along with gymnodinoid dinoflagellates, often carry out the majority of grazing on large diatoms, even in coastal systems (Strom et al. 2001), contrary to earlier

[1] Material from this chapter is being prepared for publication: Kahn P, Herfort L, Crump BC, Zuber P, and Peterson TD. Spatiotemporal distribution of heterotrophic protists in the Columbia River coastal margin, USA.

assumptions that a transition from microzooplankton to macrozooplankton (e.g. copepods) accompanies a shift to larger phytoplankton (Moloney & Field 1991).

Protist populations often exhibit dramatic responses to changes or fluctuations in environmental conditions, generally manifested by changes in the composition of their assemblages. For example, a decrease in the abundance of primary consumers and the subsequent decrease of top-down controls on primary producers has been noted in communities subject to warming (Petchey et al. 1999). These types of trophic level changes could alter the transfer of carbon in aquatic ecosystems and decrease the resilience of a community to environmental changes by reducing biodiversity (Loreau et al. 2001). Thus, tracking fluctuations in the assemblages of protists may offer clues about ecosystem processes, including trophic status and susceptibility of an ecosystem to perturbation (Odum 1985, Zinger et al. 2012, Bradford et al. 2013).

Heterotrophic protists comprise a diverse group characterized by great morphological variability and broad taxonomic representation, which poses challenges for estimating the diversity (including functional diversity and genetic potential) of a given assemblage. The advent of culture-independent molecular-based techniques such as traditional DNA sequencing (Sanger 1977) through small subunit rRNA gene sequence clone libraries has greatly advanced estimates of protist diversity across a wide range of environments (López-García & Rodríguez-Valera 2001, Staay et al. 2001, Stoeck & Epstein 2003, Bass & Cavalier-Smith 2004). However, these studies have often been limited by biases inherent to the method, such as PCR amplification biases, primer selectivity, and multiple rRNA gene copy numbers (Vargas et al. 2009). In addition, it is difficult to obtain sufficient coverage of the number of different sequences present in an

environment to assess protist diversity and species richness using low-throughput sequencing (Terrado et al. 2009, Shokralla et al. 2012).

Major advances in molecular phylogenetics have been possible because of the massive amount of sequencing information captured through next-generation sequencing (NGS) technology. These advances have allowed for far more extensive sequence-based characterization of microbial communities, leading to very high estimates of species richness and biodiversity (Stock et al. 2009, Heywood et al. 2010, Medinger et al. 2010, Eiler et al. 2013, Bachy et al. 2013, Santoferrara et al. 2014). The large number of sequences provided by NGS have begun to provide new insights into contested yet important concepts surrounding protist biogeography and diversity, particularly in debates engaged over the idea that 'everything is everywhere' (Finlay 2002) and in the characterization of the microbial 'rare biosphere' (Sogin & Morrison 2006). Estimates of protist diversity based on rRNA sequence divergence are challenging, however, due to highly variable rRNA gene copy numbers across eukaryotes (Gong et al. 2013) and sequence datasets often containing large discrepancies when compared to traditional morphological analysis (Medinger et al. 2010).

Estuaries often harbor particularly diverse microbial assemblages due to the extensive mixing of water masses along the river-to-ocean continuum. In addition, there are often multiple sources of organic matter, including terrestrial detritus, to fuel heterotrophic microbial growth. Heterotrophic protists play important roles as grazers and food sources to the metazoan food web (Bazin et al. 2013), and they display rapid and seasonal shifts in response to the highly dynamic environmental conditions inherent to coastal margins (Vigil & Countway 2009). In the San Francisco Bay, heterotrophic

protists have been shown to consume as much as 73% of the phytoplankton standing stock in the spring and 15% in the summer (Rollwagen-Bollens et al. 2011). At least some of the variability in heterotrophic protist populations is driven by prey availability (Kiss et al. 2009).

This study focuses on the Columbia River coastal margin, where the population dynamics of heterotrophic microbes (particularly eukaryotes) is poorly characterized. Previous studies in the estuary have focused on the autotrophs (Haertel et al. 1969, Frey et al. 1984, Small et al. 1990, Sullivan et al. 2001) and have shown that the algal assemblages of the Columbia River estuary are dominated by freshwater diatoms. However, the combination of high turbidity and short residence time of the estuary (2-5 days) (Neal 1972) tends to suppress primary productivity and keep phytoplankton standing stocks low relative to freshwaters or to the river plume (Frey et al. 1984). Furthermore, the Columbia River estuary contains large amounts of allochthonous detritus, mainly originating from river inputs, that drive ecosystem processes (Simenstad et al. 1990). One study (Crump & Baross 1996) revealed that nanoflagellates and oligotrich ciliates are the most common form of heterotrophic protists in the estuary, but their diversity and variability, both seasonally and across the river-to-ocean continuum, as well as their role as primary consumers, is virtually unknown.

In this study, we used an NGS metagenomic approach (Illumina HiSeq) to uncover the spatiotemporal distribution of heterotrophic protists in the Columbia River coastal margin. In particular, we analyzed small subunit 18S rRNA gene sequences (hereafter referred to as SSU) to determine the distribution of two major heterotrophic protist groups, the ciliates (Phylum Ciliophora) and dinoflagellates (Class Dinophyceae),

in the Columbia River estuary (including tidal freshwater areas) and its plume during the spring and summer. This dataset was then compared to a similar analysis of nearly full-length SSU heterotrophic protist sequences that were amplified through Sanger sequencing methods, presented in further detail in **Chapter Three**. In order to validate our DNA sequence data and provide more quantitative information, microscopic cell counts of ciliates and dinoflagellates were also performed for each sample. These approaches were employed to answer the following research questions: How well do two molecular methods, metagenomic and amplicon sequencing, capture the diversity of two major heterotrophic protist groups, the ciliates and heterotrophic dinoflagellates, and are they in agreement with morphological assessment of cell abundance? What are the changes in heterotrophic protist assemblage composition across a river-to-ocean continuum over the spring and summer?

2.2. METHODS

2.2.1. Study area

The Pacific Northwest coast is strongly influenced by the Columbia River, which is the second largest river in the U.S. by flow (Simenstad et al. 1990). The Columbia River estuary consists of both a tidal brackish water region (from river and ocean water mixing) and a freshwater tidal region that extends further upstream. The volume of freshwater discharge from the Columbia River strongly influences the river-to-ocean continuum, mainly through the modulation of the salinity intrusion into the estuary (Chawla et al. 2008). The river culminates in an expansive plume that delivers river-borne dissolved and particulate matter to the Washington/Oregon coastal waters (Frey et

al. 1984, Sullivan et al. 2001). The Columbia River plume enters into the northeastern Pacific, which is characterized by wind-driven upwelling in the late spring and summer, followed by relaxation and downwelling in the fall (Huyer 1983, Hickey et al. 2005). During seasonal upwelling, nutrient-rich water is brought to the surface, which fuels diatom blooms (Small & Menzies 1981, Frame & Lessard 2009).

2.2.2. Sample collection

Water samples for SSU sequence analyses and examination by light microscopy were collected 17 times in spring and summer between 2007 and 2012 at sites in the Columbia River coastal margin along the river-to-ocean continuum from sites with distinct salinities (0, 15, >30) (Fig. 2.1). Table 2.1 provides the details of location, salinity, temperature, and depth for all samples used in the SSU sequence analysis. Freshwater (salinity = 0) and brackish water (salinity = 15) samples were collected within the Columbia River estuary. Plume water (defined as salinity = 28-31) was collected outside the Columbia River bar. Water was collected either from Niskin bottles attached to a Seabird 911plus CTD (conductivity-temperature-depth) rosette system or with a high volume-low pressure centrifugal pump attached to a PVC hose lowered alongside a Seabird 911plus CTD system.

2.2.3. Nucleic acid extraction

After collection, water was immediately filtered through 0.2 μm-pore-size Sterivex filters (PES, ESTAR, Millipore) using a peristaltic pump set to low speed until it became clogged (1-5 L). One Sterivex filter was collected per sample. We chose not to pre-filter in order to capture particle-attached unicellular microorganisms. Water was

manually forced gently out of each filter using an air-filled syringe and 2 mL of the fixative RNAlater (Ambion) was added to the Sterivex before freezing at -80°C onboard the ship. DNA was extracted from the particulate material from each sample using a phenol-based extraction as described in Herfort et al. (2011). Extraction was performed twice for each Sterivex filter and the two total extracts were pooled.

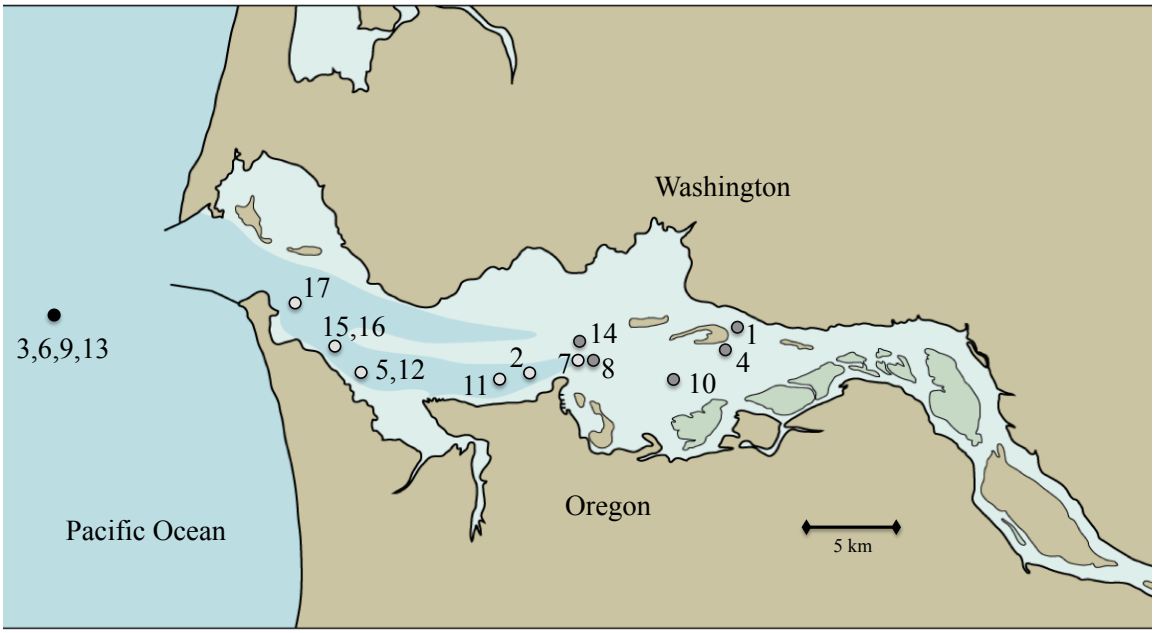


Figure 2.1. Locations used for samples used in this study. Numbers correspond to Sample Number in Table 1. Black circle = plume (salinity of ~28-31), Light gray circle = brackish (salinity of ~15), Dark gray circle = freshwater (salinity of 0).

Table 2.1. Sample number, date, salinity, temperature, and depth for samples used in Illumina HiSeq, Sanger sequencing, and light microscopy. Location of samples is indicated in Fig. 1. Water was collected aboard several vessels, including M/V *Forerunner* (estuary April 2007 & 2009, May 2012), R/V *Barnes* (estuary August 2007) and R/V *Wecoma* (all other samples).

Sample Number	Date	Salinity	Temperature (° C)	Depth (m)	Illumina HiSeq	Sanger	Light Microscopy
1	Apr 2007	0.3	8.9	2.0	X	X	X
2	Apr 2007	14.2	9.7	10.0		X	X
3	Apr 2007	28.0	10.1	2.0	X	X	X
4	Aug 2007	0.1	20.6	2.0	X	X	X
5	Aug 2007	14.0	17.9	12.0	X	X	X
6	Aug 2007	28.3	15.1	2.0	X	X	X
7	Apr 2008	14.4	8.9	17.0		X	X
8	Apr 2008	0.2	8.9	2.0	X	X	X
9	Apr 2008	30.0	8.8	2.0	X	X	X
10	Jul 2008	0.1	19.6	2.0	X	X	X
11	Jul 2008	14.7	14.5	2.0	X	X	X
12	Sep 2008	13.6	14.8	2.0	X	X	X
13	Sep 2008	31.2	11.3	2.0	X	X	X
14	Sep 2008	0.3	14.8	2.0	X	X	X
15	Apr 2009	14.8	9.1	13.0	X		
16	Apr 2010	16.2	9.8	6.0	X		
17	May 2012	15.6	10.4	2.0	X		

2.2.4. SSU sequence analysis of Sanger sequencing amplicon dataset

Nearly full-length heterotrophic protist SSU sequences were amplified in a Bio-Rad DYAD PCR thermocycler, cloned and sequenced as described in Kahn et al. (Kahn et al. 2014) using the eukaryote-specific primers EukA (5'-AACCTGGTTGATCCTGCCAGT-3'; position ~1-20) and EukB (5'-TGATCCTTCTGCAGGTTCACCTAC-3'; position ~1780-1800) (Díez et al. 2001). For each SSU clone, contigs (continuous consensus sequences) were assembled using

Geneious v5.3 software (Drummond et al. 2010) from the sequencing reads of the EukA/ EukB primer set as well as an internal primer, 528f (5'-CCGCGGTATTCCAGCTC-3'; position ~528-548) (Elwood et al. 1985) to generate a ~1800 bp sequence. All SSU heterotrophic protist sequences generated from the Euka/ EukB primer set have been deposited to NCBI with accession numbers KJ925152-KJ926058 and have already been published in Kahn et al. (Kahn et al. 2014).

2.2.5. Illumina sequencing

Nucleic acid samples were sent for Illumina sample preparation and for two lanes of Illumina HiSeq 2000 paired-end 2x151 bp sequencing at the Oregon State University Center for Genome Research & Biocomputing. Resulting raw output files were then processed by the HiSeq 2000 CASAVA pipeline where they were demultiplexed and returned as .fastq files. Sequences were then submitted to the MG-RAST 3.0 pipeline (Meyer et al. 2008), where poor quality bases (below phred score of 20) were trimmed, paired-end files were combined and overlapping paired-end reads were merged for automated annotation. Filtered and annotated sequence data are available through MG-RAST accession numbers 4533700-4533806.

2.2.6. SSU sequence analysis of metagenomic dataset

Sequences from the Illumina HiSeq dataset that were annotated as eukaryotic SSU sequences were submitted to SILVA-NGS (Quast et al. 2013) for operational taxonomic unit (OTU) clustering, defined as ≥ 98% sequence identity, which has been used previously for approximate genus level analysis of protist assemblages (Bachy et al.

2013, Bazin et al. 2013). SSU Sequences were classified taxonomically against the SILVA SSU database, release 115 (Quast et al. 2013) and sorted according to their class and genus. Diversity of ciliate and dinoflagellate sequences was estimated as OTU richness using the non-parametric Chao 1 richness estimator (Chao 1984). Rarefaction values were calculated at the genus level in R using the iNEXT package (Hsieh et al. 2013) for ciliate and dinoflagellate sequences.

2.2.7. Light microscopy

For microscopic cell counts, 40 mL of water was fixed with formalin (final concentration: 4%) at room temperature for 1 h and stored at -80°C. Samples were then slowly thawed and cells within 25 mL were allowed to settle overnight using the Utermöhl method (Utermöhl 1958). While several taxa of protists were identified and counted (including diatoms, chlorophytes, cryptophytes), only ciliates and dinoflagellates were used in the comparison to our SSU sequence analysis, since they were the dominant heterotrophic forms observed. Using an inverted microscope (Leica DMIL) at 200x-400x magnifications, an average of 300 cells were identified and grouped according to their taxonomy. The limit of detection was estimated as 1.3 cells mL^{-1}.

2.3. RESULTS

2.3.1. Morphological and sequence-based assessment of ciliates and dinoflagellate population dynamics

Partial SSU sequences (~200 bp) and nearly complete SSU sequences (~1800 bp) were retrieved using metagenomic (Illumina HiSeq) and amplicon-based Sanger sequencing approaches, respectively, from freshwater (salinity of 0), brackish (salinity of ~15), and plume (salinity of ~28-31) samples collected during the spring and summer in the Columbia River coastal margin (Table 2.1, Fig. 2.1). Sequences attributable to ciliates and dinoflagellates were dominant, together accounting for > 65% of total heterotrophic protist sequences in all samples (Fig. 2.2). During the spring, the highest proportions of heterotrophic protist sequences in the metagenomic dataset were attributable to ciliate classes throughout the system (i.e., at all salinities; Fig. 2.3A). In the summer, however, ciliate sequences dominated the freshwater and estuarine sites, but comprised only 24% of heterotrophic protist sequences in the plume samples. Instead, dinoflagellate sequences dominated in the plume (62% of heterotrophic protist SSU sequences).

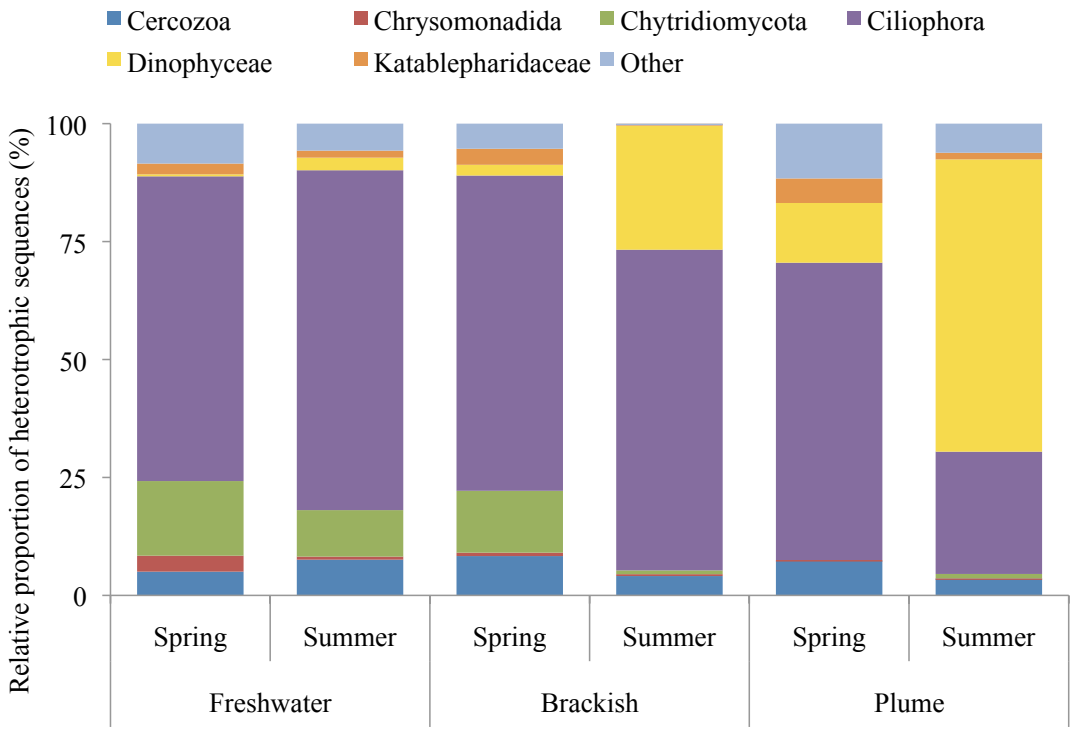


Figure 2.2. Relative proportions of heterotrophic protist sequences detected in the metagenomic (Illumina HiSeq) dataset for water collected in the Columbia River estuary and its plume. Freshwater = salinity of 0; Brackish = salinity of ~15; Plume = salinity of 28-31. 'Other' category designates sequences associated with the following protist groups: Apicomplexa, Bicosoecida, Centroheliozoa, Diplonemea, Kinetoplastea, Labyrinthulomycetes, Protalveolata, and uncultured marine stramenopile (MAST) clades. Sequences related to Ciliophora and Dinophyceae forms combined for greater than 65% of the total heterotrophic protist sequences in all samples.

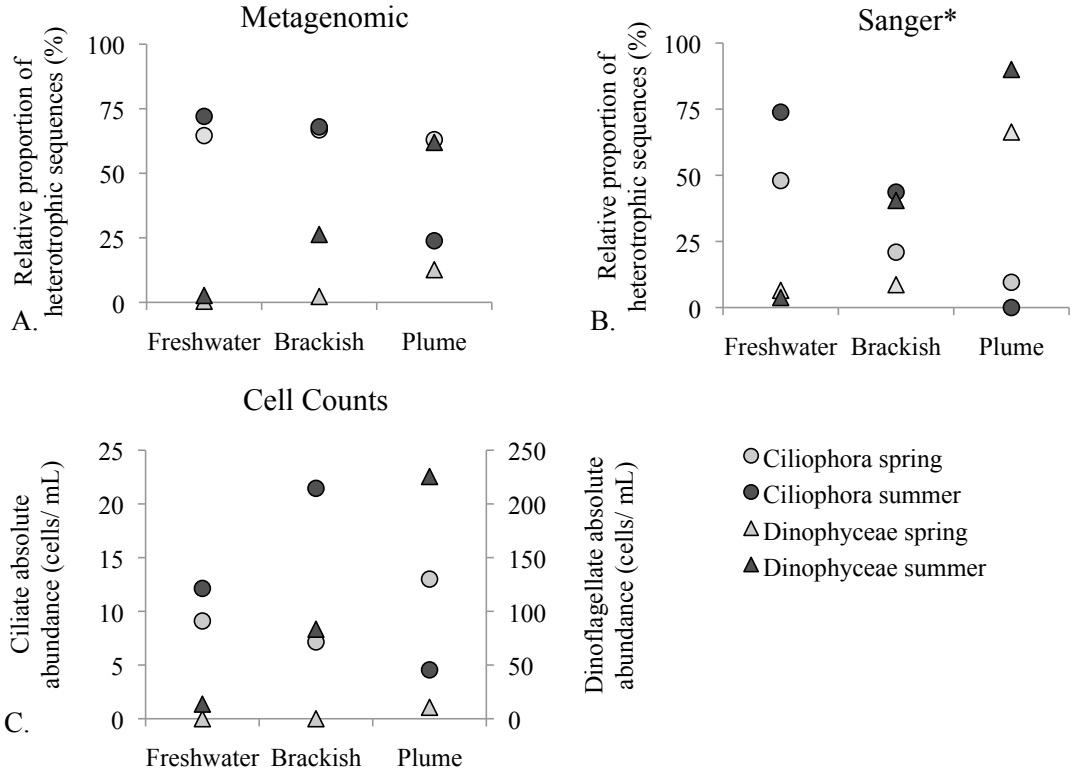


Figure 2.3. A. Proportion of total ciliate and dinoflagellate SSU sequences relative to total heterotrophic protist SSU sequences based on the metagenomic (Illumina HiSeq) dataset; B. Proportion of total ciliate and dinoflagellate SSU sequences relative to total heterotrophic protist SSU sequences based on the Sanger sequencing amplicon dataset; C. Absolute abundance of ciliates and dinoflagellates based on light microscopy. * Data adapted from Chapter Three.

Sequences related to the ciliate class Spirotrichea were present at high proportions in all samples, most of which were associated with the genera *Rimostrombidium* and *Tintinnidium* (subclass Choreotrichia) as well as *Strombidium* (subclass Oligotrichia) (Fig. 2.4A, C). Litostomatea and Oligohymenophorea sequences also accounted for a large proportion of ciliate sequences in freshwater and brackish samples, which consisted

primarily of *Didinium*, *Vorticella* and *Stokesia* genera, respectively. *Vorticella* sequences were numerous in the spring freshwater samples, while *Stokesia* sequences were particularly numerous in the summer freshwater and brackish samples, where they represented 39% and 58% of ciliate sequences, respectively. Heterotrophic dinoflagellate sequences consisted primarily of *Gymnodinium* and *Gyrodinium* genera, while *Peridinium* (subclass Peridiniphycidae*)* and *Amphidinium* (subclass Gymnodiniphycidae) sequences accounted for most of the other dinoflagellate sequences (Fig. 2.4B, D).

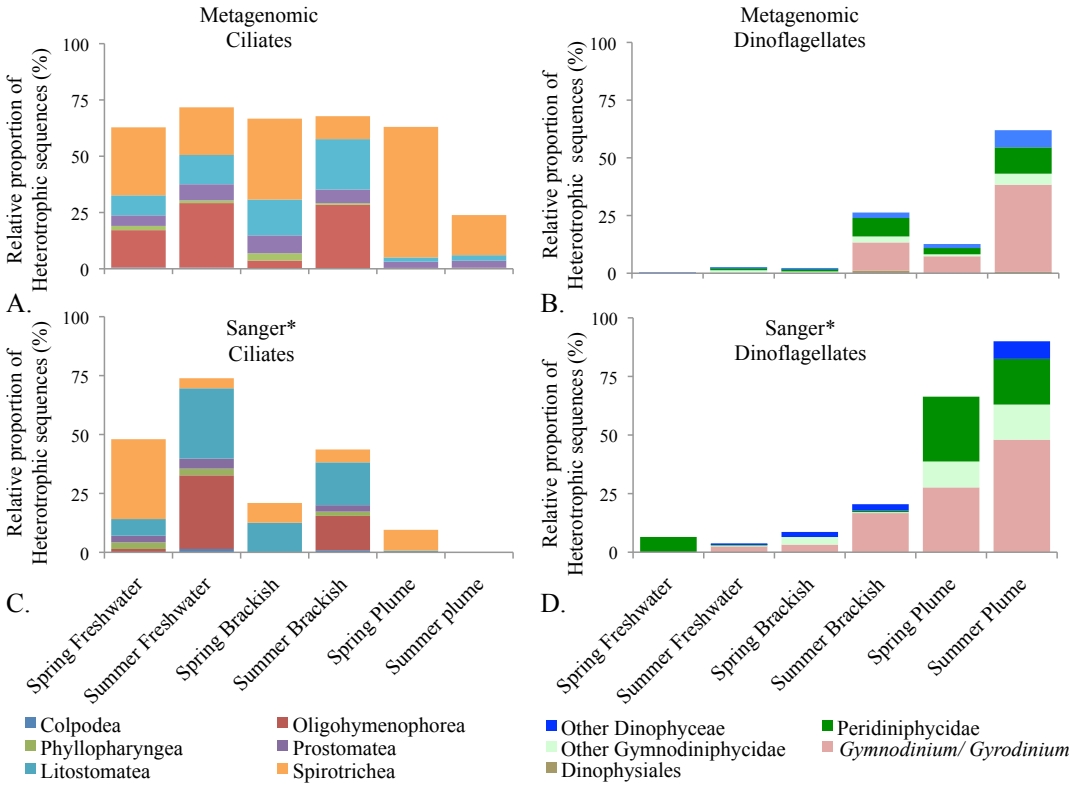


Figure 2.4. A. Proportion of ciliate classes relative to total heterotrophic protist SSU sequences based on the metagenomic (Illumina HiSeq) dataset; B. Proportion of dinoflagellate subclasses relative to total heterotrophic protist SSU sequences based on the metagenomic dataset; C. Proportion of ciliate classes relative to total heterotrophic

protist SSU sequences based on the Sanger sequencing amplicon dataset; D. Proportion of dinoflagellate subclasses relative to total heterotrophic protist SSU sequences based on the Sanger sequencing amplicon dataset. * Data adapted from Chapter Three.

In the Sanger dataset, ciliate sequences accounted for a much smaller proportion of heterotrophic protist sequences in all spring samples compared to the metagenomic dataset (Fig. 2.3B). Instead, the brackish water samples collected in the spring contained high proportions of *Katablepharis* sequences, a heterotrophic nanoflagellate, while the heterotrophic fraction of protists in the spring plume samples was dominated by dinoflagellate sequences.

To relate metagenomic and Sanger datasets to cell abundance, absolute abundances of ciliates and dinoflagellates were determined. Dinoflagellates were present at relatively low abundances in all spring samples and increased between the spring and summer (83 cells mL^{-1} and 225 cells mL^{-1} in the summer brackish and plume samples, respectively; Fig. 2.3C). Greater variability in ciliate abundances was observed among the different salinities in summer (21 cells mL^{-1} in the brackish sample versus 4 cells mL^{-1} in the plume sample; Fig 2.3C) compared to the spring (7-13 cells mL^{-1} across all salinities). There was a strong linear positive correlation between the relative proportion of sequences retrieved in the metagenomic dataset and cell abundance for ciliates (r = 0.75, P = 0.08) and dinoflagellates (r = 0.98, $P < 0.0001$) (Table 2.2). Although there were linear positive relationships between the proportion of amplicon sequences attributable to ciliates and dinoflagellates and their cell abundances, it was not significant for either (r = 0.52, P = 0.44 for ciliates, and r = 0.77, P = 0.16 for dinoflagellates).

Table 2.2. Correlation (Pearson's product moment) of ciliate and dinoflagellate cell abundance with proportion of ciliate and dinoflagellate sequences for metagenomic (Illumina HiSeq) and Sanger sequencing amplicon datasets. Significant correlations indicated in bold (a = 005).

	Dataset	*r*	*P*
Ciliates	**Metagenomic**	**0.752**	**0.084**
	Sanger	0.519	0.439
Dinoflagellates	**Metagenomic**	**0.987**	**0.001**
	Sanger	0.774	0.158

2.3.2. Richness estimates and rarefaction analysis of metagenomic and Sanger SSU sequence datasets

Estimates of richness for metagenomic and Sanger datasets were based on OTUs of 98% sequence similarity groups, which has been used previously for approximate genus level analysis of protist assemblages (Bachy et al. 2013, Bazin et al. 2013). Rather than species richness, we refer to these as "OTU richness." Rarefaction curves were generated and OTU richness estimated with the non-parametric Chao 1 index (Chao 1984) for the subset of ciliate and dinoflagellate sequences from each sequencing approach. Ciliate OTU richness was highest in the freshwater samples and decreased across the river-to-ocean continuum in the spring and summer using both sequence datasets (Fig. 2.5). Dinoflagellate OTU richness showed the opposite pattern, in which richness was lowest in the freshwater samples in both spring and summer (Fig. 2.5). Rarefaction analyses of ciliate sequences suggest that neither metagenomic nor Sanger datasets fully sampled ciliate diversity, since neither curve reached an asymptote (Fig. 2.6

A, C). The Sanger dataset was particularly poor at capturing diversity, as an extrapolation to 2n (two times the number of sequences) does not suggest a flattening of its slope. Rarefaction analyses suggest that dinoflagellate diversity was captured well in the metagenomic and Sanger datasets, as both curves are close to saturation (Fig. 2.6 B, D).

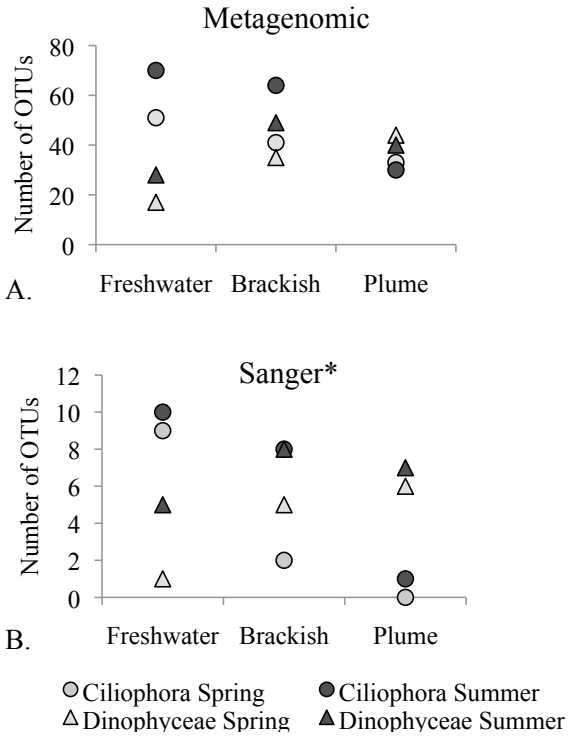


Figure 2.5. A. Number of ciliate and dinoflagellate OTU's (98%) detected in the metagenomic (Illumina HiSeq) dataset; B. Number of ciliate and dinoflagellate OTU's (98%) detected in the Sanger sequencing amplicon dataset. *Data adapted from Chapter Three.

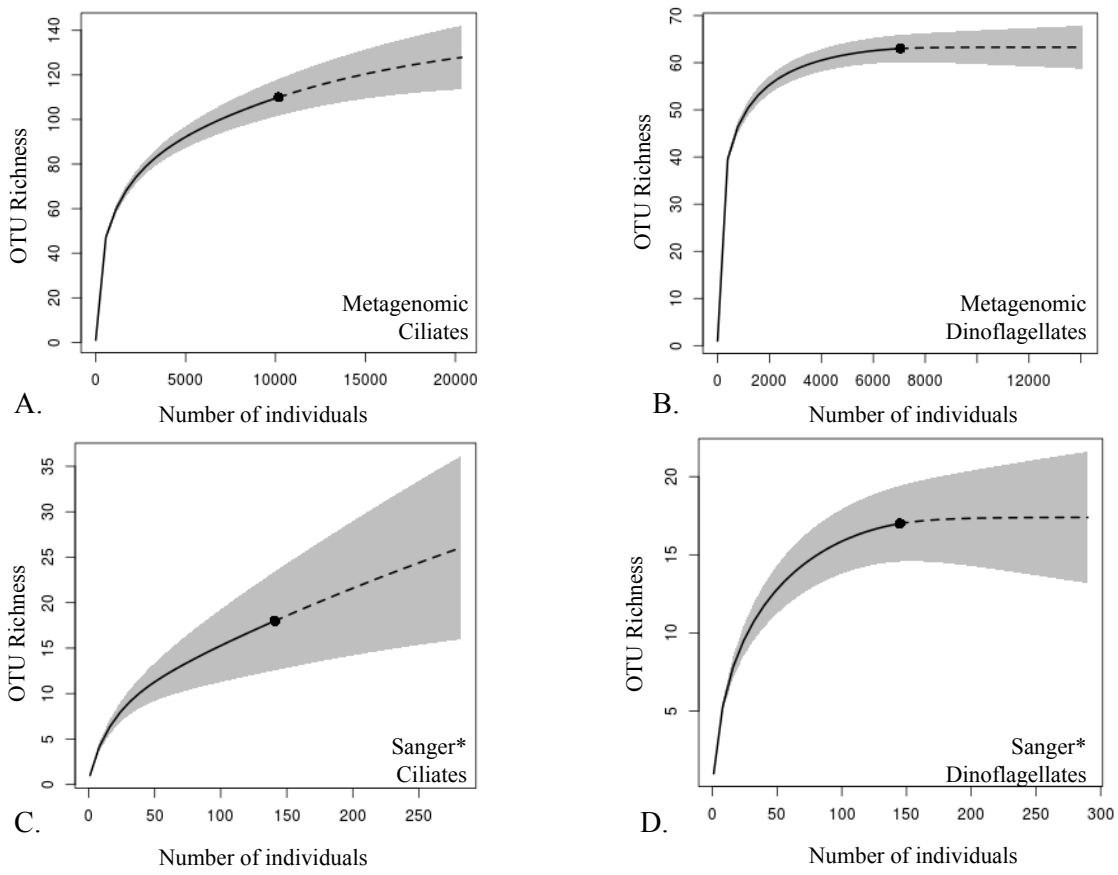


Figure 2.6. Rarefaction analysis with OTU's defined as 98% with richness estimated by the Chao 1 index with 95% confidence interval indicated by shaded area. Dashed lines indicate extrapolated values carried out to 2n, or two times the numbers of individual sequences (Chao & Jost 2012) for: A. Total ciliate SSU sequences from the metagenomic (Illumina HiSeq) dataset; B. Total dinoflagellate SSU sequences from the metagenomic dataset; C. Total ciliate SSU sequences from the Sanger sequencing amplicon dataset; D. Total dinoflagellate SSU sequences from the Sanger sequencing amplicon dataset. *Data adapted from Chapter Three.

2.4. DISCUSSION

2.4.1. Ciliate and heterotrophic dinoflagellate diversity estimated through Sanger sequencing and metagenomic datasets

SSU sequence analysis with NGS methods, as well as a traditional cloning/ Sanger sequencing approach, have been used to uncover unexpectedly high levels of protist diversity and taxon richness over the past several years in a wide range of environments (Stock et al. 2009, Heywood et al. 2010, Medinger et al. 2010, Eiler et al. 2013, Bachy et al. 2013, Santoferrara et al. 2014). Highly variable rRNA gene copy number amongst protist taxa have made it challenging to quantify protist abundance through environmental DNA sequence analysis (Vargas et al. 2009), and it is common to detect discrepancies between taxonomic assignations made using DNA sequence-based estimates and morphological characteristics (Zhu et al. 2005). For example, although ciliate and dinoflagellate sequences often dominate the heterotrophic protist assemblages according to SSU sequence estimates of diversity, these two groups tend to be poorly represented in assessments based on morphological features (Savin et al. 2004). However, significant correlations between microscopic cell abundance (not proportion of cells) and NGS data of broad protist taxa (phytoplankton) (Eiler et al. 2013) as well as ciliate species (*Balonion planktonicum*) (Medinger et al. 2010) has been noted in other comparative studies. We found that the relative proportions of both ciliate and dinoflagellate sequences from the metagenomic dataset sequenced with Illumina HiSeq technology were strongly positively correlated with their absolute cell abundances, suggesting that our NGS data estimates absolute abundance of ciliate and dinoflagellate

cells well and is suitable for characterizing transitions of these heterotrophic protist groups across seasons and across the river-to-ocean continuum.

There was a weak positive correlation between proportions of ciliate sequences from the Sanger sequencing dataset and ciliate absolute abundance. Lower proportions of ciliates were observed in samples within the Sanger dataset compared to the metagenomic dataset and very low OTU richness was estimated, particularly in the plume samples. Rarefaction analysis of ciliate sequences from the Sanger dataset also failed to reach an asymptote, suggesting that ciliate diversity was undersampled. This undersampling is not uncommon for Sanger sequencing datasets of protist assemblages (Terrado et al. 2009). However, this undersampling, coupled with potential PCR biases and variable gene copy number (Kahn et al. 2014), suggests that heterotrophic protist diversity is not adequately captured by analysis of sequences retrieved by the Sanger method, but that analysis of sequences produced using NGS methods provides a reasonable estimate of protist diversity and abundance across a wide range of aquatic environments. Sanger sequencing, although not always truly representative of the microorganisms that we can count under the microscope (as shown in the present paper), can provide important characterization of the heterotrophic protists that are not easily observed via microscopy, such as *Katablepharis* CRE (Kahn et al. 2014). More quantitative methods (qPCR, FISH) can then be employed to determine their spatial and temporal patterns in absolute abundances.

2.4.2. Seasonal variation of heterotrophic protist assemblages across the river-to-ocean continuum in the Columbia River coastal margin

The Columbia River estuary has previously been classified as a diatom-dominated system (Hobson 1963, Haertel et al. 1969, Frey et al. 1984, 1990, Sullivan et al. 2001). However, these observations were based primarily on classification of autotrophic taxa, and relatively little is known about heterotrophic protist diversity in the system. Ciliates were the dominant heterotrophic taxa throughout the estuarine samples based on Illumina amplicon sequences, with a transition from *Tintinnidium* and *Rimostromidium* (Spirotrichea) sequences in the spring to *Stokesia* (Oligohymenophorea) sequences in the summer. *Stokesia* is a mixotrophic ciliate that preys on chlorophytes (Berninger 1986). An increase in *Stokesia* abundance during the summer has been noted in other estuarine and freshwater systems, primarily driven by an increase in temperature and algal biomass, specifically chlorophyte (Kiss et al. 2009). The increase of *Stokesia* sequences in the Columbia River estuary during the summer could potentially be linked to increased prey biomass, because an increase in chlorophyte abundance from spring to summer was observed through our microscopic analysis with freshwater cell concentrations increasing from 10 cells mL^{-1} to 51 cells mL^{-1} and brackish cell concentrations increasing from 8 cells mL^{-1} to 33 cells mL^{-1}, respectively (data not shown).

While heterotrophic protist assemblages of the freshwater portion of the estuary consisted primarily of ciliates during both spring and summer, a transition towards dinoflagellates was evident in the plume samples between the spring and the summer. The most numerous heterotrophic dinoflagellate sequences were attributed to *Gyrodinium* and *Gymnodinium*, which have been shown to be dominant primary consumers during

upwelling off the Oregon coast (Neuer & Cowles 1994) and North Pacific (Strom et al. 2001). This succession from ciliates to heterotrophic dinoflagellates has been observed in other coastal systems (Löder et al. 2012) and the early elevation in ciliate abundance has been attributed to their higher growth rates compared to dinoflagellates (Strom & Morello 1998), which allows ciliates to respond more rapidly to enhanced phytoplankton biomass during the spring. However, heterotrophic dinoflagellates are generally able to consume a more diverse range of prey than ciliates, which likely results in greater persistence of their populations than those of ciliates when prey items become less abundant (Sherr & Sherr 2007). A similar transition occurs among the autotrophs in the Columbia River plume, as diatoms are replaced by dinoflagellates and more flagellated forms when nutrients become depleted (Frame & Lessard 2009).

Heterotrophic dinoflagellate sequences also increased in abundance within the estuary during the summer, likely as a result of their increased proportions in the plume and the strong seasonality of river discharge in the Columbia River. As river flow decreases late into the summer and early autumn, saline water is able to intrude further upstream (Sherwood & Jay 1990). In turn, more marine protists are likely transported further into the estuary. The mixing of freshwater and marine protist forms has also been observed within diatom populations of Columbia River estuary turbidity maxima, and implies multiple sources of organic matter being delivered into the estuary (Herfort et al. 2011).

This study, coupled with Kahn et al. (Kahn et al. 2014), provides one of the first characterizations of the seasonal distribution of heterotrophic protist assemblages across the river-to-ocean continuum in the Columbia River coastal margin. Through DNA

sequence analysis and traditional morphological techniques we determined the major heterotrophic protists and described the spatiotemporal succession of these assemblages. Given the large amounts of particulate organic matter in the system, heterotrophic protists likely play an important role in determining the fate of particulate organic matter in the Columbia River estuary. Addressing this critical gap in knowledge can aid in identification of trophic linkages within the aquatic food web and be useful for monitoring and management purposes. Continued spatial and temporal monitoring of heterotrophic protists will improve our estimates of protist diversity and our understanding of biogeochemical and ecosystem function in the coastal margin systems.

Chapter Three: Discovery of a *Katablepharis* sp. in the Columbia River estuary that is abundant during the spring and bears a unique large ribosomal subunit sequence element [2]

3.1. INTRODUCTION

Heterotrophic protists play significant roles in pelagic food webs as bacterivorous and herbivorous consumers (Pomeroy 1974, Azam 1983), as food sources for organisms at higher trophic levels such as metazoans (Gifford 1991), and as remineralizers of essential nutrients such as nitrogen and phosphorus (Caron et al. 1990). Heterotrophic protists, particularly small cells (<20 mm), are often difficult to assign taxonomically using light or electron microscopy, as many of them lack distinctive morphological features or characteristic pigments. As a result, they have often been placed into broad groups, such as "heterotrophic nanoflagellates" (2-20 mm), which could include a wide range of organisms of different taxonomic groups bearing different metabolic potentials and that play varied roles in aquatic food webs. The collection of established cultures of heterotrophic protists is also not likely to be representative of the dominant cells in the environment, as organisms that are easily cultured are often found at low abundance in

[2] Material from this chapter has been published: Kahn P, Herfort L, Peterson TD, Zuber P (2014) MicrobiologyOpen 3(5): 764-776

natural assemblages (Lim et al. 1999). Since the advent of culture-independent molecular-based techniques, numerous studies (Diez et al. 2001, Lopez-Garcia et al. 2001, Moon-van der Staay et al. 2001, Stoeck et al. 2003, Bass & Cavalier-Smith 2004) have reported unexpectedly high levels of diversity within the autotrophic, mixotrophic, and heterotrophic protist populations in diverse aquatic environments. For example, these studies have uncovered several novel lineages within the bacterivorous marine stramenopiles that can comprise a large proportion of protist populations and can be responsible for up to 10% of bacterivory and nutrient remineralization in the upper ocean (Massana et al. 2006).

The Columbia River coastal margin, which includes freshwater, brackish, and saline environments along a river-to-ocean continuum, is an ideal place to investigate the diversity of heterotrophic protists because of the broad range of environmental conditions that exist over a relatively narrow geographical region. The Columbia River coastal margin provides habitats for a wide variety of ecologically and economically important species, such as salmonids and various types of shellfish (Roegner et al. 2011). The Columbia River estuary is characterized by large amounts of allochthonous detritus from the adjacent river and ocean that is the primary source of organic matter driving ecosystem processes, with allochthonous organic matter supporting up to 84% of secondary production by the estuarine microbial populations (Simenstad et al. 1990). Yet, the overwhelming majority of research conducted on protist assemblages in the Columbia River (including the saline and freshwater reaches of the estuary) thus far has been focused on the autotrophic fraction, which is dominated by freshwater diatoms (Haertel et al. 1969, Frey et al. 1984, Lara-Lara et al. 1990, Small et al. 1990, Sullivan et al. 2001).

Previous studies on heterotrophs in the Columbia River estuary have shown that the mesozooplankton (0.2-2 mm) are composed primarily of freshwater, oligohaline, and polyhaline forms (Haertel et al. 1969, Simenstad et al. 1990), while particle-attached bacteria accounted for up to 90% of heterotrophic bacterial activity and are an important part of the estuarine food web (Crump et al. 1998). One study (Crump & Baross 1996) suggested that nanoflagellates and oligotrich ciliates are the most common form of heterotrophic protist in the estuarine turbidity maximum, but that study was limited in spatial and temporal extent. Relatively little is known about the composition and ecological role of heterotrophic protists in the river, estuary, and plume environments, despite their likely importance in organic matter transformations within aquatic food webs that link microbial activity and higher trophic levels (Sherr & Sherr 1994, Arndt et al. 2000). Given the large amounts of particulate organic matter in the system, heterotrophic nanoplankton could play an important role in the fate of particulate organic matter in the Columbia River estuary. This critical gap in knowledge prevents the identification of trophic linkages within the aquatic food web, which could be used to inform management decisions. Higher spatial and temporal resolution of monitoring for taxa of interest, as well as improved estimates of protist diversity, are necessary in order to determine the drivers of population structure as well as biogeochemical and ecosystem function.

A molecular approach can offer valuable new insights into protist assemblage structure and diversity in the Columbia River coastal margin, particularly for the heterotrophs. To evaluate heterotrophic protist assemblages, small subunit (SSU) rRNA gene clone libraries were analyzed for water samples collected at salinities of ~0 and ~15

in the Columbia River estuary, and salinities of ~28-31 in the plume in April and August 2007 and in April, July and September 2008. Further analysis of the large subunit (LSU) rRNA gene sequences was conducted for the highly dominant heterotrophic flagellate in the mid-salinity water SSU rRNA gene clone libraries in both April 2007 and 2008, a unique katablepharid henceforth referred to as *Katablepharis* CRE (Columbia River Estuary). This analysis uncovered a 332 base pair unique sequence element (USE) within the D2 region of the LSU that shows no significant similarity to any LSU sequences in the National Center for Biotechnology Information (NCBI) database and displays an elevated GC content compared to its associated SSU and LSU rRNA sequences (data retrieved on January 10, 2014). The presence and diversity of this element were further examined to answer the following research questions:

1. What is the spatial and temporal distribution of organisms bearing this unique element amongst *Katablepharis* CRE and other katablepharids in the Columbia River coastal margin?
2. Is this unique element found in any other organisms in the Columbia River coastal margin and/or elsewhere?
3. Can the unique element be used as a taxonomic marker to facilitate ecological studies of *Katablepharis* CRE?

3.2. METHODS

3.2.1. Sample acquisition

Samples for SSU sequence analysis were collected in the Columbia River coastal margin along the river-to-ocean gradient from sites with three distinct salinities in April 2007 and 2008. Fig. 3.1 and Table 3.1 provide the details of location, salinity,

temperature, and depth for all samples used for SSU sequence analysis. Water was collected from the Columbia River estuary and its plume during April and August 2007, as well as April, July, and September 2008 aboard several vessels [M/V *Forerunner* (estuary April 2007), R/V *Barnes* (estuary August 2007) and R/V *Wecoma* (all other samples)]. The Columbia River estuary consists of both a tidal brackish water region (from river and ocean water mixing) and a tidal freshwater region that extends further upstream. Freshwater and mid-salinity water samples were collected within the Columbia River estuary and were defined as having salinity values of 0 and 15, respectively. Plume water was collected outside the Columbia River bar and was defined as having a salinity of 28-31 (Barnes et al. 1972). In addition, samples for quantitative PCR and fluorescence in-situ hybridization (FISH) were collected once a month from April-June 2013 aboard the M/V *Forerunner* in surface and bottom waters throughout the estuary at five sites: near the SATURN-04 observatory station (Baptista et al. 2008) in the south shipping channel of the estuary, near the SATURN-03 observatory station, in the estuary mouth, and in the north channel of the estuary (Fig. 3.2). Surface samples were also collected near SATURN-03, south channel, and the north channel in July 2013. Figure 3.2 and Table 3.2 provide the details of location, salinity, temperature, depth, and turbidity for all samples used for qPCR analysis. Water was collected either from Niskin bottles attached to a Seabird 911plus CTD (conductivity- temperature-depth) rosette or with a high volume-low pressure centrifugal pump used to collect water using a PVC hose lowered alongside a Seabird 911plus CTD system.

3.2.2. Nucleic acid extraction

After collection, water was immediately filtered through 0.2-µm-pore-size Sterivex filters (PES, ESTAR, Millipore) using a peristaltic pump set to low speed until it became clogged (1-5 L). One Sterivex filter was collected per sample taken. We chose not to pre-filter as we did not want to omit particle-attached unicellular microorganisms. Water was manually forced gently out of each filter using an air-filled syringe and 2 mL of the fixative RNAlater (Ambion) was added to the Sterivex before freezing at -80°C aboard the ship. DNA was extracted from the particulate material from each sample using a phenol-based extraction as described in Herfort et al. (2011). Extraction was performed twice for each Sterivex filter and the two total extracts were pooled.

3.2.3. PCR conditions

SSU rRNA gene DNA sequences were amplified using the eukaryote-specific primers EukA (5'-AACCTGGTTGATCCTGCCAGT-3'; position ~1-20) and EukB (5'-TGATCCTTCTGCAGGTTCACCTAC-3'; position ~1780-1800) (Díez et al. 2001) in a Bio-Rad DYAD PCR thermocycler. The PCR mixture, reaction, and cleanup followed Herfort et al. (2011). PCR products were stored at -20°C.

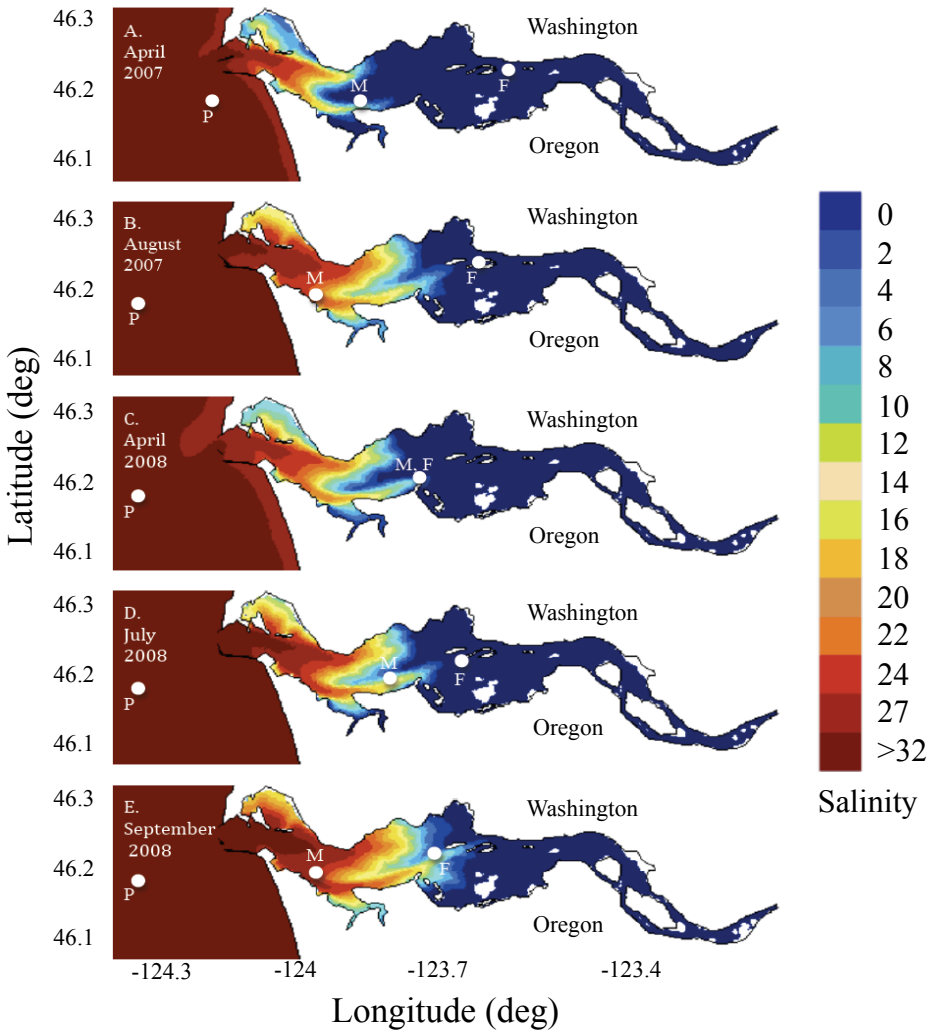


Figure 3.1. Sampling locations in the Columbia River estuary and plume for water collected in A. April and B. August 2007, as well as C. April, D. July, and E. September 2008. Samples were taken in freshwater (F; salinity of 0), mid-salinity water (M; salinity of 15), and the plume (P; salinity of 28-31). In April 2008, freshwater and mid-salinity samples were taken at the same location but at different depths (2 and 17 m, respectively). Color gradient indicates maximum bottom salinity intrusion simulations taken from the DB14 river-to-shelf simulation database (http://www.stccmop.org/datamart/virtualcolumbiariver/simulationdatabases). Note the increase in salinity intrusion (more saline water reaching further upstream) from April to

August in 2007, and from April to September in 2008. Sampling sites of mid-salinity water from August 2007 and September 2008 were not located further upstream than those for April 2007 and 2008 samples, however, because August and September samples were taken towards the end of ebb tides while April samples were obtained at the end of flood tides.

Table 3.1. Physical characteristics of water samples collected for this study. Sampling locations are given in Fig. 3.1.

	Freshwater					Mid-Salinity					Plume				
	Apr-07	Aug-07	Apr-08	Jul-08	Sep-08	Apr-07	Aug-07	Apr-08	Jul-08	Sep-08	Apr-07	Aug-07	Apr-08	Jul-08	Sep-08
Salinity (PSU)	0.3	0.1	0.2	0.1	0.3	14.2	14.0	14.4	14.7	13.6	28.0	283	30.0	29.4	31.2
Temperature (˚C)	8.9	20.6	8.9	19.6	18.5	9.7	17.9	8.9	14.5	14.8	10.1	15.1	8.8	10.8	11.3
Depth (m)	2.0	2.0	2.0	2.0	2.0	10.0	12.0	17.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Table 3.2. Salinity (Sal), temperature (Temp), depth, and turbidity (Turb) for samples used in qPCR analysis. NC refers to samples collected from the North Channel, EM in the estuary mouth, S03 near the SATURN-03 observatory station, SC in the South Channel, and S04 near the SATURN-04 observatory station.

	4-Apr-13				23-May-13				20-Jun-13				18-Jul-13		
	Sal.	Temp. (˚C)	Depth (m)	Turb. (NTU)	Sal.	Temp. (˚C)	Depth (m)	Turb. (NTU)	Sal.	Temp. (˚C)	Depth (m)	Turb. (NTU)	Sal	Temp. (˚C)	Depth (m)
NC	1.0	9.8	0.8	4.4	2.9	13.3	1.0	3.0	4.3	17.4	1.3	23.9	9.0	16.0	1.0
	26.2	9.4	20.1	9.5	28.2	9.8	23.9	11.1	26.5	14.9	20.2	24.4			
EM	5.8	11.0	0.8	1.3	9.9	12.4	1.1	2.3	19.1	15.7	1.2	22.9			
	18.9	9.8	8.4	5.4	30.1	9.3	8.1	3.7	30.2	13.8	15.2	21.9			
													8.0	16.0	1.0
S03	3.5	10.1	0.8	2.7	3.3	13.2	1.0	4.0	2.1	17.8	1.0	1.9			
	25.7	9.5	15.2	6.4	28.6	9.9	15.0	6.4	26.9	14.8	15.2	4.2			
													4.0	17.0	1.0
SC	2.2	10.0	0.8	2.1	0.1	13.8	1.0	5.2	0.1	18.0	1.0	14.6			
	24.2	9.5	16.4	2.2	0.1	13.8	10.2	5.2	0.4	17.9	11.2	15.2			
S04	0.2	10.1	1.0	2.9	0.1	13.8	1.0	4.0	0.1	17.9	0.8	13.0			
	0.7	10.0	7.1	3.8	0.1	13.8	7.3	4.4	0.0	17.9	8.4	11.9			

Figure 3.2. Sampling locations for qPCR and FISH analyses conducted for April-July 2013. Surface and bottom samples were taken monthly for five sites: SATURN-04, South Channel, SATURN-03, Estuary Mouth, and the North Channel.

3.2.4. Cloning & Sequencing

The purified PCR products were ligated with a TOPO vector (pCR 2.1, Life Technologies) and the resulting constructs were then introduced by transformation into One Shot Top 10 chemically competent *Escherichia coli* cells from the TA cloning kit (Life Technologies), plated, and inoculated into two 2 mL 96-deep well plates (Thermo-Fisher Matrix) following Herfort et al. (2011). Plates were stored at -80°C until sent to the Genome Sequencing Center at Washington University in St. Louis to be sequenced using the BigDye® Terminator protocol (Life Technologies) with the EukA/ EukB primer set, as well as an internal primer, 528f (5'-CCGCGGTATTCCAGCTC-3') (Elwood et al. 1985). Two 96-well plates were sequenced per sample for a total of 192 raw sequences per sample.

3.2.5. SSU Sequence analysis

For each SSU clone, contigs (continuous consensus sequences) were assembled using Geneious v5.3 software (Drummond et al. 2010) from the sequencing reads of the forward primer (EukA: position ~1-20), reverse primer (EukB: position ~1780-1800) and an internal primer (528f: position ~528-548) to generate a ~1800 bp sequence. Poor sequence reads (Phred score < 20) and vector sequences were excluded from further analysis. Contig sequences were then searched against the NCBI non-redundant nucleotide database for homologous sequences, and those sequences that had at least 1000 bp aligned, an expectation value ≤ 1e-80, and percent identity ≥ 97% were used for further analysis. Sequences resembling multicellular organisms (for example, copepods) were removed from the data set, while those most closely related to known protist sequences were grouped according to their metabolism (autotrophic or heterotrophic), class and genus. An average of 30 sequences per sample related to heterotrophic protist taxa were used for further analysis, and an average of 14 metazoan and 27 autotrophic sequences per sample were removed from the dataset, respectively. A replicate 500 bp dataset constructed with sequencing reads from the forward primer (EukA: position ~1-20) was also used for class-level analysis (average of 42 heterotrophic sequences per sample), and confirms trends seen with the 1800 bp dataset (Fig. S3.1). All SSU sequences have been deposited to NCBI with accession numbers KJ925152-KJ926058.

3.2.6. LSU sequence analysis of Katablepharis CRE

A full sequence of the internal transcribed spacer (ITS) 1, 5.8S, and ITS 2 region and a partial sequence of the large subunit (LSU) of the rRNA gene were obtained from *Katablepharis* CRE (accession number KJ925151). A katablepharid-specific SSU forward primer (Kj3F: 5'-TGGATCGAAAGGTCTGGGTA-3', position:1451-1471) was designed and tested for specificity against the NCBI non-redundant nucleotide database. The Kj3F primer was used with a general eukaryotic LSU reverse primer (LR9: 5'-AGAGCACTGGGCAGAAA-3', position 2188-2204) available from the Vilgalys laboratory web site (http://www.biology.duke.edu/fungi/mycolab/primers.html). This sequencing revealed a 332 bp unique sequence element (USE) detected in the LSU region of the rRNA gene of *Katablepharis* CRE.

A suite of primers were designed to examine the presence and diversity of the unique sequence element (USE) detected in the LSU region of the rRNA gene of *Katablepharis* CRE. These PCR primers were tested for specificity against the NCBI non-redundant nucleotide database and are detailed in Table 3 and in the Results section. For all PCR reactions, PCR reaction mixtures contained 1x PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM each of dATP, dCTP, dGTP, and dTTP, 0.2 µM of each primer, 2 units per reaction of Platinum *Taq* polymerase (Life Technologies), and 1 µL template DNA (~100 ng) in a final 25 µL volume. The following PCR steps were performed: Initial denaturation at 94 °C for 2 min, 30 cycles of 30 s of denaturation at 94 °C, 30 s of annealing at 55 °C, and 1 min of extension at 72 °C. 5 µL of the reaction product were run in 1.5% (w/v) agarose gel stained with GelRed (Biotium). Positive amplicons were cleaned using UltraClean PCR Clean-up Kit (MO BIO), cloned using a TOPO TA

cloning kit (Life Technologies), and transformants were plated as described above. Positive, white colonies were picked and inoculated into 2 mL 2x Yeast extract/ Tryptone and Ampicillin (0.05 mg mL^{-1}) and grown overnight at 37 °C with shaking at 50 rpm. Plasmids were purified using a FastPlasmid mini kit (5 Prime) and sequenced at the Oregon National Primate Research Center with M13F (5'- TGT AAAACGACGGCCAGT-3') and M13R (5'- AGGAAACAGCTATGACCAT-3').

3.2.7. Quantitative PCR analysis of estuarine samples for the Katablepharis CRE USE

Quantification of the *Katablepharis* CRE unique sequence element (USE) was performed for samples collected from March to July 2013. To do so, forward (K28VF6: 5'- GGAATTAGGCCAGCATCAGA-3') and reverse (K28VR6: 5'- CCAACGGCAACAATTGACTA-3') primers were designed to amplify a 144 bp sequence unique for the USE of *Katablepharis* CRE. Primer specificity was tested through end-point PCR followed by TOPO cloning and sequence analysis (described above), and PCR conditions were optimized to minimize primer-dimer formation. All sequences recovered from this PCR were highly related (>99%) to the USE of *Katablepharis* CRE. qPCR reactions were performed in a final volume of 20 µL containing 10 µL SYBR Green I PCR Master Mix (Life Technologies), 8 µL water, 0.25 µM of each primer, and 1 µL template. All reactions were performed on a StepOnePlus Real-Time PCR system (Life Technologies), with an initial denaturation step (94 °C, 2 min) followed by forty cycles of 15 s of denaturation at 94 °C, annealing-extension at 60°C for 1 min, and 15 s of data collection at 81°C. Melting curve analysis was performed to assess non-specific amplification of primer-dimers, as SYBR green I binds

to all double-stranded DNA without specificity. The dissociation curve from 60 to 95°C was measured after the last qPCR cycle and the melting temperature (T_m) of potential primer-dimers and the specific PCR products was obtained. The majority of samples contained a single melting temperature peak with no evidence of primer-dimer formation; however, in order to suppress fluorescence caused by primer-dimer formation (T_m ~65 °C), the temperature of the detection step was set above that of primer-dimers but below that of the specific PCR product (~85 °C).

A plasmid bearing cloned USE for *Katablepharis* CRE was constructed and linearized to use as standards for qPCR. Six standard reactions with concentrations ranging from $4.07 \times 10^0 - 4.07 \times 10^5$ gene copies μL^{-1} were used to construct standard curves. The concentration of genomic insert DNA from linearized plasmids was measured fluorometrically using a Qubit® 2.0 flourometer (Life Technologies). The number of gene copies in the standard was calculated as:

$$\text{Molecules } \mu L^{-1} = a/((4080 \times 660) \times 6.022 \times 10^{23}),$$

where a is the plasmid DNA concentration (g μL^{-1}), 4080 is the plasmid length, including the vector (3931 bp) and inserted PCR fragment (149 bp)), 660 is the average molecular weight of one base pair, and 6.022×10^{23} is the number of molecules in a mole of a substance. Environmental samples, standards and blanks (water as template) were run in triplicate and 1 µL DNA sample were used per reaction. The concentration of targeted LSU rDNA (copies mL^{-1}) was calculated from the following formula:

$$\text{Copies } mL^{-1} = cda/b,$$

where *c* is the LSU rDNA concentration estimated by qPCR (copies μL^{-1}), *d* is the sample dilution factor, *a* is the volume of solution the DNA extract was resuspended (μL), and *b* is the volume of water filtered (mL). qPCR data is available in Table S3.1.

3.2.8. FISH analysis of estuarine samples for Katablepharis CRE

In order to estimate cell abundance and gain a sense of cell morphology of *Katablepharis* CRE, an oligonucleotide probe labeled with Alexa Fluor 555 (Life Technologies) specific for the unique sequence element of *Katablepharis* CRE was designed (VR113: 5'-GGAATTAGGCCAGCATCAGA-3'). Fluorescence In Situ Hybridization (FISH) was conducted for a subset of samples from the 2013 time series that represented a mix of different depths, locations, and sampling dates (indicated by asterisks in Fig. 5B). For each sample, aliquots were fixed with paraformaldehyde (4% final concentration) and stored at 4°C. Protocols for detection of specific protist taxa by FISH have been reported previously (Pernthaler et al. 2001, Massana et al. 2006). Briefly, 10 mL fixed aliquots were filtered on 0.6 μm pore size polycarbonate filters and hybridized for 3 h at 46°C in the appropriate buffer (with 30% formamide) (Pernthaler et al. 2001, Massana et al. 2006), washed at 48°C in a second buffer (Pernthaler et al. 2001, Massana et al. 2006), counter-stained with 4',6-diamidino-2-phenylindole (DAPI; 5 μg μL^{-1}) and proflavin (Sherr et al. 1993), and mounted in a slide. Cells were then observed and counted by epifluorescence microscopy (Axiovert 200M, Zeiss) on 400x magnification with oil immersion under a Cy3 filter. Cell concentration was calculated with the following formula:

$$\text{Cell concentration} = (c / f \times 817)/v,$$

where *c* is the total number of cells counted, *f* is the total number of fields of view used, 817 is the number of field of views per filter on 400x magnification, and *v* is the volume filtered. A minimum of 20 fields of view was inspected.

3.3. RESULTS

3.3.1. Heterotrophic protist assemblages in the Columbia River coastal margin

Nucleotide sequence analysis of SSU rRNA gene DNA from the clone libraries was performed to examine seasonal and inter-annual variations of heterotrophic protist assemblages in the Columbia River estuary and its plume during the spring, summer, and autumn of 2007 and 2008. The most abundant heterotrophic protist sequences detected in the freshwater samples were attributable to ciliates (with 97-100% similarities at the genus level; Fig. 3.3). Ciliate sequences represented a large proportion of all freshwater samples, with April samples consisting primarily of *Rimostrombidium*, which was also present in August 2007 along with the tintinnid genus *Tintinnopsis*. The genus *Stokesia* of the order Peniculida was particularly frequent in July and September 2008, while *Pelagodileptus* was also abundant in September 2008. Heterotrophic chrysomonad sequences were detected in all freshwater samples, with sequences related to *Paraphysomonas* accounting for high proportions of total sample clones in April 2008, while katablepharid sequences related to *Katablepharis japonica*, a small (~5 μm) heterotrophic flagellate, were detected in both April samples as well as August 2007.

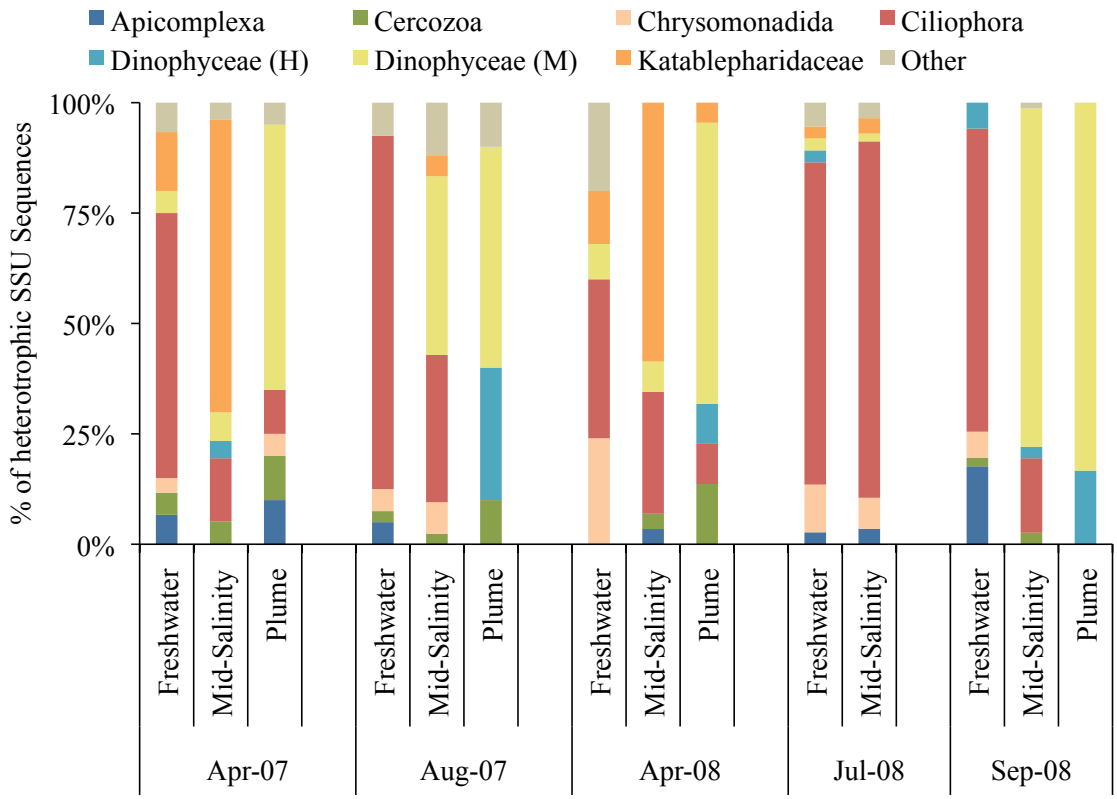


Figure 3.3. Percent composition of heterotrophic protists at the class level based on analysis of SSU sequence data for water collected in the Columbia River estuary and its plume in April and August 2007, and in April, July and September 2008. Freshwater = salinity of 0; Mid-Salinity = salinity of 15; Plume = salinity of 28-31. "H" refers to putative heterotrophic dinoflagellates, while "M" indicates putative mixotrophic dinoflagellates. "Other" category designates sequences associated with the following protist taxa: Bicosoecida, Centroheliozoa, Choanoflagellatea, Ichthyosporea, Labyrinthulida, Stramenopile MAST-12 group, Oomycetes, Pirsonia, and Telonemida. The dominance of Katablepharid sequences in April 2007 and 2008 mid-salinity waters denotes the genus *Katablepharis*.

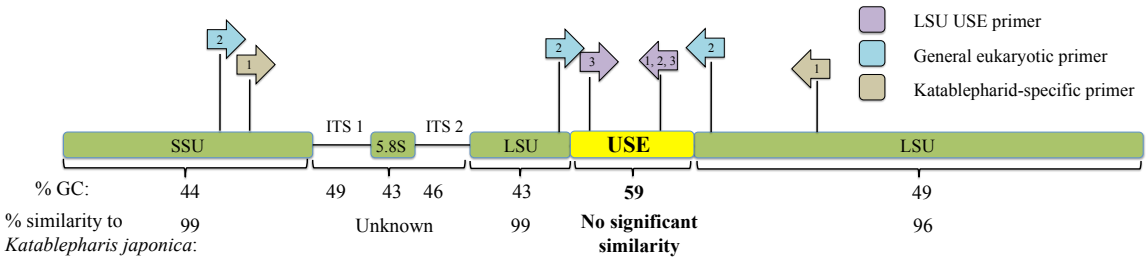
In both April 2007 and 2008, mid-salinity (salinity ~ 15) water samples were taken in the south channel of the estuary (Fig. 3.1). SSU sequences were dominated by those belonging to the Katablepharidaceae, a class of colorless heterotrophic flagellate (Fig. 3.3). The overwhelming majority of sequences relating to the class Katablepharidaceae in the April mid-salinity samples most closely resembled the SSU (18S) sequence of *Katablepharis japonica* (98% identity). These sequences also represented a large proportion of all sequences collected, including both heterotrophic and autotrophic protists in the April mid-salinity samples. Similar to the freshwater sample, ciliates were the dominant heterotrophs in the July 2008 mid-salinity water sequences. The majority of the retrieved ciliate sequences were related to *Stokesia*, which was found in comparable relative abundance (25 and 26%) in the July and September 2008 freshwater samples (Fig. 3.3). In August 2007 and September 2008 mid-salinity samples, dinoflagellate sequences relating to the dinoflagellate *Gyrodinium*, a diverse genus that can be autotrophic, mixotrophic, or heterotrophic, were particularly frequent, ranging from 20 to 35% of heterotrophic sequences (Fig. 3.3). *Gyrodinium* sequences were also the dominant dinoflagellate sequences in the plume samples from August 2007 and September 2008, while *Alexandrium*, *Pentapharsodinium*, and *Peridinium* represented most of the dinoflagellate sequences found in the other plume samples.

3.3.2. LSU sequence analysis of Katablepharis CRE

To provide higher resolution of the genetic variability of the highly dominant katablepharid referred to herein as *Katablepharis* CRE, a full sequence of the internal transcribed spacer (ITS) 1, 5.8S, and ITS 2 region and a partial sequence of the large

subunit (LSU) of the rRNA gene were obtained from *Katablepharis* CRE (accession number KJ925151). The ITS1-5.8S-ITS2 region of *Katablepharis* CRE is a 667 bp sequence that is ~93% similar to an ITS1-5.8S-ITS2 sequence of a cultured strain of *Katablepharis japonica* (CCMP 2791) from the Neuse River estuary in North Carolina (Fig. 3.4). A 332 bp Unique Sequence Element (USE) was uncovered in the *Katablepharis* CRE LSU gene at the 138th nucleotide from the 5' end of the sequence. This region showed no significant similarity to any LSU rRNA sequence in the NCBI database (last checked January 3, 2014). However, the sequences down- and upstream of this USE region show 99% and 96% similarity to the *K. japonica* sequence in the NCBI database (accession number FJ973371), respectively. Another intriguing feature of the USE region is the difference in GC content of the variable region compared to its flanking regions. The USE has a GC content of 59%, while the down- and upstream regions are both <50%. To assess the diversity and distribution of the USE in *Katablepharis* CRE, and potentially other organisms, a suite of PCR primers were used (Table 3.3), and positive amplifications were cloned and sequenced. To determine how widespread the USE is amongst *Katablepharis* CRE and other katablepharids in the Columbia River coastal margin, PCR was performed using katablepharid-specific SSU forward Kj3F and LSU reverse (Kj281R: 5'-TCCTCTGACTTCACCCTGCT-3', position:947-966) primers flanking the USE (Fig. 3.4). All amplified products were cloned and sequenced, and all recovered sequences were related to one of two katablepharids: *Katablepharis* CRE or *Leucocryptos*. All sequences related to *Katablepharis* CRE recovered from the estuary contained the D2 region USE. The only other katablepharid detected in the system (*Leucocryptos*) did not have this element, but

rather contained an LSU sequence that aligned well (>97% similarity) with the *K. japonica* (accession number FJ973371) and *L. marina* (accession number DQ980471) sequences in the NCBI database in this region.



A.

Katablepharis CRE (Columbia River Estuary)

GAGACCGATAGCGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAAAGTGCTTGAAATTGTTAAGAGGAAAAAC
GATTGACTCCAATACTGTATCCACGGTAATTCAGCTGGGGCTTATGGATTCTGGGCGTTGTGACACCGTGAAGGGTCCTGCGTTCGGGC
GAGAAGTAAGCACCAGTGCACTTTGCCGGGGAATTAGGCCAGCATCAGATACTTCCGCGGGAGATGGCGGCCGGTAGTGGGTGGGTTT
CGATCTGCTTTGCCGGTTGCGGCACCGGGGGAGTGTCTGAGGAATGAAGGGCGCGCTTTAGCGGGACGCCTAGTCAATTGTTGCCGTT
GGACGGGATGCACTGCTTTGCGGTGTCCCTGGCCAGAGGTCTCAGTTGGTTAGAACGCACCTGCGCCCTGAACATGCTTGGCGAAATGG
TATCAACCGACCCGTCTTGAAACACGGACCAAGGAGTCTAACACGTATGCGAGTAGCTGGGTGACAAACCTAGATGCGTAATGAAAG

B.

Figure 3.4. A. Sequencing of the rRNA gene of *Katablepharis* CRE revealed a 332 bp region of the 28S rRNA gene that is unique to the CRE strain. This region is GC-rich compared to the rest of the gene and shows no significant similarity to other katablepharids in the NCBI database, while the rest of rRNA gene aligns well with *K. japonica* and other sequenced katablepharids. Colored arrows indicate PCR primers designed in the 28S variable region (purple), to be general for all eukaryotes (teal), or katablepharid-specific (light brown). Numbers within the arrows refer to the primer sets that were used to answer the research questions discussed in the introduction.

B. Nucleotide sequence of the *Katablepharis* CRE USE (bold) with flanking LSU sequence (green). ORF primers, general eukaryotic primers used to amplify additional D2 USE, are underlined.

To analyze the diversity and extent of USE in other organisms, or other systems outside the Columbia River coastal margin, a primer set consisting of a general eukaryotic forward primer (528f), and a USE-specific reverse primer (K28VR6) were employed to amplify sequences from all eukaryotic organisms that contained this region. Locations of water samples used for attempted amplification included the Columbia River coastal margin as well as Puget Sound, Grays Harbor Washington, Amazon River plume, Beaufort Sea, and several Russian rivers (see Crump et al. 2009 for sampling details). A total of 32 samples were analyzed but positive PCR amplification of the USE sequences occurred only in the Columbia River estuary samples, with all recovered sequences being identical to *Katablepharis* CRE. Several Columbia River estuary locations upriver of SATURN-04, in the freshwater tidal zone of the estuary, were also tested. However, all of these samples failed to amplify. This element also failed to align with the NCBI and Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) metagenomic databases. However, a similar sequence (96% identity) was detected in a metagenome collected from surface water off the Delaware coast during May 2010 (source IMG-MER).

To determine if other similar elements exist in other organisms, general forward (ORF1: 5'-GACTCCTTGGTCCGTGTTTCAAGA-3') and reverse (ORF2: 5'-CGAACAAGTACCGTGAGGGAAAGATGCAAA-3') primers that flank the USE were designed and used in PCR reactions (Fig. 3.4). Resulting amplicons revealed unique elements in other protists at about the same location and of the same approximate size (294-400 bp). For example, a sequence related to the parasitic dinoflagellate

Eudubosquella sp. Ex *Favella arcuata* (accession number JN934989) amplified from the Beaufort Sea lagoon in Alaska contained a 294 bp region. A sequence, related to *Eudubosquella* sp. Ex *Favella arcuata*, was also detected from the Columbia River coastal margin and contained a 292 bp region. The sequences flanking the elements align to each other, but the elements themselves do not align with each other or any other sequence in the NCBI database (analysis done on January 3, 2014).

Table 3.3 Primers used for analysis of the unique sequence element (USE) detected in *Katablepharis* CRE. Research question refers to the questions discussed in the introduction.

Primer name	Sequence (5'-> 3')	Target group	rRNA gene region	Research question	Reference
Kj3F	TGGATCGAAAGGTCTGGGTA	*Katablepharidaceae*	SSU	1	This study
Kj281R	TCCTCTGACTTCACCCTGCT	*Katablepharidaceae*	LSU	1	This study
528f	CCGCGGTATTCCAGCTC	Eukaryota	SSU	2	Elwood et al. 1985
K28VR6	CCAACGGCAACAATTGACTA	*Katablepharis* CRE	LSU-USE	2, 3	This study
K28VF6	GGAATTAGGCCAGCATCAGA	*Katablepharis* CRE	LSU-USE		This study
ORF1	GACTCCTTGGTCCGTGTTTCAAGA	Eukaryota	LSU	2	This study
ORF2	CGAACAAGTACCGTGAGGGAAAGATGCAAA	Eukaryota	LSU	2	This study

3.3.3. Distribution of Katablepharis CRE assessed through qPCR

While the use of SSU sequence libraries provides a snapshot of the protist assemblage and suggests that *Katablepharis* CRE is one of the dominant heterotrophic protists in the Columbia River estuary during spring, it is difficult to quantify protist populations using this method due to potential PCR biases and differences in gene copy numbers among protists (Heywood et al. 2011). To determine the distribution of *Katablepharis* CRE in the Columbia River coastal margin, USE-specific primers were designed and qPCR was performed for samples collected monthly between March and July 2013. The sampling period captured the annual spring freshet, with outflow from Bonneville Dam beginning to increase around the beginning of April (Fig. 3.5A). During this period *Katablepharis* CRE was detected at all five sites within the estuary (salinities ranging from 0 to 26) with USE gene copy numbers as high as 1.06×10^4 copies mL^{-1} measured (Fig. 3.5B). These values are comparable to those found in samples from mid-salinity waters in April 2007 and 2008, where *Katablepharis* CRE USE gene copy numbers were present at 9.89×10^3 and 1.09×10^4 copies mL^{-1}, respectively. For the May 2013 samples, USE gene copy numbers were lower at each site than they were for the April 2013 samples, with the highest USE gene copy numbers measured at the bottom estuary mouth sample (8.1 m depth) with 2.09×10^3 copies mL^{-1} (Fig. 3.5B). However, USE gene copy numbers did increase again in three June samples, with the highest gene copy numbers of all samples (1.52×10^4 copies mL^{-1}) occurring in the bottom sample waters at the North Channel site characterized by a salinity of 26.5 and turbidity of 24.4 NTU, which was more than double the turbidity observed in the April or May samples at

the same site and depth (9.5 and 11.1 NTU, respectively). There was an increase in river outflow as well during early June (Fig. 3.5A), likely caused by rain events during late May and early June that brought 89.5 mm of precipitation to Astoria, Oregon between May 24th, 2013 and June 19th, 2013 (source: NOAA National Climatic Data Center). All copy numbers for the July samples were below 1.00 x 10^2 copies mL^{-1}. By this sampling date, river outflow had nearly returned to pre-freshet levels (Fig. 3.5A).

3.4.4. FISH of natural samples for Katablepharis CRE

Fluorescence in situ hybridization (FISH) was conducted for a subset of samples (Table 3.3 and samples marked by an asterisk on Fig. 3.5B) in order to gain a sense of cell size and morphology and to estimate LSU gene copy number per cell of *Katablepharis* CRE. The fluorescent probe hybridized to ribosomal DNA within a ~7 μm organism (Fig. 3.6), which is the approximate size of katablepharids in other systems (Okamoto & Inouye 2005). The abundance of Katablepharis CRE estimated by qPCR was plotted against that obtained by FISH counts, with the slope of the regression provided to estimate the gene copy number for *Katablepharis* CRE (Fig. 3.7) (Zhu et al. 2005). There was a weak positive correlation between qPCR and FISH estimates of abundance, and the slope of the regression suggests a ratio of 2.5 gene copies per cell. Similarly to other comparisons of FISH and qPCR using environmental samples (Rodríguez-Martínez et al. 2009), the R^2 value (0.58) for this correlation is low, likely because biases in DNA extraction and qPCR optimization, as well as loss of cell integrity from fixation, may cause discrepancies between the two methods (Zhu et al. 2005).

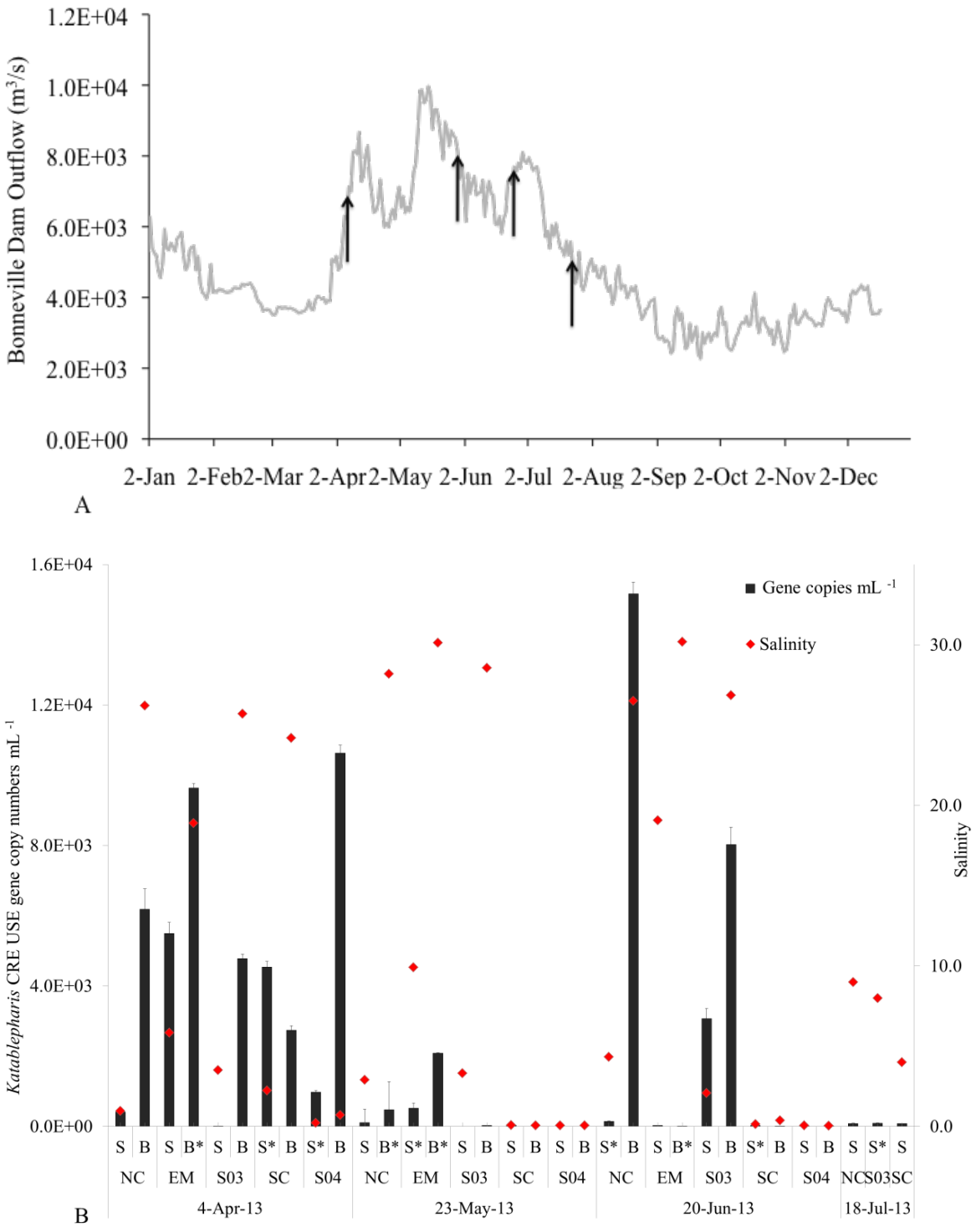
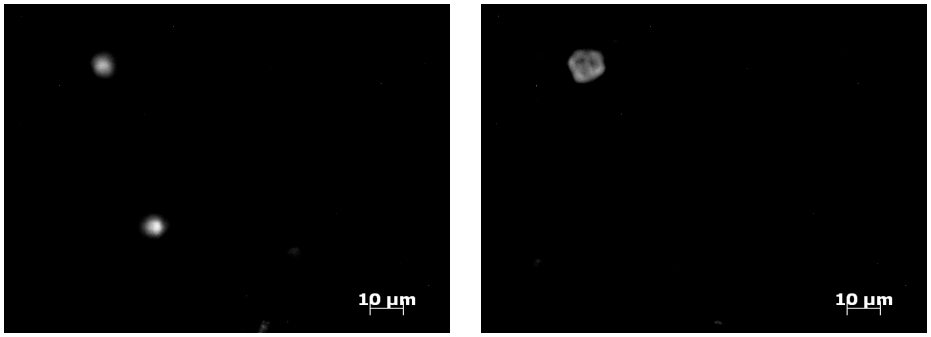


Figure 3.5. A. Annual Columbia River discharge (m^3/s) measured at the outflow of Bonneville Dam for 2013 (daily mean). Gray arrows indicate sampling dates for qPCR and FISH analyses. B. Distribution of *Katablepharis* CRE USE in the Columbia River

estuary estimated by qPCR from March-July 2013. NC refers to samples collected from the North Channel, EM in the estuary mouth, S03 near the SATURN-03 observatory station, SC in the South Channel, and S04 near the SATURN-04 observatory station (see Fig. 3.2 for exact locations). S = surface water; B = bottom water. Asterisks denote samples which were also analyzed with FISH. Error bars indicate standard deviation for each sample. Red dots indicate salinity.



A. B.

Figure 3.6. Epifluorescence micrographs of microorganisms larger than 0.6 µm (10 mL filtered) in Columbia River estuary water collected at the surface of the North Channel station on May 23, 2013. A. DAPI-stained cells and B. the corresponding microscopic field using the FISH *Katablepharis*-specific probe. Probe hybridized to a ~7 µm organism, the approximate size of katablepharids found in other systems.

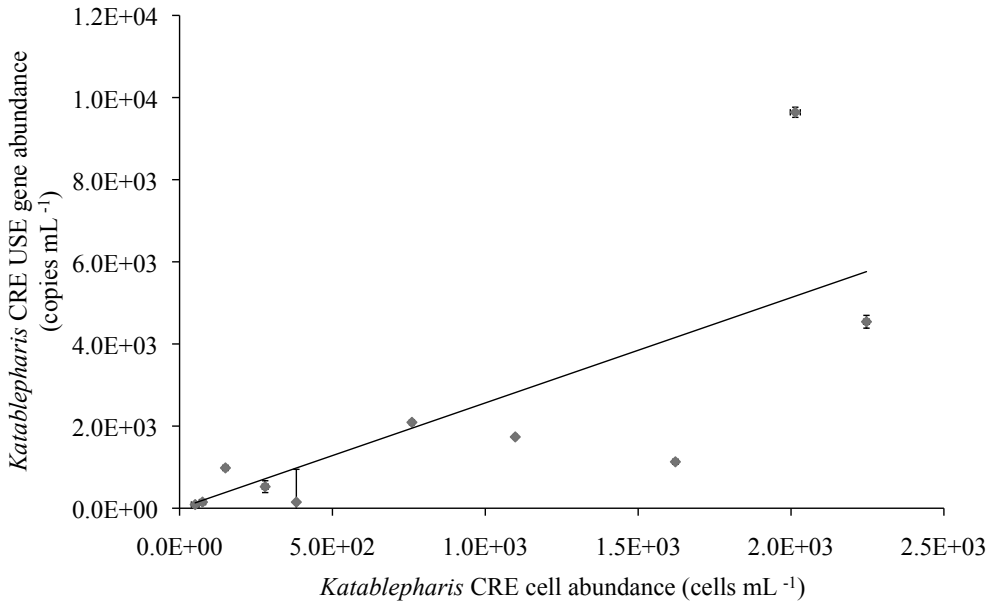


Figure 3.7. Comparison of *Katablepharis* CRE abundance estimated by FISH and qPCR using probes specific for its Unique Sequence Element (USE) within the large subunit for both approaches. The *x* axis corresponds to numbers of cells estimated through FISH and the *y*-axis to LSU copies measured by qPCR. Regression with y-intercept set to zero was used to estimate gene copy number per cell, with a slope of 2.6 and an R^2 value of 0.59. Horizontal and vertical error bars indicate standard deviation for abundance estimates by FISH and qPCR, respectively.

3.4. DISCUSSION

A major goal of this study was to offer new insight into the heterotrophic protist assemblages of the Columbia River coastal margin through the use of molecular methods. The most notable, and unexpected, finding from our initial survey study was the dominance of SSU sequences resembling that of the heterotrophic flagellate class Katablepharidaceae in the mid-salinity (~salinity = 15) samples in both April 2007 and

2008. Katablepharid SSU sequences comprised > 40% of all the SSU sequences observed in each of these two samples. No other individual sequence, heterotrophic or autotrophic, during the two-year time series was found in such high relative proportions. While it is not uncommon for autotrophic taxa to reoccur in high abundance, such as annually recurring spring blooms (Sverdrup 1953), this same phenomenon is less commonly observed in heterotrophic taxa.

In our study, *Katablepharis* sequences were most prominent in the estuarine clone libraries in April, with only a few sequences retrieved from the Columbia River plume libraries (Fig. 3.3). Abundance estimates of these plume samples determined through qPCR using the *Katablepharis* CRE USE-specific primers were also relatively low compared to those in the estuary, with only 0.3 and 34.6 gene copies mL^{-1} observed in April 2007 and 2008, respectively (data not shown). *Katablepharis* has been found to be an important primary consumer and bacterivore associated with particles in freshwater and estuarine environments (Ploug et al. 2002, Domaizon et al. 2003, Šlapeta et al. 2006). Events occurring in the spring, such as upriver diatom blooms and spring runoff, deliver organic matter to the estuary (Sullivan et al. 2001) that could fuel katablepharid proliferation. Indeed, as freshwater phytoplankton enter more saline waters they tend to lyse due to osmotic stress, releasing organic matter (Frey et al. 1984). *Katablepharis* CRE is likely an essential and yet previously undetected link between the microbial and herbivorous food webs in the Columbia River estuary. The USE marker identified herein provides a tool with which to investigate temporal and spatial dynamics of this organism that can be used to investigate potential trophic linkages.

Grazing by heterotrophic protists such as *Katablepharis* CRE can be an important force in shaping bacterial populations, transferring prey carbon to higher trophic levels, and remineralization of nutrients such as nitrogen and phosphorus (Caron et al. 1990). This information is particularly important in the Columbia River estuary, as it is often classified as a detritus-driven system, fueled by allocthonous organic matter with high levels of bacterial growth and productivity (Crump & Baross 1996). Through their role as bactivorous grazers, *Katablepharis* and other heterotrophic protists provide a critical link between bacterial production and higher trophic levels, including invertebrates and fish. The identification of *Katablepharis* as an abundant taxon of heterotrophic protist within the SSU rRNA gene sequence data in the estuary during April 2007 and 2008 exemplifies the strength of molecular tools in illuminating taxonomic diversity in microbial assemblages, as this flagellate is a small, non-pigmented, non-descript eukaryote that could easily go unidentified using microscopic cell counting methods and undetected through pigment analyses.

The discovery of the unique sequence element (USE) within the LSU of *Katablepharis* CRE provided an excellent reagent to track the distribution of *Katablepharis* CRE in the system through USE-specific probes combined with quantitative and qualitative methods. Unlike ITS (internal transcribed spacer), another commonly used discriminating genetic marker, the USE sequence is an integral part of the LSU rRNA and is not removed by pre-rRNA processing. Hence, it presents a high copy target for probing abundance and distribution of protist variants. This yielded a specific qPCR assay that provided quantitative data and allowed for visualization of this uncultured katablepharid through FISH. This qPCR assay confirmed the wide distribution

of *Katablepharis* CRE in the Columbia River estuary with respect to salinity (0-26), but with overall higher abundances measured in bottom waters at each site. Salinity is often a key determinant of protist populations as halotolerance differs among species (Frey et al. 1984), but other heterotrophic flagellates, such as marine strains of *Bodo designis*, can survive and grow at a wide range of salinities (Koch et al. 2005). Molecular diversity studies of heterotrophic flagellate communities from other brackish systems such as the Baltic Sea also indicate a dominance of putative marine species that can survive brackish to fully marine salinity ranges (Weber et al. 2012). While *Katablepharis* CRE was detected at a wide range of salinities, its geographic range in the estuary seems to be constrained to the portion of the estuary affected by salinity intrusion. To verify this, additional freshwater samples collected in April 2013 as far upriver as the SATURN-05 observatory station near Longview, Washington were also assayed by qPCR using the *Katablepharis* CRE USE-specific primers, but *Katablepharis* CRE was not detected any further upriver than SATURN-04 (data not shown), which is located near the limit of salinity intrusion during April (Fig. 3.1).

The presence of the USE in *Katablepharis* CRE and other protists has several potential important evolutionary and ecological implications. It is found within the D2 region of the LSU, which can contain variable length and be more divergent than the rest of the LSU gene (Hassouna et al 1984). However, the unique element of *Katablepharis* CRE confers extreme variability compared to the rest of the D2 region, with no significant similarity to any sequences in the NCBI database. It also contains elevated GC content compared to the rest of the associated rRNA genes, possibly as a result of lateral gene transfer, a process that may be more prevalent in phagotrophic protists than

previously thought (Andersson 2005). The USE sequences that have been uncovered can greatly increase taxonomic resolution of LSU protist diversity, be utilized for qualitative and quantitative monitoring through strain-specific probes, and increase the taxonomic resolution of ecogenomic technologies such as the environmental sample processor (Scholin 2010).

Chapter Four: High resolution aquatic protist biogeography through identification and application of LSU rRNA unique sequence elements [3]

4.1. INTRODUCTION

Protists, including autotrophs, mixotrophs, and heterotrophs, play essential roles in energy transfer in aquatic food webs and in global biogeochemical cycles (Mallin 1994). Cultivation-independent molecular techniques revealed a vast underestimation of protist diversity in aquatic systems (Diez et al. 2001, Lopez-Garcia et al. 2001, Moon-van der Staay et al. 2001, Stoeck et al. 2003, Bass & Cavalier-Smith 2004). Despite these new estimates of protist diversity, consensus on whether protists exhibit biogeographic patterns has been heavily debated (Finlay & Fenchel, 1999; Foissner, 2006). This debate contrasts two opposing views regarding the global distribution of protists. The first view, the "ubiquity theory," asserts that protist biogeography does not exist and that protists are globally distributed due to their high abundances, widespread dispersal abilities, and negligible rate of allopatric speciation (Finlay 2002). The idea that "everything is everywhere, but the environment selects" became an early paradigm in microbial ecology (Beijerinck 1913) that has received support in global surveys of cosmopolitan protist morphospecies, a species distinguished from others based on distinct morphological

[3] This chapter is being prepared for submission: Kahn P, Voorhees I, McAteer D, Koberstein J, Herfort L, Crump BC, Peterson TD, Zuber P

characteristics (Finlay & Fenchel 2004). The second theory, the "moderate endemicity distribution theory" puts forth that at least some protists have a restricted geographic distribution, which is driven primarily by the limited viability of cysts in certain species, which limits dispersal. Evidence supporting the second idea has come from the use of "flagship species", those that have conspicuous morphologies and whose presence or absence can be easily determined during microscopic inspection of water samples (Foissner 2006).

The debate between ubiquity and endemicity has important implications for estimates of global protist biodiversity (Caron 2009). If protists do not exhibit restricted biogeography, then global protist species richness should be low (Finlay & Clarke, 1999). However, global protist diversity might be high if some species have a restricted biogeography that is not the result of their narrow ecological requirements (Foissner 1999). Furthermore, if protist distribution is not globally ubiquitous, the conservation of endemic protist populations likely need to be integrated into ecosystem management of local biodiversity that has traditionally targeted vascular plants and animals (Cotterill et al. 2007), along with necessary mitigation strategies for potentially detrimental anthropogenic influences on protist biogeography, such as the introduction of invasive or harmful algal species through ballast water (Hallegraeff & Bolch 1992).

A significant impediment in addressing the biogeographic distribution of protists is that many field surveys likely undersample diversity (Finlay & Fenchel 2004; Foissner 2006). Reported morphological or low-throughput DNA sequence assessments of protist diversity [i.e., based on Sanger sequencing (Sanger 1977)] are based on counts of dozens to hundreds of cells or on dozens to hundreds of DNA sequences per sample, which may

adequately enumerate the most abundant taxa but fail to detect and quantify the rare taxa (Shokralla et al. 2012; Terrado et al. 2009). Recently, however, advances in molecular phylogenetics through high-throughput next-generation sequencing (NGS) technologies have resulted in extensive sequence-based characterization of microbial communities, and led to very high estimates of species richness and biodiversity (Bachy et al. 2013; Eiler et al. 2013; Heywood et al. 2010; Medinger et al. 2010; Santoferrara et al. 2014; Stock et al. 2009). Analysis of the large number of sequences provided by NGS has begun to yield new information relevant to the ubiquity versus endemicity debate (Nolte et al. 2010) and has uncovered the previously undetected members of the protistan 'rare biosphere' (Sogin & Morrison 2006).

Another key issue in the determination of protist diversity and biogeography centers on the variable level of resolution applied to the taxonomic assignments used in assessing protist diversity (Mitchell & Meisterfeld 2005). Biogeography is directly linked to the taxonomic units included in an investigation, since the geographical distribution of a given taxon may appear to shrink as finer taxonomic resolution is applied (Bass & Boenigk 2011). In many cases, surveys of protist geographical distributions that are based on either the morphospecies concept or on highly conserved genes such as the small subunit rRNA gene (SSU) may underestimate protist diversity (Weisse 2008), thus producing broad distributional ranges of limited ecological significance. This inadequate taxonomic resolution, coupled with an undersampling in field surveys, could result in the incorporation of several cryptic species into a single taxonomic unit with an artificially large distribution (Knowlton 1993). More discriminating markers such as the internal transcribed spacer (ITS) regions of the rRNA gene (Bass et al. 2007) or divergent

domains (D1-D12) of the large subunit (LSU) rRNA gene (Wylezich et al. 2010) provide perhaps more powerful tools for determining the biogeographic distribution that more accurately reflect functional diversity of protist taxa, provided that a relationship between genotype and phenotype is found (Bass & Boenigk 2011). For example, several ITS-defined lineages can be found within a single SSU-type of cercomonad species (*Eocercomonas* and *Paracercomonas*), many of which have restricted distributions and differ significantly in physiological characteristics, such as their salinity tolerance and propensity to form cysts (Bass et al. 2007).

In a recent study focused on the Columbia River coastal margin, a 332 bp unique sequence element (USE) insertion was discovered in the D2 region of the LSU rRNA gene of *Katablepharis* CRE, an estuarine heterotrophic flagellate found abundantly in the spring (Kahn et al. 2014). This element is characterized by extreme variability compared to the rest of the D2 region, and shows no homology to other katablepharids or other eukaryotes despite the close alignment (96-99% similarity) of the SSU and the rest of the LSU of *Katablepharis* CRE to other katablepharids in the National Center for Biotechnology Information (NCBI) sequence database. This USE was utilized as a taxonomic marker of *Katablepharis* CRE, which was found to be biogeographically restricted to the Columbia River coastal margin.

In the present study we identified USE—regions that show no homology to other protist LSU sequences—among other protist taxa found primarily in the Columbia River coastal margin but also in several other coastal margin environments (Amazon River, the Chesapeake Bay, the Susquehanna River and the Beaufort Sea Lagoon) through sequence analysis of LSU D2 region sequences generated through Sanger sequencing and NGS

technology (Illumina HiSeq). These approaches were employed to determine the distribution of USE amongst protists; if the USEs can be broadly utilized as taxonomic markers to track the biogeographic distribution of specific strains within protist taxa; and if the USEs that are detected in the Columbia River coastal margin are constrained to this region or also found elsewhere. We tested the utilization of USEs to study the biogeography within the Columbia River coastal margin of three protist groups that contained multiple USE phylotypes: the parasitic dinoflagellate genus *Euduboscquella*, the deep water heterotrophic class Diplonemea, and the heterotrophic flagellate phylum Cercozoa. Our findings are discussed in the context of the ubiquity versus endemicity debate, and we propose that the USEs presented here can be an effective tool for determining protist biogeography due to their highly specific nature and ability to detect phenotypic differences amongst closely related strains with putatively restricted biogeography.

4.2. MATERIALS & METHODS

4.2.1. Columbia River Coastal Margin

The Pacific Northwest coast is strongly influenced by the Columbia River, which is the second largest river in the U.S. by flow (Simenstad et al. 1990). The Columbia River estuary consists of both a tidal brackish water region (from river and ocean water mixing) and a freshwater tidal region that extends further upstream. The volume of freshwater discharge from the river strongly influences the river-to-ocean gradient, mainly through the modulation of the salinity intrusion into the estuary (Chawla et al. 2008). The estuary transitions from a salt wedge estuary with strong stratification

between the upper freshwater layer and intruding salty bottom layer during high discharge to a partially mixed estuary during low discharge (Hughes & Rattray 1980).

4.2.2. Sample collection

Water samples from the Columbia River coastal margin were collected at various times of the year between 2007 and 2013. Water was collected either from Niskin bottles attached to a Seabird 911plus CTD (conductivity- temperature-depth) rosette system or with a high volume-low pressure centrifugal pump attached to a PVC hose lowered alongside a Seabird 911plus CTD system. Table S4.1 in the Appendix contains a list of all Columbia River coastal margin samples used in this study, as well as their depth, salinity, and analyses performed. Figure 4.1 maps the location of all these samples.

Figure 4.1 Sampling locations from the Columbia River coastal margin.

4.2.3. Nucleic acid extraction

After collection, water was immediately filtered through 0.2 µm-pore-size Sterivex filters (PES, ESTAR, Millipore) using a peristaltic pump set to low speed until the filter became clogged (1-5 L). One Sterivex filter was collected per sample taken. We chose not to pre-filter in order to capture particle-attached unicellular microorganisms. Water was manually forced gently out of each filter using an air-filled syringe and 2 mL of the fixative RNAlater (Ambion) was added to the Sterivex before freezing at -80°C aboard the ship. DNA was extracted from the particulate material from each sample using a phenol-based extraction as described in Herfort et al. (2011). Extraction was performed twice for each Sterivex filter and the two total extracts were pooled.

4.2.4. PCR conditions, cloning, & sequencing

A suite of primers was designed to determine the presence and diversity of the USE detected in the LSU region of the rRNA gene of protists. These PCR primers were tested for specificity against the NCBI non-redundant nucleotide database and are detailed in Table 4.1. For all PCR reactions the following were combined: 1x PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM each of dNTP, 0.2 µM of each primer, 2 units per reaction of Platinum *Taq* polymerase (Life Technologies), and 1 µL template DNA (~100 ng) in a final 25 µL volume. The following PCR steps were performed: initial denaturation at 94 °C for 2 min, 30 cycles of 30 s of denaturation at 94 °C, 30 s of annealing at 55 °C, and 1 min of extension at 72 °C. Five µL of the reaction product were run in 1.5% (w/v) agarose gel stained with GelRed (Biotium). Positive amplicons were cleaned using UltraClean PCR Clean-up Kit (MO BIO), cloned using a TOPO TA cloning kit (Life Technologies), and transformants were plated onto LB agar containing bromo-chloro-

indolyl-galactopyranoside (XGAL) and 50 μg/mL ampicillin before incubating overnight at 37°C. Positive, white colonies were picked and inoculated into 2 mL of 2x Yeast extract/ Tryptone and Ampicillin (0.05 mg mL^{-1}) and grown overnight at 37 °C with shaking at 50 rpm. Plasmids were purified using a FastPlasmid mini kit (5 Prime) and sequenced at the Oregon National Primate Research Center with M13F (5'- TGT AAAACGACGGCCAGT-3') and M13R (5'- AGGAAACAGCTATGACCAT-3').

Table 4.1. PCR primers used in the identification of unique sequence elements. The LR6 primer sequence is a general eukaryotic primer and was obtained from the Vilgalys laboratory website (http://www.biology.duke.edu/fungi/mycolab/primers.html).

Target group	Primer name	Sequence (5'-3')	rRNA subunit	Reference
Eukaryota	ORF2	CGAACAAGTACCGTGAGGGAAAGATGCAAA	LSU	This study
	ORF1	GACTCCTTGGTCCGTGTTTCAAGA	LSU	This study
	LR6	CGCCAGTTCTGCTTACC	LSU	Vilgalys lab
Euduboscquella	Dub2F	GCGCCTGAAACCACTGTATTACAAGCA	LSU	This study
	Dub2R	TTTCAAGACGGGTCATTGAAACCTT	LSU	This study
Diplonemea	DiploF	GATATCTAAACCTGTC	SSU	Lara et al. 2009

4.2.5. Quantitative PCR analysis of USE

Quantification of the USE was performed on samples collected from the Columbia River coastal margin (Table S4.1). All PCR primers used in quantitative PCR (qPCR) assays are detailed in Table 4.2. Primer specificity was tested through end-point PCR followed by TOPO cloning and sequence analysis (described above), with PCR conditions being optimized to minimize primer-dimer formation. All sequences recovered from these PCR amplifications were closely related (>99%) to their target USE. qPCR

reactions were performed in a final volume of 20 µL containing 10 µL SYBR Green I PCR Master Mix (Life Technologies), 8 µL water, 0.25 µM of each primer, and 1 µL template. All reactions were performed on a StepOnePlus Real-Time PCR system (Life Technologies), with an initial denaturation step (94 °C, 2 min) followed by forty cycles of 15 s of denaturation at 94 °C, annealing-extension at 60 °C for 1 min, and 15 s of data collection at 70 °C. Melting curve analysis was performed to assess non-specific amplification of primer-dimers, as SYBR green I binds to all double-stranded DNA without specificity. The dissociation curve from 60 to 95 °C was measured after the last qPCR cycle and the melting temperature (T_m) of potential primer-dimers and the specific PCR products was obtained. The majority of samples contained a single melting temperature peak with no evidence of primer-dimer formation; however, in order to suppress fluorescence caused by primer-dimer formation (T_m ~65 °C), the temperature of the detection step was set above that of primer-dimers but below that of the specific PCR product (~85 °C). Plasmids bearing each cloned USE were constructed and linearized to use as standards for qPCR. The concentration of genomic insert DNA from linearized plasmids was measured using a Nanodrop 2000 (Thermo Scientific). The number of gene copies in the standard was calculated as:

$$\text{Molecules } \mu L^{-1} = a/(((3931 + b) \text{ x } 660) \text{ x } 6.022 \text{ x } 10^{23}),$$

where a is the plasmid DNA concentration (g μL^{-1}), 3931 is the vector length and b is the inserted PCR fragment of the target USE (54-151 bp), 660 is the average molecular weight of one base pair, and 6.022 x 10^{23} is the number of molecules in a mole of a substance. Environmental samples, standards and blanks (water as template) were run in

triplicate. The concentration of targeted LSU rDNA (copies mL^{-1}) was calculated from the following formula:

$$\text{Copies } mL^{-1} = cda/b,$$

where *c* is the LSU rDNA concentration estimated by qPCR (copies μL^{-1}), *d* is the sample dilution factor, *a* is the volume of solution the DNA extract was resuspended (μL), and *b* is the volume of water filtered (mL).

Table 4.2. PCR primers used in qPCR assays of *Euduboscquella*, diplonemid, and cercozoan USEs. Target product size is indicated in parentheses.

Target USE	Primer name	Sequence (5'-3')
Euduboscquella CRE USE 1 (151 bp)	DubC1F	GCGACGTGAACCGTCAGGGA
	DubC1R	AGGGAATCCGGCTCTCTTGGAAA
Euduboscquella CRE USE 2 (85 bp)	Dub3F	TTGGCGGGCTTCCATTGTGT
	Dub3R	GCACGGTTGATTGGCAGCTCCTT
Diplonemid Ubiq USE 1 (77 bp)	DipS1F	GGGCTACGCGGATAGGGAGG
	DipS1R	AAGCCCTTGCCCAATCGGAGA
Diplonemid Deep USE 1 (160 bp)	DipD1F	CCCAATCCTGTTGTGGTTTC
	DipD1R	GAACTTACCCAGCCACGAGA
Diplonemid Deep USE 2 (127 bp)	DipD2F	TGAGCCGCCCTTGCTCAACA
	DipD2R	GGTCTGCGCAGGCTAGCTCTAA
Cercozoan USE 4 (144 bp)	Cerf4F	CGGCCAAGCGATCGAGCAGA
	Cerc4R	CGATCGCTTGGCCGGGTACTG
Cercozoan USE 5 (54 bp)	Cerc5F	GCCAGAGTACCCCAGTACACGCT
	Cerc5R	GCTTGGAGATCAAAGATCTGCGGTGG

4.2.6. FISH analysis of Columbia River estuarine samples

An Alexa Fluor 555 (Life Technologies) fluorescently labeled oligonucleotide probe homologous to the unique sequence element of *Euduboscquella* CRE 1 was

designed (DubC1R: 5'- AGGGAATCCGGCTCTCTTGGAAA-3'). FISH was conducted for a sample collected from Youngs Bay in September 2012 (YB Sept-12, Table S1) during a bloom of the mixotrophic ciliate *Mesodinium* sp.. Aliquots of water samples for FISH were fixed with paraformaldehyde (4% final concentration) and stored at 4°C. Protocols for detection of specific protist taxa by FISH have been reported previously (Pernthaler et al. 2001, Massana et al. 2006). Briefly, 10 mL fixed aliquots were filtered on 0.6 μm pore size polycarbonate filters and hybridized for 3 h at 46 °C in the appropriate buffer (with 30% formamide) (Pernthaler et al. 2001, Massana et al. 2006), washed at 48°C in a second buffer (Pernthaler et al. 2001, Massana et al. 2006), counter-stained with 4',6-diamidino-2-phenylindole (DAPI; 5 μg μL^{-1}) and proflavin (Sherr et al. 1993), and mounted on a slide. Cells were then observed by epifluorescence microscopy (Axiovert 200M, Zeiss) on 400x magnification with oil immersion under a Cy3 filter.

4.2.7. Illumina sequencing

Nucleic acid samples were sent for Illumina sample preparation and two lanes of Illumina HiSeq 2000 paired-end 2x151 bp sequencing at the Oregon State University Center for Genome Research & Biocomputing. Resulting raw output files were then processed by the HiSeq 2000 CASAVA pipeline where they were demultiplexed and returned as .fastq files. Sequences were then submitted to MG-RAST 3.0 pipeline (Meyer et al. 2008), where poor quality bases (below phred score of 20) were trimmed, paired-end files were combined and overlapping paired-end reads were merged for automated annotation. Filtered and annotated sequence data are available through MG-RAST accession numbers 4533700-4533806. Sequences that were annotated as eukaryotic LSU

sequences were downloaded from the MG-RAST server and used for analysis of unique sequence elements detailed in the Results.

Analysis of LSU D2 sequences generated through an NGS approach (Illumina HiSeq) was undertaken in an effort to provide a more in-depth examination of unique sequence element diversity amongst protists and other eukaryotes in the Columbia River coastal margin. Illumina HiSeq libraries were constructed for freshwater (Salinity of 0), brackish (Salinity of ~15), and plume (salinity of ~28-31) samples from the spring and summer of 2007-2012. To isolate D2 region sequences flanking the unique sequence elements, an *in silico* search with the ORF1 reverse primer was conducted against annotated eukaryotic LSU sequences. LSU sequences that aligned with the ORF1 primer were annotated as D2 sequences and extracted from the total LSU dataset.

4.3. RESULTS

4.3.1. Identification of Unique Sequences Elements

A search of the LSU sequence of *Katablepharis* CRE for open reading frames (ORF) that might provide insight to the origin of the USE sequence within the D2 region revealed the remnants of two protein-coding genes on the antisense strand of the LSU sequence flanking the USE. An ORF specifying a segment of TAR1p of *Saccharomyces cerevisiae* (Accession CAY81395) was detected flanking the 5' end of the USE while an ORF corresponding to HLY5mc1 of *Entamoeba histolytica* (Accession CAA82858) was detected flanking the 3' end. A forward (ORF2: 5'-CGAACAAGTACCGTGAGGGAAAGATGCAAA-3') and reverse (ORF1: 5'-GACTCCTTGGTCCGTGTTTCAAGA-3') primer set designed to the TAR1p and

HLY5mc1 ORF sequences, respectively, were used in PCR amplification. Subsequent cloning and sequence analysis uncovered USEs in the LSU genes of other protist taxa in water samples from the Columbia River coastal margin. The primer sequences were also used to uncover USE sequences in metagenomic collections from other locations.

From 9 samples collected in coastal margin waters (8 from the Columbia River coastal margin and one from the Beaufort Sea), a total of 113 nucleotide sequences from the LSU D2 region were retrieved using the ORF1/ORF2 primer set (NCBI Accession numbers to be determined). While the majority (95%) of the resulting amplicons showed significant homology (e-value < $1e^{-6}$) to existing sequences in the NCBI database, 5 sequences attributable to protist taxa contained regions of the same approximate size and location (148-294 bp), but showed no significant similarity (e-value > $1e^{-6}$) to any sequence in the NCBI database (analysis done August 1, 2014) (Fig. 4.2).

4.3.2. Unique sequence elements uncovered through a next-generation sequencing approach

A total of 5792 LSU D2 sequences were retrieved from samples collected in the Columbia River coastal margin (including freshwater, brackish, and plume samples) and subjected to a BLAST (Basic Local Alignment Search Tool) search against sequences in the NCBI database. 43 USE were discovered in the D2 sequence dataset. None of the 43 contained regions with significant similarity (e-value > $1e^{-6}$) to any sequences in the NCBI database. The sequences also failed to align to annotated LSU sequences, or any other sequences, when searched against metagenomes outside of the Columbia River coastal margin, including CAMERA and IMG-MER. The USE detected in Columbia

River coastal margin metagenomes belonged primarily to heterotrophic flagellates, including several cercozoan sequences and the previously identified *Katablepharis* CRE USE (Kahn et al. 2014), as well as dinoflagellates and parasitic chytrid fungi (Table 4.3). Several of the USEs were related to multicellular taxa, including sequences attributed to annelids, rotifers, and bryozoans. Chytrid and cercozoan USE detected in the Illumina HiSeq dataset were validated through successful amplification (using specific primers), cloning, and sequence analysis. PCR amplicon sequences aligned well (>97%) with the corresponding USE detected in the Illumina HiSeq dataset.

Euduboscquellae CRE USE 1 (Columbia River Estuary)
CGAACAAGTACCGTGAGGGAAAGATGCAAAGTACTTTGAAAAGAGGGTTATATGCGCCTGAA
ACCACTGTATTACAAGCACTTGGATCCTGTTTTTTCGTTGGATTAG**CTGATTTAACAGGCGAC**
GTGAACCGTCAGGGAGTCCTCACGGAGTCTCACCGAGTCAATTGAACCAGTGGAATCTA
CCGTCAATCACCAATGTAGTTAAAGGGGCTGTCAATTCCTCATGTGGGAGCGTTCTCTG
AACACTTCGCATGACGATAAGTTGGCACCAAGTTTTGCTTTCCAAGAGAGCCGGATTCC
CTGACACAGCATATTCTGATTTTTTCCAAAAGCGCATGCGTTGAGGTCAGATCAAAATGT
GCTGTCAGGTACGGATTCTCTTTGGCATTGGTACCAAAAGGTTTCAAATGACCCGTCTTGA
AACACGGACCAAGGAGTC

Euduboscquellae BSL USE 1 (Beaufort Sea lagoon)
CGAACAAGTACCGTGAGGGAAAGATGCAAAGTACTTTGAAAAGAGGGTTATATGCGCCTGAA
ACCACTGTATTACAAGCACTTGGATCCTGTTTTTTCGTTGGATTAG**GCTGGTTTACGCAGCGA**
CATGTTCCGTCGATGAAATCTCACGGGATCATAACGCGTCACACAATCCAGCAGAATCA
AACGACAAGCTCCAATGCTATCAAAGGAGCTACCAACAAACCATAGAGACGAGTTCCTT
GAACACCTTGTATGGTAATACGAAAGCATCGAATTTTGACTAACAAGAAACTCCTACTCC
ATATATCAGTGTATTCTCGTTCTATCTGCAGCCGCGCGCGCTGGAGTCTGTTTAAGGTAC
ACTTTCTGGTGTTATATTTCCCTTGCACTGGAAGCAAAAAGGTTTCAAATGACCCGTCTTG
AAACACGGACCAAGGAGTC

Uncultured deep-water diplonemid (NE Pacific Coast)
CGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAAGTCCCTGAA
ACCGATGAGTGGTAAG**AGTACGTGACGCGCCCAATTCTCGTAATTGCGAAAGCTGTTGCG**
AGTCGTCTGCACTAGGTAACTTCGACTACTGGCTGTTCCAGATTCTGGTTCGGTTGTGCA
GATCGAGGATGCCCAGTAACAGGTCTAGAAGGCAACCGGACTGCGCTTCGCCGCTAGGT
AAGGGCAATTCACTTCGGTGGAGTGTTCTTCCTGGTGTCGTGGGTGTACAAGGTTGTTT
CTCGATGGCCTGGTTGTAGACACAGCGATCACGCACTACCCGTCTTGAAACACGGACCAA
GGAGTC

Katablepharis CRE (Columbia River Estuary)
GAGACCGATAGCGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTA
AAAAGTGCTTGAAATTGTTAAGAGGAAAAACGATTGA**CTCCAATACTGTATCCACGGTAAT**
TCAGCTGGGGCTTATGGATTCTGGGCGTTGTGACACCGTGAAGGGTCCTGCGTTCGGGC
GAGAAGTAAGCACCAGTGCACTTTGCCGGGGAATTAGGCCAGCATCAGATACTTCCGCG
GGAGATGGCGGCCGGTAGTGGGTGGGTTTCGATCTGCTTTGCCGGTTGCGGCACCGGG
GGAGTGTCTGAGGAATGAAGGGCGCGCTTTAGCGGGACGCCTAGTCAATTGTTGCCGTT
GGACGGGATGCACTGCTTTGCGGTGTCCCTGGCCAGAGGTCTCAGTTGGTTAGAACGCA
CCTGCGCCCTGAACATGCTTGGCGAAATGGTATCAACCGACCCGTCTTGAAACACGGACCA
AGGAGTCTAACACGTATGCGAGTAGCTGGGTGACAAACCTAGATGCGTAATGAAAG

Uncultured stramenopile (Columbia River estuary)
AGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGGGTGAAAAGCGCCTGAAATTGCT
GAAAGGGAAGGGATTGGTAGCAATAC**CACCGACCTCCGTAGCCGAAAGGTGCAAGGGGC**
CGGGTGTCAGTATCAGTTCATGCCGGAGGATATCGCTGATGAGGAGGTACGTGCTTCGG
CACCGTTTAACTCATTGGTTGTGCTCTGGCGTGGACTGAGGAAAATTGTTCTCAGGATAC
TGACGAAATGCTACTAATCCACCCGTCTTGAAACACGGACCAAGGAGTC

Figure 4.2. Unique sequence elements (USE) detected with the ORF1/ ORF2 primer set. DNA sequences in bold and highlighted green indicate USE, regions that show no significant similarity to LSU sequences in sequence databases.

4.3.3. Unique Sequence Elements highlight restricted biogeography of closely related taxa

4.3.3.1 Case study #1: the parasitic dinoflagellate, Euduboscquella, includes estuarine and coastal phylotypes

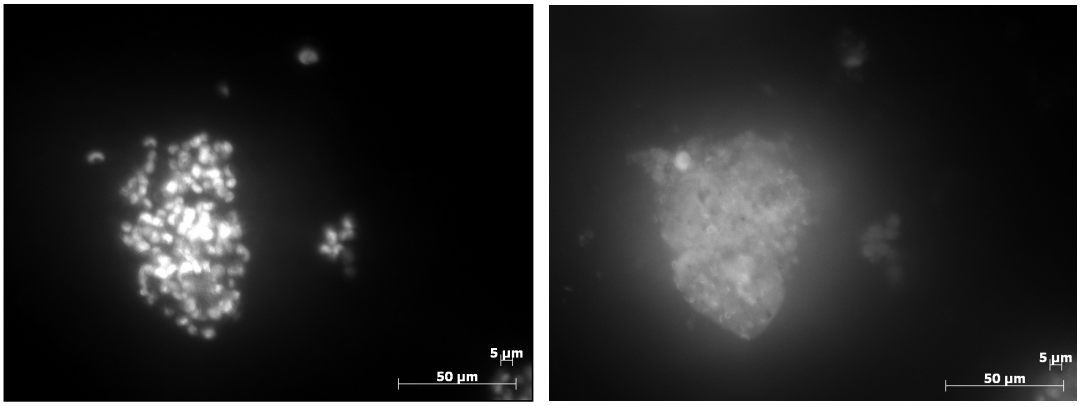
We discovered evidence of closely related, yet sufficiently different USE belonging to a protist taxon found in samples originating from different geographical locations. Two different sequences retrieved using ORF1/ORF2 primers contained highly variable LSU D2 regions that were closely related to the parasitic dinoflagellate *Euduboscquella* sp. Ex *Favella arcuata* isolate OC20 (accession number JN934989) based on the sequences flanking the elements. The ORF primers retrieved one *Euduboscquella* sequence from the Beaufort Sea lagoon (294 bp) and a different one from the Columbia River coastal margin (292 bp). The sequences flanking the elements aligned closely to each other (93-98% similarity), yet the elements themselves did not align with each other or with any other sequence in the NCBI database. These data were corroborated when a second set of primers designed to target USE from *Euduboscquella* were employed in PCR reactions (Table 1 with Dub2F, Dub2R). Out of 27 amplicons sequenced with the second primer set, 22 aligned ~100% with the USE retrieved using the ORF1/ ORF2 primers. The remaining five sequences contained a second USE (*Euduboscquella* USE 2, Accession number XX), which consisted of a 292 bp region with no significant similarity to any sequences in the NCBI database. The USE 2 contained a 225 bp region with 70% similarity to *Euduboscquella* USE 1, suggesting that by using species-specific primers, fine-scale taxonomic resolution can be achieved when USE are present.

In order to demonstrate the potential utility of USE for ecological studies, we performed quantitative PCR studies using samples from the Columbia River coastal margin, including the estuary and adjacent coastal ocean (see Table S4.1). 18S small subunit rRNA sequences attributable to the genus *Eudubosc*quella were previously detected at high abundance during *Mesodinium* blooms in this estuary in late summer and early autumn (unpubl. data). Analysis by FISH of a *Mesodinium*-rich sample revealed that *Eudubosc*quella were associated with tintinnid ciliates, since they were observed within tintinnid lorica (Fig. 4.3).

*Eudubosc*quella USE 1 was detected in greatest abundance during *Mesodinium* bloom periods with low abundances of USE 1 copies detected in the estuary before the onset of the blooms (Fig. 4.4). *Eudubosc*quella USE 1 abundance displayed a strong positive correlation to *Mesodinium* cell abundance ($r = 0.94$, $p < 0.01$) (data not shown) as well as chlorophyll fluorescence ($r = 0.86$, $p < 0.01$) (data not shown). In contrast, *Eudubosc*quella USE 2 did not display any clear seasonal trend or correlation to *Mesodinium* abundance. Instead, *Eudubosc*quella USE 2 was detected at greatest abundance in coastal samples (Fig. 4.5), with relatively low copy numbers in the estuarine samples. The number of USE 2 amplicons was moderately positively correlated to salinity ($r = 0.44$, $p < 0.01$), suggesting that the phylotype bearing USE 2 may be confined to coastal rather than estuarine waters. Furthermore, neither USE were detected in samples outside of the Columbia River coastal margin (i.e. Amazon River plume, Chesapeake Bay, Susquehanna River, and Beaufort Sea Lagoon), or other available metagenomic datasets, suggesting that both phylotypes have a restricted biogeographic distribution.

Table 4.3. LSU USE detected in Columbia River coastal margin metagenomes.

USE name	Taxa	Closest LSU sequence in NCBI	Number non-aligned bp to closest NCBI hit
Alve1	Alveolata	Alveolata sp. CCMP3155	92
Anne1	Annelida	*Protodrilus ciliatus*	70
Basi1	Basidiomycota	*Antherospora vaillantii* voucher HAI 2857	72
Bico1	Bicosoecida	*Bicosoeca* sp. HFCC85	101
Bico2	Bicosoecida	*Nerada mexicana* strain ATCC 50061	90
Bodo1	Bodonidae	*Neobodo designis* HFCC95	94
Boli1	Bolidophyceae	DH114_3A83	62
Bryo1	Bryozoa	*Plumatella* sp. ZHY-2005	66
Cent1	Centroheliozoa	*Chlamydaster sterni* strain HFCC65	56
Cerc1	Cercozoa	*Bodomorpha minima* strain ATCC 50339	115
Cerc2	Cercozoa	*Cercomonas* sp. HFCC556	80
Cerc3	Cercozoa	*Cercomonas* sp. HFCC564	68
Cerc4	Cercozoa	Cercozoa sp. Brady Beach 2007	160
Cerc5	Cercozoa	*Cryothecomonas* sp. APCC MC5-1Cryo	64
Cerc6	Cercozoa	*Gymnophrys* sp. ATCC 50923	139
Cerc7	Cercozoa	*Gyromitus* sp. HFCC94	52
Choa1	Choanoflagellida	*Didymoeca costata*	116
Choa2	Choanoflagellida	*Monosiga brevicollis*	90
Choa3	Choanoflagellida	*Monosiga* sp. ATCC 50635	86
Chyt1	Chytridiomycota	*Alphamyces chaetifer* isolate ARG003	69
Chyt2	Chytridiomycota	Chytridiomycota sp. LLSA5_1 PML-2011s	75
Chyt3	Chytridiomycota	Chytridiomycota sp. LLSF2_1 PML-2011a	174
Chyt4	Chytridiomycota	*Rhizophlyctis rosea* isolate JEL205	69
Chyt5	Chytridiomycota	*Rhizophydium chlorogonii*	126
Cili1	Ciliophora	*Coleps hirtus* isolate CoHi	99
Cili2	Ciliophora	*Paramecium putrinum*	61
Cryp1	Cryptophyta	*Falcomonas daucoides*	133
Dino1	Dinophyceae	*Amphidinium semilunatum*	130
Dino2	Dinophyceae	*Borghiella dodgei*	172
Dino3	Dinophyceae	*Gymnodinium corollarium* clone K-0983	146
Dino4	Dinophyceae	*Gymnodinium pygmaeum* strain K-0968	136
Dino5	Dinophyceae	*Heterocapsa niei* isolate IFR10-193	70
Dino6	Dinophyceae	*Karlodinium* sp. KAMS0708	68
Dino7	Dinophyceae	*Pelagodinium beii*	110
Dino8	Dinophyceae	*Peridinella catenata*	64
Dino9	Dinophyceae	*Peridiniella* sp. NC-2011	97
Dino10	Dinophyceae	*Prorocentrum sigmoides*	157
Dino11	Dinophyceae	*Protodinium* sp. CCMP419	66
Dino12	Dinophyceae	*Woloszynskia tenuissima*	162
Fung1	Fungi incertae sedis	*Olpidium brassicae*	73
Kata1	Katablepharidaceae	*Katablepharis japonica* strain NIES 1334	149
Oomy1	Oomycota	*Plasmopara* sp. 1 HV-2013	105
Roti1	Rotifera	*Ploesoma truncatus*	89



A. B.

Figure 4.3. Epifluorescence micrographs of Columbia River estuary water stained with A. a USE-specific probe for the *Eudubosquella* CRE USE 1 strain; and B. the corresponding DAPI-stained microscopic field.

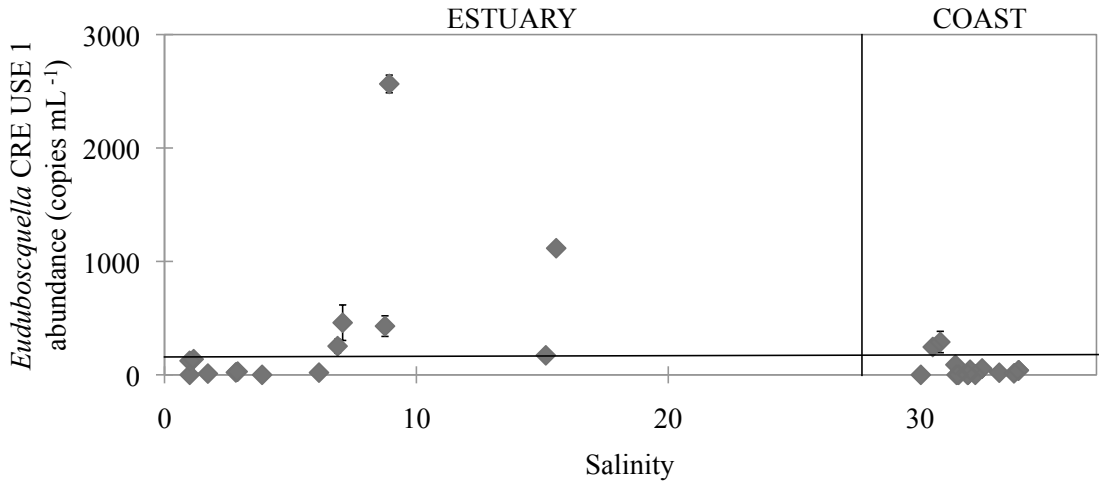


Figure 4.4. *Eudubosquella* CRE USE 1 gene abundance estimated by qPCR for Columbia River coastal margin samples. Coastal samples were collected from 2007-2013 at a variety of locations, depths, and dates (Table S1). Error bars represent standard deviation of triplicate qPCR samples. Horizontal black line indicates three times the

average standard deviation. USE 1 gene abundance had a very strong positive correlation with chlorophyll concentration and *Mesodinium* cell abundance.

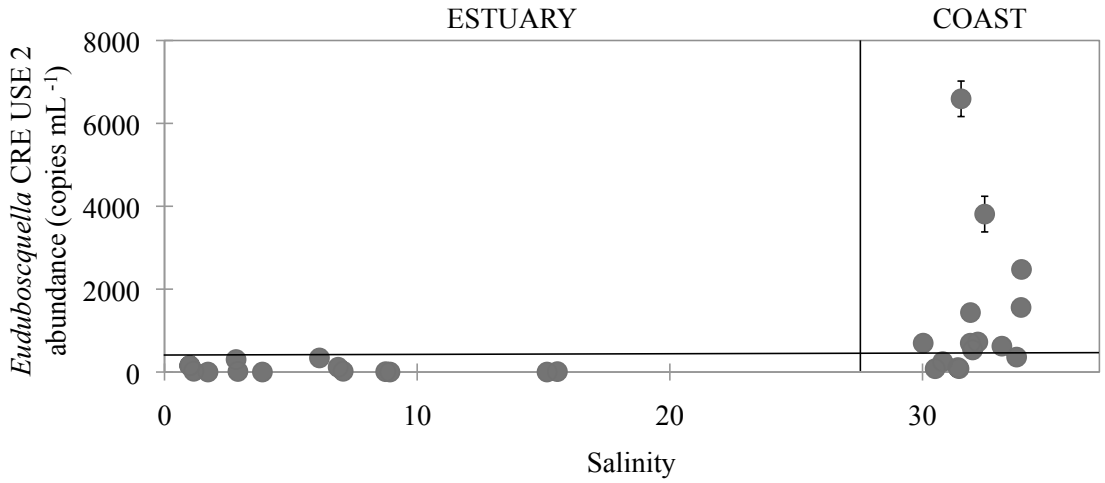


Figure 4.5. *Eudubosquella* CRE USE 2 gene abundance estimated by qPCR for Columbia River coastal margin samples. Coastal samples were collected from 2007-2013 at a variety of locations, depths, and dates (Table S1). Error bars represent standard deviation of triplicate qPCR samples. Horizontal black line indicates three times the average standard deviation. USE 2 gene abundance had a moderate positive correlation with salinity.

4.3.3.2 Case study #2: Extensive USE diversity in Diplonemea reveals depth-specific differences in phylotype

Using the ORF1/ORF2 primer set on a deep-water sample 40 km off the Oregon coast, we uncovered a USE within the LSU rRNA gene of diplonemids (Diplonemea), a group of heterotrophic flagellates found in deep water environments (Lara et al. 2009).

When more specific primer sets were used to generate PCR amplicons (i.e., DiploF/ ORF1 and DiploF/ LR6 in Table 4.1) from samples collected throughout the Columbia River coastal margin, a total of 53 different USE related to diplonemids (based on their flanking sequences) were uncovered (Fig. 4.6). The set of diplonemid USE showed little homology to each other, or to any other sequences in the NCBI database.

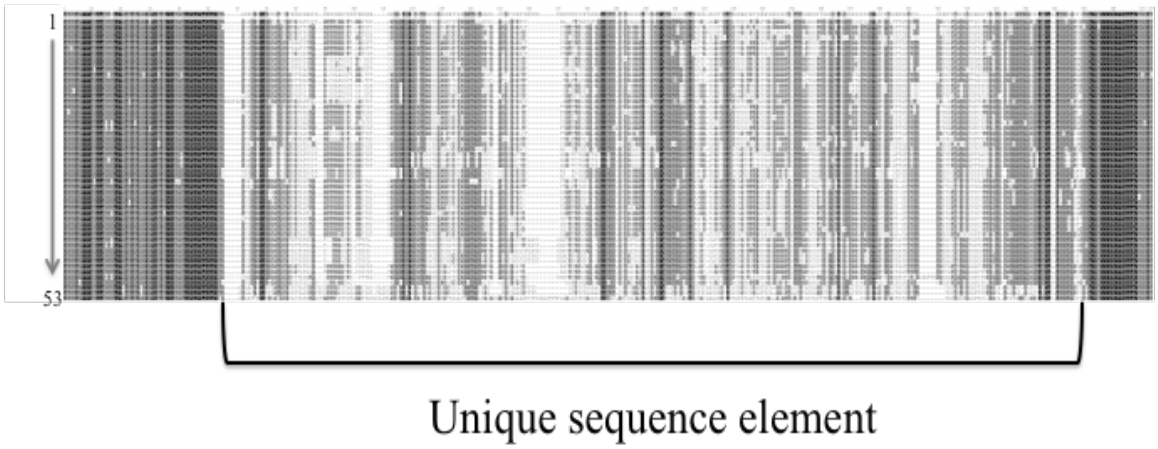


Figure 4.6. Alignment of diplonemid unique sequence elements detected off the Oregon coast. 53 USEs related to diplonemids were detected. Alignment score is indicated by grayscale, with higher score indicated by darker shade. The drop in homology denoted in brackets is indicative of the unique sequence elements.

While many of the 53 diplonemid USE were detected at multiple locations throughout the Columbia River coastal margin, several were associated with specific depth ranges and would therefore make potentially useful biogeographical indicators. For example, two USE were consistently detected in waters deeper than 25 meters (phylotype Dip 2 and Dip 3), while one USE was observed at a wider range of depths in the water colum (phylotype Dip 1), (Table 4.4). The USE ascribed to each of the three phylotypes

displayed low homology (particularly with respect to the level of homology typically observed within the rRNA operon), bearing only 61-69% similarity to each other. Using qPCR primers specific for the three phylotypes, we characterized the vertical distribution of the three USE at a site 40 km off the Oregon coast (CR-40). The highest USE copy numbers here associated with Dip 1 in waters shallower than 25 m, while the deep phylotypes occupied deeper depths as indicated by elevated numbers of Dip 2 and Dip 3 USE copies, with little overlap in depth range of the different phylotypes (Fig. 4.7). Low oxygen waters entering the Columbia River estuary during upwelling periods also contained elevated numbers of Dip 2 copies compared to normoxic estuarine waters (Fig. 4.8), suggesting that this phylotype could also provide a genetic marker for estuarine intrusion of hypoxic coastal waters. The USE appears to be a more discriminating marker than the internal transcribed spacer (ITS) region (ITS1-5.8S-ITS2): a similarity of 69% was observed between Dip 2 and Dip 3 USE compared to 84% between the ITS region of both phylotypes.

Table 4.4. Endpoint PCR analysis of Dip 1, Dip 2, and Dip 3 diplonemid USE phylotypes with primer sets specific to each USE. Amplification is indicated by '+' while no amplification is indicated by '-'.

Site ID	Sampling Date	Sampling depth (m)	Dip 1 phylotype	Dip 2 phylotype	Dip 3 phylotype
CR40	Apr-14	1.0	+	-	-
		20.0	+	-	-
		100.0	-	+	+
		150.0	-	+	+
		200.0	-	+	+
		400.0	-	+	+
		800.0	+	+	+
GH41	Sep-13	1.0	+	-	-
		250.0	-	+	+
		800.0	-	+	+
LP6	Sep-09	2.0	+	-	-
		10.0	+	+	-
		50.0	-	+	+
LP17	Sep-09	2.0	+	-	-
		25.0	+	+	-
		125.0	-	+	+
LP32	Sep-09	2.0	+	-	-
		12.0	+	+	-
		750.0	-	+	+
LP52	Sep-09	2.0	+	-	-
		19.0	+	+	-
		1545.0	-	+	+

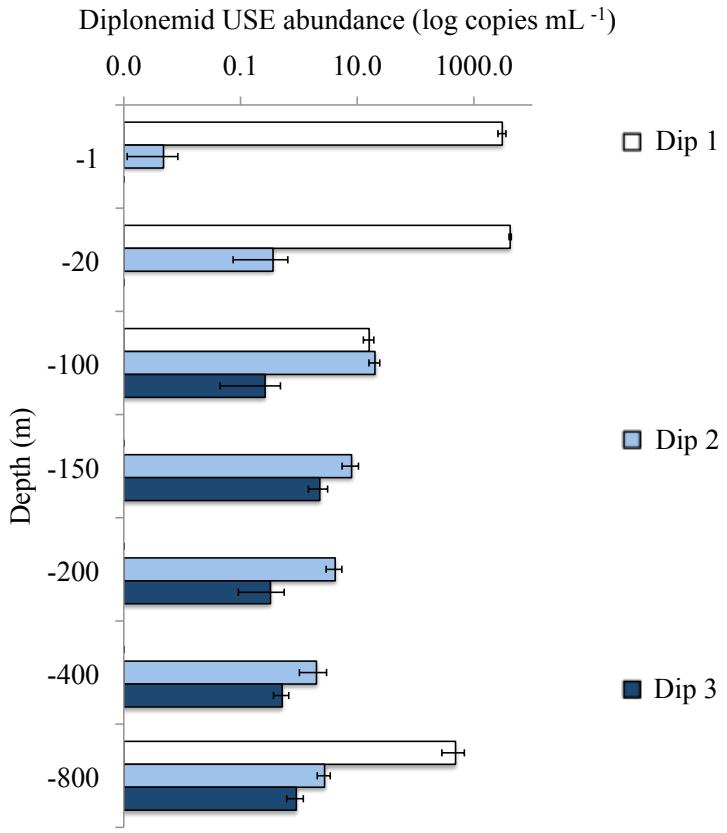


Figure 4.7. Abundance of diplonemid phylotypes Dip 1, Dip 2, and Dip 3 in a water sample collected in April 2014 at a site (CR40) located 40 km off the Oregon coast. Primers used in amplification were designed to anneal to the USE of each phylotype. The Dip 1 phylotype is most abundant within the first 50 meters, while Dip 2 and 3 were abundant below 100 m.

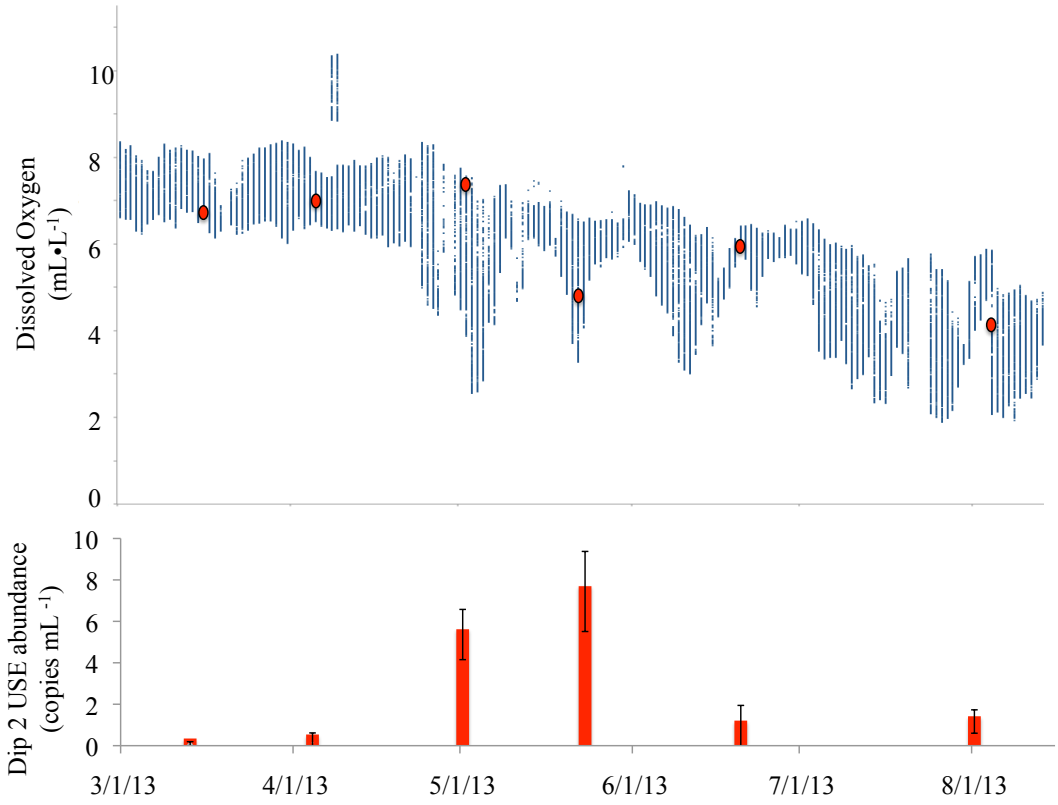


Figure 4.8. Abundance of Dip 2 diplonemid phylotype (bottom panel) as well as dissolved oxygen concentrations (top panel) of the bottom waters (13 m) of the estuarine monitoring station SATURN-03 in Mar-Jul 2013. Location of SATURN-03 is provided in Fig. 1. Dip 2 abundance was greatest in low oxygen waters, suggesting it could be utilized as a genetic marker for hypoxic water intrusion into the estuary.

4.3.3.3 Case study #3: USE suggests cosmopolitan vs. geographically limited Cercozoan phylotypes

The geographic distribution of two USE found in our Illumina dataset, Cerc4 and Cerc5 (Table 4.3), was also characterized through a qPCR assay using samples from the Columbia River coastal margin and other environments (Amazon River, Chesapeake Bay, Susquehanna River, and Beaufort Sea Lagoon). There was a strong correlation

between USE 4 amplicon copies and chlorophyll fluorescence ($r = 0.72$, $p < 0.01$) in the estuary (but not the coast) in samples from September 2012 (Fig. 4.9). In contrast, relatively high and low abundances of USE 5 amplicons were observed on the coast and in the Columbia River estuary, respectively (Fig. 4.10). There was a moderate positive correlation between the abundance of USE 5 amplicons and salinity ($r = 0.57$, $p < 0.01$). Although the USE 5 sequence was not detected in any of the samples collected outside the Columbia River coastal margin, the USE 4 sequences were detected in a broad range of samples from the Amazon River, Chesapeake Bay, Susquehanna River, and Beaufort Sea Lagoon, suggesting a more cosmopolitan distribution.

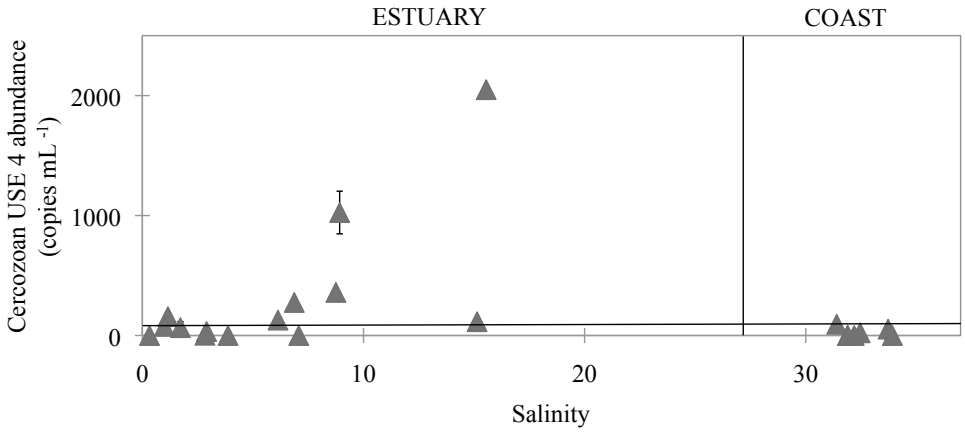


Figure 4.9. Cercozoan USE 4 gene abundance estimated by qPCR for Columbia River coastal margin samples. Coastal samples were collected from 2007-2013 at a variety of locations, depths, and dates (Table S1). Error bars represent standard deviation of triplicate qPCR samples. Horizontal black line indicates three times the average standard deviation.

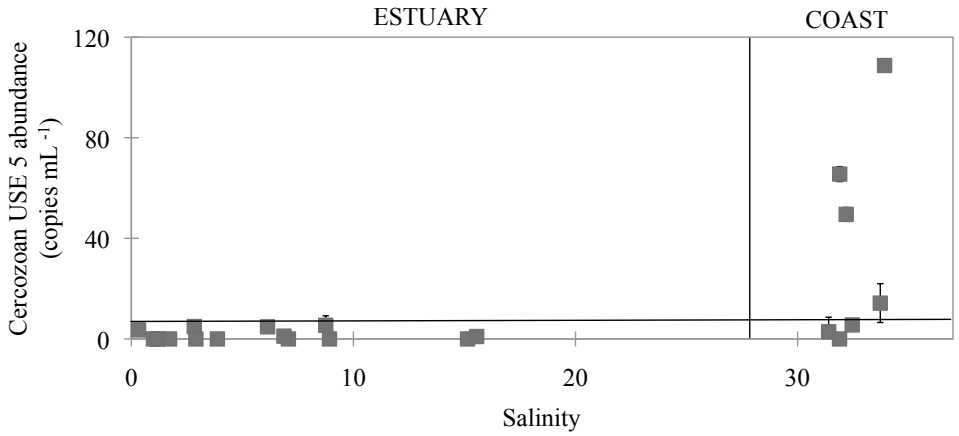


Figure 4.10. Cercozoan USE 5 gene abundance estimated by qPCR for Columbia River coastal margin samples. Coastal samples were collected from 2007-2013 at a variety of locations, depths, and dates (Table S4.1). Error bars represent standard deviation of triplicate qPCR samples. Horizontal black line indicates three times the average standard deviation.

4.3.4. Unique sequence elements uncovered from a Delaware Coast metagenomic study

In order to determine the USEs from a system outside of the Columbia River coastal margin, we surveyed a metagenomic dataset from the Delaware Coast (GOLD Project ID in IMG-MER database: Gm00309). An *in silico* search using the ORF1 and ORF 2 primers was conducted on all metagenomic sequences without first retrieving the annotated LSU sequences. All 67 sequences retrieved were homologous to the D2 LSU region. Out of these 67 sequences from the D2 region, 23 (34%) contained USE with no significant similarity (e-value > $1e^{-6}$) to any sequences in the NCBI database. Furthermore, the USE detected from the Delaware Coast dataset were not detected in the Columbia River coastal margin dataset. The majority of the USE were related to ciliates, while a USE related to *Eduboscquella* was also detected (Table 4.5). This Delaware

Coast *Euduboscquella* USE was also 292 bp in length, the same size as the two Columbia River coastal margin USEs, but was < 50% similar to either the Columbia River USEs or to the Beaufort Sea lagoon USE.

Table 4.5. LSU unique sequence elements detected in a Delaware Coast metagenome. Elements that were detected multiple times are indicated in parentheses. The USE related to *Euduboscquella* sp. ex *Favella arcuata* isolate OC20 shows little homology to the *Euduboscquella* USEs detected in Columbia River coastal margin and Beaufort Sea lagoon samples.

Taxa	Closest LSU sequence in NCBI	Number non-aligned bp to closest NCBI hit
Chlorophyceae	*Dunaliella tertiolecta* isolate PL1	237
Choanoflagellida	*Diaphanoeca grandis* strain ATCC 50111	367
Ciliophora	*Pseudocohnilembus hargisi* isolate JJM2010031301 (2)	178
Ciliophora	*Strombidium crassulum* isolate SNK09121701	138
Ciliophora	*Strombidium rassoulzadegani* (4)	197
Ciliophora	Uncultured ciliate clone E06_757_MCM2_2	193
Ciliophora	Uncultured marine ciliate clone DH18_2A74 (2)	119
Copepoda	*Acartia tonsa*	143
Diatoma	*Plagiogramma atomus*	229
Dinophyceae	*Amoebophrya* sp. ex *Akashiwo sanguineaum*	342
Dinophyceae	*Euduboscquella* sp. ex *Favella arcuata* isolate OC20 (2)	292
Dinophyceae	*Scrippsiella rotunda* isolate D232	321
Oomyceta	Uncultured oomycete clone OTU-59	157
Picozoa	*Picomonas judraskeda*	288
Telonemida	*Telonema subtile* strain RCC 404 (2)	293

4.4. DISCUSSION

The discovery of the unique sequence elements (USE) within the D2 region of LSU rRNA gene, presented first in Kahn et al. 2014 and further expanded in this study,

has several important implications for the current debate over protist biogeography. This study demonstrates that the scale of taxonomic resolution is crucial when assessing protist biogeography. The USEs provide a more discriminating genetic marker than either the SSU, the rest of the LSU, or in the case of the diplonemids, the ITS regions, other commonly used discriminating genetic markers. For example, analysis of the biogeography of USE-bearing taxa, such as *Katablepharis* CRE or *Eudubosquellae* CRE 1 or 2, based on SSU or LSU sequences might suggest they have a ubiquitous distribution, with nearly identical (>99%) sequence similarity to several other protists sequenced from a variety of environments around the world. In contrast, analysis of the USEs associated with these microorganisms suggested that they possess a more restricted biogeographical distribution. Additionally, unlike ITS, the USE is an integral part of the LSU rRNA gene, providing a high copy target that is not removed by pre-rRNA processing. It has long been known that the D2 region of the LSU rRNA gene is a region undergoing an elevated rate of evolution (Hassouna et al. 1984), and that it shows high sequence polymorphism, even among members of the same species (Beszteri et al. 2005).The ubiquity and high copy number of the LSU allows for a quantitative analysis of the distribution of protist variants based on D2 sequences, for example using qPCR or FISH assays. Opportunities to exploit the D2 sequences exist even when USEs are not present, since it is possible to take advantage of the extremely high polymorphism in this region to design highly specific primers or probes.

The USEs specific for organisms from the Columbia River coastal margin are rare within the Illumina HiSeq collection analyzed in this study. However, a survey of the metagenome of the Delaware Coast GOLD project using the ORF1 and ORF2 primers

uncovered a higher percentage of USEs within the D2 region. USEs are readily identified in amplicons generated by the two ORF primers. ORF1 corresponds to a remnant of a coding sequence found in the *TAR1p* gene of *S. cerevisiae* as well as in a gene (locus tag: MTR_5g051040) annotated as encoding an rRNA intron homing endonuclease of *Medicago truncatula.* The Tar1p protein is encoded by a transcript that is anti-sense with respect to the 28S rRNA sequence, and is localized to mitochondria where it resides in the inner membrane. It is believed to interact with an enzyme responsible for coenzyme Q biosynthesis (Bonawitz & Chatenay-Lapointe 2008). The putative assignment of the *Medicago truncatula* gene showing homology to Tar1p and ORF1 sequence as a homing endonuclease has not been functionally validated, and the amino acid sequence shows little if any homology to known intron homing endonucleases. At present, we do not know why these PCR primers, specifying ORF remnants in the 28S rRNA gene frequently give rise to amplicons that bear USEs.

The origins of the USEs are a mystery, as they do not resemble the products of nucleotide substitution, addition or deletion that would yield sequence polymorphisms. Instead they resemble the products of sequence translocation, transposition, or horizontal gene transfer. The previously describe USE of *Katablepharis* CRE showed an elevated GC content compared with the surrounding SSU and LSU sequence, suggestive of some mechanism of horizontal gene transfer responsible for the delivery of the element to the D2 region (Andersson 2005). Apparently, the function of the LSU rRNA can tolerate the insertion of a USE in the D2 region, or, alternatively, the USE-bearing LSU rRNA allele encodes a product the function of which is comprised, but is tolerated because the USE is found in a small percentage of genes within the nucleolus. The R2 class of retroposons of

Drosophila simulans (Zhou & Eickbush 2009) and in notostracan crustaceans (Luchetti et al. 2012) are examples of transposition events targeting LSU rRNA genes, which is advantageous for transposon propagation since interruption of a small percentage of the rRNA gene copies in the nucleolus will not likely impair rRNA expression and growth of the host.

As demonstrated here, the USEs have the potential to be an important tool for determining protist biogeography, both for tracking specific protist taxa distributions within a system and globally. Differences in the presence and abundance among *Euduboscquella* CRE USEs, among cercozoan USEs, and among diplonemid USEs suggest that these elements can be used to characterize fine-scale biogeographic distributions within a system, even between closely related taxa, which could in turn shed light on the functional diversity of protist taxa. For example, the extreme diplonemid USE diversity observed within a single water sample and the coexistence of several phylotypes at the same depth may indicate that the diplonemids are highly specialized grazers that derive resources from different prey, feeding behavior, or parasitized hosts (Lara et al. 2009). The depth-specific differences of diplonemid USE phylotypes identified in this study could then reflect differences in their ecological specializations, with diplonemid distributions dictated by the stratification of their preferred prey or hosts (Lara et al. 2009).

Many of the elements detected in the Columbia River coastal margin have not been found in other systems tested in the present study, suggesting a possible restricted biogeographic distribution that would support the moderate endemicity theory of protist biogeography. Previous support for this theory has included use of "flagship species",

species that are so morphologically discriminating that they can easily be detected in an environmental sample (Foissner 2006). Likewise, these USE can be employed in a similar fashion as “flagship sequences”, sequences that are so discriminating that highly specific probes can be designed for them and thus be used to qualitatively and quantitatively monitor their biogeographic distribution. In addition to its application for protist biogeography, the potential utility of the USE is wide-reaching, and could be employed for a wide variety of environmental applications, such as harmful algal bloom monitoring, parasitology, probe development for ecogenomic monitoring platforms like the Environmental Sample Processor sandwich hybridization system (Scholin 2010), or in the tracking of ballast water if it is found to have a distinct genetic signature.

Chapter Five: Concluding Remarks and Future Directions

Estimation of protist diversity and classification of its taxonomy has been a challenging undertaking for over two centuries, due to the enormous morphological and physiological variability and broad taxonomic representation. All approaches that are used for characterization of protist assemblages have biases that affect our interpretation of the real extent of protist diversity and abundance. However, successful estimation of protist diversity must include all protists regardless of cell size, morphology, or physiology. Traditional techniques such as light microscopy that utilize the morphospecies concept to identify protist taxa are increasingly deemed insufficient in the characterization of protist diversity (McManus & Katz 2009). Over the last few decades molecular biology has provided a powerful array of alternative approaches to diversity assessments, such as clone libraries and NGS technologies. One genetic marker, the SSU rRNA gene, has frequently been used for whole and partial sequence analysis in protist diversity studies. These studies have greatly enhanced our ability to characterize protist assemblages, particularly by refining the taxonomic resolution of small, morphologically non-descript cells and the rare biosphere, and increasing richness estimates of important protist groups, such as ciliates, by an order of magnitude (Adl et al. 2007).

There have been concerns over the interpretation of diversity assessments using rRNA gene sequence analysis as a single approach due to the highly variable rRNA gene copy number in protists (Gong et al. 2013) as well as PCR amplification, primer selectivity, and cloning biases inherent to Sanger sequencing methods (Vargas et al.

2009). The problem of multiple rRNA gene copy number in protists currently requires a second method to support the DNA sequence data. In **Chapter Two** of this dissertation I analyzed SSU rRNA gene sequences using an NGS approach (Illumina HiSeq) and traditional Sanger sequencing, as well as microscopic cell counts to determine the distribution of two major heterotrophic protist groups, the ciliates and dinoflagellates, in the Columbia River estuary (including tidal freshwater areas) and its plume during the spring and summer. Overall, the good accordance in diversity and abundance estimates of the Illumina HiSeq and light microscopy datasets for ciliates and heterotrophic dinoflagellates, two dominant heterotrophic groups, leads to the conclusion that this approach can generate reliable protist surveys and is suitable for characterizing transitions in protist diversity and abundances across a wide range of aquatic environments. Furthermore, these are in agreement with previous work indicating that NGS approaches can greatly increase taxa richness estimates compared to light microscopy or traditional Sanger sequencing approaches. PCR-amplified sequences retrieved by the Sanger method, in contrast, did not adequately capture heterotrophic protist diversity, and correlations between cell abundance and Sanger sequences of ciliates and heterotrophic dinoflagellates were weak.

While Sanger sequencing provides increased taxonomic resolution of protist diversity compared to Illumina HiSeq because of longer read lengths allowed by this method, the cloning bias, potential for PCR and primer biases, and small throughput, confine its application for diversity analyses. As read length continues to increase, and cost continues to decline, for NGS tools such as Illumina HiSeq, or 454-pyrosequencing, their applicability in protist diversity studies will increasingly constitute the

methodological standard. Combining these approaches with traditional methods, such as microscopy and cultivation, and studies of biogeochemical rate processes, will extend the link between DNA sequence datasets and morphological, physiological, and functional information, a crucial step for a deepened understanding of the interactions and relationships of heterotrophic protists in food webs and biogeochemical cycles.

So far, the majority of protist surveys in the Columbia River coastal margin were conducted by traditional approaches like light microscopy and HPLC analysis of photosynthetic pigments, and focused primarily on autotrophs. The research presented in this dissertation is the first to characterize in detail the seasonal distribution of heterotrophic protist assemblages across the river-to-ocean continuum in the Columbia River coastal margin. In **Chapters Two** and **Three**, taxonomically distinct assemblages were observed from the freshwater portion of the estuary to the river plume, with seasonal shifts in assemblage composition occurring in both the estuary and plume. Ciliates were the dominant heterotrophic protists in the estuary during both spring and summer according to analysis of the Illumina amplicon sequences, with a seasonal transition in abundant taxa occurring from the spring (primarily *Tintinnidium* and *Rimostrombidium*) to summer (primarily *Stokesia*), driven potentially in part by changes in preferred prey availability. Ciliates were also the dominant heterotrophic protists in the plume during the spring; however, a transition towards dominance of heterotrophic dinoflagellates (*Gyrodinium* and *Gymnodinium*) between the spring and the summer was evident. The seasonal transition from ciliates to dinoflagellates has been attributed to the higher growth rates of ciliates compared to dinoflagellates (Strom & Morello 1998), which allows a more rapid response by ciliates to enhanced spring phytoplankton

biomass. The more generalist nature of heterotrophic dinoflagellates as consumers, however, likely results in their greater persistence when ciliates preferred prey items become less abundant (Sherr & Sherr 2007).

The discovery of the unique sequence element (USE) within the LSU of *Katablepharis* CRE in **Chapter Three** provided an excellent template to determine the spatial and temporal patterns in absolute abundances of *Katablepharis* CRE in the system through USE-specific probes combined with quantitative and qualitative methods, such as through a specific qPCR assay. Importantly, the USE also allowed for visualization of this uncultured katablepharid through FISH. Through these specific quantitative assays I was able to determine that *Katablepharis* CRE is an abundant spring heterotrophic flagellate in the Columbia River estuary, and that it is biogeographically restricted to the Columbia River coastal margin. Events occurring in the spring, such as upriver diatom blooms and spring runoff, deliver organic matter to the estuary (Sullivan et al. 2001) that could fuel katablepharid proliferation. Given their high abundance and repeatable temporal patterns of occurrence, *Katablepharis* CRE is likely an essential and yet previously undetected link between the microbial and herbivorous food webs in the Columbia River estuary.

Chapters Two and **Three** of this dissertation lay the groundwork for future studies to further investigate the ecological roles of heterotrophic protists in the Columbia River coastal margin. A characterization of heterotrophic protist diversity is critical for enhancing our knowledge of ecosystem functioning, as it provides a necessary first step in determining their impact in food webs and biogeochemical cycles. Future research that focuses on the specific response of single protists under different abiotic and biotic

factors, such as in grazing experiments and microcosm experiments, can help further elucidate the drivers of protist assemblage structure and ecological function. The USE marker discovered in **Chapter Three** provides a potentially powerful tool for determining the physiological and ecological roles of specific taxa. For example, it could be utilized as a highly discriminating marker in bacterivory and herbivory studies of *Katablepharis* CRE to determine its grazing impact and role in the microbial loop, a necessary step to a more comprehensive characterization of organic matter cycling in the Columbia River coastal margin.

Recently, a FISH protocol was employed (Massana et al. 2009) that labeled the predator, the uncultured marine stramenopile (MAST) lineages, with an oligonucleotide probe to study grazing rates of the uncultured MAST lineages on fluorescently labeled bacteria (FLB). In this study both the live bacterial prey and MAST predators were labeled with fluorescence oligonucleotide probes. This protocol is appealing in the case of *Katablepharis* CRE, or other heterotrophic flagellates such as the diplonemids and cercozoans presented in **Chapter Four**, as they are difficult to distinguish morphologically but contain USE that has allowed for specific probe development. The lack of knowledge concerning heterotrophic protist physiology and food web interactions in the Columbia River coastal margin also constitutes an important drawback for ecosystem modeling. Filling these gaps will greatly improve predictions concerning the influence of anthropogenic and environmental change on the Columbia River coastal margin system.

Whether or not the debate over protist biogeography will be resolved depends greatly on our ability to reconcile morphospecies descriptions with DNA sequence

information and agree on an appropriate level of taxonomic resolution that accurately reflects the functional biogeographic distribution of protist taxa. As demonstrated in **Chapter Four**, the USEs also have the potential to be an powerful tool for determining protist biogeography, both for tracking specific protist taxa distribution within a system and globally. Many of the USE detected in the Columbia River coastal margin have not been detected in other systems, suggesting a possible restricted biogeography that would support the moderate endemicity theory. USEs can be applied to the characterization of fine scale biogeographic distribution of USE-bearing protists within a system, as well as for determining their global biogeographic distribution, as demonstrated by differences in the presence and abundance amongst *Euduboscquella* CRE USEs, amongst cercozoan USEs, and amongst diplonemid USEs. This in turn can shed light on the functional diversity and ecological niches of closely related protist taxa. For example, the extreme diplonemid USE diversity observed within a single water sample, and the coexistence of several phylotypes at the same depth, may indicate that the diplonemids are highly specialized grazers that derive resources from different prey, feeding behavior, or parasitized hosts (Lara et al. 2009). The depth specific differences of diplonemid USE phylotypes identified in this study could then reflect differences in their ecological specializations, with diplonemid distribution dictated by the stratification of their preferred prey or hosts (Lara et al. 2009). Likewise, differences in the spatiotemporal distribution of *Euduboscquella* CRE USEs might indicate differences in their preferred parasitized tintinnid hosts, which could in turn be an important factor in shaping heterotrophic protist assemblages, as *Euduboscquella* species are known to produce mass

lethal infections of its host and helps faciliate population decline and primary production recycling in the microbial loop (Coats & Heisler 1989).

Importantly, the distinct set of USEs found in other environments, such as the Delaware Coast, suggest that the USEs can be employed to characterize a distinct genetic signature for a system. Further characterization of USE diversity, variability and distribution from a variety of other environments, such as other coastal margin systems, the open ocean, freshwater lakes, and extreme environments (e.g. hypersaline lakes), would help further determine the utility of the USE region in biogeographic assessments of protist taxa. For example, *Euduboscquella* CRE USE 1 and 2 are more closely related than the *Euduboscquella* USE detected from either the Beaufort Sea lagoon or the Delaware Coast. Future studies examining the degree of *Euduboscquella* USE diversity from a wide range of environments and geographic location would help determine the drivers of USE variability, such as geographic distance or other factors (e.g. differences in host specificity or physiology). In addition to its application for protist biogeography, the USE could be utilized to increase the taxonomic resolution of other environmental applications, such as ecogenomic monitoring platforms like the Environmental Sample Processor sandwich hybridization system (Scholin 2010). Through USE-specific probes, they could be utilized for qualitative or quantitative monitoring at fine scale temporal resolution, which would further illuminate the drivers of population dynamics of specific protist taxa.

In summary, the chapters of this dissertation have helped unravel the phylogenetic diversity and spatiotemporal dynamics of heterotrophic protists in the Columbia River coastal margin, and can serve as status quo for future changes of heterotrophic protist

assemblages due to changing anthropogenic or environmental forcings. It also presents new tools to further our knowledge of protist biogeography at a fine-scale taxonomic resolution. Further research focused on heterotrophic protists role in the pelagic food web and integration into ecosystem modeling will allow for more accurate predictions of ecosystem processes.

References

Adl SM, Leander BS, Simpson AGB, Archibald JM, Anderson OR, Bass D, Bowser SS, Brugerolle G, Farmer MA, Karpov S, Kolisko M, Lane CE, Lodge DJ, Mann DG, Meisterfeld R, Mendoza L, Moestrup Ø, Mozley-Standridge SE, Smirnov A V, Spiegel F (2007) Diversity, nomenclature, and taxonomy of protists. Syst Biol 56:684–689

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov A V, Spiegel FW, Taylor MFJR (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451

Andersson JO (2005) Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62:1182–97

Arndt H, Dietrich D, Auer B, Cleven EJ, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In: The flagellates: unity, diversity and evolution, 240–268.

Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F (1983) The Ecological Role of Water-Column Microbes in the Sea. Mar Ecol Prog Ser 10:257–263

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. ISME J 7:244–255

Baptista A, Howe B, Freire J, Maier D, Silva CT (2008) Scientific exploration in the era of ocean observatories. Comput. Sci. Eng. 10:53–58

Barnes CA, Duxbury AC, Morse BA (1972) Circulation and selected properties of Columbia River effluent at sea. In: Pruter AT, Alverson DL (eds) The Columbia River Estuary and Adjacent Ocean Waters. University of Washington Press Seattle WA, Ul–80

Bass D, Boenigk J (2011) Everything is everywhere: a twenty-first century de-/reconstruction with respect to protists. In: Fontaneto D (ed) Biogeography of Microscopic Organisms: Is Everything Small Everywhere? Cambridge University Press United Kingdom, p 88-110

Bass D, Cavalier-Smith T (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). Int J Syst Evol Microbiol 54:2393–2404

Bass D, Richards TA, Matthai L, Marsh V, Cavalier-Smith T (2007) DNA evidence for global dispersal and probable endemicity of protozoa. BMC Evol Biol 7:162

Bazin P, Jouenne F, Deton-Cabanillas A-F, Pérez-Ruzafa Á, Véron B (2013) Complex patterns in phytoplankton and microeukaryote diversity along the estuarine continuum. Hydrobiologia 726:155–178

Becking L (1934) Geobiologie, of Inleiding Tot de Milieukunde: Met Literatuurlijst en Ind. Van Stockum The Netherlands, p 263

Beijerinck M (1913) De infusies ende ontdekking der bacteriën. Jaarboek van de Koninklijke Akademie v. Wetenschappen Muller The Netherlands, p 119-140

Bent SJ, Forney LJ (2008) The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. ISME J 2:689–695

Berninger (1986) The Spatial Distribution and Ecology of Zoochlorellae Bearing Ciliates in a Productive Pond. J Protozool 33:557–563

Beszteri B, Acs E, Medlin LK (2005) Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. Protist 156:317–333

Bonawitz N, Chatenay-Lapointe M (2008) Expression of the rDNA-encoded mitochondrial protein Tar1p is stringently controlled and responds differentially to mitochondrial respiratory demand and dysfunction. Curr Genet 54(2): 83-94

Bradford TM, Morgan MJ, Lorenz Z, Hartley DM, Hardy CM, Oliver RL (2013) Microeukaryote community composition assessed by pyrosequencing is associated with light availability and phytoplankton primary production along a lowland river. Freshw Biol 58:2401–2413

Calbet A (2008) The trophic roles of microzooplankton in marine systems. ICES J Mar Sci 65:325–331

Calbet A, Landry MR (1999) Mesozooplankton influences on the microbial food web : Direct and indirect trophic interactions in the oligotrophic open ocean. Limnol Oceanogr 44:1370–1380

Calbet A, Landry M (2004) Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. Limnol Oceanogr 49:51–57

Caron DA, Goldman JC, Fenchel T (1990) Protozoan respiration and metabolism. In: Ecology of marine protozoa. Oxford University Press, New York, NY, p 307–322

Caron DA (1991) Evolving Role of Protozoa in Aquatic Nutrient Cycles. In: Reid PC, Turley CM, Burkill PH (eds) Protozoa and Their Role in Marine Processes. Springer Berlin Heidelberg, Berlin, Heidelberg, p 387–415

Caron DA (2009a) Past President's address: protistan biogeography: why all the fuss? J Eukaryot Microbiol 56:105–112

Caron DA (2009b) New Accomplishments and Approaches for Assessing Protistan Diversity and Ecology in Natural Ecosystems. Biosci 59:287–299

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012) Marine Protistan Diversity. Ann Rev Mar Sci 4:467–493

Chao A (1984) Nonparametric estimation of the number of classes in a population. Scand J Stat 11:265–270

Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93:2533–2547

Chawla A, Jay D, Baptista A (2008) Seasonal variability and estuary–shelf interactions in circulation dynamics of a river-dominated estuary. Estuaries and Coasts 31:269–288

Cho B, Azam F (1990) Biogeochemical significance of bacterial biomass in the ocean's euphotic zone. Mar Ecol Prog Ser Oldend 63:253–259

Cleven E, Weisse T (2001) Seasonal succession and taxon-specific bacterial grazing rates of heterotrophic nanoflagellates in Lake Constance. Aquat Microb Ecol 23:147–161

Coats DW, Heisler JJ (1989) Spatial and temporal occurrence of the parasitic dinoflagellate Duboscquella cachoni and its tintinnine host Eutintinnus pectinis in Chesapeake Bay. Mar Biol 101:401–409

Cotterill FPD, Al-Rasheid K, Foissner W (2007) Conservation of protists: is it needed at all? Biodivers Conserv 17:427–443

Crump B, Baross J (1996) Particle-attached bacteria and heterotrophic plankton associated with the Columbia River estuarine turbidity maxima. Mar Ecol Prog Ser 138:265–273

Crump BC, Baross JA, Simenstad CA (1998) Dominance of particle-attached bacteria in the Columbia River estuary, USA. Aquat Microb Ecol 14:7–18

Crump BC, Peterson BJ, Raymond PA, Amon RM, Rinehart A, McClelland JW, Holmes RM (2009) Circumpolar synchrony in big river bacterioplankton. Proceedings of the National Academy of Sciences, 106(50):21208–21212

Darling KF, Wade CM (2008) The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. Mar Micropaleontol 67:216–238

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

Doherty M, Costas, BA, McManus GB, Katz, LA (2007) Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. Aquatic microbial ecology, 48(2): 141.

Dolan JR (1991) Microphagous ciliates in mesohaline Chesapeake Bay waters: estimates of growth rates and consumption by copepods. Marine Biology, 111(2): 303-309

Dolan JR (1997) Phosphorus and ammonia excretion by planktonic protists. Mar Geol 139:109–122

Dolan JR, & Gallegos CL (2001) Estuarine diversity of tintinnids (planktonic ciliates). Journal of Plankton Research, 23(9): 1009-1027

Domaizon I, Viboud S, Fontvieille D (2003) Taxon-specific and seasonal variations in flagellates grazing on heterotrophic bacteria in the oligotrophic Lake Annecy - importance of mixotrophy. FEMS Microbiol Ecol 46:317–329

Drummond A, Ashton B, Buxton S (2010) Geneious v5. 3. Biomatters, Ltd.

Eiler A, Drakare S, Bertilsson S, Pernthaler J, Peura S, Rofner C, Simek K, Yang Y, Znachor P, Lindström ES (2013) Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. PLoS One 8:e53516

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

Fenchel T, Finlay B (2004) The ubiquity of small species: patterns of local and global diversity. Bioscience 54(8): 777-784

Fenchel T, Finlay B (2006) The diversity of microbes: resurgence of the phenotype. Philosophical Transactions of the Royal Society B: Biological Sciences 361(1475): 1965-1973

Finlay B (2002) Global dispersal of free-living microbial eukaryote species. Science 296(5570):1061–1063

Finlay B, Clarke K (1999) Ubiquitous dispersal of microbial species. Nature 400(6747): 828-828.

Finlay BJ, Fenchel T (1999) Divergent perspectives on protist species richness. Protist 150:229–233

Finlay BJ, Fenchel T (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 155:237–244

Foissner W (1999) Protist diversity: estimates of the near-imponderable. Protist 150(4): 363-368

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool 45(2): 111-136

Foissner W (2007) Dispersal and biogeography of protists: recent advances. Jpn J Protozool 40: 1-16

Foissner W, Chao A, Katz L (2009) Diversity and geographic distribution of ciliates (Protista: Ciliophora). In: Foissner W, Hawksworth D (eds) Protist Diversity and Geographical Distribution. Springer Netherlands, p 111–129

Fontaneto D, Brodie J (2011) Why biogeography of microorganisms? In: Fontaneto D (ed) Biogeography of Microscopic Organisms: Is Everything Small Everywhere? Cambridge University Press United Kingdom, p 3-10

Fontaneto D, Hortal J (2013) At least some protist species are not ubiquitous. Mol Ecol 22:5053–5055

Frame E, Lessard E (2009) Does the Columbia River plume influence phytoplankton community structure along the Washington and Oregon coasts? J Geophys Res: Oceans (1978–2012): 114(C2)

Frey B, Lara-Lara R, Small L (1984) Water column primary production in the Columbia River estuary. Columbia River Estuary Data Development Program, Astoria OR, p 133

Gifford DJ (1991) The Protozoan-Metazoan Trophic Link In Pelagic Ecosystems. J Protozool 38:81–86

Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. Protist 164:369–379

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Beattie AJ (2004) Spatial scaling of microbial eukaryote diversity. Nature, 432(7018): 747-750

Haertel L, Osterberg C, Curl H, Park PK (1969) Nutrient and Plankton Ecology of the Columbia River Estuary. Ecology 50:962-978

Hallegraeff GM, Bolch CJ (1992) Transport of diatom and dinoflagellate resting spores in ships' ballast water: implications for plankton biogeography and aquaculture. J Plankton Res 14:1067–1084

Hambright K, Zohary T, Gude H (2007) Microzooplankton dominate carbon flow and nutrient cycling in a warm subtropical freshwater lake. Limnol Oceanogr 52:1018–1025

Hassouna N, Michot B, Bachellerie J, Narbonne D (1984) The complete nucleotide sequence of mouse 28S rRNA gene: Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucl Acids Res 12(8):3563–3583

Heger T, Mitchell E (2009) The curse of taxonomic uncertainty in biogeographical studies of free living terrestrial protists: a case study of testate amoebae from Amsterdam Island. J Biogeogr 36(8): 1551-1560

Herfort L, Peterson TD, Campbell V, Futrell S, Zuber P (2011) *Myrionecta rubra* (*Mesodinium rubrum*) bloom initiation in the Columbia River estuary. Estuar Coast Shelf Sci 95:440–446

Herfort L, Peterson T, McCue L, Zuber P (2011) Protist 18S rRNA gene sequence analysis reveals multiple sources of organic matter contributing to turbidity maxima of the Columbia River estuary. Mar Ecol Prog Ser 438:19–31

Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R (2010) Capturing diversity of marine heterotrophic protists : one cell at a time. ISME J 5:674–684

Hickey B, Geier S, Kachel N, MacFadyen A (2005) A bi-directional river plume: The Columbia in summer. Cont Shelf Res 25:1631–1656

Hickey BM, Pietrafesa LJ, Jay DA, Boicourt WC (1998) The Columbia River Plume Study: Subtidal variability in the velocity and salinity fields. J Geophys Res 103(C5):10339-10368

Hobson LA (1963) Some influences of the Columbia River effluent on marine phytoplankton during January 1961. Universty of Washington Press Seattle WA 367:223–234

Horner RA, Postel JR, Halsband-Lenk C, Pierson JJ, Pohnert G, Wichard T (2005) Winter-spring phytoplankton blooms in Dabob Bay, Washington. Progress in Oceanography, 67(3): 286-313.

Hsieh T, Ma K, Chao A (2013) iNEXT online: interpolation and extrapolation (Version 1.0)

Hughes FW, Rattray M (1980) Salt flux and mixing in the Columbia River Estuary. Estuar Coast Mar Sci 10:479–493

Huyer A (1983) Coastal upwelling in the California Current system. Prog Oceanogr 12:259–284

Jakobsen HH, Hansen PJ (1997) Prey size selection, grazing and growth response of the small heterotrophic dinoflagellate Gymnodinium sp. and the ciliate Balanion comatum-a comparative study. Marine ecology progress series, 158: 75-86

Jeong, HJ, Yoo YD, Kim ST, Kang NS (2004) Feeding by the heterotrophic dinoflagellate Protoperidinium bipes on the diatom Skeletonema costatum. Aquatic microbial ecology, 36(2): 171-179.

Jeong HJ, Du Yoo Y, Kim JS, Seong KA, Kang NS, Kim TH (2010) Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. Ocean science journal, 45(2): 65-91.

Jürgens K, Massana R, Kirchman D (2008) Protist grazing on marine bacterioplankon. Kirchman DL (ed) Microbial ecology of the oceans, 2nd edn. Wiley-Liss, New York, 383-442

Jürgens K, Matz C (2002) Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. Antonie Van Leeuwenhoek 81:413–34

Jurgens K, Stolpe G (1995) Seasonal dynamics of crustacean zooplankton, heterotrophic nanoflagellates and bacteria in a shallow, eutrophic lake. Freshw Biol 33:27–38

Kahn P, Herfort L, Peterson TD, Zuber P (2014) Discovery of a *Katablepharis* sp. in the Columbia River estuary that is abundant during the spring and bears a unique large ribosomal subunit sequence element. MicrobiologyOpen 3(5): 764–776

Katz LA (2012) Origin and diversification of eukaryotes. Annu Rev Microbiol 66:411–27

Kim E, Wilcox L, Graham L, Graham J (2004) Genetically distinct populations of the dinoflagellate Peridinium limbatum in neighboring Northern Wisconsin lakes. Microbial Ecology, 48(4): 521-527

Kiss AK, Acs E, Kiss KT, Török JK (2009) Structure and seasonal dynamics of the protozoan community (heterotrophic flagellates, ciliates, amoeboid protozoa) in the plankton of a large river (River Danube, Hungary). Eur J Protistol 45:121–38

Knowlton N (1993) Sibling species in the sea. Annu Rev Ecol Syst 24: 189-216

Koch T, Ekelund F (2005) Strains of the heterotrophic flagellate *Bodo designis* from different environments vary considerably with respect to salinity preference and SSU rRNA gene composition. Protist 156:97–112.

Lacap DC, Lau MC, Pointing BS (2011) Biogeography of prokaryotes. In: Fontaneto D (ed) Biogeography of Microscopic Organisms: Is Everything Small Everywhere? Cambridge University Press United Kingdom, p 35-42.

Lara E, Heger T (2011) COI gene and ecological data suggest size dependent high dispersal and low intra specific diversity in free living terrestrial protists (Euglyphida: *Assulina*). J Biogeogr 38(4): 640-650

Lara E, Moreira D, Vereshchaka A, López-García P (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. Environ Microbiol 11:47–55

Lara-Lara JR , Frey BE, Small LF (1990) Primary production in the Columbia River estuary I. Spatial and temporal variability of properties. Pacific Sci 44:17–37.

Lehrter JC, Pennock JR, McManus GB (1999) Microzooplankton Grazing and Nitrogen Excretion across a Surface Estuarine-Coastal Interface. Estuaries 22:113-125

Lessard EJ, Swift E (1986) Dinoflagellates from the North Atlantic classified as phototrophic or heterotrophic by epifluorescence microscopy. Journal of Plankton Research, 8(6): 1209-1215

Lessard E, Murrell M (1998) Microzooplankton herbivory and phytoplankton growth in the northwestern Sargasso Sea. Aquat Microb Ecol 16:173–188

Lim E, Dennett M, Caron D (1999) The ecology of *Paraphysomonas imperforata* based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. Limnol Oceanogr 44:37–51

Löder M, Kraberg A, Aberle N (2012) Dinoflagellates and ciliates at Helgoland Roads, North Sea. Helgol Mar Res 66:11–23

López-García P, Rodríguez-Valera F (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409:603–607

Loreau M, Naeem S, Inchausti P (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. Science 294(5543):804–808

Luchetti A, Mingazzini V, Mantovani B (2012) 28S junctions and chimeric elements of the rDNA targeting non-LTR retrotransposon R2 in crustacean living fossils (Branchiopoda, Notostraca). Genomics 100:51–56

Lundholm N, Bates SS, Baugh KA, Bill BD, Connell LB, Léger C, Trainer VL (2012) Cryptic and pseudo-cryptic diversity in diatoms-with descriptions of *Pseudo-nitzschia Hasleana* sp. nov. and *P. Fryxelliana* sp. nov 1. J Phycol 48:436–454

Mallin MA (1994) Phytoplankton Ecology of North Carolina Estuaries. Estuaries 17:561

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes X V, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach A-L, Smith VH, Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112

Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Älios C (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environ. Microbiol. 8:1515–1522

Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J, Not F (2009) Grazing rates and functional diversity of uncultured heterotrophic flagellates. ISME J 3:588–596

Mayr E (1942) Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press Cambridge MA, p 334

McGrady-Steed J, Morin P (2000) Biodiversity, density compensation, and the dynamics of populations and functional groups. Ecology 81(2): 361-373

McManus GB, Katz LA (2009) Molecular and morphological methods for identifying plankton: what makes a successful marriage? J Plankton Res 31:1119–1129

Medinger R, Nolte V, Pandey RAMV, Jost S (2010) Diversity in a hidden world: potential and limitation of next- generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol ecol 19(S1):32–40

Meyer F, Paarmann D, D'Souza M (2008) The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386

Mitchell EAD, Meisterfeld R (2005) Taxonomic Confusion Blurs the Debate on Cosmopolitanism versus Local Endemism of free-living protists. Protist 156:263–267

Moloney C, Field J (1991) The size-based dynamics of plankton food webs. I. A simulation model of carbon and nitrogen flows. J Plankton Res 13:1003–1038

Monchy S, Grattepanche J-D, Breton E, Meloni D, Sanciu G, Chabé M, Delhaes L, Viscogliosi E, Sime-Ngando T, Christaki U (2012) Microplanktonic community

structure in a coastal system relative to a Phaeocystis bloom inferred from morphological and tag pyrosequencing methods PLoS One 7(6):e39924

Moreira D, López-García P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol 10:31–38

Müller H (2000) Evidence of dormancy in planktonic oligotrich ciliates. Verh Internat Verein Limnol 27:3206–3209

Müller H (2002) Laboratory study of the life cycle of a freshwater strombidiid ciliate. Aquat Microb Ecol 29:189–197

Muylaert K, Sabbe K, Vyverman W (2000) Spatial and temporal dynamics of phytoplankton communities in a freshwater tidal estuary (Schelde, Belgium). Estuarine, Coastal and Shelf Science, 50(5): 673-687

Nassonova E, Smirnov A, Fahrni J, Pawlowski J (2010) Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. Protist 161:102–115

Neal V (1972) Physical aspects of the Columbia River and its estuary. In: The Columbia River estuary and adjacent coastal waters, University of Washington Press Seattle WA. p, 19–40

Neuer S, Cowles T (1994) Protist herbivory in the Oregon upwelling system. Mar Ecol Prog Ser Oldend 113:147–162

Nolte V, Pandey RV, Jost S, Medinger R, Ottenwälder B, Boenigk J, Schlötterer C (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Mol Ecol 19:2908–2915

Odum E (1985) Trends expected in stressed ecosystems. Bioscience 35:419–422

Okamoto N, Inouye I (2005) A secondary symbiosis in progress? Science, 310(5746):287-287

Paerl H, Peierls B (2008) Ecological responses of the Neuse River–Pamlico Sound estuarine continuum to a period of elevated hurricane activity: Impacts of individual storms and longer. Am Fish Soc Symp 64:101–116

Paerl HW, Rossignol KL, Hall SN, Peierls BL, Wetz MS (2009) Phytoplankton Community Indicators of Short- and Long-term Ecological Change in the Anthropogenically and Climatically Impacted Neuse River Estuary, North Carolina, USA. Estuaries and Coasts 33:485–497

Patterson DJ (1999) The Diversity of Eukaryotes. Am Nat 154:S96–S124

Pernthaler J, Glockner FO, Schonhuber W, Amann R (2001) Fluorescence in situ hybridization (FISH) with rrna-targeted oligonucleotide probes. Methods Microbiol. 30:207–226.

Petchey OL, McPhearson PT, Casey TM, Morin PJ (1999) Environmental warming alters food-web structure and ecosystem function. 402:69–72

Pfandl K, Chatzinotas A (2009) SSU rRNA gene variation resolves population heterogeneity and ecophysiological differentiation within a morphospecies (Stramenopiles, Chrysophyceae). Limnol Oceanogr 54(1): 171-181

Ploug H, Zimmerman-Timm H, Schweitzer B (2002) Microbial communities and respiration on aggregates in the Elbe Estuary, Germany. Aquat Microb Ecol 27:241–248

Pomeroy L (1974) The ocean's food web, a changing paradigm. Bioscience 24:499 – 504

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nuc Acids Res 41:D590–596

Rodríguez-Martínez R, Labrenz M, Campo J del, Forn I, Jürgens K, Massana R (2009) Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. Environ Microbiol 11:397–408

Roegner GC, Needoba JA, Baptista AM (2011) Coastal upwelling supplies oxygen-depleted water to the Columbia River estuary. PLoS One 6(4):e18672

Rollwagen-Bollens G, Bollens SM, Penry DL (2006) Vertical distribution of micro- and nanoplankton in the San Francisco Estuary in relation to hydrography and predators. Aquat Microb Ecol 44:143–163

Rollwagen-Bollens G, Gifford S, Bollens S (2011) The Role of Protistan Microzooplankton in the Upper San Francisco Estuary Planktonic Food Web: Source or Sink? Estuaries and coasts 34:1026–1038

Rodríguez-Martínez R, Labrenz M, del Campo J, Forn I, Jürgens K, Massana R (2009) Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. Environ. Microbiol. 2009:397–408.

Sanger F (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Santoferrara LF, Grattepanche J-D, Katz LA, McManus GB (2014) Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic

ciliates and comparison with traditional methods. Environ Microbiol 16(9): 2752-2763

Savin MC, Martin JL, LeGresley M, Giewat M, Rooney-Varga J (2004) Plankton diversity in the Bay of Fundy as measured by morphological and molecular methods. Microb Ecol 48:51–65

Schlegel M, Meisterfeld R (2003) The species problem in protozoa revisited. Eur J Protistol 39:349–355

Scholin C (2010) What are“ ecogenomic sensors?” A review and thoughts for the future. Ocean Sci 6(1): 51-60.

Sharma N, Rai A (2010) Biodiversity and biogeography of microalgae: progress and pitfalls. Environ Rev 19:1-15.

Sherr EB, Caron DA, Sherr BF (1993) Staining of heterotrophic protists for visualization via epifluorescence microscopy. Handbook of methods in aquatic microbial ecology. Lewis Publishers Boca Raton FL, p 213–227

Sherr EB, Sherr BF (1994) Bacterivory and herbivory: Key roles of phagotrophic protists in pelagic food webs. Microb Ecol 28:223–235

Sherr EB, Sherr BF (2002) Significance of predation by protists in aquatic microbial food webs. Antonie Van Leeuwenhoek 81:293–308

Sherr EB, Sherr BF, Wheeler PA (2005) Distribution of coccoid cyanobacteria and small eukaryotic phytoplankton in the upwelling ecosystem off the Oregon coast during 2001 and 2002. Deep Sea Research Part II: Topical Studies in Oceanography, 52(1): 317-330

Sherr EB, Sherr BF (2007) Heterotrophic dinoflagellates: a significant component of microzooplankton biomass and major grazers of diatoms in the sea. Mar Ecol Prog Ser 352:187–197

Sherwood C, Jay D (1990) Historical changes in the Columbia River estuary. Prog Oceanogr 25:299–352

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21:1794–1805

Simenstad CA, Small LF, David McIntire C, Jay DA, Sherwood C (1990) Columbia river estuary studies: An introduction to the estuary, a brief history, and prior studies. Prog Oceanogr 25:1–13

Simenstad C, Small L, McIntire CD (1990) Consumption processes and food web structure in the Columbia River estuary. Prog Oceanogr 25:271–297

Škaloud P, Rindi F Ecological differentiation of cryptic species within an asexual protist morphospecies: a case study of filamentous green alga *Klebsormidium* (Streptophyta). J Eukaryot Microbiol 60:350–362

Slapeta J, López-García P, Moreira D (2006a) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol Biol Evol 23:23–29

Slapeta J, López-García P, Moreira D (2006b) Present status of the molecular ecology of kathablepharids. Protist 157:7–11

Small LF, McIntire CD, MacDonald KB, Lara-Lara JR, Frey BE, Amspoker MC, Winfield T (1990) Primary production, plant and detrital biomass, and particle transport in the Columbia River Estuary. Prog Oceanogr 25:175–210

Small L, Menzies D (1981) Patterns of primary productivity and biomass in a coastal upwelling region. Deep Sea Res Part A 28:123–149

Smith H, Bobrov A, Lara E (2008) Diversity and biogeography of testate amoebae. Biodiversity and Conservation 17(2): 329-343

Sogin M, Morrison H (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci 103:12115–12120

Staay SM, Wachter R De, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409:607–610

Stock A, Jürgens K, Bunge J, Stoeck T (2009) Protistan diversity in suboxic and anoxic waters of the Gotland Deep (Baltic Sea) as revealed by 18S rRNA clone libraries. Aquat Microb Ecol 55:267–284

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

Stoeck T, Jost S, Boenigk J (2008) Multigene phylogenies of clonal *Spumella*-like strains, a cryptic heterotrophic nanoflagellate, isolated from different geographical regions. Int J Syst Evol Microbiol 58:716–724

Stoeck T, Stock A (2009) The protistan gap in the eukaryotic tree of life. Paleodiversity 3:151–154

Strom SL (2000) Bacterivory: Interactions between bacteria and their grazers. Microb Ecol Ocean:351–386

Strom S, Brainard M, Holmes J, Olson M (2001) Phytoplankton blooms are strongly impacted by microzooplankton grazing in coastal North Pacific waters. Mar Biol 138:355–368

Strom S, Morello T (1998) Comparative growth rates and yields of ciliates and heterotrophic dinoflagellates. J Plankton Res 20:571–584

Sullivan B., Prahl F., Small L., Covert P. (2001) Seasonality of phytoplankton production in the Columbia River: A natural or anthropogenic pattern? Geochim Cosmochim Acta 65:1125–1139

Sverdrup H (1953) On conditions for the vernal blooming of phytoplankton. ICES J. Mar. Sci. 18:287–295

Terrado R, Vincent W, Lovejoy C (2009) Mesopelagic protists: diversity and succession in a coastal Arctic ecosystem. Aquat Microb Ecol 56:25-40

Utermöhl H (1958) Zur vervollkommnung der quantitativen phytoplankton-methodik. Mitt int Ver theor angew Limnol 9:1–38

Vargas C De, Massana R, Not F, Campo J, Balague V (2009) New Insights into the Diversity of Marine Picoeukaryotes. PLoS One 4:e7143

Vigil P, Countway P (2009) Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. Aquat Microb Ecol 54:83-100

Urrutxurtu I, Orive E, Sota ADL (2003) Seasonal dynamics of ciliated protozoa and their potential food in an eutrophic estuary (Bay of Biscay). Estuarine, Coastal and Shelf Science, 57(5): 1169-1182.

Watts PC, Martin LE, Kimmance SA, Montagnes DJ, Lowe CD (2010) The distribution of *Oxyrrhis marina*: a global disperser or poorly characterized endemic? Journal of plankton research 33(4): 579-589

Weber F, del Campo J, Wylezich C, Massana R, Jürgens K (2012) Unveiling trophic functions of uncultured protist taxa by incubation experiments in the Brackish Baltic Sea. PLoS ONE 7:e41970

Weisse T (2008) Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. Biodivers Conserv 17(2):243-259

Wörner U, Zimmerman-Timm H, Kausch H (2000) Succession of Protists on Estuarine Aggregates. Microb Ecol 40:209–222

Wylezich C, Nies G, Mylnikov AP, Tautz D, Arndt H (2010) An evaluation of the use of the LSU rRNA D1-D5 domain for DNA-based taxonomy of eukaryotic protists. Protist 161:342–352

Zhou J, Eickbush TH (2009) The pattern of R2 retrotransposon activity in natural populations of *Drosophila simulans* reflects the dynamic nature of the rDNA locus. PLoS Genet 5:e1000386

Zhu F, Massana R, Not F, Marie D, Vaulot D (2005) Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol Ecol 52:79–92

Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. Mol Ecol 21:1878–96

Appendix

Table S3.1. Abundance of *Katablepharis* CRE USE in the Columbia River estuary estimated by qPCR from March-July 2013. NC refers to samples collected from the North Channel, EM in the estuary mouth, S03 near the SATURN-03 observatory station, SC in the South Channel, and S04 near the SATURN-04 observatory station (see Fig. 4.2 for exact locations). S = surface water; B = bottom water. S.D = standard deviation.

Date	Location	Depth	*Katablepharis* CRE gene copies mL $^{-1}$	S.D.
4-Apr-13	NC	S	4.2E+02	1.2E+02
		B	6.2E+03	5.8E+02
	EM	S	5.5E+03	3.2E+02
		B	9.6E+03	1.2E+02
	S03	S	5.8E+00	5.3E-01
		B	4.8E+03	1.2E+02
	SC	S	4.5E+03	1.6E+02
		B	2.7E+03	1.3E+02
	S04	S	9.8E+02	4.1E+01
		B	1.1E+04	2.2E+02
23-May-13	NC	S	1.1E+02	3.8E+02
		B	4.8E+02	8.0E+02
	EM	S	5.3E+02	1.4E+02
		B	2.1E+03	3.0E+00
	S03	S	0.0E+00	0.0E+00
		B	2.5E+01	3.0E+00
	SC	S	3.8E+01	4.3E+01
		B	0.0E+00	0.0E+00
	S04	S	0.0E+00	0.0E+00
		B	0.0E+00	0.0E+00
20-Jun-13	NC	S	1.5E+02	1.9E+00
		B	1.5E+04	3.3E+02
	EM	S	2.7E+01	4.2E-01
		B	2.6E-01	9.5E-03
	S03	S	3.1E+03	2.8E+02
		B	8.0E+03	4.9E+02
	SC	S	2.7E+01	2.4E+01

Table S3.1
continued

Date	Location	Depth	*Katablepharis* CRE gene copies mL $^{-1}$	S.D.
		B	5.7E-01	8.2E-04
18-Jul-13	NC	S	8.2E+01	8.6E+00
	S03	S	8.9E+01	7.2E+00
	SC	S	8.0E+01	2.3E+00

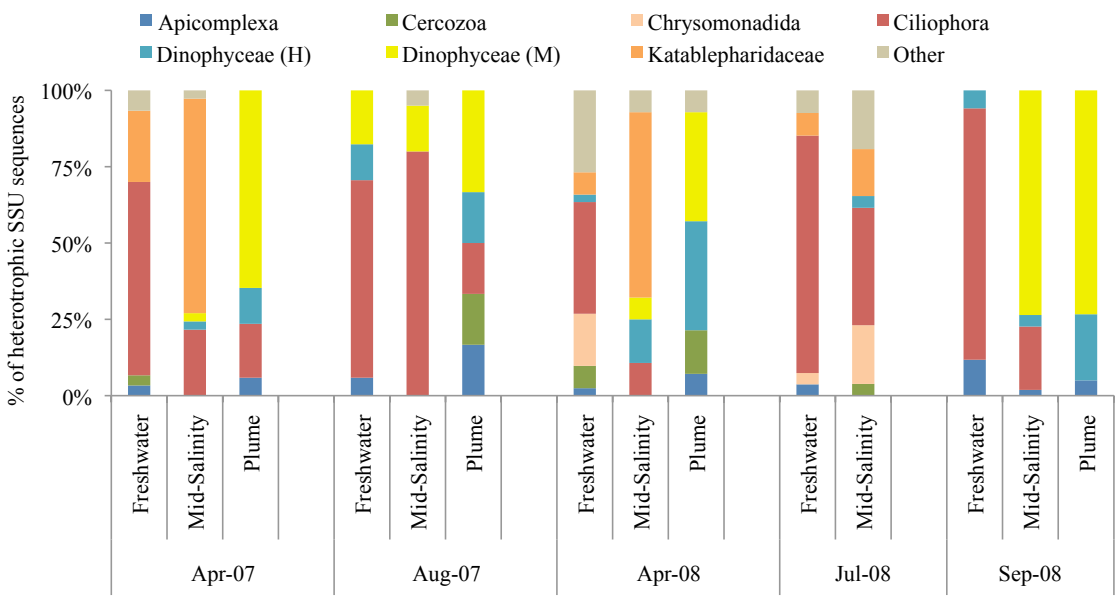


Figure S3.1. Percent composition of heterotrophic protists at the class level based on analysis of 500 bp SSU sequence data for water collected in the Columbia River estuary and its plume in April and August 2007, and in April, July and September 2008. Freshwater = salinity of 0; Mid-Salinity = salinity of 15; Plume = salinity of 28-31. "H" refers to putative heterotrophic dinoflagellates, while "M" indicates putative mixotrophic dinoflagellates. "Other" category designates sequences associated with the following protist taxa: Bicosoecida, Centroheliozoa, Choanoflagellatea, Ichthyosporea, Labyrinthulida, Stramenopile MAST-12 group, Oomycetes, Pirsonia, and Telonemida. The dominance of Katablepharid sequences in April 2007 and 2008 mid-salinity waters denotes the genus *Katablepharis*.

Table S4.1. Columbia River coastal margin samples used in this study. Locations of sample sites are detailed if Figs. 1 and 2. 'X' refers to samples used for each analysis.

Location	Site ID	Date	Sampling Depth (m)	Salinity	ORF1/ ORF2 PCR	Dub2F/ Dub2R PCR	Diplo. USE PCR	DubC1F/ DubC1R qPCR	Dub3F/ Dub3R qPCR	Cerc4F/ Cerc4R qPCR	Cerc5F/ Cerc5R qPCR	Diplo. USE qPCR
Columbia River estuary	262	Apr-08	17.0	14.4	X							
	340	Jun-08	11.0	15.8	X							
	B124	Jul-08	2.0	14.7	X							
	BB	May-12	1.0	6.1	X			X	X	X	X	
	BB	Aug-12	1.0	4.1		X						
	BB	Sep-12	0.6	8.7				X	X	X	X	
	YB	May-12	0.8	1.7	X			X	X	X	X	
	YB	Sep-11	1.0	1.3	X	X						
	YB	Sep-12	2.7	8.9				X	X	X	X	
	SC08	May-12	1.1	2.8				X	X	X	X	
	SC08	Sep-12	0.8	7.0				X	X	X	X	
	SC02	May-12	1.0	9.9				X	X	X	X	
	SC02	Sep-12	1.4	1.2				X	X	X	X	
	ML03	May-12	1.2	3.9				X	X	X	X	
	ML03	Sep-12	1.4	15.1				X	X	X	X	
	SAT03	May-12	1.0	2.9				X	X	X	X	
	SAT03	Sep-12	1.8	15.5				X	X	X	X	
	SAT01	May-12	0.9	1.0								
	SAT01	Sep-12	1.7	6.9				X	X	X	X	
NE Pacific Coast	CR7	Apr-08	2.0	30.0	X	X						
	CR7	Jul-08	2.0	29.4	X							
	CR7	Sep-08	2.0	31.2		X						
	CR7	Sep-12	2.0	31.9				X	X	X	X	
	CR7	Apr-13	7.5	29.0		X						
	CR15	Jul-12	9.0	32.0			X	X	X	X	X	
	CR15	Sep-13	2.0	31.4				X	X			
	CR30	Sep-13	2.0	30.5				X	X			
	CR40	May-13	1.3	30.0				X	X			

Table S4.1 continued

Location	Site ID	Date	Depth	Salinity	ORF1/ ORF2 PCR	Dub2F/ Dub2R PCR	Diplo. USE PCR	DubC1F/ DubC1R qPCR	Dub3F/ Dub3R qPCR	Cerc4F/ Cerc4R qPCR	Cerc5F/ Cerc5R qPCR	Diplo. USE qPCR
NE Pacific Coast	CR40	Sep-13	2.0	30.8			X	X	X			X
	CR40	Sep-13	100.0	33.2			X					
	CR40	Sep-13	150.0	33.5			X					
	CR40	Sep-13	300.0	33.9			X					
	CR40	Apr-14	1.0	32.2			X					X
	CR40	Apr-14	20.0	32.5			X					X
	CR40	Apr-14	100.0	33.1			X					X
	CR40	Apr-14	150.0	33.8			X					X
	CR40	Apr-14	200.0	34.0			X					X
	CR40	Apr-14	400.0	34.2			X					X
	CR40	Apr-14	800.0	34.4			X					X
	NH3	Jul-12	42.0	33.7				X	X	X	X	
	NH25	Apr-11	2.0	32.5				X	X	X	X	
	NH55	Sep-13	1000.0	34.5			X					
	GH41	Sep-13	250	32.8			X					
	SH50	Nov-07	2.0	33.1				X	X			
	SH70	Jul-12	2.0	33.9				X	X	X	X	
	SH100	Apr-07	2.0	31.5				X	X			
	LP6	Nov-07	50.0	31.9			X	X	X	X	X	
	LP6	Sep-09	2.0	31.2			X					
	LP6	Sep-09	10.0	31.4			X					
	LP6	Sep-09	50.0	31.8			X					
	LP17	Sep-09	2.0	31.9			X					
	LP17	Sep-09	10.0	32.0			X					
	LP17	Sep-09	125.0	33.8			X					
	LP17	Jul-12	100.0	33.9				X	X			
	LP32	Sep-09	2.0	32.1			X					
	LP32	Sep-09	12.0	32.2			X					
	LP32	Sep-09	750.0	32.7			X					

Table S4.1 continued

Location	Site ID	Date	Depth	Salinity	ORF1/ ORF2 PCR	Dub2F/ Dub2R PCR	Diplo. USE PCR	DubC1F/ DubC1R qPCR	Dub3F/ Dub3R qPCR	Cerc4F/ Cerc4R qPCR	Cerc5F/ Cerc5R qPCR	Diplo. USE qPCR
NE Pacific Coast	LP32	Jul-12	2.0	31.4				X	X			
	LP52	Nov-07	2.0	33.1				X	X	X	X	
	LP52	Sep-09	2.0	32.0			X			X		
	LP52	Sep-09	19.0	33.8			X			X		
	LP52	Sep-09	1545	34.5								

S6. Linking Columbia River coastal margin protist assemblages to environmental variables

S6.1. INTRODUCTION

In addition to the characterization of heterotrophic protist assemblages presented in **Chapter Three**, autotrophic protist assemblages were also evaluated through analysis of small subuint (SSU) rRNA gene clone libraries for water collected at ~0 and ~15 PSU in the Columbia River estuary, and at ~28-31 PSU in the plume in April and August 2007 as well as April, July, and September 2008. A suite of environmental parameters (biotic and abiotic) were also collected and correlated to our assemblages through linear regression multivariate analysis (BIO-ENV; (Clarke & Ainsworth 1993)). Through analysis of the clone libraries and multivariate analyses, we detected inter-annual variation of protist communities in spring and summer across the river-to-ocean gradient and determined which environmental factors play essential roles in shaping the protist assemblages.

S6.2. METHODS

S6.2.1. Nutrient and pigment analyses

For each sample, nitrate+nitrite (hereafter referred to as nitrate), ammonium, dissolved silica, and phosphate were determined using standard continuous segmented flow autoanalyzer techniques as described in (Gordon et al. 1994). Photosynthetic

pigments (chlorophyll *a*, phaeophytin) were also determined for each sample. Briefly, 100-300 mL of water was filtered through 25 mm GF/F (Whatman) filters. Filters were flash frozen in liquid nitrogen. Samples were cold-extracted (-20°C) in polypropylene centrifuge tubes using a fixed volume of 90% acetone in water (v/v). April 2007 samples were analyzed with the fluorometric method (Holm-Hansen et al. 1965) using a Trilogy Laboratory Fluorometer (Turner Designs) calibrated with pure chlorophyll *a* from *Anacystis nidulans* (Sigma). All other samples were analyzed using high performance liquid chromatography (Wright 1991). Chromatographic separations were made using a C8 reverse-phase column and diode array detection at 436 nm. Pigments were quantified by comparison of integrated peak area for chlorophyll *a* with response factors for authentic standards.

S6.2.2. Statistical analyses

To link environmental variables to the observed protist assemblages, the BIO-ENV procedure (Clarke & Ainsworth 1993) was performed using Primer v6 software (PRIMER-E Ltd.). BIO-ENV ranks each individual environmental variable as well as sets of variables (up to 5 in our study) using a Spearman rank coefficient (ρ) to calculate the level of association between similarity matrices of protist assemblage (Bray-Curtis similarity calculated at the genus level) and environmental data (Euclidean distance). With this method a coefficient of one denotes that all assemblage variability is explained. BIO-ENV analysis was performed for matrices of the full dataset, as well as matrices calculated from the same salinity group (freshwater, mid-salinity, or plume). Environmental data used in BIO-ENV analysis included salinity, temperature, phosphate,

nitrate, silicic acid, ammonium, chlorophyll *a*, phaeophytin, river discharge (source: US Army Corps of Engineers, data from flow gauges at Bonnevillle Dam), and Coastal Upwelling Index at 45°N (source: National Oceanic and Atmospheric Administration Pacific Fisheries Environental Laboratory), calculated from Ekman's theory of mass transport to determine strength of wind forcing on the ocean (Schwing et al. 2006). These variables were plotted pairwise against each other with a draftsman plot, and those that were markedly skewed were log transformed. This transformation removed skewness and approximates a multivariate-normal distribution, increasing the efficiency of the Euclidean distance matrix (Clarke & Ainsworth 1993). Pairwise correlation coefficients were calculated for the full dataset as well as each salinity group to ensure environmental variables did not correlate with each other. Silicic acid was strongly correlated with salinity ($\rho>90$) and was therefore removed from analysis of the full dataset to allow for more interpretable results. However, silicic acid was included, and salinity removed, from analysis of each salinity group, as values of the latter variable were nearly identical within each group.

S6.3. RESULTS

S6.3.1. Physical characteristics during 2007 and 2008

Peak river flow associated with the spring freshet occurred earlier in the year during 2007 compared to 2008 (Fig. S6.1A). The maximum river discharge was approximately 20% lower in 2007 compared to 2008. In 2007 the discharge ranged from ca. 3500 m^3s^{-1} (summer) to ca. 4000 m^3s^{-1} (spring). However, in 2008 the late summer discharge rate on the day of sampling (ca. 2000 m^3 s^{-1} in September) was 3-fold lower

than in early summer (July) and 3.5-fold lower than in spring (April). Coincident with a decrease in river flow, the salinity intrusion length, the maximum distance that saltwater (defined as >1PSU) reaches upstream, increased from 27 to 36 km between spring and summer 2007 and 31 to 39 km between spring and summer 2008 (Fig. 3.1) (data from http://www.stccmop.org/datamart/virtualcolumbiariver and based on variation of salinity simulation model presented in (Baptista et al. 2005)). The salinity intrusion length is related to the areal extent of saline waters present in the estuary and – while it did not influence the location of sampling sites in this study – it yields an estimate of the spatial extent over which our observations are applicable. The lack of a relationship between salinity intrusion length and our sample sites reflects the variety of factors that result in the salinity distributions in the estuary. For example, in August 2007 and September 2008 the mid-salinity (15 PSU) samples were not located further upstream than in April or July (Fig. 3.1); however, this was caused partly by variations in tidal input during the time of sampling. Whereas in April the mid-salinity samples were taken after a large flood tide in both 2007 and 2008 (when saltwater reached furthest upstream), the mid-salinity samples from August 2007 and September 2008 were obtained at the end of ebb tides, when saltwater recedes towards the river mouth.

In addition to river discharge, upwelling is an important variable to consider, as the direction, thickness, and volume of the Columbia River plume are influenced by upwelling and downwelling favorable winds (Burla et al. 2010). The balance between upwelling and downwelling in the Columbia River plume can affect nutrient concentrations and in turn the relative abundance of diatoms versus dinoflagellates (Frame & Lessard 2009). During our study, only one sample (July 2008) was taken

during strong upwelling (Fig. S6.1B). During the other sampling periods the Coastal Upwelling Index (at 45°N) was either negative or near zero (indicating downwelling favorable or neutral winds) when the plume sample was taken.

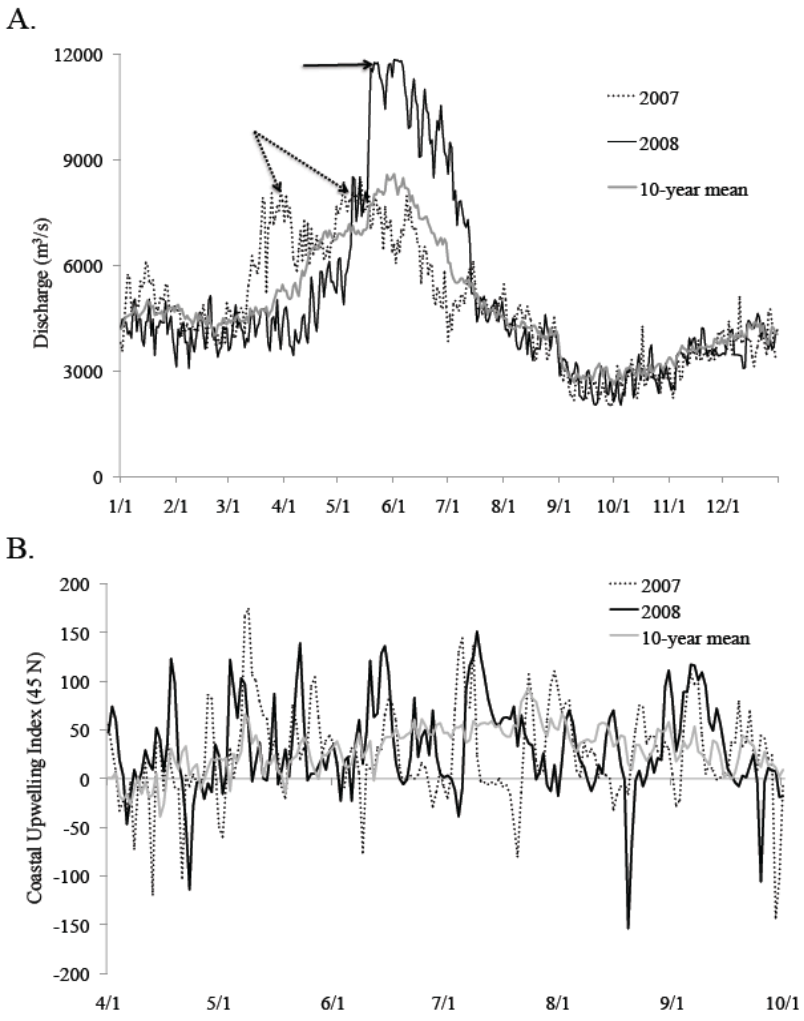


Figure S6.1. A. Annual Columbia River discharge (m^3/s), measured at the outflow of Bonneville Dam for 2007 and 2008 (daily mean), as well as 10-year daily mean from 1999-2008. Source: US Army Corps of Engineers. Dashed arrows = periods of maximum river discharge for 2007; Solid arrow= period of maximum river discharge for 2008. Note that in 2008 a greater discharge volume occurred later in the year compared to 2007.

B. Coastal upwelling index measured at 45 °N for 2007 and 2008 (daily mean), as well as 10-year daily mean from 1999-2008. The July 2008 plume sample was collected during strong upwelling favorable winds, while all other samples were collected were collected during downwelling or neutral winds. Source: NOAA Pacific Fisheries Environmental Laboratory.

S6.3.2. Nutrients and pigments along the salinity gradient

Dissolved silicic acid concentrations were strongly correlated with salinity, ranging from 19 to 180 μmol L^{-1} in the freshwater samples, 64 to 110 μmol L^{-1} in the mid-salinity samples, and 10 to 50 μmol L^{-1} in the plume samples (Table S6.1). Phosphate concentrations in the estuarine samples ranged from 0.1 μmol L^{-1} in July 2008 to 0.5 μmol L^{-1} in September 2008, while compared to our other samples, the plume sample from July 2008 had an elevated phosphate concentration of 1.9 μmol L^{-1} that was related to the upwelling event (Fig. S6.1B). Plume nitrate concentrations were also highest during July 2008 (23.8 μmol L^{-1}) but were in the estuarine samples typically higher in April (23.7 and 10.5 μmol L^{-1} in April and August 2007 freshwater samples, respectively) (Table S6.1). Ammonium showed a similar variation in our estuarine samples, with lowest values occurring in July 2008 and highest concentrations measured in April 2008 (Table S6.1).

For freshwater samples, chlorophyll *a* concentrations were strongest during July 2008, at 10.7 μg L^{-1} (Table S6.1). The mid-salinity sample from July had a similar concentration (9.2 μg L^{-1}, however September 2008 had the highest chlorophyll concentrations at 14.9 μg L^{-1}. In the plume samples, August 2007 had by far the highest chlorophyll concentrations at 22.5 μg L^{-1}.

S6.3.3. Autotrophic protist assemblages

Analysis of SSU rRNA gene (hereafter referred to as SSU) sequences was performed to examine seasonal and inter-annual variations of protist assemblages in the Columbia River estuary and its plume. The Columbia River estuary has previously been classified as a diatom-dominated system (Hobson 1966, Haertel et al. 1969, Frey et al. 1984, Small et al. 1990). These observations were based primarily on microscopic examination of protist morphology as well as pigment analysis using HPLC (Sullivan et al. 2001). According to SSU sequence analysis, Bacillariophyceae (diatoms) was the dominant autotrophic protist class in all of the freshwater samples collected (Fig. S6.2). Sequences resembling the cosmopolitan centric diatom *Stephanodiscus* were prevalent in each freshwater sample, and the pennate *Asterionella* and centric *Aulacoseira* (both freshwater genera) were found to be prevalent in April of 2007 and 2008. However, sequence analysis indicated that the dominant diatom genera were *Stephanodiscus* and the centric marine genus *Cyclotella* in August 2007. The halotolerant centric genus *Skeletonema* was also present in fairly high proportions in July and September 2008. Other autotrophic taxa such as cryptophytes, chrysophytes, and chlorophytes were also detected consistently throughout freshwater samples. While these other protist groups were generally present in lower proportions, there were some exceptions. The cryptophyte *Rhodomonas* was detected in August 2007 and chlorophyte sequences related to *Desmodesmus* and *Volvox* were found in nearly all the freshwater samples.

Similar to the freshwater assemblages, diatoms were the most abundant autotrophic taxa in the majority of the mid-salinity samples (Fig. S6.2). The only

exception is August 2007, where nearly all of the sequences recovered were related to *Mesodinium (Myrionecta rubra)*, a bloom forming mixotrophic ciliate. The diatom genera *Stephanodiscus* and *Asterionella* were present in both April 2007 and 2008, while the centric marine genus *Thalassiosira* was also evident. In August 2007, marine genera such as *Asterionellopsis*, *Thalassiosira* were present in higher proportions than freshwater genera. A large proportion of the halotolerant *Skeletonema* was uncovered in July 2008, along *Stephanodiscus* and the marine genus *Cyclostephanos*, while the September 2008 diatom SSU sequences consisted almost entirely of *Thalassiosira*. Similar to the mid-salinity assemblages, diatom sequences were the most prevalent in four of the five plume samples (Fig. S6.2). In all five samples, the diatom sequences recovered were mostly related to the centric diatom, *Thalassiosira*. *Asterionellopsis* and *Skeletonema* sequences were also detected in April 2007 and August 2007. Sequences resembling *Pseudo-nitzschia* were recovered in April 2008. Unfortunately, the SSU library from the July 2008 plume sample had to be discarded, as it was comprised of nearly all copepod sequences, with only 18 sequences related to protists. However, microscopic cell counts of the July sample were dominated by marine diatoms, again mostly *Thalassiosira* as well as members of the genus *Chaetoceros*. *Thalassiosira* and *Chaetoceros* were also present in high proportions in the other four plume samples as well. *Asterionellopsis* was also detected in April 2007, August 2007, and April 2008, but not in July or September 2008 plume samples. Dinoflagellates comprised a small proportion of SSU sequences in April 2007 and 2008 as well as August 2007, however they dominated the protist sequences in September 2008 (Fig. S6.2). Similar to the mid-salinity sample, these sequences were mostly related to *Gyrodinium* (a possible

autotroph, mixotroph, or heterotroph), while *Alexandrium*, *Pentapharsodinium*, and *Peridinium* were also detected in high proportions.

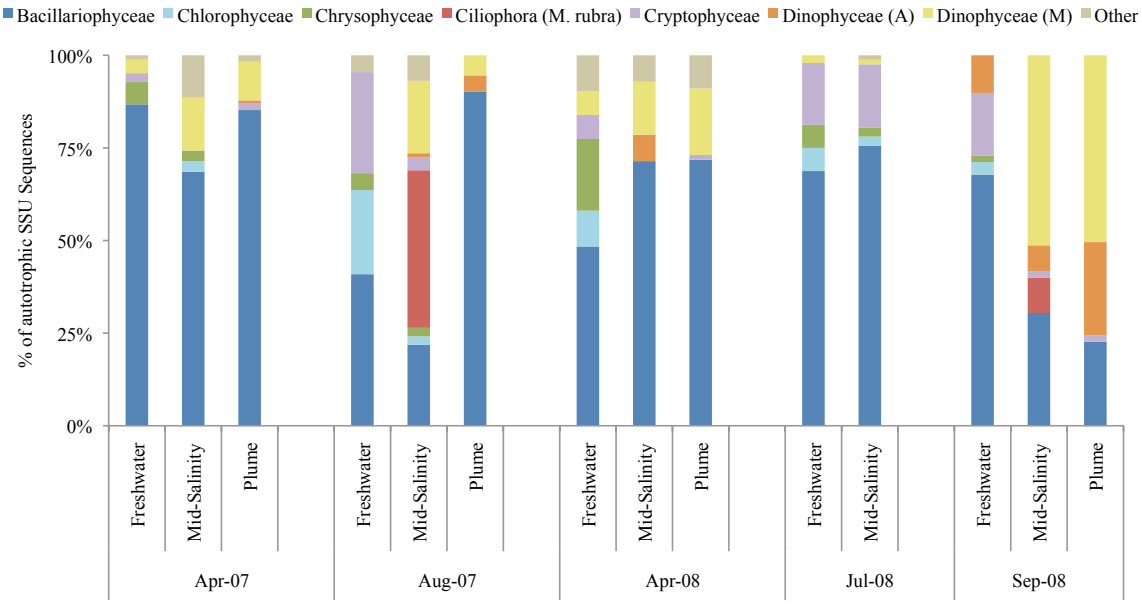


Figure S6.2. Percent composition of autotrophic protists at the class level, based on analysis of SSU sequence data for water collected in the Columbia River estuary and its plume in April and August 2007 and April, July, and September 2008. F= Freshwater (0 PSU); M=Mid-Salinity (15 PSU); P= Plume (28-31 PSU). "Other" category refers to sequences associated with the protist classes Bolidophyceae, Dictyochophyceae, Haptophyceae, Prasinophyceae, and Synurophyceae.

		Freshwater					Mid-Salinity					Plume				
		Apr-07	Aug-07	Apr-08	Jul-08	Sep-08	Apr-07	Aug-07	Apr-08	Jul-08	Sep-08	Apr-07	Aug-07	Apr-08	Jul-08	Sep-08
Silicate	Dissolved silicate ($\mu mol\ L^{-1}$)	19.4	141.3	18.5	143.4	166.1	71.0	64.7	84.1	110.7	92.3	12.1	10.8	12.9	50.7	15.3
Phosphorus	Phosphate ($\mu mol\ L^{-1}$)	0.3	0.4	0.3	0.1	0.4	0.3	0.4	0.2	0.1	0.5	0.7	0.1	0.1	1.9	0.5
Nitrogen	Ammonium ($\mu mol\ L^{-1}$)	0.9	1.4	1.2	0.4	1.2	2.2	1.6	2.4	0.4	1.9	0.3	0.0	1.3	0.3	0.7
	Nitrate+ Nitrite ($\mu mol\ L^{-1}$)	23.7	10.5	25.7	0.3	9.7	15.1	6.9	14.1	0.1	10.7	5.1	0.1	1.3	23.8	2.8
Biotic	Chlorophyll *a* ($\mu g\ L^{-1}$)	4.2	3.4	10.7	3.7	2.6	5.6	7.5	9.2	6.3	14.9	4.6	22.5	10.0	0.7	11.3
	Phaeophytin ($\mu g\ L^{-1}$	0.8	4.4	0.3	2.8	3.0	1.8	2.2	2.6	1.0	5.3	6.0	0.6	0.6	1.2	1.9

Table S6.1. Biogeochemical and biological characteristics of water samples used in this study. Freshwater=0 (PSU); Mid-Salinity=14-15 PSU; Plume=28-31 PSU.

S6.4. CONCLUSIONS

S6.4.1. Linking assemblages to environmental variables

Nutrient availability plays a key role in determining protist assemblage composition in the Columbia River system (Haertel et al. 1969, Lara-Lara et al. 1990, Kudela & Peterson 2009). For example, a transition from diatoms towards motile taxa (including flagellates and dinoflagellates) has been previously observed in the Columbia River plume (Frame & Lessard 2009), the Pacific Northwest coast (Sherr et al. 2005), and other upwelling areas (Smith et al. 1983), as nutrient depletion that often follows upwelling relaxation leads to a shift towards motile organisms. Nutrient depletion can also promote a shift from assemblages of autotrophs towards mixotrophs and heterotrophs (Nygaard & Tobiesen 1993). Previous studies along the Pacific Northwest coast have found lower nutrient concentrations during non-upwelling periods in the summer/autumn than during the spring (Anderson 1964, Bruland et al. 2008). Indeed, nutrient analyses conducted during this study do show a decrease in nutrient concentration from spring to non-upwelling periods in the summer months (Table S6.1). Only one of our plume samples (July 2008) was collected during a strong upwelling period. This sample had elevated nutrient levels compared to other plume samples, and the protist assemblage was dominated by marine diatoms (as determined by cell counts because SSU sequences were contaminated with copepod sequences). While this data set is quite small, similar analyses of nutrients were conducted for more than 500 samples collected throughout the estuary and coastal zone which confirm the trends observed in this study (data available online at http://www.stccmop.org/datamart/campaigndata).

Nitrate concentrations in the freshwater samples dropped dramatically between spring and summer, coincident with a shift from diatoms to motile autotrophic (cryptophytes), heterotrophic or mixotrophic (ciliates) protists. Multivariate analyses were conducted to help determine which environmental parameters best explained the variability observed in our assemblage data. BIO-ENV multivariate analysis supports nitrate as a key variable in determining seasonal change of protist assemblages, as most of the freshwater and mid-salinity variability at the genus level could be explained by nitrate concentrations (rank coefficient of ρ=0.891 for freshwater assemblages; ρ=0.685 for mid-salinity assemblages). A combination of nitrate, temperature, ammonium, and river discharge together increased the rank coefficient of the mid-salinity assemblage to ρ=0.867.

Similar analysis to our plume samples suggested that silicic acid concentrations can explain most of the variability of protist assemblages at the genus level, with a rank coefficient of ρ=0.733. Silicic acid is strongly correlated with salinity and is derived from the Columbia River, and therefore is tied to river discharge (Stefánsson & Richards 1963, Bruland et al. 2008). Although the supply of silicic acid from the Columbia River to the coast decreases to a minimum in late summer, nitrate has previously been described as the dominant limiting nutrient in plume-influenced waters (Lohan & Bruland 2006, Kudela & Peterson 2009). In our study, nitrate concentrations did decrease dramatically from April to August 2007 and July to September 2008 (Table S6.1).

While nutrient availability can explain much of the observed seasonal variation in protist assemblages, salinity intrusion and river discharge are key driving forces in determining protist assemblages in the estuary. The freshwater (0 PSU) assemblages were

dominated by freshwater protists while marine protists were more prevalent in the plume (28-31 PSU). In the mid-salinity (15 PSU) samples, the amount of freshwater input influenced the balance between marine and freshwater genera. The spring mid-salinity samples were composed primarily of freshwater diatoms and *Katablepharis*, a heterotrophic flagellate that has been found in marine, estuarine, and freshwater environments (Ploug et al. 2002, Domaizon et al. 2003, Šlapeta et al. 2006).

As river flow decreases late into the summer and early autumn, saline water is able to intrude further upstream (Sherwood et al. 1990). This is easily visualized in the salinity intrusion length estimates, which indicated longer salinity intrusion length distances in August and September than in April or July (Fig. 3.1). In turn, marine protists were transported further into the estuary. An indicator of this is the marine diatom *Thalassiosira*. It was rarely detected in the April mid-salinity samples while it comprised 25% of sequences in the mid-salinity sample from September when salinity intrusion length was greater. Similarly, dinoflagellates such as *Gymnodinium* spp. that were found in the plume samples were also detected in the mid-salinity samples in September. Our BIO-ENV multivariate analysis of all samples confirmed that salinity was the factor that best explained the variability (ρ=0.46), while a combination of salinity and chlorophyll *a* only slightly increased that coefficient (ρ=0.476).

Appendix References

Anderson GC (1964) The Seasonal and Geographic Distribution of Primary Productivity Off the Washington and Oregon Coasts. Limnology and Oceanography 9:284-302

Baptista AM, Zhang Y, Chawla A, Zulauf M, Seaton C, Myers Iii EP, Kindle J, Wilkin M, Burla M, Turner PJ (2005) A cross-scale model for 3D baroclinic circulation in estuary, plume, and shelf systems: II. Application to the Columbia River. Recent Developments in Physical Oceanographic Modelling: Part II 25:935-972

Bruland KW, Lohan MC, Aguilar-Islas A, Smith GJ, Sohst B, Baptista A (2008) Factors influencing the chemistry of the near-field Columbia River plume: Nitrate, silicic acid, dissolved Fe, and dissolved Mn. J Geophys Res 113:C00B02

Burla M, Baptista AM, Zhang Y, Frolov S (2010) Seasonal and interannual variability of the Columbia River plume: A perspective enabled by multiyear simulation databases. J Geophys Res 115:C00B16

Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series 92:205-219

Domaizon I, Viboud S, Fontvieille D (2003) Taxon-specific and seasonal variations in flagellates grazing on heterotrophic bacteria in the oligotrophic Lake Annecy-importance of mixotrophy. Federation of European Microbiological Societies Microbiology Ecology 46:317-329

Frame ER, Lessard EJ (2009) Does the Columbia River plume influence phytoplankton community structure along the Washington and Oregon coasts? JGeophysRes 114:C00B09

Frey BE, Lara-Lara R, Small LF (1984) Water column primary production in the Columbia River Estuary. Columbia River Estuary Data Development Program:133 pp

Gordon LI, Jennings JC, Ross AA, Krest JM (1994) A suggested protocol for continuous flow analysis of seawater nutrients (phosphate, nitrate, nitrite, and silicic acid) in the WOCE Hydrographic Program and the Joint Global Ocean Fluxes Study. WHP Office Report 91

Haertel L, Osterberg C, Curl H, Jr., Park PK (1969) Nutrient and Plankton Ecology of the Columbia River Estuary. Ecology 50:962-978

Hobson LA (1966) Some Influences of the Columbia River Effluent on Marine Phytoplankton During January 1961. Limnology and Oceanography 11:223-234

Kudela RM, Peterson TD (2009) Influence of a buoyant river plume on phytoplankton nutrient dynamics: What controls standing stocks and productivity? JGeophysRes 114:C00B11

Lara-Lara JR, Frey BE, Small LF (1990) Primary production in the Columbia river estuary. I. spatial and temporal variability of properties. Pacific Science 44:17-37

Lohan MC, Bruland KW (2006) Importance of vertical mixing for additional sources of nitrate and iron to surface waters of the Columbia River plume: Implications for biology. Marine Chemistry 98:260-273

Nygaard K, Tobiesen A (1993) Bacterivory in Algae: A Survival Strategy During Nutrient Limitation. Limnology and Oceanography 38:273-279

Ploug H, Zimmermann-Timm H, Schweitzer B (2002) Microbial communities and respiration on aggregates in the Elbe Estuary, Germany. Aquatic Microbial Ecology 27:241-248

Schwing FB, Bond NA, Bograd SJ, Mitchell T, Alexander MA, Mantua N (2006) Delayed coastal upwelling along the US West Coast in 2005: A historical perspective. Geophys Res Lett 33:L22S01

Sherr EB, Sherr BF, Wheeler PA (2005) Distribution of coccoid cyanobacteria and small eukaryotic phytoplankton in the upwelling ecosystem off the Oregon coast during 2001 and 2002. Deap-Sea Research II Topical Studies in Oceanography 52:317-330

Sherwood CR, Jay DA, Bradford Harvey R, Hamilton P, Simenstad CA (1990) Historical changes in the Columbia River Estuary. Progress in Oceanography 25:299-352

Šlapeta J, López-García P, Moreira D (2006) Present status of the molecular ecology of kathablepharids. Protist 157:7-11

Small LF, McIntire CD, MacDonald KB, Lara-Lara JR, Frey BE, Amspoker MC, Winfield T (1990) Primary production, plant and detrital biomass, and particle transport in the Columbia River Estuary. Progress in Oceanography 25:175-210

Smith WO, Heburn GW, Barber RT, O'Brien JJ (1983) Regulation of phytoplankton communities by physical processes in upwelling ecosystems. Journal of Marine Research 41:539-556

Stefánsson U, Richards FA (1963) Processes Contributing to the Nutrient Distributions Off the Columbia River and Strait of Juan de Fuca. Limnology and Oceanography 8:394-410

Sullivan BE, Prahl FG, Small LF, Covert PA (2001) Seasonality of phytoplankton production in the Columbia River: A natural or anthropogenic pattern? Geochimica et Cosmochimica Acta 65:1125-1139

Wright SW (1991) Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. Marine Ecology Progress Series 77:183-196