Fall 12-4-2015

Expansion of the Chlorovirus Genus by Studies on Virus Natural History and Chlorella Host Metabolism

Cristian F. Quispe
University of Nebraska-Lincoln, quispecristian@gmail.com

EXPANSION OF THE *CHLOROVIRUS* GENUS BY STUDIES ON VIRUS NATURAL HISTORY AND *CHLORELLA* HOST METABOLISM

by

Cristian F. Quispe

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Biological Sciences
(Genetics, Cell, and Molecular Biology)

Under the Supervision of Professor James L. Van Etten

Lincoln, Nebraska
December, 2015

EXPANSION OF THE *CHLOROVIRUS* GENUS BY STUDIES ON VIRUS NATURAL HISTORY AND *CHLORELLA* HOST METABOLISM

Cristian F. Quispe, Ph.D.
University of Nebraska, 2015
Adviser: James L. Van Etten

Inland waters cover about 2.5 percent of our planet and harbor huge numbers of known and unknown microorganisms including viruses. Viruses likely play dynamic, albeit largely undocumented roles in regulating microbial communities and in recycling nutrients in the ecosystem. Phycodnaviruses are a genetically diverse, yet morphologically similar, group of large dsDNA-containing viruses (160- to 560-kb) that inhabit aquatic environments. Members of the genus *Chlorovirus* are common in freshwater. They replicate in eukaryotic, single-celled, chlorella-like green algae that normally exist as endosymbionts of protists in nature. Very little is known about the natural history of the chloroviruses and how they achieve high-titer and long-term persistence in nature. To study their natural history, we examined chloroviruses over a three-year period to determine their abundance, prevalence, and genetic diversity in a small lake in Nebraska (Chapter II). These studies indicated that the amount of infectious virus particles was seasonal and both host- and site-dependent. *Chlorovirus* populations persisted year-round, suggesting that the viruses are either very stable or that viral production occurs in an unknown natural host(s). During this study, a new viral group was discovered and characterized, expanding the *Chlorovirus* genus

(Chapter III). This group, designated as *Only Syngen virus*es (OSy), replicates in *Chlorella variabilis* (Syngen 2-3) cells. Furthermore, OSy viruses also have non-permissive features in two phylogenetically related *C. variabilis* sub-species and constitute the first report of a post-infection host mechanism that results in resistance against infection. In Chapter IV, five symbiotic-virus suceptible and four free-living *Chlorella* species were evaluated for their capabilities to assimilate nutrients. Hierarchical clustering reveals a clear distinction of both groups based on their assimilation of galactose, nitrate, asparagine, proline, and serine. Additionally, genomic and differential expression analyses of symbiotic algae confirm an abundance of amino acid transporter genes, some of which are constitutively expressed when the symbiotic algae either grow axenically or as an endosymbiont within their host. Such similarities indicate a parallel coevolution of shared metabolic pathways across multiple independent symbiotic events and suggest that physiological changes driving the *Chlorella* symbiotic phenotype also contribute to their natural fitness.

To the mysterious and invisible forces that have made everything work in the past, the present and the future…

ACKNOWLEDGMENTS

I would like to acknowledge the wonderful people with whom I have interacted through my life; because, directly or indirectly, they have helped me in shaping my personality. I do not have enough time and space to name everyone, but I want to include my family, friends and professors that have shared with me the much-needed support, friendship, advice and encouragement throughout these years. You have changed my life for the better, and all of my achievements are shared with you. Overall, these six years at the University of Nebraska-Lincoln (UNL) have been a great and invaluable experience, although tough at moments, I could never have done it without all of you.

I would like to express my utmost appreciation and thankfulness to my adviser and mentor, Dr. James L. Van Etten. His guidance and patience during these years have been invaluable. He has allowed me to pursue my own scientific interests while imparting his advice and wisdom. He has contributed to my professional and personal growth, for which I will always be indebted to him. Dr. Van Etten, thank you for letting me be a part of the fascinating world of giant viruses.

I am further grateful to all the professors in the Nebraska Center for Virology, especially Drs. Clinton Jones, Matthew Wiebe, Asit Pattnaikt and Qingsheng Li, for imparting their knowledge, which has helped improved my experiences at UNL. I would also like to thank the members of my supervisory committee: Drs. Kenneth Nickerson, Amit Mitra, Zhang Lwen, and Wayne Riekhof. Thank you for your comments and collaborative effort to improve the work presented in this dissertation. Additionally, I would like to thank Dr. James Steadman at the Department of Plant Pathology for connecting me with the opportunity to begin my journey at UNL.

A special thanks to all current and past members and students in the Van Etten Lab; it has been a great joy to share these years with you. I was very lucky to work with such a unique and supportive group of people. A special mention

goes to Anya Seng, Olivia Sonderman and Michelle McQuinn for their constant support and understanding, especially in difficult moments during my Master and Ph.D. programs. Also, I extend my highest gratitude to Dr. Les Lane for believing in my “new prospective” to study *Chlorovirus*. Additionally, thanks to James Gurnon, who taught me all necessary protocols and tips to work with the viral-host system.

In a wider context, none of this work would have been possible without the support of my family. All your love, care and wisdom significantly contributed to the final completion of this work. I am very grateful to my mother, Ligia, who raised me to pursue dreams, value knowledge and honor education. I also want to offer my appreciation to all of my friends in the US, who have made my experiences here so unique and unforgettable. Additional appreciation goes to my friends in Ecuador and around the world, who are the best friends anyone could ask for; thanks for staying close to me all these years.

This work was supported by the UNL School of Biological Sciences Teaching Assistantships including Special Funds, the Nebraska Center for Virology, the Department of Plant Pathology, the undergraduates funded by the UCARE and ARD scholarships, the NSF-EPSCoR, Stanley Medical Research Institute, and the COBRE grants.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES

Literature Review

Three-year Survey of Abundance, Prevalence and Genetic Diversity of Chlorovirus Populations in a Small Urban Lake

Characterization of a New Chlorovirus Type with Permissive and Non-Permissive Features on Phylogenetically Related Algae Strains

Comparative Genomics, Transcriptomics and Metabolism Distinguish Symbiotic from Free-living Chlorella

TABLES

Literature Review

Three-year Survey of Abundance, Prevalence and Genetic Diversity of Chlorovirus Populations in a Small Urban Lake

Characterization of a New Chlorovirus Type with Permissive and Non-Permissive Features on Phylogenetically Related Alga

Comparative Genomics, Transcriptomics and Metabolism Distinguish Symbiotic from Free-living Chlorella

CHAPTER I

INTRODUCTION

Literature Review

Viruses: cosmopolitan, abundant and important entities in nature

Viruses are ubiquitous members of the biosphere as they are found in essentially every ecosystem on the planet (Short, 2012). For example, analyses of aquatic environmental samples indicate that high concentrations of viruses (10^5 to 10^9 particles/ml) that infect microorganisms, primarily bacteria, are present in marine and inland waters (e.g., Lim et al., 2013; Rodriguez-Brito et al., 2010; Short, 2012; Yau et al., 2011). The virus number typically exceeds that of cellular organisms by at least an order of magnitude; thus, the number of different viruses within a community is huge. Their functions of predation and gene transfer make viruses key drivers in the dynamics of microbial ecosystems (Mokili et al., 2012; Suttle, 2007). Furthermore, viruses play important roles in the global biogeochemical cycling of carbon and nutrients (Bratbak et al., 1990; Rohwer & Thurber, 2009).

Studies from diverse biomes show that different environments possess distinct viral community structures. Even small and individual ecosystems, such as human feces, contain around 1,000 viral genotypes, whereas viral communities in seawater, although they are more diverse, contain around 5,000 genotypes (Breitbart et al., 2002; Breitbart et al., 2003). In both environments, the predominant viral type accounted for at least 1% of the total population. In contrast, samples collected from near-shore marine sediments were highly diverse, hosting between 10,000 and one million viral genotypes, with the most

abundant type representing less than 0.01% of the community. Thus, our biosphere is abundant with genetic information, which mainly is of viral origin and we have not yet been able to assign a role, function or evolutionary origin (Cortez et al., 2009).

"Virus" misleading us for decades

Virology officially started in 1898 (Beijerinck & Johnson, 1964), and for decades, viruses were defined by what they were not: very small entities (ultra filterable microbes) not visible under the microscope and not culturable in the absence of a host. Viruses were first considered as possible intermediate forms between mineral and true cellular life (Witzany, 2012). Additionally, at the end of the 20th century, the first viruses known to the public were those causing malignant phenotypes in clinical or agricultural organisms such as yellow fever in humans, mosaic disease in tobacco, and foot-and-mouth disease in livestock. Not surprisingly, virus is the Latin term for "poison, venom, or slimy fluid," which reflects its common strategy to survive (Witzany, 2012). Although at the time, the depiction of viruses as malicious killers was very appealing, new information about viruses emphasizes the important roles they possess, not only in the evolution of all life, but also as symbionts or co-evolutionary partners of host organisms (Witzany, 2012). Thus, with the advances in sequencing and technology of modern science, we are just now able to rediscover and understand what "viruses" really encompass.

Viruses are unlimited sources for diversity in their composition and shape

Viral genomes are the major source of genetic information in the biosphere, as they are known to evolve rapidly (Koonin et al., 2015). Despite being the most diverse biological entities, viruses are also the least characterized microbe in terms of their genetic, taxonomic, and functional diversity; for example, they often contain unique genes for which no homologues exist (Witzany, 2012).

There are countless unique genes in viruses with the potential to have unique and completely unexpected functions; in such a diverse pool, genes can produce structurally and functionally conserved proteins that have no apparent cellular ancestors. For example, novel proteins are able to generate unlimited structures, as evidenced by the various shapes seen in prokaryotic viruses: lemon-shaped viruses, tulip-shaped viruses, bottle-shaped viruses, stick-shaped viruses with hooks and pleomorphic-viruses along with others with globular, icosahedral and filamentous shapes (Pietila et al., 2014) (Figure 1).

Viruses also lack a universal gene (Rohwer & Edwards, 2002) such as the ribosomal RNA genes that are used to assess microbial diversity. Some genes, however, are conserved within particular taxonomic groups, as evidenced in the sequenced genomes of viral isolates; thus, their sequences are similar enough at the nucleotide level to facilitate taxonomic identifications (Mokili et al., 2012). Taxonomically, viruses are also classified by the nature of their nucleic acids following the Baltimore classification (Baltimore, 1971).

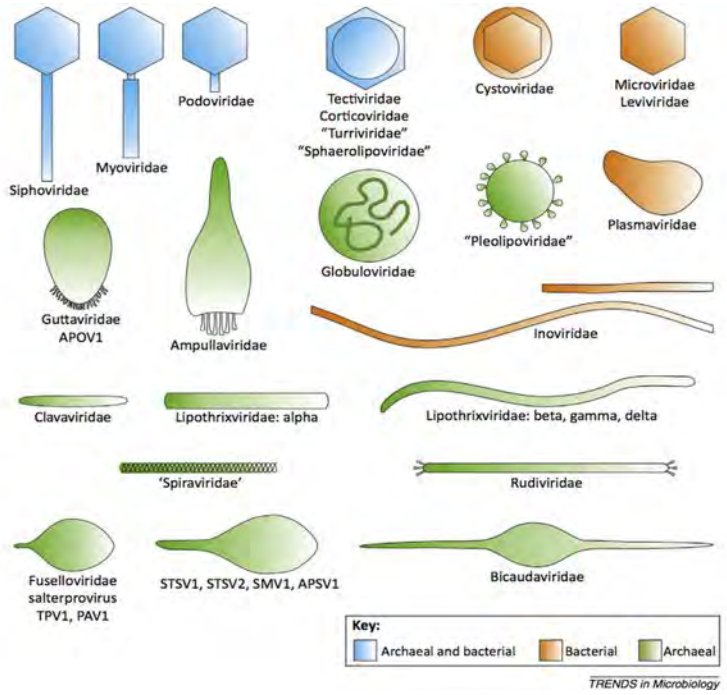


Figure 1. Diversity of virion morphotypes of prokaryotic viral structures generated by novel proteins. Virions are not drawn to scale. Abbreviations: APOV1, Aeropyrum pernix ovoid virus 1; APSV1, Aeropyrum pernix spindle-shaped virus 1; PAV1, Pyrococcus abyssi virus 1; SMV1, Sulfolobus monocaudavirus 1; STSV1, Sulfolobus tengchongensis spindle-shaped virus 1; STSV2, Sulfolobus tengchongensis spindle-shaped virus 2; TPV1, Thermococcus prieurii virus 1. Illustration taken from Pietila et al. 2014.

This classification system organizes viruses into one of seven groups depending on a combination of their genetic material (DNA or RNA), strandedness (single-stranded or double-stranded), sense (positive or negative), and replication approach.

Order "Megavirales" a breakthrough in the virology world

Generally, viral genomes are small compared to those of cellular organisms. In recent years however, the discovery of several groups of giant viruses has

dramatically changed this paradigm (Abergel et al., 2015; Koonin et al., 2015; Van Etten & Dunigan, 2012). Currently, viral genome sizes range from about 2 kilobases (kb) to more than 2.5 megabases (Mb). This expansion blurs the differences between cells and viruses in terms of genome size and complexity. The genomes of some giant viruses are even larger than numerous bacteria, archaea, and a few eukaryotic organisms (Koonin et al., 2015; Koonin et al., 2015a).

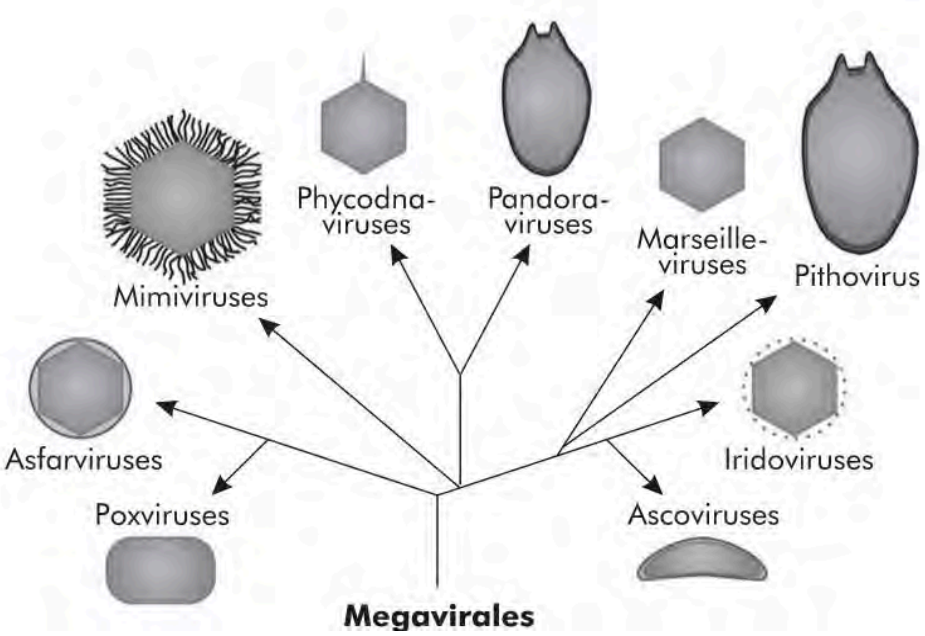


Figure 2. Nine viral families within the Megavirales order. Illustration taken from Koonin, Dolja and Krupovic 2015 with modifications.

The order “Megavirales” unites diverse families of giant viruses. The genome size of members of this group ranges from 100-kb to 2.5-Mb. Viruses in this order are believe to have a monophyletic lineage based on evolutionary genomic analysis, and they include nine large dsDNA virus families: Phycodnaviridae,

Poxviridae, Asfarviridae, Iridoviridae, Ascoviridae, Mimiviridae, Marseillevirus, Pandoravirus, and Pithovirus (Chen & Suttle, 1996; Koonin et al., 2015) (Figure 2).

Collectively these viruses are designated as nucleo-cytoplasmic large dsDNA viruses (NCLDV). They infect animals and diverse unicellular eukaryotes, and they replicate either exclusively in the cytoplasm of the host cells or possess both cytoplasmic and nuclear stages in their life cycle (Koonin et al., 2015; Van Etten et al., 2010). Intriguingly, NCLDVs have not yet been reported in any higher plants.

Phycodnaviruses are cosmopolitan in marine and freshwater environments

The Phycodnaviridae family represents icosahedral dsDNA viruses that infect marine and freshwater eukaryotic algae. Phycodnaviruses are key elements in aquatic ecosystems with important roles in the regulation of algal microbial habitats such as communities of red and brown algae (Coll et al., 2010; Kaiser, 2000; Short, 2012).

The family Phycodnaviridae is divided into six genera based on their host range: *Chlorovirus, Coccolithovirus, Phaeovirus, Prasinovirus, Prymnesiovirus, and Raphidovirus*. These divisions are supported not only by phylogenetic analysis, but also by sequence identity and structural conformation of their major capsid proteins. Their genomes range in size from 100- to 550-kb (Larsen et al., 2008; Van Etten et al., 2002).

Genus Chlorovirus

Members of the genus *Chlorovirus*, are ubiquitous in nature and have been isolated from inland waters collected throughout the world (Yamada, et al., 2006) including North and South America, Europe, Asia and Australia (Cho et al., 2002; Short & Short, 2009; Van Etten et al.,1985a; Van Etten, et al.,1985b; Zhang et al., 1988) (Figure 3). Chloroviruses infect certain unicellular, eukaryotic, ex-symbiotic chlorella-like green algae, often referred to as zoochlorellae (Meints, et al.,1984; Reisser et al.,1991).

Typically, chlorovirus titers in native waters fluctuate between 1-100 plaque-forming units (PFU) per ml; however, titers as high as 100,000 PFU/ml of indigenous water have been observed. Titers fluctuate with the seasons, with the highest titers occurring in the spring; however, the mechanism(s) in nature that allows long-term chlorovirus persistence and distribution in freshwater is still unknown (Cho, et al.,2002; Reisser, et al.,1988; Van Etten, 1995; Yamada, et al.,1991; Yamada, et al.,1993; Zhang, et al.,1988).

Common endosymbiosis between zoochlorellae and protist species

Green algae are some of the most abundant and ancient organisms on the planet. They have emerged as significant contributors in global energy and biogeochemical recycling (Grossman, 2005). Algae form a group of diverse photosynthetic organisms, ranging from multicellular to single-celled genera such as *Chlorella* (Proschold et al., 2011). Members in the genus *Chlorella* (Phylum

Chlorophyta) are small (2 to 10 μm in diameter), coccoid, nonmotile, unicellular green algae with a rigid cell wall and a single chloroplast, that exist as one of the most widely distributed algae in freshwater throughout the world. They reproduce by mitotic division in a simple developmental cell cycle. Vegetative cells increase in size and divide into two, four, eight, or more progeny depending on the species and environmental conditions. The progeny is then released by rupture or enzymatic digestion of the parental walls (Shihira & Krauss, 1965; Van Etten & Meints, 1999). Although most *Chlorella* species are free-living, *C. variabilis* is a species that exists as an endosymbiont of ciliated protozoan *P. bursaria* in nature (Table 1). They are often referred to as ex-symbiotic chlorellae or zoochlorellae (Proschold et al., 2011, Jolley & Smith, 1978, Siegel, 1960) (Figure 3).

Chlorella variabilis Algal strain	*P. bursaria* strain	*P. bursaria* collection site
SAG 211-6		USA
ATCC 50258/CCAP 211/84 (NC64A)		North Carolina, USA
ATCC 30562 (Syngen 2-3)		Ohio, USA
N-1-A		USA
NIES-2541 (OK1-ZK)	OK1	Aichi, Japan
So13-ZK	So13	Nagano, Japan
NIES-2540 (F36-ZK)	F36	(cross breed, Japan-Japan)
KM2-ZK/pbKM2	KM2	Shimane, Japan
Dd1-ZK	Dd1	Ibaraki, Japan
Bnd1-ZK	Bnd1	Hiroshima, Japan
HB2-2-1	HB2-2	Hiroshima, Japan
shiP-7-A4	shiP-7	Miyazaki, Japan
takaP-3-A2	takaP-3	Oita, Japan
(uncultured)†	Cs2	Shanghai, China
(uncultured)†	MRBG1	Melbourne, Australia

Table 1. List of names and collection site of the *Chlorella variabilis* algae strains isolated from their respective *P. bursaria* hosts. Table taken from Fujishima et al., 2010 with modifications.

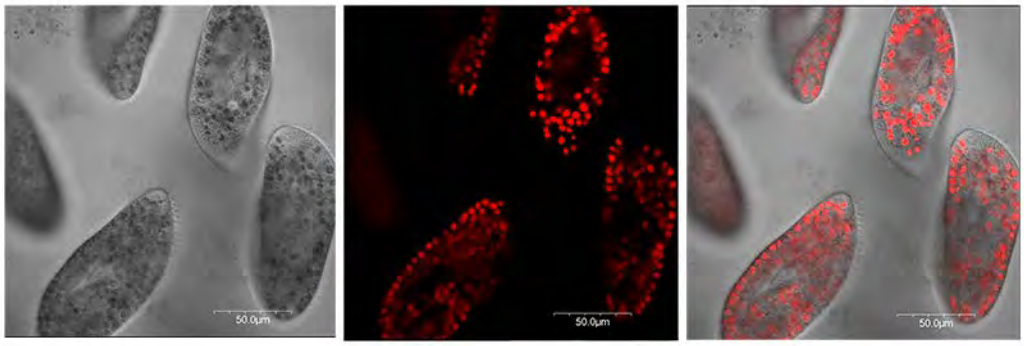


Figure 3. *Paramecium bursaria* shown in white light, ultraviolet light, and merged lights highlighting the red-fluorescing chlorophyll of the green algae housed within the symbiont.

C. variabilis is extremely variable in response and sensitive to small differences in culture conditions, thus its name *variabilis* (Shihira & Krauss, 1965).

In *P. bursaria*, hereditarily intracellular zoochlorellae inhabit the gastrodermal symbiosomes (perialgal vacuoles) of the protist and transfer an important amount of their photosynthetically fixed carbon (e.g. maltose, fructose) and amino acids to the non-photosynthetic partner (Cernichiari et al., 1969; Fujishima, 2010; Karakashian, 1975; Matzke, et al.,1990; Ziesenisz, et al.,1981) (Figure 3). Additionally, ex-symbiotic algae produce three times more oxygen than their free-living counterparts at low light intensity rates (Cronkite & van den Brink, 1981). Probably, high rate oxygen release in low light intensities is a special adaptive feature stemming from endosymbiotic interactions. Some individuals also differ in their uptake of nutrients. It has been suggested that ex-symbionts posses an efficient system to import and metabolize many organic nitrogen sources, while they can not utilize inorganic components, such as nitrate or nitrite, as their only

nitrogen source (Kamako et al., 2005; McAuley, 1987; Yellowlees, 2008).

All features stated above indicate that symbiotic algae strains, including *C. variabilis*, are not common free-living organisms in nature, more likely due to specific algal-nutrient requirements that are only provided in their symbiotic stages (Johnson, 2011). In fact, attempts to isolate *Chlorella* symbionts from nature, as free-living organisms, have generally been unsuccessful.

Despite the constraints of culture-based techniques, some attempts to isolate intact algae free of the protist host have been successful such as those involving *Paramecium bursaria* (Albers, et al.,1982a; Fujishima, 2010; Karakashian, 1975; Karakashian & Karakashian, 1965), *Acanthocystis turfacea* (Bubeck & Pfitzner, 2005) *and Hydra viridis* (Cernichiari, et al.,1969; Pardy, 1976; Park, et al.,1967). After disruption of the intracellular interaction symbiotic algal strains can be cultured axenically (Proschold et al., 2011). Isolates from different hosts are not identical; however, most belong to the genus *Chlorella*. Biodiversity and taxonomic analyses show that ex-symbiotic algae are morphologically indistinguishable based on structural comparisons. However, they can be distinguished by their ribosomal RNA (rRNA) gene sequences (Proschold et al., 2011) (Figure 4).

Axenic chlorella cultures allow scientists to perform studies to understand the mechanism for the uptake of the alga by the host cells (Park, et al., 1967), the persistence of such interactions (Pardy & Muscatine, 1973), their heritability during cell division of the protist (Pardy, 1974), the ability of the algae to benefit

the host (Karakashian, 1975) and their susceptibility to lytic viral infections (Meints et al., 1981).

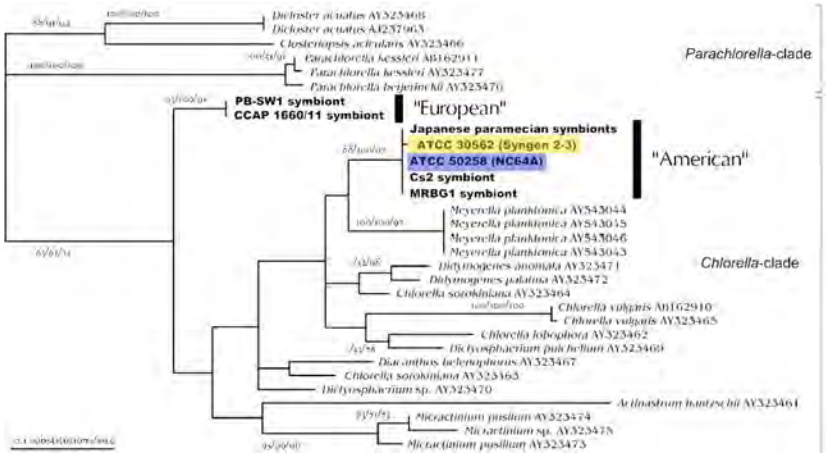


Figure 4. Phylogenetic analysis using 18S rRNA gene sequences of *Chlorella* algae including zoochlorellae species. *Chlorella variabilis* NC64A (ATCC 50258) and Syngen 2-3 (ATCC 30562) are highlighted. Illustration taken from Proschold et al. 2011.

Serendipitous discovery of chloroviruses

Previous characterizations of a chlorella-infecting particle were limited to a 1965 report of a small lytic agent (41 nm) infecting free-living *Chlorella pyrenoidosa.* The particle was described as a "chlorellophage" that showed polygonal shape and structural organization that resembled bacterial phage viruses (Van Etten, et al., 1991).

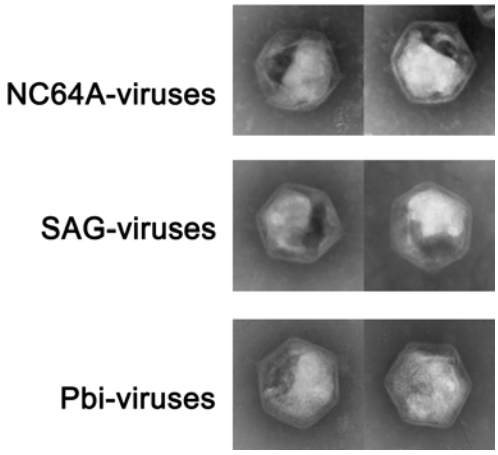


Figure 5. Electron microscopy images of all *Chlorovirus* types exhibiting the icosahedral structure characteristic of the viral group.

In 1978 (Kawakami & Kawakami, 1978) the appearance of 180 nm diameter lytic virus-

like particles (VLPs) was described in zoochlorellae after the algae were released from *P. bursaria*. However, they were not detected in algal cells growing symbiotically inside the protist. The observation of these VLPs although serendipitous was also novel, surprising and interesting. Electron microscopic studies clearly showed the first evidence that icosahedral VLPs (Figure 5) attached to freshly isolated endosymbiotic algae from *P. bursaria.* The paramecium was isolated from a tiny freshwater pond in Japan (Kawakami & Kawakami, 1978; Sherman et al., 1978).

Typically, large VLPs were detected outside the *P. bursaria* cells in depressions of the pellicle and in their food vacuole. Intriguingly, upon zoochlorella isolation, similar particles appeared almost immediately in the cytoplasm of the extracellular chlorella cells in a crystalline array (Kawakami & Kawakami, 1978). Therefore, in this chlorella:protist symbiotic interaction the host of the alga seems to function as both the vector and the protector of the algae against the virus (Kawakami & Kawakami, 1978; Meints et al., 1981; Sherman et al., 1978). Usually VLPs attached to the external surface of algae and a channel was formed which penetrated the cell wall. Rapidly some cells showed considerable gaps within the plasma membrane and the cell walls. One-hour (1h) post isolation, infected cells showed an atypical shrunk-nucleus and the mitochondria displayed irregular shapes; however, the Golgi apparatus and the chloroplast retained their normal lamellar structure and thylakoidal membrane. At around 3h PI, the number of dense-icosahedral VLPs increased, and the Golgi apparatus,

mitochondria, and chloroplast were displaced against one side. Although infected cells still conserved cytoplasmic organelles with normal appearance, the nucleus had completely disappeared. Finally, 5h PI cells burst and release viral progeny (Kawakami & Kawakami, 1978; Meints et al., 1981).

More protists, more zoochlorellas and more chloroviruses

Later in 1981, similar lytic viruses were also observed in zoochlorellae isolated from the green coelenterate *Hydra viridis* (Meints et al. 1981; Van Etten et al., 1981).

Chorella-like green algae are also endosymbionts in the digestive cells of the Florida hydra strain, *H. viridis.* Typically, isolated algae cannot be kept at room temperature for more than 3h and retain the ability to reconstitute their endosymbiotic relationship with the hydra. However, if the freshly isolated algae were maintained at 4^{o} C for periods under 24h, some cells could retain their endosymbiotic ability.

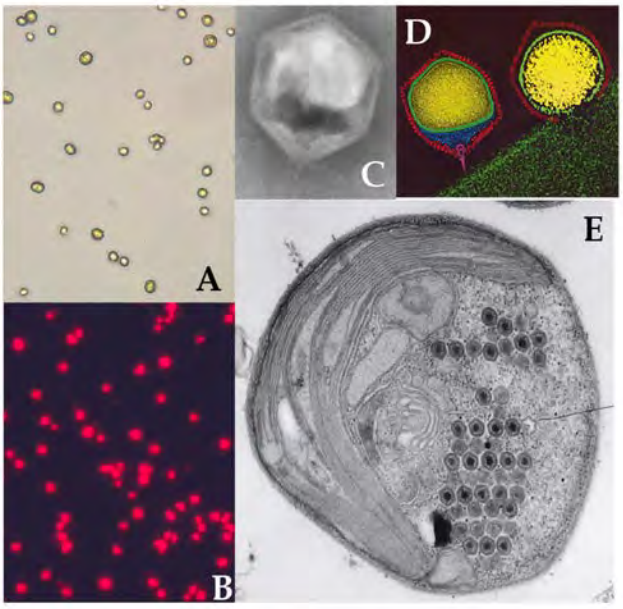


Figure 6. White light (A) and ultraviolet light (B) microscopy images of chlorella cells cultured axenically in MBBM. Electron microscopy image of the icosahedral structure of a *Chlorovirus* (ATCV-1*).* Note the spike structure present on the bottom vertex (C). PBCV-1 cryo-electron microscopy reconstruction during the initial phase of attachment (D). Formation of viral factories within zoochlorellae cells occurs approximately 3-4h pi (E).

These ex-symbiotic *Chlorella* species have a single cup-shaped chloroplast, mitochondria, and nucleus but lack a pyrenoid in the chloroplast (Van Etten et al., 1981); however, attempts to isolate and axenically culture the hydra-algae free of the host were unsuccessful (Meints et al., 1981).

Ultra structural analysis of algae isolated from the hydra and incubated at room temperature showed evidence that VLPs begin to appear and infect the algae shortly after they are isolated from the hydra. Virus attachment is accompanied by degradation of the host cell wall, followed by the injection of the viral genome inside the host. Then, a large population of icosahedral VLPs rapidly populate the ex-symbiotic algae cells shortly after they are isolated from their natural host (Figure 6). Viruses lead to a rapid lysis of the cells. Thus, the first VLP that infects ex-symbiotic chlorella-like green algae isolated from hydra was named *Hydra viridis* chlorella virus (HVCV). Particles were large at around 185 nm in size (Van Etten et al., 1982).

Inside the infected cells, small numbers of empty and filled icosahedral VLPs were observed 2h post algae isolation; by 6h many algae were filled with these icosahedral VLPs, and by 24h, algae cells were disrupted and lost all photosynthetic

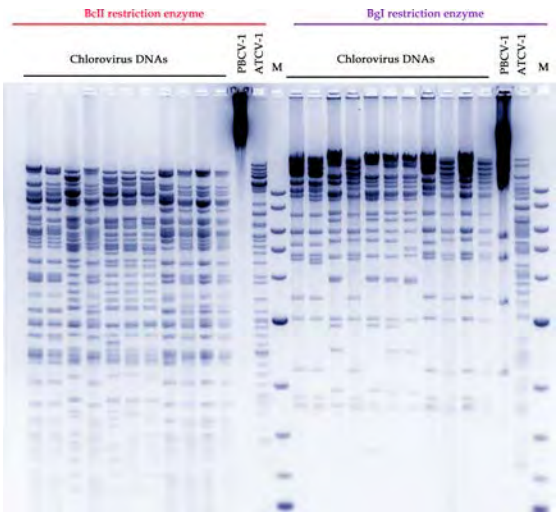


Figure 7. DNA restriction patterns of chloroviruses after digestion with restriction endonucleases BclI and BglI

capabilities. An intact nuclear membrane was never observed through the course of infection; therefore, these observations suggest that the virus might replicate in the nucleus since most organelles were intact (Meints et al., 1981).

Purified virus particles were isolated from mass cultures of *Hydra viridis* using linear 10-40% sucrose gradients that showed a sharp single band. Virion protein profiles showed at least 19 polypeptides, ranging in size from 10.3 to 82 kDa. The major capsid protein was about 46 kDa in size and represented the greatest percentage of the total protein content.

The viral genome consisted of double stranded DNA (dsDNA). This was confirmed by DNAse treatment and digestion via restriction endonucleases (Figure 7). HVCV particles were stable in pH solution ranges (4 to 10 pH) and disrupted at extreme values. Virion particles were stable in some detergents and sodium dodecyl sulfate (SDS) incubated at room temperature for up to 15 min. However, viruses were disrupted with 2% SDS at room temperature for longer time and/or temperatures higher than 30^{o}C (Van Etten et al., 1981).

By 1984, the presence of large VLPs in symbiotic systems, such as the green hydra and paramecium, were recognized as a widespread phenomenon. Even more, it was suggested that ciliates might serve as a source for the isolation of strains of eukaryotic-algae viruses (Kvitko, 1984).

Chlorovirus types and hosts

As of today, chloroviruses can replicate in large quantities only in certain unicellular, eukaryotic, ex-symbiotic chlorella-like green algae (zoochlorellae), that in nature are associated with the protozoan *P. bursaria*, the coelenterate *Hydra viridis* or the helizoon *Acanthocystis turfacea* (Van Etten & Dunigan, 2012). Four such zoochlorellae isolates can be cultured axenically and are susceptible to lytic virus infections, allowing for plaque assays (Figure 8).

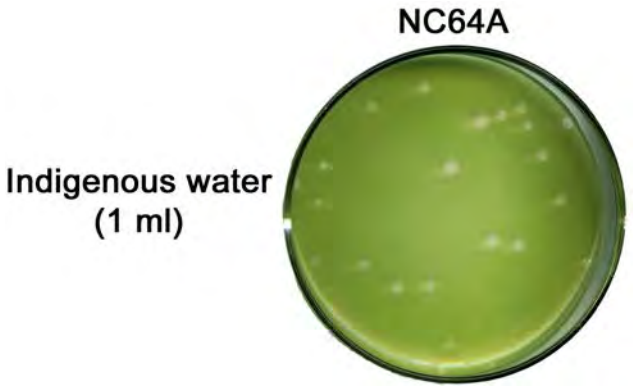


Figure 8. Plaque assay on *Chlorella variabilis* strain NC64A lawn using 1 ml of indigenous water sample collected in Lincoln NE.

These zoochlorellae recently named *Chlorella variabilis* (NC64A), *Chlorella heliozoae* (SAG 3.83), and *Micratinium conductrix* (Pbi) (Hoshina, et al., 2004; Proschold, et al., 2011) (Figure 4). Viruses infecting these three zoochlorellae are referred to as NC64A- (Van Etten et al., 1983), SAG- (Bubeck & Pfitzner, 2005), and Pbi-viruses (Reisser et al., 1988), respectively (Figure 5). As described in Chapter III of this thesis, a new group of chloroviruses, named OSy viruses, was discovered that infect *Chlorella variabilis* (Syngen 2-3).

The Chlorovirus model NC64A:PBCV-1

Paramecium bursaria chlorella virus 1 (PBCV-1) is the type member of the genus *Chlorovirus*. PBCV-1 infects and forms plaques on two *C. variabilis* strains, NC64A and Syngen 2-3. Both can grow axenically permitting plaque assay of the virus and study of its life cycle in deeper detail (Figure 9) (Van Etten et al., 1983).

At the time (1983), it was assumed that both *C. variabilis* strains might be identical; consequently, for the past 30 years the study of algae-virus interactions focused on PBCV-1 and NC64A alone.

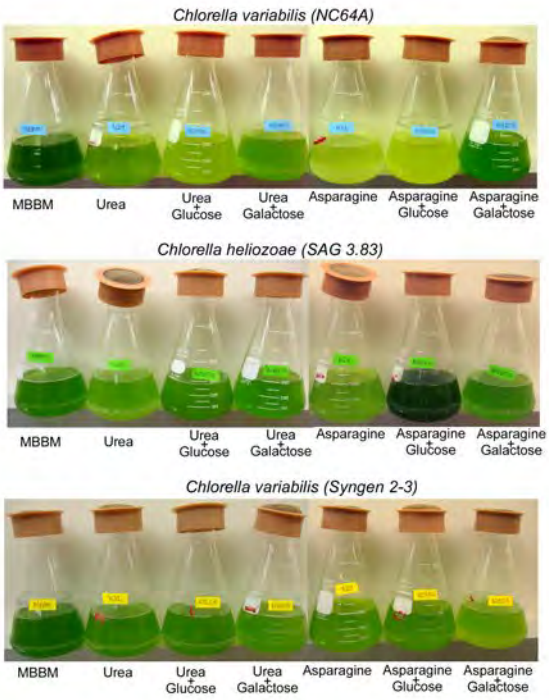


Figure 9. Different growth patterns of three zoochlorellae strains on chemically defined medium.

Similar to some bacteriophages, PBCV-1 virions display an icosahedral shaped-head with a distinctive spike-like structure at one vertex (Figure 10). The head encloses the genetic information of the virus whereas the spike serves as a tool for attaching to the NC64A host cells.

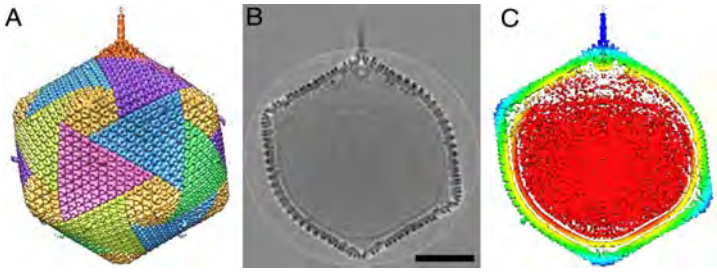


Figure 10. *Chlorovirus* prototype, *Paramecium bursaria* chlorella virus 1 (PBCV-1), as observed via 5-fold-averaged cryoEM (A), a central cross-section of the cryoEM (B), and a radially colored view of the cryoEM (C). Illustrations taken from Cherrier et al. 2009.

The architecture of PBCV-1 is composed of a complex mixture of proteins and includes a major outer glycoprotein capsid (major capsid protein MCP A430L) that surrounds a single lipid bilayered membrane and a dsDNA genome. The MCP (Vp54) is the major component of the capsid, represents about 40% by weight of the total protein content in the virus and is present in approximately 5,000 copies per virion (Nandhagopal et al., 2002; Yan et al., 2000). Cryoelectron microscopy observations, assuming icosahedral symmetry, revealed that the virion has a diameter ranging from 1,650 Å, measured along the 2-fold and 3-fold axes, and 1,900 Å, measured along the 5-fold axes. PBCV-1 has a triangulation number (T) of 169 quasi-equivalent lattice.

Five fold three dimensional reconstruction analyses showed that the icosahedral symmetry of PBCV-1 has a unique vertex containing a spike structure. The external portion of the cylindrical spike structure is 340-Å-long, and the part of the spike structure that is outside the capsid has an external diameter of about 35 Å at the tip expanding to about 70 Å at the base (Cherrier et al., 2009). The spike is

predicted to aid in puncturing the host cell wall, similar to structures characterized in bacteriophages.

The spike is too narrow to deliver DNA and so it must be moved aside during DNA release into the cell. This unique spike vertex also contains an internal pocket adjacent to the cylindrical spike structure, predicted to house enzymes involved in the initial stages of infection (e.g. cell wall degrading enzymes). Thus, the viral DNA located inside the envelope is packaged nonuniformly in the particle (Cherrier et al., 2009).

The PBCV-1 host: Chlorella variabilis NC64A

C. variabilis NC64A first appeared in a report by Karakashian and Karakashian in 1965 (Proschold et al., 2011). NC64A was isolated from *P. bursaria* syngen 1 collected in North Carolina, USA, (Table 1) and resides at the American Type Culture Collection (ATCC 50258). It is particularly intron-rich, containing eight group I introns in the nuclear ribosomal DNA (rDNA) (Hoshina et al., 2004). NC64A cells are single, planktonic, spherical or ovoid, solitary cells without mucilaginous covering, and 2–7 mm in size (Hoshina et al., 2004). Their chloroplast is a single cup- or girdle-shape that fills more than half of the adult cell, with a pyrenoid covered by grains of starch. NC64A reproduction is asexual by autospores that release no more than four autospores from a mother cell in normal growth conditions. However, some characters such as cell size and thickness of the cell walls can vary with nutritional conditions (Shihira & Krauss,

1965).

The NC64A genome

NC64A has a 46.2-Mb genome that was recently sequenced (Blanc et al., 2010) (Figure 11). The genome had 9X coverage with 89% of the genome on 413 scaffolds (46 Mb). While the overall GC content is high (67.2%), there are genomic islands with significantly lower GC content that have greater expressed sequence target coverage. NC64A contains at least 9,791 protein-encoding genes (CDSs), and the gene annotation reveals adaptation signatures of endosymbiosis. Specifically, it contains an expansion of some protein families (PFAM) that could have participated in adaptation to symbiosis. Intriguingly, a subset of PFAM domains was found overrepresented also in organisms that have intracellular or symbiotic life styles. Thus, the corresponding proteins in NC64A could potentially also play roles in the *C. variabilis* intracellular interaction with the protozoan *P. bursaria.*

These PFAM domains include proteins containing protein–protein interaction motifs (F-box and MYND), adhesion domains (fasciclin), Cys-rich GCC2_GCC3 signatures, trypsin-like proteases domains, class 3 lipase motifs and amino acid (Aa) transporters domain (Blanc et al., 2010).

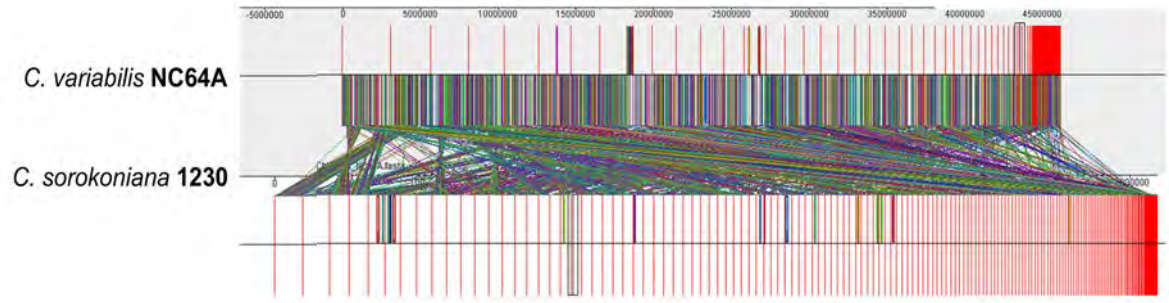


Figure 11. Full genome comparison of ex-symbiotic *Chlorella variabilis* NC64A with the free-living strain *Chlorella sorokoniana* 1230.

PBCV-1 life cycle: effective and successful replication

PBCV-1 attaches rapidly and specifically to NC64A cells by viral surface proteins, such as spike- and fiber-like structures (Figure 12). The host receptor for PBCV-1 is probably a polysaccharide-like component (Meints et al., 1988). Immediately after PBCV-1 attachment, a viral-packaged cell wall-degrading enzyme(s) digests the wall at the point of attachment (Meints et al., 1984).

After cell wall degradation, the viral internal membrane fuses with the host membrane causing host membrane depolarization (Frohns et al., 2006), potassium ion efflux (Neupartl et al., 2008), and an increase in the cytoplasmic pH. These events probably facilitate entry of the viral DNA and virion-associated proteins into the cell. The empty capsid remains on the surface of the alga. Viral DNA and proteins subsequently must be actively transported to the nucleus where early transcription can be detected 7 min after mixing the virus and the host cells. PBCV-1 infection rapidly inhibits host RNA, protein syntheses, CO_2 fixation and photosynthesis (Van Etten et al., 1983). By 5 min PI, host DNA and

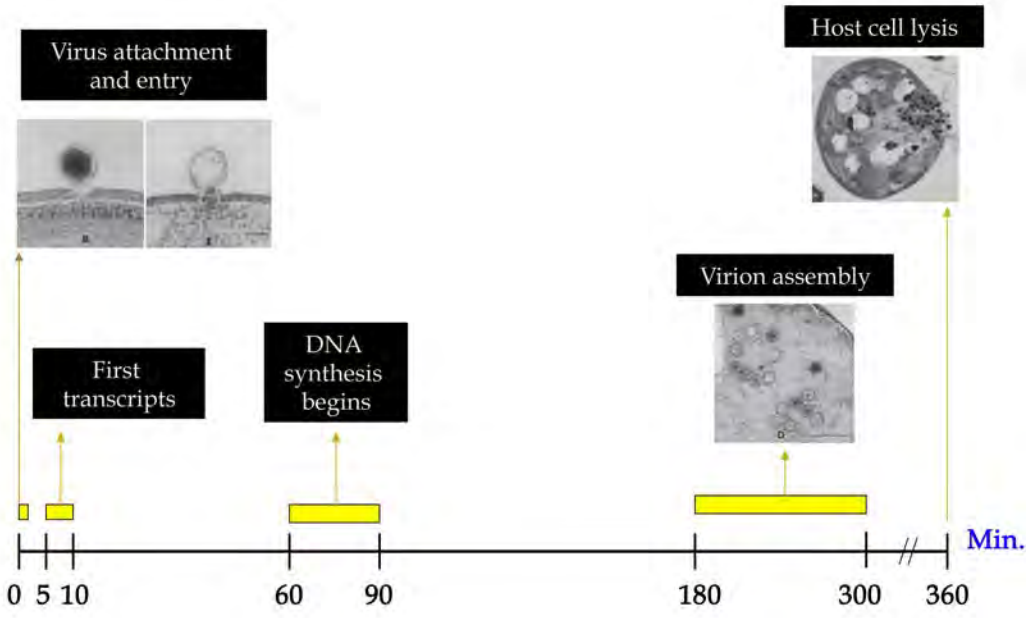


Figure 12. PBCV-1 virus replication cycle. First transcripts are made 5-10 minutes post infection (pi). Early and early/late transcripts are processed up to 1h pi when DNA synthesis begins. Viral factories begin to form 3h pi, and host cell lysis occurs at 6-8h pi. Illustration taken from Yanai and Van Etten 2009.

chromatin begin to be degraded (Agarkova et al., 2006). Within 5 to 10 min PI the synthesis of early viral transcripts begins, presumably after sequestering the cellular transcriptional machinery because the virus lacks a recognizable RNA polymerase gene (Fitzgerald et al., 2007a; Fitzgerald et al., 2007b). Additionally, no polymerase activity was detected in virion extracts reinforcing the sequestering hypothesis (J. Rohozinski and J. Van Etten, unpublished results). Chloroplast ribosomal RNAs (rRNA), but not cytoplasmic rRNAs, are degraded beginning at 30 min PI (Van Etten et al., 1984). Virus DNA synthesis starts 60 to 90 min PI and is followed by transcription of late virus genes.

The nucleus, mitochondria, and Golgi apparatus become appressed to the chloroplast, leaving finely granulated electron-translucent areas in the cytoplasm where PBCV-1 virions ultimately assemble in viral factories (VFs) (Meints et al., 1986; Milrot et al., 2015). PBCV-1 VFs are considerably different from those generated by other NCLDV relatives such as Vaccinia and Mimivirus. PBCV-1 VFs exhibit a complex membrane network composed of host cisternae and open membrane sheets. Cisternae are ER membranes that are derived from host outer nuclear membranes and act as precursors of PBCV-1 internal membranes. Subsequently, cisternae membranes are ruptured into abundant single bilayer membrane sheets that accumulate in the center of the PBCV-1 factories (Milrot et al., 2015).

Initial PBCV-1 DNA replication occurs at specific regions in the periphery of the host nuclei which might allow PBCV-1 infection to employ the host RNA polymerase. At 1 h PI the host nuclei lose their spherical shape and assume elongated morphologies that reveal enhanced heterochromicity (Milrot et al., 2015). At 2 h PI, VFs are detected in the host cytoplasm in a rosette-like organization (Meints et al., 1986; Milrot et al., 2015). At 2–3 h PI factory generation is accompanied by massive accumulation of double bilayer membrane cisternae that partially surround the VFs in contrast to open single-bilayer membrane sheets that accumulate in the center of the VFs. These open sheets interact with a capsid protein to form pre-capsids. Thus, VFs contain a network of single membrane bilayers acting as capsid templates in the central

region, and *de novo* viral genomes spread throughout the host cytoplasm but are excluded from the membrane-containing sites.

Infected cells have VFs with viral particles at various maturation stages and mature virions appear to be forced away from the VF core by the continuous generation of new virus progeny (Milrot et al., 2015). During genome encapsidation into pre-assembled capsids, both internal membrane and capsid remain incomplete with a large aperture that enables efficient DNA packaging. DNA molecules in the cell increase 4- to 10-fold by 4h PI (Van Etten et.al.,1984). Progeny PBCV-1 begin to be released 5h PI, and by 6 to 8h PI the majority of infectious virus particles are liberated. The final step involves lysis of the cell membrane and wall, presumably by late viral gene products (Van Etten et al., 1983). Like most bacteriophages, PBCV-1 replicates most efficiently in actively growing log-phase cells and poorly in stationary-phase cells. The typical PBCV-1 burst size is 200–350 plaque-forming units (PFU), although around 1,000 total virus particles are produced per cell (Van Etten et al., 1983). Thus, PBCV-1 infection transforms the cell into a very efficient viral factory. Assuming such successful replication in nature, the influence of chloroviruses turnover might be significant in driving diversity and evolution in microbial communities.

PBCV-1 physical and chemical properties

PBCV-1 has a sedimentation coefficient of about 2300 S in sucrose density gradients (Van Etten et al., 1983) and an estimated molecular mass of 1×10^9 Daltons (Yonker et al., 1985) (Figure 14).

Around 25% to 30% of the viral particles following sucrose density gradients are infectious and form plaques in NC64A cells (Van Etten et al., 1983). The PBCV-1 viral particle consists of 64% protein, 21-25% DNA, and 5-10% lipid (Skrdla et al., 1984). One distinct characteristic of *Chlorovirus* virions is the presence of a single bilayered membrane located underneath the outer capsid shell that is required for virus infectivity (Skrdla et al., 1984).

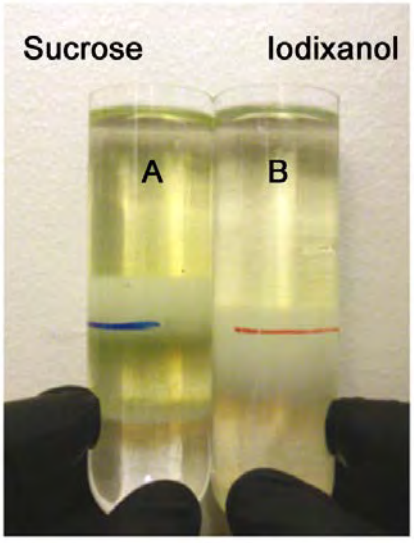


Figure 14. PBCV-1 density gradient using sucrose (A) or iodixanol (B).

PBCV-1 transcription hastily overrides the majority of highly expressed host genes

PBCV-1 transcription occurs rapidly by reprogramming the host transcription and mRNA processing machinery (Blanc et al., 2014; Yanai-Balser et al., 2010). A SET domain containing histone lysine methyltransferase (vSET) encoded by and packaged in PBCV-1 specifically methylates histone H3 at lysine 27 (H3K27), causing quick initial inhibition of host transcription and probably also aids the take over of the host transcription machinery (Mujtaba et al., 2008; Qian et al., 2006; Wei & Zhou, 2010a). vSET activity has been demonstrated both *in vitro* and *in vivo* and the family of vSET-like lysine methyltransferases are probably encoded by all chlorella viruses (Wei & Zhou, 2010b).

The virus-encoded and virion-packaged DNA restriction endonucleases also initiate rapid degradation of host chromatin, which aids in the virus take over of the host transcription machinery (Agarkova et al., 2006).

The initiation of viral transcription implies a tight regulation; as a result PBCV-1 transcription can be divided into early and late stages based on the initiation of virus DNA synthesis and the incorporation of adenine into polyadenylate-containing RNAs (Blanc et al., 2014; Yanai-Balser et al., 2010). Additionally, some genes labeled as early-late, are expressed before DNA synthesis begins, but expression is still detected after 60 min PI. No viral transcripts are detected within PBCV-1 virions (Blanc et al., 2014).

As early as 7 min PI, 50 viral genes are transcribed, and by 20 min PI most PBCV-1 genes are transcribed. Consequently, early genes are probably under the control of transcription factors that are immediately active upon entry into the cell. Some of the most active early transcribed PBCV-1 genes include an RNAse III, a SWI/SNF family helicase, transcription factor TFIIB and a mRNA capping enzyme (Blanc et al., 2014).

Transcript levels continued to increase globally up to 60 min PI even at higher levels than most greatly expressed host genes. Thus by 60 min PI 41% of the poly (A+) tail containing RNAs in the infected cells are coded by PBCV-1 (Blanc et al., 2014). Proteome analysis of PBCV-1 virions suggests that 62% of late-genes are detected in the mature particle, while early and early-late genes accounted for 9% and 29% of the virion-associated proteins, respectively (Dunigan et al., 2012).

PBCV-1 untranscribed viral regions are very short, and contain many overlapping open reading frames (Li et al., 1997). For example, the sum of the sizes of the mRNAs that hybridize to viral DNA probes are often 40% to 60% larger than the probe, thus PBCV-1 probably contains overlapping genes, transcribes both strands of DNA, significantly processes RNA after transcription and/or contains some polycistronic transcripts (Schuster et al., 1986; Schuster et al., 1990).

The PBCV-1 virion proteome

PBCV-1 particles contain polypeptides which range in size from 10- to 280-kDa (Que et al., 1994; Skrdla et al., 1984). The PBCV-1 virion proteome analysis detected 148 virion-associated proteins that are included in 11 functional categories. Most proteins appear to be structural; however, others suggest enzymatic, chromatin modification, and signal transduction functions. Although, the majority of the proteins (72%) are placed in the unknown-function category, some protein functions were inferred by sequence similarity analyses (Dunigan et al., 2012). For instance, 13 out of the 148 proteins potentially function in DNA binding, cell signaling via phosphorylation, DNA degradation, virus structure, cell attachment, and polyamine biosynthesis. Additionally, other identified proteins were restriction endonucleases probably responsible for host DNA degradation early in the infection cycle. The majority of the proteome is transcribed by genes that are dispersed throughout the virus genome and usually are expressed either early-late or late during the viral replication cycle. Interestingly, the PBCV-1 proteome contains one protein (101 aa) derived from the host. BLAST analysis suggests that the protein might have nucleosome binding abilities (Dunigan et al., 2012).

PBCV-1 and Chlorovirus genomes

The PBCV-1 genome is a linear, non-permuted, 330-kb (330,805 nt), dsDNA molecule with covalently closed hairpin termini. The termini of the PBCV-1

genome consist of two 35-nucleotide partially paired terminal loops, and an identical 2222-bp inverted repeat is adjacent to each hairpin end (Zhang et al., 1994; Strasser et al., 1991). The remainder of the genome contains primarily single-copy DNA with around 802 open reading frames (ORFs) of at least 40 codons (Li et al., 1997). Four hundred and sixteen ORFs (92.8%) of the genome have an average protein size of 249 amino acids and are classified as major coding DNA sequences (CDSs) and the remaining minor 386 ORFs have an average size of 86 amino acids and are probably not CDSs. PBCV-1 also encodes 11 tRNAs (Dunigan et al., 2012).

Currently 41 chloroviruses genomes have been characterized (Jeanniard et al., 2013). Gene predictions identified 319 to 416 (CDSs) in each genome, of which 48% were given a functional annotation. All genomes were predicted to contain at least 5 and up to 16 tRNA genes (Fitzgerald et al., 2007a; Fitzgerald et al., 2007b; Jeanniard et al., 2013) and some also encode introns and inteins (Grabherr et al., 1992).

One hundred and fifty-five protein families are shared by all chloroviruses and comprise the *Chlorovirus* core protein family set. The majority (66%) have an annotated function. They include proteins such as DNA polymerase B, major capsid protein, primase-helicase, packaging ATPase and transcription factor TFII. Additional functions include proteins associated with the degradation of the host cell-wall (alginate lyase, chitinase and chitosanase), DNA replication, transcription, protein maturation, cell-wall glycan metabolism, protein

glycosylation, ion channels and transporters, polyamine metabolism, DNA methytransferases and DNA restriction endonucleases, ankyrin repeat domain-containing proteins, glycosyltransferases and additional capsid proteins (Jeanniard et al., 2013).

The chlorovirus sequence analysis also indicates that gene order (colinearity), nucleotide conservation and phylogenetic affinity are highly conserved among viruses infecting the same eukaryotic host, with only a few localized rearrangements (Jeanniard et al., 2013). Additional phylogenetic studies show that viruses infecting the same host clustered in monophyletic clades. The within-clade average protein sequence identity is 93%, 95% and 97% identity for NC64A-, SAG- and Pbi-viruses, respectively (Jeanniard et al., 2013).

Thesis approach

The following dissertation evaluates the natural history of the chloroviruses through a weekly three-year analysis of water samples from a small pond in Nebraska to determine viral abundance, prevalence, and genetic diversity (Chapter II). This work also expands the genus *Chlorovirus* by the discovery and characterization of a new virus type (OSyNE-5) with permissive and non-permissive features in phylogenically related algal species (Chapter III), and the final chapter IV includes the evaluation of metabolic hallmarks unique to the *Chlorvirus* hosts', probably related to their symbiotic life style and susceptibility to virus infections.

References

Abergel, C., Legendre, M., & Claverie, J. M. (2015). The rapidly expanding universe of giant viruses: Mimivirus, pandoravirus, pithovirus and mollivirus. *FEMS Microbiology Reviews, 39*(6), 779-796.

Agarkova, I., Dunigan, D., Gurnon, J., Greiner, T., Barres, J., Thiel, G., et al. (2008). Chlorovirus-mediated membrane depolarization of chlorella alters secondary active transport of solutes. *J Virol, 82*(24), 12181-90.

Agarkova, I., Dunigan, D. D., & Van Etten, J. L. (2006). Virion-associated restriction endonucleases of chloroviruses. *J Virol, 80*(16), 8114-23.

Albers, D., Reisser, W., & Wiessner, W. (1982). Studies on the nitrogen supply of endosymbiotic chlorellae in greem paramecium bursaria. *Plant Science Letters, 25*(1), 85-90.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews, 35*(3), 235–241.

Beijerinck, M. W., & Johnson, J. (1964; 1942). Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., et al. (2010). The chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell, 22*(9), 2943-55.

Blanc, G., Mozar, M., Agarkova, I. V., Gurnon, J. R., Yanai-Balser, G., Rowe, J. M., et al. (2014). Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in paramecium bursaria chlorella virus 1. *PloS One, 9*(3), e90989.

Bratbak, G., Heldal, M., Norland, S., & Thingstad, T. F. (1990). Viruses as partners in spring bloom microbial trophodynamics. *Applied and Environmental Microbiology, 56*(5), 1400-1405.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., et al. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology, 185*(20), 6220-6223.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14250-14255.

Bubeck, J. A., & Pfitzner, A. J. (2005). Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic chlorella strain of the heliozoon acanthocystis turfacea. *J Gen Virol, 86*(Pt 10), 2871-7.

Cernichiari, E., Muscatine, L., & Smith, D. (1969). Maltose excretion by the symbiotic algae of hydra viridis. *Proceedings of the Royal Society of London.Series B.Biological Sciences, 173*(1033), 557-576.

Chen, F., & Suttle, C. A. (1996). Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology, 219*(1), 170-178.

Cherrier, M. V., Kostyuchenko, V. A., Xiao, C., Bowman, V. D., Battisti, A. J., Yan, X., et al. (2009). An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proc Natl Acad Sci U S A, 106*(27), 11085-9.

Cho, H. H., Park, H. H., Kim, J. O., & Choi, T. J. (2002). Isolation and characterization of chlorella viruses from freshwater sources in korea. *Molecules and Cells, 14*(2), 168-176.

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., et al. (2010). The biodiversity of the mediterranean sea: Estimates, patterns, and threats. *PloS One, 5*(8), e11842.

Cortez, D., Forterre, P., & Gribaldo, S. (2009). A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biology, 10*(6), R65-2009-10-6-r65. Epub 2009 Jun 16.

Cronkite, D., & van den Brink, S. (1981). The role of oxygen and light in guiding photoaccumulation in the paramecium bursaria-chlorella symbiosis. *Journal of Experimental Zoology, 217*(2), 171-177.

Dunigan, D. D., Cerny, R. L., Bauman, A. T., Roach, J. C., Lane, L. C., Agarkova, I. V., et al. (2012). Paramecium bursaria chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. *J Virol, 86*(16), 8821-34.

Fitzgerald, L. A., Graves, M. V., Li, X., Feldblyum, T., Nierman, W. C., & Van Etten, J. L. (2007). Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect chlorella NC64A. *Virology, 358*(2), 472-84.

Fitzgerald, L. A., Graves, M. V., Li, X., Hartigan, J., Pfitzner, A. J., Hoffart, E., et al. (2007). Sequence and annotation of the 288-kb ATCV-1 virus that infects an

endosymbiotic chlorella strain of the heliozoon acanthocystis turfacea. *Virology, 362*(2), 350-61.

Frohns, F., Kasmann, A., Kramer, D., Schafer, B., Mehmel, M., Kang, M., et al. (2006). Potassium ion channels of chlorella viruses cause rapid depolarization of host cells during infection. *J Virol, 80*(5), 2437-44.

Fujishima, M. (2010). *Endosymbionts in paramecium* Springer Berlin Heidelberg.

Grabherr, R., Strasser, P., & Van Etten, J. L. (1992). The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. *Virology, 188*(2), 721-31.

Grossman, A. R. (2005). Paths toward algal genomics. *Plant Physiology, 137*(2), 410-427.

Hoshina, R., Kamako, S. I., & Imamura, N. (2004). Phylogenetic position of endosymbiotic green algae in paramecium bursaria ehrenberg from japan. *Plant Biology (Stuttgart, Germany), 6*(4), 447-453.

Jeanniard, A., Dunigan, D. D., Gurnon, J. R., Agarkova, I. V., Kang, M., Vitek, J., et al. (2013). Towards defining the chloroviruses: A genomic journey through a genus of large DNA viruses. *BMC Genomics, 14*, 158.

Johnson, M. D. (2011). The acquisition of phototrophy: Adaptive strategies of hosting endosymbionts and organelles. *Photosynthesis Research, 107*(1), 117-132.

Jolley, E., & Smith, D. C. (1978). The green hydra symbiosis. I. isolation, culture and characteristics of the chlorella symbiont of 'european' hydra viridis. *New Phytologist, 81*(3), 637-645.

K. Kvitko, B. V. G. (1984). New finding of a titratable infectious zoochlorella virus. *Dokl. Akad. Nauk. SSSR,* (279), 998-999.

Kaiser, J. (2000). ECOLOGY: California algae may be feared european species. *Science (New York, N.Y.), 289*(5477), 222b-3b.

Kamako, S., Hoshina, R., Ueno, S., & Imamura, N. (2005). Establishment of axenic endosymbiotic strains of japanese paramecium bursaria and the utilization of carbohydrate and nitrogen compounds by the isolated algae. *European Journal of Protistology, 41*(3), 193-202.

Karakashian, M. W. (1975). Symbiosis in paramecium bursaria. *Symposia of the Society for Experimental Biology, (29)*(29), 145-173.

Karakashian, S. J., & Karakashian, M. W. (1965). Evolution and symbiosis in the genus chlorella and related algae. *Evolution, 19*(3), 368-377.

Kawakami, H., & Kawakami, N. (1978). Behavior of a virus in a symbiotic system, paramecium bursaria-zoochlorella. *The Journal of Protozoology, 25*(2), 217-225.

Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology, 479-480*, 2-25.

Koonin, E. V., Krupovic, M., & Yutin, N. (2015)a. Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences, 1341*, 10-24.

Larsen, J. B., Larsen, A., Bratbak, G., & Sandaa, R. A. (2008). Phylogenetic analysis of members of the phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. *Applied and Environmental Microbiology, 74*(10), 3048-3057.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L., et al. (1997). Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology, 237*(2), 360-77.

Lim, Y. W., Schmieder, R., Haynes, M., Willner, D., Furlan, M., Youle, M., et al. (2013). Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. *Journal of Cystic Fibrosis : Official Journal of the European Cystic Fibrosis Society, 12*(2), 154-164.

Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A. W., Sachchidanand, et al. (2003). A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nature Structural Biology, 10*(3), 187-196.

Matzke, B., Schwarzmeier, E., & Loos, E. (1990). Erratum : Maltose excretion by the symbiotic chlorella of the heliozoan acanthocystis turfacea. *Planta, 182*(2), 312.

McAuley, P. J. (1987). Nitrogen limitation and amino-acid metabolism of chlorella symbiotic with green hydra. *Planta, 171*(4), 532-538.

Meints, R. H., Burbank, D. E., Van Etten, J. L., & Lamport, D. T. (1988). Properties of the chlorella receptor for the virus PBCV-1. *Virology, 164*(1), 15-21.

Meints, R. H., Lee, K., Burbank, D. E., & Van Etten, J. L. (1984). Infection of a chlorella-like alga with the virus, PBCV-1: Ultrastructural studies. *Virology, 138*(2), 341-6.

Meints, R. H., Lee, K., & Van Etten, J. L. (1986). Assembly site of the virus PBCV-1 in a chlorella-like green alga: Ultrastructural studies. *Virology, 154*(1), 240-5.

Meints, R. H., Van Etten, J. L., Kuczmarski, D., Lee, K., & Ang, B. (1981). Viral infection of the symbiotic chlorella-like alga present in hydra viridis. *Virology, 113*(2), 698-703.

Milrot, E., Mutsafi, Y., Fridmann-Sirkis, Y., Shimoni, E., Rechav, K., Gurnon, J. R., et al. (2015). Virus-host interactions: Insights from the replication cycle of the large paramecium bursaria chlorella virus. *Cellular Microbiology*.

Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology, 2*(1), 63-77.

Mujtaba, S., Manzur, K. L., Gurnon, J. R., Kang, M., Van Etten, J. L., & Zhou, M. M. (2008). Epigenetic transcriptional repression of cellular genes by a viral SET protein. *Nat Cell Biol, 10*(9), 1114-22.

Nandhagopal, N., Simpson, A. A., Gurnon, J. R., Yan, X., Baker, T. S., Graves, M. V., et al. (2002). The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci U S A, 99*(23), 14758-63.

Neupartl, M., Meyer, C., Woll, I., Frohns, F., Kang, M., Van Etten, J. L., et al. (2008). Chlorella viruses evoke a rapid release of K+ from host cells during the early phase of infection. *Virology, 372*(2), 340-8.

Pardy, R. L. (1974). Some factors affecting the growth and distribution of the algal endosymbionts of hydra viridis. *Biological Bulletin Lancaster, 147*(1), 105.

Pardy, R. L. (1976). The production of aposymbiotic hydra by the photodestruction of green hydra zoochlorellae. *The Biological Bulletin, 151*(1), 225-235.

Pardy, R. L., & Muscatine, L. (1973). Recognition of symbiotic algae by hydra viridis. A quantitative study of the uptake of living algae by aposymbiotic H. viridis. *Biological Bulletin, 145*(3), 565-579.

Park, H. D., Greenblatt, C. L., Mattern, C. F. T., & Merril, C. R. (1967). Some relationships between chlorohydra, its symbionts and some other chlorophyllous forms. *Journal of Experimental Zoology, 164*(2), 141-161.

Pietila, M. K., Demina, T. A., Atanasova, N. S., Oksanen, H. M., & Bamford, D. H. (2014). Archaeal viruses and bacteriophages: Comparisons and contrasts. *Trends in Microbiology, 22*(6), 334-344.

Proschold, T., Darienko, T., Silva, P. C., Reisser, W., & Krienitz, L. (2011). The systematics of zoochlorella revisited employing an integrative approach. *Environmental Microbiology, 13*(2), 350-364.

Qian, C. M., Wang, X. Q., Manzur, K., Sachchidanand, Farooq, A., Zeng, L., et al. (2006). Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. *Journal of Molecular Biology, 359*(1), 86-96.

Que, Q., Li, Y., Wang, I. N., Lane, L. C., Chaney, W. G., & Van Etten, J. L. (1994). Protein glycosylation and myristylation in chlorella virus PBCV-1 and its antigenic variants. *Virology, 203*(2), 320-7.

Reisser, W., Burbank, D. E., Meints, R. H., Becker, B., & Vanetten, J. L. (1991). Viruses distinguish symbiotic chlorella spp of paramecium-bursaria. *Endocytobiosis and Cell Research, 7*(3), 245-251.

Reisser, W., Burbank, D. E., Meints, S. M., Meints, R. H., Becker, B., & Van Etten, J. L. (1988). A comparison of viruses infecting two different chlorella-like green algae. *Virology, 167*(1), 143-9.

Reisser, W., Klein, T., & Becker, B. (1988). Studies on phycoviruses. On the ecology of viruses attacking chlorellae exsymbiotic from an european strain of paramecium-bursaria. *Archiv Fur Hydrobiologie, 111*(4), 575-583.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., et al. (2010). Viral and microbial community dynamics in four aquatic environments. *The ISME Journal, 4*(6), 739-751.

Rohwer, F., & Edwards, R. (2002). The phage proteomic tree: A genome-based taxonomy for phage. *Journal of Bacteriology, 184*(16), 4529-4535.

Rohwer, F., & Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature, 459*(7244), 207-212.

Schuster, A. M., Girton, L., Burbank, D. E., & Van Etten, J. L. (1986). Infection of a chlorella-like alga with the virus PBCV-1: Transcriptional studies. *Virology, 148*(1), 181-9.

Schuster, A. M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D., et al. (1990). Transcription and sequence studies of a 4.3-kbp fragment from a ds-DNA eukaryotic algal virus. *Virology, 176*(2), 515-523.

Sherman, LouisA., Brown, R.Malcolm, Jr. (1978). *Newly characterized protist and invertebrate viruses* - Newly Characterized Protist and Invertebrate Viruses SE- (Comprehensive Virology. Cyanophages and Viruses of Eukaryotic Algae ed.) Springer US.

Shihira, I., & Krauss, R. W. (1965). *Chlorella: Physiology and taxonomy of forty-one isolates*. College Park: University of Maryland.

Short, S. M. (2012). The ecology of viruses that infect eukaryotic algae. *Environmental Microbiology, 14*(9), 2253-2271.

Short, S. M., & Short, C. M. (2009). Quantitative PCR reveals transient and persistent algal viruses in lake ontario, canada. *Environmental Microbiology, 11*(10), 2639-2648.

Siegel, R. W. (1960). Hereditary endosymbiosis in paramecium bursaria. *Experimental Cell Research, 19*(2), 239-252.

Skrdla, M. P., Burbank, D. E., Xia, Y., Meints, R. H., & Van Etten, J. L. (1984). Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. *Virology, 135*(2), 308-15.

Strasser, P., Zhang, Y. P., Rohozinski, J., & Van Etten, J. L. (1991). The termini of the chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology, 180*(2), 763-9.

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews.Microbiology, 5*(10), 801-812.

Van Etten, J. L., Burbank, D. E., Kuczmarski, D., & Meints, R. H. (1983). Virus infection of culturable chlorella-like algae and dlevelopment of a plaque assay. *Science, 219*(4587), 994-6.

Van Etten, J. L. (2003). Unusual life style of giant chlorella viruses. *Annu Rev Genet, 37*, 153-95.

Van Etten, J. L., Burbank, D. E., Schuster, A. M., & Meints, R. H. (1985)a. Lytic viruses infecting a chlorella-like alga. *Virology, 140*(1), 135-43.

Van Etten, J. L., Burbank, D. E., Xia, Y., & Meints, R. H. (1983)a. Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. *Virology, 126*(1), 117-25.

Van Etten, J. L., & Dunigan, D. D. (2012). Chloroviruses: Not your everyday plant virus. *Trends Plant Sci, 17*(1), 1-8.

Van Etten, J. L., Graves, M. V., Muller, D. G., Boland, W., & Delaroque, N. (2002). Phycodnaviridae--large DNA algal viruses. *Arch Virol, 147*(8), 1479-516.

Van Etten, J. L., Lane, L. C., & Dunigan, D. D. (2010). DNA viruses: The really big ones (giruses). *Annu Rev Microbiol, 64*, 83-99.

Van Etten, J. L., Lane, L. C., & Meints, R. H. (1991). Viruses and viruslike particles of eukaryotic algae. *Microbiol Rev, 55*(4), 586-620.

Van Etten, J. L., & Meints, R. H. (1999). Giant viruses infecting algae. *Annu Rev Microbiol, 53*, 447-94.

Van Etten, J. L., Meints, R. H., Burbank, D. E., Kuczmarski, D., Cuppels, D. A., & Lane, L. C. (1981). Isolation and characterization of a virus from the intracellular green alga symbiotic with hydra viridis. *Virology, 113*(2), 704-11.

Van Etten, J. L., Meints, R. H., Kuczmarski, D., Burbank, D. E., & Lee, K. (1982). Viruses of symbiotic chlorella-like algae isolated from paramecium bursaria and hydra viridis. *Proc Natl Acad Sci U S A, 79*(12), 3867-71.

Van Etten, J. L., Van Etten, C. H., Johnson, J. K., & Burbank, D. E. (1985)b. A survey for viruses from fresh water that infect a eucaryotic chlorella-like green alga. *Appl Environ Microbiol, 49*(5), 1326-8.

Van Etten, J. L., Xia, Y., Narva, K. E., & Meints, R. H. (1986). Chlorella algal viruses. *Basic Life Sci, 40*, 337-47.

Van Etten, J. L. (1995). Giant chlorella viruses. *Molecules and Cells, 5*(2), 99-106.

Wei, H., & Zhou, M. M. (2010a). Dimerization of a viral SET protein endows its function. *Proceedings of the National Academy of Sciences of the United States of America, 107*(43), 18433-18438.

Wei, H., & Zhou, M. M. (2010b). Viral-encoded enzymes that target host chromatin functions. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms, 1799*(3-4), 296-301.

Witzany, G. (2012). *Viruses: Essential agents of life*. Dordrecht: Springer.

Yamada, T., Higashiyama, T., & Fukuda, T. (1991). Screening of natural-waters for viruses which infect chlorella cells. *Applied and Environmental Microbiology, 57*(12), 3433-3437.

Yamada, T., Onimatsu, H., & Van Etten, J. L. (2006). Chlorella viruses. *Adv Virus Res, 66*, 293-336.

Yamada, T., Shimomae, A., Furukawa, S., & Takehara, J. (1993). Widespread distribution of chlorella viruses in japan. *Bioscience Biotechnology and Biochemistry, 57*(5), 733-739.

Yan, X., Olson, N. H., Van Etten, J. L., Bergoin, M., Rossmann, M. G., & Baker, T. S. (2000). Structure and assembly of large lipid-containing dsDNA viruses. *Nat Struct Biol, 7*(2), 101-3.

Yanai, G. M., & Van Etten, J. L. (2009). *Transcription analysis of the chlorovirus paramecium bursaria chlorella virus-1*

Yanai-Balser, G. M., Duncan, G. A., Eudy, J. D., Wang, D., Li, X., Agarkova, I. V., et al. (2010). Microarray analysis of paramecium bursaria chlorella virus 1 transcription. *J Virol, 84*(1), 532-42.

Yau, S., Lauro, F. M., DeMaere, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., et al. (2011). Virophage control of antarctic algal host-virus dynamics. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6163-6168.

Yellowlees,. (2008). Metabolic interactions between algal symbionts and invertebrate hosts. *Plant Cell and Environment, 31*(5), 679.

Yonker, C. R., Caldwell, K. D., Giddings, J. C., & Van Etten, J. L. (1985). Physical characterization of PBCV virus by sedimentation field flow fractionation. *J Virol Methods, 11*(2), 145-60.

Zhang, Y. P., Burbank, D. E., & Van Etten, J. L. (1988). Chlorella viruses isolated in china. *Appl Environ Microbiol, 54*(9), 2170-3.

Ziesenisz, E., Reisser, W., & Wiessner, W. (1981). Evidence of de novo synthesis of maltose excreted by the endosymbiotic chlorella from paramecium bursaria. *Planta, 153*(5), 481-485.

CHAPTER II

CHLOROVIRUS NATURAL HISTORY

Three-year Survey of Abundance, Prevalence and Genetic Diversity of *Chlorovirus* Populations in a Small Urban Lake

Short title: Chlorovirus Natural History in a Small Urban Lake

Cristian F. Quispe[1,2,3], Olivia Sonderman[2], Anya Seng[2], Brenna Rasmussen[2], Garrett Weber[1,2], Claire Mueller[2], David D. Dunigan[1,2], and James L. Van Etten[1,2]

[1]Department of Plant Pathology, [2]Nebraska Center for Virology and [3]School of Biological Science, University of Nebraska-Lincoln, NE 68583-0900

Correspondence: Cristian Quispe. School of Biological Sciences. University of Nebraska-Lincoln. 204 Morrison Center. Lincoln, NE 68583. Fax: 402.472.3323. Phone: 402 472 5776. Email: quispecristian@gmail.com.

Author contributions: C.F.Q. and J.L.V.E. designed research; C.F.Q., O.S., A.S., B.R., G.W., D.D.D. and C.M. performed the experiments; C.F.Q., D.D.D. and J.L.V.E. analyzed data; and C.F.Q., O.S. and J.L.V.E. wrote the paper.

Conflict of Interest

The authors declare no conflict of interest.

Three-year Survey of Abundance, Prevalence and Genetic Diversity of *Chlorovirus* Populations in a Small Urban Lake

Abstract

Inland water environments cover about 2.5 percent of our planet and harbor huge numbers of known and still unknown microorganisms. In this report we examined water samples for the abundance, prevalence, and genetic diversity of a group of viruses (*Chloroviruses*) that infect symbiotic chlorella-like green algae. Samples were collected on a weekly basis for a period of 24 to 36 months from a recreational freshwater lake in Lincoln, Nebraska and assayed for infectious viruses by plaque assay. The numbers of infectious virus particles were both host- and site-dependent. The consistent fluctuations in numbers of viruses suggest their impact as key factors in shaping microbial community structures in the water surface. Even in low-viral abundance months, chlorovirus populations were maintained, suggesting that the viruses are either very stable or that there is ongoing viral production in a natural host(s). Recently, the presence of some chlorovirus DNA sequences in the human oropharyngeal virome was associated with a modest decrease in certain cognitive functions, thus the potential impact of human exposure to these viruses raises the importance of the current surveillance work.

Keywords: Chloroviruses, algal viruses, freshwater, Nebraska

Introduction

Viruses are ubiquitous members of the biosphere as they are found in essentially every ecosystem on Earth (e.g., Breitbart and Rohwer, 2005; Mokili et al., 2012). For example, microscopic and metagenomic analyses of aquatic environmental samples indicate that high concentrations of viruses (10^5 to 10^9 particles/ml) that infect microorganisms, primarily bacteria, are present (e.g., Bergh et al., 1989; Proctor and Fuhrman, 1990; Breitbart and Rohwer, 2005; Suttle, 2005). The virus number typically exceeds that of cellular organisms by at least an order of magnitude; thus, the number of different viruses within a community is huge. Their functions of predation and gene transfer make viruses key drivers in the dynamics of microbial ecosystems (Suttle, 2007; Mokili et al., 2012). Furthermore, viruses play important roles in the global biogeochemical cycling of carbon and nutrients (Bratbak et al., 1994; Fuhrman, 1999; Suttle, 2005; Rohwer and Thruber, 2009). Although most studies have been conducted on marine environments, large numbers of viruses infecting microorganisms also exist in freshwater environments (e.g., Maranger and Bird, 1995; Weinbauer et al., 2003). In addition to bacterial viruses, viruses that infect eukaryotic algae are also common in both terrestrial and marine waters throughout the world, including members of the family *Phycodnaviridae* (Short, 2012). Phycodnaviruses are a genetically diverse, yet morphologically similar, group of large dsDNA-containing viruses (160 to 560 kb) that infect eukaryotic algae (Wilson et al., 2009, 2011). Phycodnaviruses are classified into six genera.

Members of one genus, *Chlorovirus*, are common in freshwater, occasionally reaching titers as high as thousands of plaque-forming units (PFU) per ml (e.g., Van Etten et al., 1985, 1985a; Zhang et al., 1988; Yamada et al., 1991, 1993; Cho et al., 2002).

Chloroviruses infect certain unicellular, eukaryotic, symbiotic chlorella-like green algae, often referred to as zoochlorellae. Known zoochlorellae that serve as virus hosts are associated with either the protozoan *Paramecium bursaria,* the coelenterate *Hydra viridis* or the helizoon *Acanthocystis turfacea* (Van Etten and Dunigan, 2012). Four such zoochlorellae isolates can be cultured axenically and are susceptible to lytic virus infections, allowing for plaque assays. These zoochlorellae, (recently named by Hoshina et al., 2010; Proschold et al., 2011), are *Chlorella variabilis* NC64A (Van Etten et al., 1983), *Chlorella variabilis* Syngen 2-3 (Van Etten et al., 1983), *Chlorella heliozoae* SAG 3.83 (Bubeck and Pfitzner, 2005) and *Micratinium conductrix* Pbi (Reisser et al., 1988). Viruses infecting these four zoochlorellae are referred to as NC64A-, Syngen-, SAG-, and Pbi-viruses, respectively.

The chloroviruses have been isolated from many geographic locations worldwide including North and South America, Europe, Asia and Australia (e.g., Van Etten et al., 1985, 1985a; Zhang et al., 1988; Yamada et al 1991, 1993; Cho et al., 2002; Short and Short, 2009). However, a systematic, culture-dependent, long-

term analysis of infectious chlorovirus populations in inland waters has not been conducted. As noted above, chloroviruses infect zoochlorella isolates grown in the laboratory, but there is no evidence that these symbiotic algae grow independent of their hosts in indigenous waters. Thus the chloroviruses probably play dynamic, albeit largely undocumented, roles in regulating microbial communities in the ecosystem. Intriguingly, it was recently reported that DNA sequences similar to the chlorovirus ATCV-1 were present in human throat swabs and that its presence was associated with a modest, but statistically significant, decrease in certain cognitive behaviors in humans. Furthermore, mice fed ATCV-1 infected algae also exhibited statistically significant decreases in performance on certain cognitive tests (Yolken et al., 2014).

The current report describes the first systematic, culture-dependent, 2 to 3-year study to monitor the dynamics and diversity of infectious chlorovirus populations weekly in a freshwater environment. The chloroviruses were ubiquitous throughout the year and they were both host- and site-dependent as well as seasonal. Highest titers usually appeared between the 2nd week of April to the 1st week of July and the 2nd week of October to the 2nd week of December for NC64A-, Syngen- and SAG-viruses. Chlorovirus populations present in the same sample exhibited heterogeneity in plaque size and shape indicating that they are dynamic and genetically diverse in nature. Additionally, attempts to grow the zoochlorellae hosts for the viruses in indigenous sterilized water were

unsuccessful, unless the water was supplemented with organic nitrogen sources. Therefore, the mechanism(s) in nature that allows long-term chlorovirus persistence and distribution in indigenous freshwater is still unknown. Importantly, the possible public health issue of human exposure to some chlorovirus types increases the significance of this study.

Materials and methods

Collection site

Holmes Lake is a 0.4 square kilometer inland lake located in an urban, recreational 13.5 square kilometer watershed in Lancaster County, Nebraska, USA. The lake was created primarily for flood control and it is managed by the City of Lincoln. The lake is fed by two drainages that consist of approximately 32 kilometers of open channels. Most of the stream network includes urban residential, rural residential and commercial property, so sediments and nutrients from the watershed are constantly flowing into the lake (Supplementary Figure S1). During the winter months, the lake and surrounding areas are used for ice-skating, fishing, hockey and sledding. In the warmer months recreational activities include boating, picnicking, swimming and fishing.

Sampling

Water samples were collected once per week from two sites in the lake. Weekly samples were analyzed from May 2011 to May 2014 for site 1 and from June

2012 to May 2014 for site 2. Most samples (~150 ml each) were collected during the daytime (6-7 am) on Tuesdays. Samples were taken within one meter of the shore near the surface of the lake using autoclaved 250 ml bottles. After collection, the samples were transported immediately to the laboratory and filtered through a sterile cellulose acetate 0.45 μm pore-size filter (FJ25ASCCA004FL01, GVS, Fisher Scientific) prior to assay (Figure 1). Occasionally we were unable to collect samples due to extreme weather conditions in the winter.

Cell cultures

C. variabilis NC64A, *C. variabilis* Syngen 2-3 and *C. heliozoae* SAG 3.83 strains were grown on Bold's Basal Medium (BBM) containing 5% (W/V) sucrose and 1% (W/V) peptone (Modified Bold's Basal Medium-MBBM) (Van Etten et al., 1983a). *M. conductrix* Pbi was grown on FES medium (Reisser et al., 1988). All strains were grown to early log phase which was optimal for plaque assay (4 - 7 x 10^6 cells/ml) and concentrated tenfold (4 - 7 x 10^7 cells/ml) by centrifugation for the plaque assays (Supplementary Figure S2). Cell cultures were kept under constant shaking (200 rpm) and light at 26^{o}C.

Plaque assays

Each water sample was analyzed by plaque assay with the four zoochlorellae strains. Plaque assays were performed as previously described (Van Etten et al.,

1983) with minor modifications. Three ml of MBBM or FES top agar (7 g/L agar) were mixed with 300 μl of actively growing cells (4 - 7 x 10^7 cells/ml) and the water sample. Freshly filtered water samples of 100 to 1500 μl were plated to produce significant counts (25 - 120 plaques/plate) when possible (Supplementary Figure S3). The samples were poured over solidified agar (15 g/L) containing the appropriate growth media. Plates were incubated for one week in constant light and temperature, and weekly plaque averages were determined from four plates per sample/strain (Figure 1). A few high-titer water samples were diluted up to tenfold in 50 mM Tris buffer, pH 7.5 to produce plaque numbers within the desired range.

Cell cultures on native water

The Nalgene™ Rapid-Flow™ sterile disposable 500 ml bottle top filter with polyethersulfone membrane (cat# 295-4545) was used to filter 400 ml of indigenous water followed by an autoclave cycle at 15 psi for 20 min. Then, 30 ml of water with or without the addition of organic nitrogen sources was inoculated with 1 - 5 x 10^5 cells/ml of actively growing algae cells. One molar sodium nitrate (Sigma), 1 M urea (Sigma) or 0.2 M asparagine (Sigma) stock solutions were prepared and added as the sole nitrogen sources to a final concentration of 10 mM to either water samples or BBM (control) (Supplementary Figure S4). Before inoculation, cells were washed three times with BBM. Cell cultures were kept

under constant shaking (200 rpm) and light at 26^{o}C. Visual evaluations and photographs in Figure 6 were recorded after 12 - 15 days.

Results and Discussion

Chloroviruses are ubiquitous throughout the year

To understand how spatial-temporal and ecological processes might interact to shape chlorovirus richness in nature, the seasonal variation and genetic diversity of chloroviruses were determined by analyzing indigenous water samples from Holmes Lake using a culture-dependent plaque assay method that specifically detects the four chlorovirus types referred to as NC64A-, SAG-, Syngen- and Pbi-viruses.

Three out of the four chlorovirus types (NC64A-, SAG-, and Syngen-viruses) were present throughout the year in Holmes Lake (Figure 2). Syngen-viruses were the most abundant viruses in the urban lake; in contrast, no Pbi-viruses were detected during the surveillance period. These results contrast with previous studies on the distribution of chloroviruses in inland waters in England, in which infectious SAG- and Pbi-viruses were present, but NC64A-viruses were absent (Kang et al., 1993), suggesting a non-uniform worldwide distribution of chloroviruses in inland waters. Other studies that focused only on NC64A-viruses indicated that they were common in various locations within the United States (Van Etten et al., 1985, 1985a) (Supplementary Figure S5), China (Zhang

et al., 1988), Japan (Yamada et al., 1993), Korea (Cho et al., 2002) and Australia (Van Etten, unpublished results), but absent in northern Europe (Van Etten, unpublished results). Likewise Pbi-viruses were common in various locations in Germany (Reisser et al., 1988), Russia (Kvitko and Gromov, 1984) and Canada (Van Etten, unpublished results). Taken together, these results indicate that chloroviruses are widely dispersed, and local environmental conditions enrich for certain viral types, which depends on the environmental distribution of their natural host(s).

Seasonal spatio-temporal pattern of chlorovirus populations

Typical chlorovirus titers in freshwater range from 1 - 100 PFU/ml (Van Etten et al., 1985, 1985a). However, titers as high as 100,000 and 40,000 PFU/ml of NC64A-viruses were detected in single samples from Montana (Nelson, unpublished results) and South Carolina (Van Etten et al., 1985), respectively. These observations suggest that titers in the thousands of PFU/ml do occur and that the abundance, ubiquity and potentially high diversity of these viruses might play important roles in freshwater environments. To determine if virus titers were similar within Holmes Lake, samples were collected on a single day (July 2011) from different locations around the lake. The number of SAG plaque-forming viruses varied from 1 to 335 PFU/ml and NC64A plaque-forming viruses from <1 to 168 PFU/ml (Figure 2).

For the seasonal survey, two sites at Holmes Lake were selected that represent contrasting ecological and chlorovirus abundance patterns; site 1 was a sandy bank that lacked natural vegetation and had more apparent anthropological disturbance (Figure 2 and Supplementary Figure S6). It consistently had lower chlorovirus titers (combined 3-year average of 26 PFU/ml) (Figure 3). Site 2 was characterized by more stagnant water and increased natural vegetation (Figure 2). It had relatively high chlorovirus titers throughout the year (combined 2-year average of 161 PFU/ml) (Figure 3). The highest titers for NC64A-, SAG- and Syngen-viruses were 58, 165 and 142 PFU/ml in 2013, respectively for site 1. In contrast, in 2013 the highest titers for NC64A-, SAG- and Syngen-viruses were 584, 1,313 and 980 PFU/ml, respectively for site 2 (Supplementary Figure S7). As a comparison, we occasionally sampled another small pond near Lincoln and the highest titers obtained were 3,882, 6,795, and 5,039 PFU for NC64A-, SAG- and Syngen-viruses, respectively on April 2012 (Supplementary Figure S8). Other Lincoln sites were sampled that showed similar patterns (Supplementary Figure S9). Therefore, the virus concentrations can vary considerably within a small geographical region.

Previous investigations of inland waters suggested that the highest titer for NC64A-viruses occurred during the late spring (Yamada et al., 1991; Van Etten et al., 1985a). Similarly, metagenomic studies of freshwater in Lake Ontario, Canada suggested that chloroviruses varied seasonally during the year and were

highest in the summer (Short et al., 2011). In the current 3-year study, chlorovirus populations showed two distinct peaks each year; site 1 had a peak between the 2nd week of April to the 1st week of July and the 2nd week of October to the 2nd week of December for all three-virus types. The NC64A- and Syngen-viruses had similar seasonal patterns that co-varied throughout the year (Figure 3), suggesting that they might share the same or a very similar host at the sampling sites. In contrast, SAG-viruses also had two peaks but at slightly different seasonal phases and they were more variable from year to year. This result indicates that SAG-viruses probably replicated in a different host(s) than the NC64A- and Syngen-viruses, which agrees with the results obtained in the laboratory. Samples collected at site 2 had higher plaque counts throughout the year (Figure 3). Although these samples exhibited less pronounced seasonal peaks, sporadic peaks in SAG-viruses occurred between June and October. For all virus types, the seasonal patterns in Site 2 are more variable from year to year than those in Site 1 (Supplementary Figure S7).

These results indicate that chlorovirus abundance can vary substantially within the same water body. As shown in our two representative sites, the site with low virus titers (combined 3-year average of 26 PFU/ml) consistently had seasonal patterns with two viral peaks per year and more dynamic seasonal variation, whereas the site with higher titers (combined 2-year average of 161 PFU/ml) exhibited less pronounced seasonal features but more stable virus populations

over time, likely because of constant local enrichment of some microorganism(s) that sustain virus replication (Supplementary Figure S10). Taken together, there is a seasonal spatio-temporal pattern that is host- and site-dependent, with chloroviruses emerging during the spring and disappearing in the summer, and returning at the end of the fall and beginning of the winter. These patterns might be controlled by environmental factors such as water temperature, pH, salinity, etc, which varied considerably (Supplementary Figure S11). Temperature could certainly be a factor in the chlorovirus variations (Table 1).

Genetic diversity of the chlorovirus community

The morphology of the chlorovirus plaques can vary in size and degree of clarity (Van Etten et al., 1985) (Figure 4). For example, significant differences were observed when two NC64A-viruses, NY-2A and the prototype *Chlorovirus* PBCV-1, were compared at the physiological, genomic and DNA methylation levels. NY-2A has the largest genome (370 kb) of all the characterized chloroviruses (Fitzgerald et al., 2007; Jeanniard et al., 2013) and it is heavily methylated relative to PBCV-1 (Van Etten et al., 1985a). In addition, NY-2A has a burst size that is two- to threefold lower than PBCV-1, as well as a longer replication cycle [6-8 hrs for PBCV-1 and ~18 hrs for NY-2A (Van Etten et al., 1988)]. Consequently, NY-2A produces small plaques, whereas PBCV-1 produces medium size plaques. Thus, to evaluate the genetic diversity of the chloroviruses in the lake we used plaque size and morphology as an indicator of genetic

diversity in our survey. Natural samples of the chloroviruses formed several plaque sizes on the same plate (Figure 4); large plaques that formed under these experimental conditions were defined as having a diameter greater than 4 mm, whereas small plaques were those with a diameter smaller than 1 mm. The plaque sizes from water samples collected throughout the year on the NC64A and Syngen lawns varied but medium size plaques (1-4 mm) were the predominant phenotype (Figure 5). Large and small plaques were sporadic and did not exhibit an obvious seasonal pattern at either collection site. All plaques were sharply defined and clear. SAG-viruses had predominantly medium and small plaque sizes (Figure 5). Some of the SAG plaques were irregular in shape and not necessarily defined, suggesting a more versatile genetic background in these viruses. Thus, SAG-viruses exhibited the highest heterogeneity in plaque size and shape compared to NC64A- and Syngen-viruses. Together, these results indicate that chlorovirus populations are dynamic and genetically diverse in nature (Supplementary Figure S12).

Chlorovirus algal hosts in indigenous freshwater

Chloroviruses replicate in four known zoochlorella strains isolated from symbiotic interactions with protist species and can be cultured axenically in the laboratory. To determine if the NC64A, Syngen and SAG zoochlorella strains possibly grow free of their symbiotic host in indigenous water from Holmes Lake, we filtered and autoclaved water collected during September, November and December,

2013. None of the zoochlorella strains grew in the water alone (Figure 6). However, all three zoochlorella isolates grew when an exogenous organic nitrogen source (urea or asparagine) was added to the sterile indigenous water samples (Figure 6). The growth rates were visually evaluated and varied among the strains after addition of the two nitrogen sources when compared to BBM plus nitrogen controls (Figure 6). Similar phenotypes were encountered on different dates as well as at other sites (Supplementary Figure S13). Addition of nitrate alone to the indigenous water samples did not support growth of any zoochlorella strains (results not shown). As expected, none of the three zoochlorella isolates grew in non-sterilized water as they were probably immediately infected by the residential chloroviruses (results not shown).

The inability of the three virus host zoochlorellae to grow in sterilized indigenous water is interesting because these results lead to the question: what is supporting the replication of the three groups of chloroviruses? Although very little is known about the natural history of the chloroviruses, several factors need to be considered in examining this issue. i) What is the population of green endosymbiotic protists containing zoochlorellae in nature and do they continually shed zoochlorellae or when they die, do they release zoochlorellae that can be infected by indigenous chloroviruses? Currently, we do not have an answer to either question. However, pertinent to these questions is the recent report that chloroviruses tend to accumulate and attach to *Paramecium bursaria* cells

(referred to as green paramecium) without actually infecting them (Yashchenko et al., 2012). Additionally, *Hydra* species also maintain a diverse community of eukaryotic viruses, including chloroviruses, as part of their holobiont (Grasis et al., 2014). Thus, in nature viruses would be near the zoochlorellae if green protists release their symbionts, either by death or for some other reason. Furthermore, if there is a temporary increase in an organic nitrogen source, the liberated zoochlorellae might grow, at least for a short time. Although a systematic count of green endosymbiotic protists was not conducted in the current study, sporadic microscopic observations indicated that they were rare in the water samples (Supplementary Figure S14). ii) In general, infectious bacterial viruses do not survive very long in natural environments because exposure to sunlight leads to UV-induced genetic mutations (Cottrell and Suttle, 1995). Equivalent stability studies have not been conducted on chloroviruses in a natural environment. It should be noted, however, that most chloroviruses encode a functional DNA repair enzyme, a pyrimidine dimer-specific glycosylase, which could aid in their survival (Furuta et al., 1997; Jeanniard et al., 2013). Although the DNA repair protein is not packaged in the virion, the gene is expressed early during virus infection in the laboratory (Furuta et al., 1997; Dunigan et al., 2012). iii) The host/virus concentrations necessary to support bacterial virus replication in an aqueous environment have been the subject of several studies (reviewed by Short, 2012). Most of these studies indicate that at least 10^3 to 10^4 host cells per ml are necessary to maintain a constant virus

population in nature. Although similar information is lacking for the chlorovirus/zoochlorella systems, one can make some rough calculations based on the following assumptions taken from laboratory studies with PBCV-1: a) each green paramecium harbors ~200 or more zoochlorellae (Karakashian, 1975), b) the average burst size for the chloroviruses is ~800 particles per zoochlorella (Van Etten et al., 1983a), and c) about 25% of the released virus particles are infectious (Van Etten et al., 1983a). Therefore, it would require five green paramecia per ml to release 1000 zoochlorellae, which is the minimum number of cells to support bacteria phage growth. If all 200 zoochlorellae from a single green paramecia were infected with a chlorovirus, one would obtain ~160,000 virus particles per paramecia, of which ~40,000 would be infectious. However, we would expect these numbers to be much lower in natural conditions because it is very unlikely that all of the released zoochlorellae would be infected by viruses and the average burst size would probably be less than 200. Furthermore, the specific infectivity of viral particles in nature would probably be much lower than 25%. Chromosomal analysis of *Paramecium bursaria* collected in Nebraska shows that the harbored algae have genomic features similar to those seen in cultured exosymbiotic algae (Supplementary Figure S15). These are some of the factors that need to be considered to explain how the chloroviruses are maintained in nature. Finally, we cannot discard the possibility that chloroviruses have another natural host, especially when thousands of infectious particles occur per ml of indigenous water. Over the years we have

made many attempts to infect natural free-living *Chlorella* species or related species with the chloroviruses (Supplementary Figure S16); all of these attempts have been unsuccessful because the viruses do not attach to the algae tested (Meints et al., 1984; Van Etten, unpublished results). However, if another host exists it might not be a green alga.

Conclusions

A 2 to 3-year study of an urban lake in Lincoln, Nebraska indicated that infectious chloroviruses infecting three zoochlorella hosts were present throughout the year. In this study the highest titer for one of the chloroviruses reached ~1300 PFU/ml. Typically, the values were in the 1 to 100 PFU/ml range, but they were host- and site-dependent. The viruses exhibited variations in plaque size and morphology, indicating that even viruses that infect the same host have genetic diversity in natural waters. In laboratory settings, chloroviruses infect a few zoochlorella strains; however, there is no evidence that these zoochlorellae grow free of their hosts in indigenous waters. This observation raises the question: what is supporting chlorovirus replication in native environments? Therefore, the ecological processes that enable long-term chlorovirus persistence and distribution in inland freshwaters remain to be discovered.

Acknowledgements

Funding for this work was partially provided by the NSF-EPSCoR grant EPS-1004094 (JVE), Stanley Medical Research Institute Grant 11R-0001 (JVE and DDD), and the COBRE program of the National Center for Research Resources Grant P20-RR15535 (JVE). OS and GW were supported by UNL UCARE and ARD scholarships. CM was supported by two summer fellowships from the Nebraska Center for Virology. We acknowledge James Gurnon for technical assistance. We thank Mike Archer at the Nebraska Department of Environmental Quality for providing the 2010 water analysis data.

References

Bergh, O., Borshelm, K.Y., Bratbak, G., Heldal, M. 1989. High abundance of viruses found in aquatic environments. Nature 340, 467-468.

Bratbak, G., Thingstad, F., Heldal, M. 1994. Viruses and the microbial loop. Microb. Ecol. 28, 209-221.

Breitbart, M., Rohwer, F. 2005. Here a virus, there a virus, everywhere the same virus. Trends Microbiol. 13, 278-284.

Bubeck, J.A., Pfitzner, A.J. 2005. Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic Chlorella strain of the heliozoon Acanthocystis turfacea. J. Gen. Virol. 86, 2871-2877.

Cho, H.H., Park, H.H., Kim, J.O., Choi, T.J. 2002. Isolation and characterization of chlorella viruses from freshwater sources in Korea. Mol. Cells. 14, 168-176.

Cottrell, M.T., Suttle, C.A. 1995. Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate Micromonas pusilla. Limnol. Oceanogr. 40, 730-739.

Dunigan, D.D., Cerny, R.L., Bauman, A.T., Roach, J.C., Lane, L.C., Agarkova, I.V., Wulser, K., Yanai Balser, G.M., Gurnon, J.R., Vitek, J.C., Kronschnabel, B.J., Jeanniard, A., Blanc, G., Upton, C., Duncan, G.A., McClung, O.W., Ma, F., Van Etten, J.L. 2012, Paramecium bursaria chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. J Virol, 86, 8821-34.

Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C., Van Etten, J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. Virology 358, 472-484.

Fuhrman, J.A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399, 541-548.

Furuta, M., Schrader, J.O., Schrader, H.S., Kokjohn, T.A., Nyaga, S., McCullough, A.K., Lloyd, R.S., Burbank, D.E., Landstein, D., Lane, L., Van Etten, J.L. 1997. Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV damage repair gene DenV. Appl. Environ. Microbiol. 63, 1551- 1556.

Grasis, J.A., Lachnit, T., Anton-Erxleben, F., Lim, Y.W., Schmieder, R., Fraune, S., Franzenburg, S., Insua, S., Machado, G., Haynes, M., Little, M., Kimble, R., Rosenstiel, P., Rohwer, F.L., Bosch, T.C. 2014. Species-specific viromes in the ancestral holobiont hydra. PloS One 9, e109952.

Hoshina, R., Iwataki, M., Imamura, N. 2010. Chlorella variabilis and Micractinium reisseri sp. nov. (Chlorellaceae, Trebouxiophyceae): Rediscription of the endosymbiotic green algae of Paramecium bursaria (Peniculia, Oligohymenophorea) in the 120th year. Phycol. Res. 58, 188-201.

Jeanniard, A., Dunigan, D.D., Gurnon, J.R., Agarkova, I.V., Kang, M., Vitek, J., Duncan, G., McClung, O.W., Larsen, M., Claverie, J.M., Van Etten, J.L., Blanc, G. 2013. Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses. BMC Genomics. 14, 158.

Kang, J.Y., Goulder, R., Woolston, C.J. 1993. Chlorella viruses in diverse fresh waters in North East England. Lett. Appl. Microb. 16, 214-216.

Karakashian, M.W. 1975. Symbiosis in Paramecium bursaria. Symp. Soc. Exp. Biol. 19, 145-173.

Kvitko, K., Gromov, B.V. 1984. New finding of a titratable infectious zoochlorella virus. Dokl. Akad. Nauk. SSSR, 279, 998-999.

Maranger, R., Bird, D.F. 1995. Viral abundance in aquatic systems: a comparison between marine and fresh waters. Mar. Ecol. Prog. Ser. 121, 217-226.

Meints, R.H., Lee, K., Burbank, D.E., Van Etten, J.L. 1984. Infection of a Chlorella-like alga with the virus, PBCV-1: Ultrastructure studies. Virology 138, 341-346.

Mokili, J.L., Rohwer, F., Dutilh, B.E. 2012. Metagenomics and future perspectives in virus discovery. Cur. Opin. Virol. 2, 63-77.

Proctor, L.M., Fuhrman, J.A. 1990. Viral mortality of marine bacteria and cyanobacteria. Nature 343, 60-62.

Proschold, T., Darienko, T., Silva, P.C., Reisser, W., Krientz, L. 2011. The systematics of zoochlorella revisited employing an integrative approach. Environ. Microbiol. 13, 350-364.

Reisser, W., Burbank, D.E., Meints, S.M., Meints, R.H., Becker, B. and Van Etten, J.L. 1988 A comparison of viruses infecting two different Chlorella-like green algae. Virology 167, 143-149.

Rohwer, F., Thurber, R.V. 2009. Viruses manipulate the marine environment. Nature 459, 207-212.

Short, S.M. 2012. The ecology of viruses that infect eukaryotic algae. Environ. Microbiol. 14, 2253 2271.

Short, C.M., Rusanova, O., Short, S.M. 2011. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada, ISME J. 5, 810-821.

Short, S.M., Short, C.M. 2009. Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. Environ. Microbiol. 11, 2639-2648.

Suttle, C.A. 2005. Viruses in the sea. Nature 437,356-361.

Suttle, C.A. 2007. Marine viruses - major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801-812.

Van Etten,J.L., Burbank, D.E., Kuczmarski, D., Meints, R.H. 1983. Virus infection of culturable chlorella-like algae and development of a plaque assay. Science 219, 994-996.

Van Etten, J.L., Burbank, D.E., Schuster, A.M., Meints, R.H. 1985. Lytic viruses infecting a chlorella like alga. Virology 140, 135-143.

Van Etten, J.L., Burbank, D.E., Xia, Y., Meints, R.H. 1983a. Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. Virology 126, 117-125.

Van Etten, J.L., Dunigan, D.D. 2012. Chloroviruses: not your everyday plant virus. Trends Plant Sci, 17, 1-8.

Van Etten, J.L., Schuster, A.M., Girton, L., Burbank, D.E., Swinton, D., Hattman, S. 1985a. DNA methylation of viruses infecting a eukaryotic Chlorella-like green alga. Nucleic Acids Res. 13, 3471- 3478.

Van Etten, J.L., Schuster, A.M., Meints, R.H. 1988. Viruses of eukaryotic Chlorella-like algae. In: Viruses of Fungi and Simple Eukaryotes. (Y. Koltin and M. J. Leibowitz, eds). Marcel Dekker Inc. New York. p. 411-428.

Van Etten, J.L., Van Etten, C.H., Johnson, J.K. and Burbank, D.E. 1985. A survey for viruses from fresh water that infect a eucaryotic chlorella-like green alga. Appl. Environ. Microbiol. 49, 1326-1328.

Weinbauer, M.G., Christaki, U., Nedoma, J., Simek, K. 2003. Comparing the effects of resource enrichment and grazing on viral production in a meso-eutrophic reservoir. Aquat. Microb. Ecol. 31, 137-144.

Wilson, W.H., Van Etten, J.L., Allen, M.J. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. In Lesser Known Large dsDNA Viruses. (J. Van Etten, ed). Cur. Topics Microbiol. Immun., 328, Springer, Berlin. p. 1-42.

Wilson, W.H., Van Etten, J.L., Schroeder, D.C., Nagasaki, K., Brussaard, C., Bratbak, G., Suttle, C. 2011. Phycodnaviridae. Virus Taxonomy -Ninth Report. (Ed. A.M. Q. King, M.J. Adams, E. B. Carstens and E.J. Lefkowitz) Academic Press. pp. 249-262.

Yamada, T., Higashiyama, T., Fukuda, T. 1991. Screening of natural waters for viruses which infect chlorella cells. Appl. Environ. Microbiol. 57, 3433-3437.

Yamada, T., Shimomae, A., Furukawa, S., Takehara, J. 1993. Widespread distribution of chlorella viruses in Japan. Biosci. Biotech. Biochem. 57, 733-739.

Yashchenko, V.V., Gavrilova, O.V., Rautian, M.S., Jakobsen, K.S. 2012. Association of Paramecium bursaria chlorella viruses with Paramecium bursaria cells: ultrastructural studies. Eur. J. Protistol. 48, 149-159.

Yolken, R.H., Jones-Brando, L., Dunigan, D.D., Kannan, G., Dickerson, F., Severance, E.,Sabunciyan, S., Talbot, C.C., Prandovszky, E., Gurnon, J.R., Agarkova, I.V., Leister, F., Gressitt, K.L., Chen, O., Deuber, B., Ma, F., Pletnikov, M.V., Van Etten, J.L. 2014. Chlorovirus ATCV-1 is part of the human oropharyngeal virome and is associated with changes in cognitive functions in humans and mice. Proc. Nat. Acad. Sci. 111, 16106-16111.

Zhang, Y.P., Burbank, D.E., Van Etten, J.L. 1988. Chlorella viruses isolated in China. Appl. Environ. Microbiol. 54, 2170-2173.

Figure legends

Figure 1. Schematic illustration of the experimental design. Weekly water samples were analyzed by plaque assay in four zoochlorellae strains. A freshly filtered water sample within the range of 100 to 1500 μl was plated to yield significant counts (25-120 plaques/plate). Plates were incubated for one week in constant light and temperature, and weekly plaque averages were determined from four plates per sample and strain.

Figure 2. Weekly water samples were collected from two sites within Holmes Lake located in Lincoln, Nebraska (NE). Site 1 is a sandy bank that lacks natural vegetation and has more apparent anthropological disturbance. Site 2 is characterized by stagnant water and increased natural vegetation. Numbers around the lake indicate the plaque-forming units/millimeter (PFU/ml) of NC64A and SAG-viruses respectively in samples collected around the lake on one day in July 2011. Note that there were over 2 log differences between some of the sites.

Figure 3. Plot representing the seasonal dynamics of chlorovirus populations over a 3-year period at site one and over a 2-year period at site two in Holmes Lake. At site one there were two seasonal peaks early (April–July) and late (August–December) during the year. More fluctuation occurred at site two. Symbols represent the average values over the multi-year study. The x axis indicates months and y axis indicates PFU/ml of indigenous water. Each panel represents relative abundance for NC64A- (A,B), Syngen- (C,D) and SAG-viruses (E,F) from each corresponding week and location. The Supplementary Figure S7 indicates the variability for the individual years.

Figure 4. A representative Syngen 2-3 plaque assay plate with the three plaque-size categories. Large plaques were those with a diameter greater than 4 mm, medium plaques between 1-4 mm, and small plaques had a diameter smaller than 1 mm.

Figure 5. Bar graphs of relative abundance of the three plaque sizes for each site during 2012. Abundance is based on the percentage of the three plaque categories out of the total number of plaques counted in each month. Each panel represents relative

abundance for NC64A- (A,B), Syngen- (C,D) and SAG-viruses (E,F) from each corresponding month and location.

Figure 6. In-vitro flask tests of algae growth in sterilized indigenous water. Strains were grown on autoclaved indigenous water alone and/or Bolds Basal Media (BBM) supplemented with 10 mM of urea or asparagine. Pictures were taken 12-15 days post incubation. Bottom of the flask was cropped using Adobe Photoshop CS5.

Supplementary Figure legends

Supplementary Figure S1. Weekly water samples were collected from two sites within Holmes Lake located in Lincoln, NE (A). Schematic map of sedimentation depths at the lake (B). Image by Olsson Associates.

Supplementary Figure S2. In order to optimize the reproducibility of the plaque assay, we tested the ability of cells growing at different concentrations to form plaques. Cells reaching concentrations of 1-5 x 10^6, 1-4 x 10^7 and 6-7 x 10^7 cells/ml in 4, 7 and 8 days respectively were used to test susceptibility to virus infection using indigenous water collected from sites 4 and 6. We concluded that for all strains, concentrations between 1-5 x 10^6 cells/ml reached in about 3-4 days were optimal for plaquing.

Supplementary Figure S3. In order to optimize the reproducibility of the plaque assay, we tested if the amount of the water used influence the plaque numbers per plate. Cells reaching concentrations of 1-5 x 10^6 were pelleted to 1-5 x 10^7 cells/ml and then mixed with 100 ul, 500 ul or 1000 ul of native water for plaquing. We concluded the water volume used did not affect the expected number of plaques/ml.

Supplementary Figure S4. Schematic illustration of the experimental design to determine if NC64A, Syngen 2-3 and SAG 3.83 zoochlorella strains grow free of their symbiotic host in indigenous water. Sterile disposable 500 ml bottle top filters were used to filter 400 ml of indigenous water followed by an autoclave cycle at 15 psi for 20 min. 30 ml of water with or without the addition of organic nitrogen sources was inoculated with 1 - 5 x 10^5 cells/ml of actively growing algae cells. 1M urea or 0.2 M asparagine were added as the sole nitrogen sources to a final concentration of 10 mM to either water samples or BBM. Flasks were shaking at 200 rpm and 26^{o}C for 15 days in constant light.

Supplementary Figure S5. Water samples were collected from eight rivers within the continental USA. Samples represent main watersheds in Nebraska, Delaware, Maryland, New York, Mississippi, Colorado, Florida, and Alabama. Samples were evaluated following the procedure on Figure 1.

Supplementary Figure S6. Pictures of seasonal patterns of five sites during the 2013 collection year.

Supplementary Figure S7. Plot representing the seasonal dynamics of chlorovirus populations over a 3-year period at several sites including Holmes Lake. The weekly average values of plaque-forming units/ml (PFU/ml) are plotted for 2011, 2012, 2013 and 2014. The x axis indicates months and y axis indicates PFU/ml of indigenous water. Each panel represents relative abundance for NC64A-, Syngen- and SAG-viruses from each location.

Supplementary Figure S8. Highest titers observed during the sampling period on a small pond near Holmes Lake on NC64A and SAG 3.83 lawns.

Supplementary Figure S9. Plot representing the seasonal dynamics of chlorovirus populations over a 3-year period on all sampling sites in Lincoln, NE. The numbers of infectious virus particles were both host- and site-dependent. The overall average values for each week during the multi-year study are plotted. The x axis indicates months and y axis indicates PFU/ml of indigenous water. Each panel represents relative abundance for NC64A-, Syngen- and SAG-viruses from each location, demonstrating the variability between the sites.

Supplementary Figure S10. Schematic representation of the seed bank and runoff model for Chloroviruses in small urban aquatic ecosystems.

Supplementary Figure S11. Summary of weekly water pH values collected on 5 sites in Lincoln NE. They were evaluated from March to December 2013.

Supplementary Figure S12. A representative collection of plaque assay plates with different plaque numbers and sizes.

Supplementary Figure S13. Additional sites evaluated for the in-vitro flask tests of algae growth in sterilized indigenous water. Strains were grown on autoclaved indigenous water alone and/or Bolds Basal Media (BBM) supplemented with 10 mM of urea or asparagine. Pictures were taken at 7 and 15 days post incubation.

Supplementary Figure S14. UV and light microscopic observations of chlorophyll-containing organisms from collection site 7 evaluated on April 2012.

Supplementary Figure S15. *Paramecium bursaria* shown in white light, ultraviolet light, and merged picture highlighting the red-fluorescing chlorophyll of the green algae housed within the symbiont (A). Pulse field gel electrophoresis (PFGE) comparing lab ex-symbiotic strains and *Paramecium bursaria* bearing Chlorella specimens collected in Nebraska. Chromosomes were separated using a 1% 0.5X TBE gel. Electrophoresis conditions included a pulse time ramped from 700 to 1800 seconds for 72 hours at 70 V. Lane M1-*Schizosaccharomyces pombe* chromosomal DNA size standard; lane M2-*Hansenula wingei* chromosomal DNA size standard. Note the similarity in patterns between the ex-symbiotic lab-strains and the *Paramecium bursaria* bearing Chlorella samples (B).

Supplementary Figure S16. Schematic illustration for the isolation of free-living algae from indigenous water collected in Lincoln, NE.

Table legends

Table 1. Summary of water chemistry parameters collected by the Nebraska Department of Environmental Quality at Holmes Lake in Lincoln. Monthly collections, taken near the damn, were evaluated during May 2010 to September 2010. ORD= oxygen reduction potential, DO= dissolved oxygen, C= specific conductance, T= turbidity.

Figure 1.

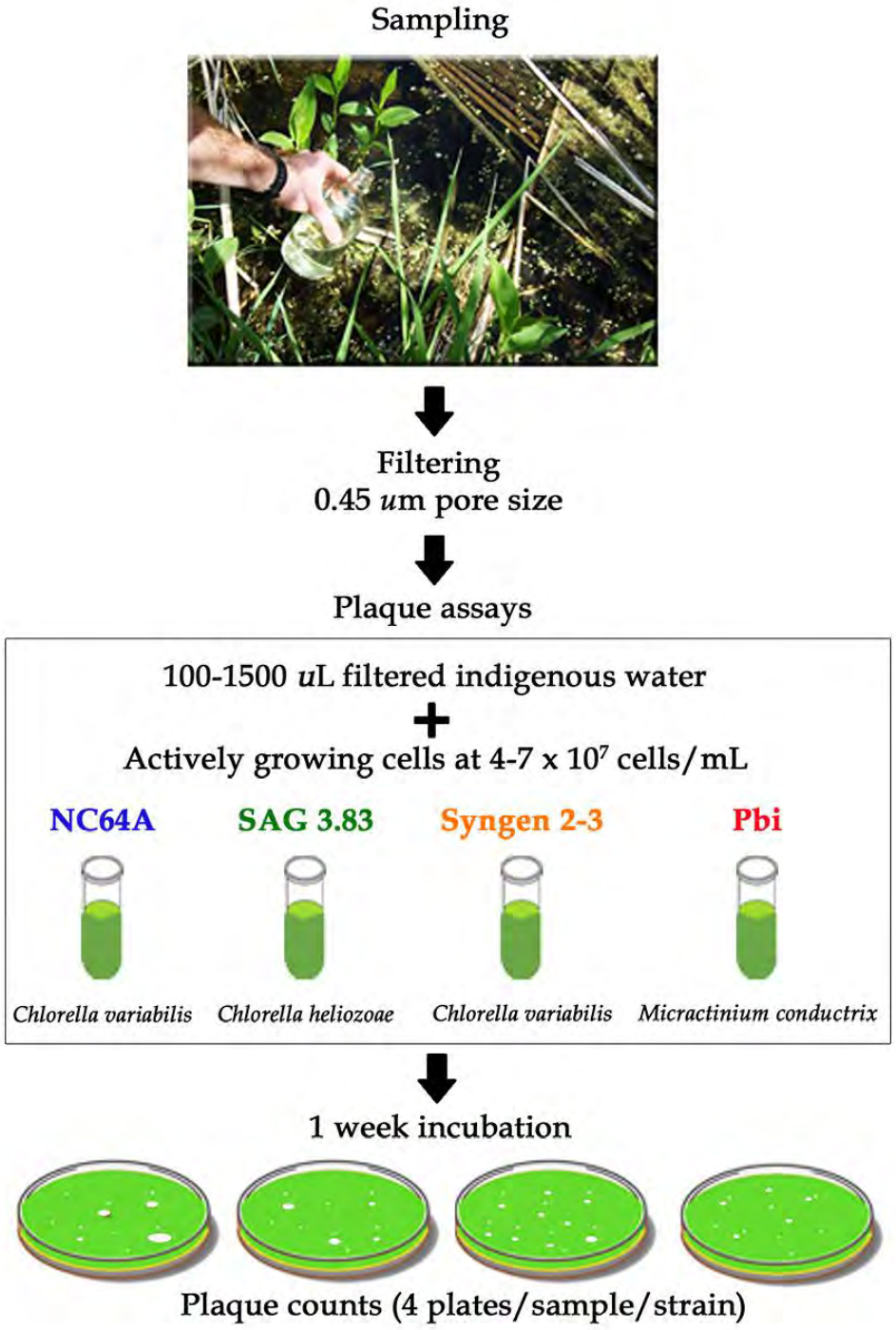

Sampling
Filtering
0.45 um pore size
Plaque assays
100-1500 uL filtered indigenous water
+
Actively growing cells at 4-7 x 10^7 cells/mL
NC64A
SAG 3.83
Syngen 2-3
Pbi
Chlorella variabilis
Chlorella heliozoae
Chlorella variabilis
Micractinium conductrix
1 week incubation
Plaque counts (4 plates/sample/strain)

Figure 2.

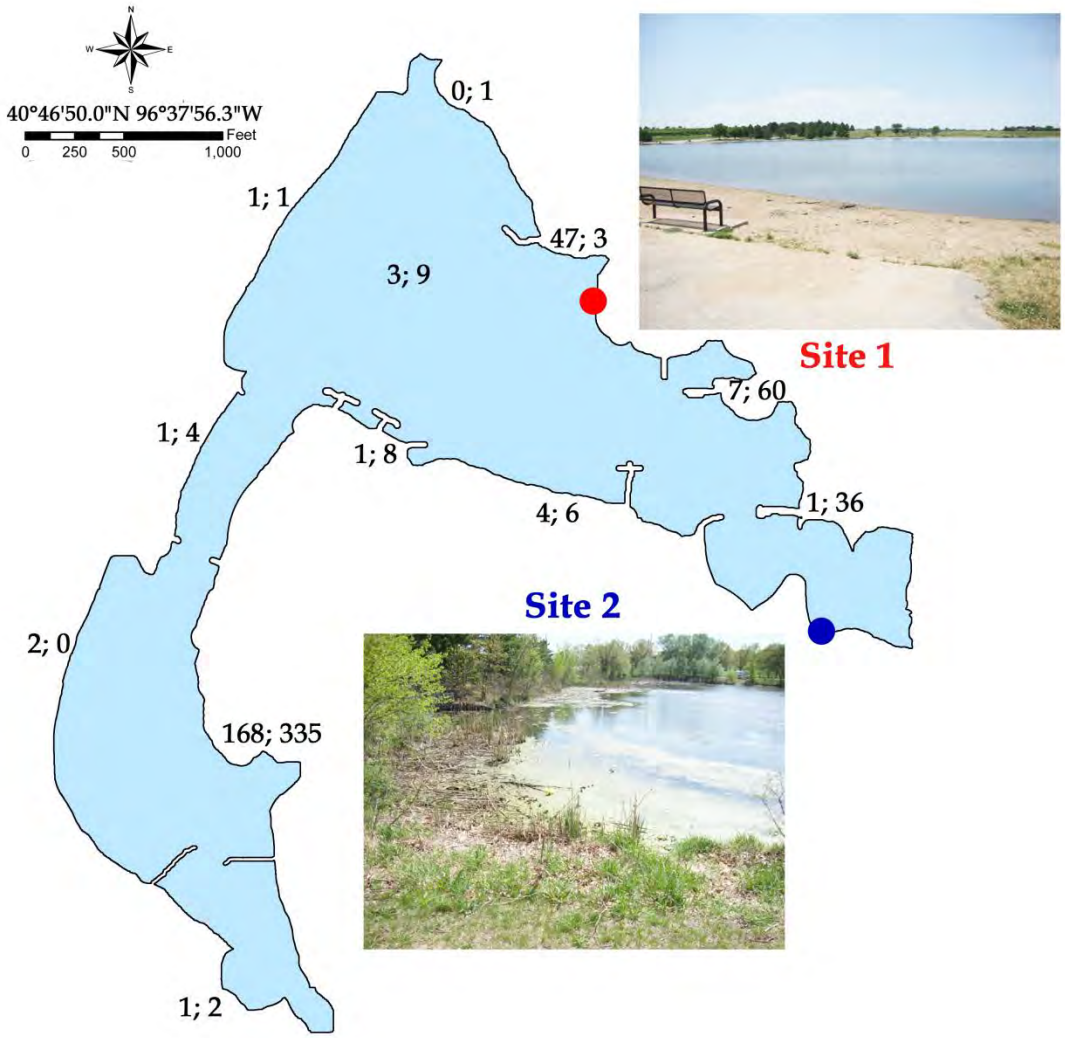

40°46'50.0"N 96°37'56.3"W
Feet
0 250 500 1,000
0; 1
1; 1
47; 3
3; 9
Site 1
7; 60
1; 4
1; 8
4; 6
1; 36
Site 2
2; 0
168; 335
1; 2

Figure 3.

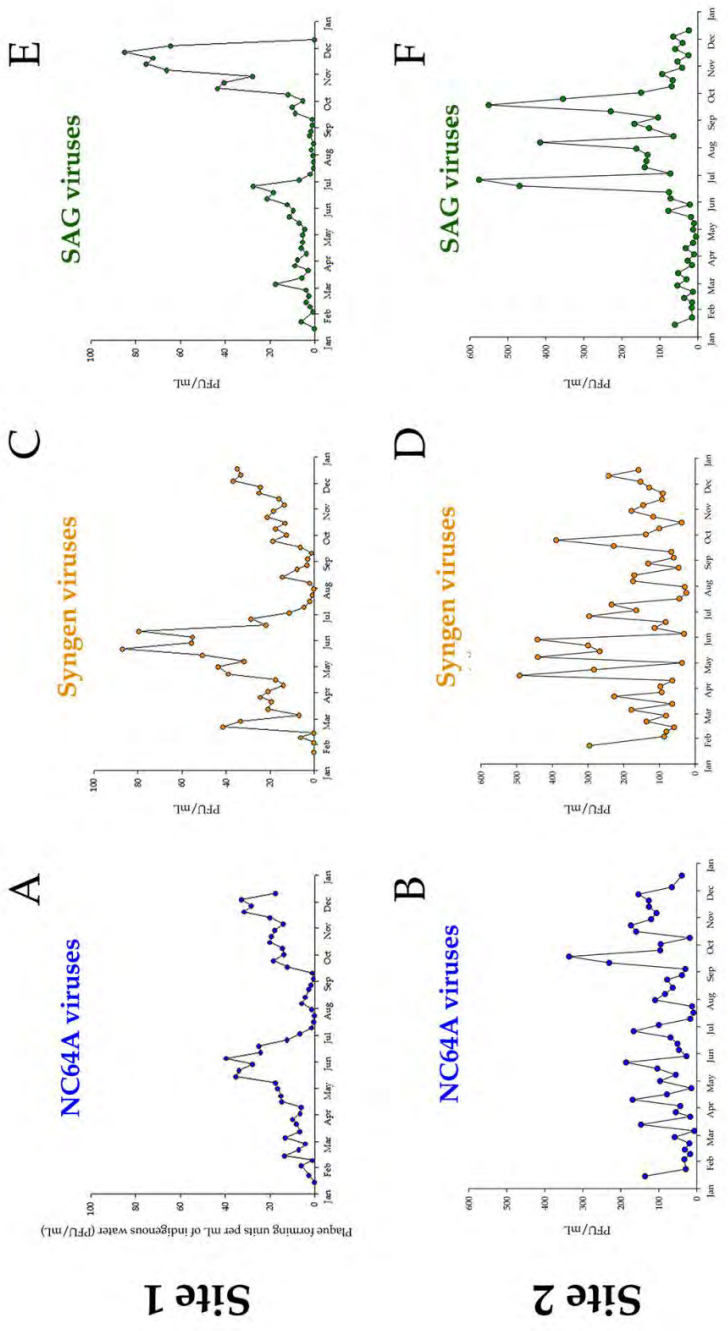

A
B
C
D
E
F
NC64A viruses
Syngen viruses
SAG viruses
Site 1
Site 2
Plaque forming units per mL of indigenous water (PFU/mL)
PFU/mL
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Figure 4.

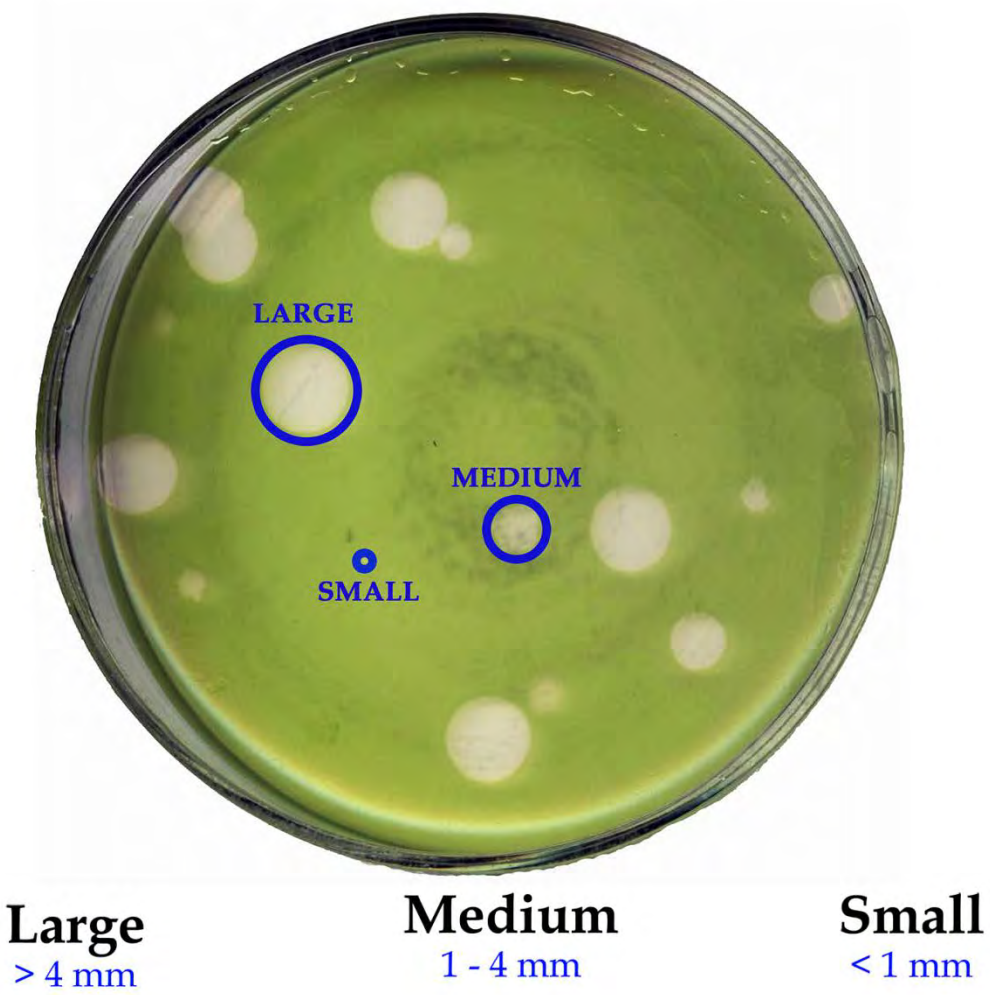


Figure 5.

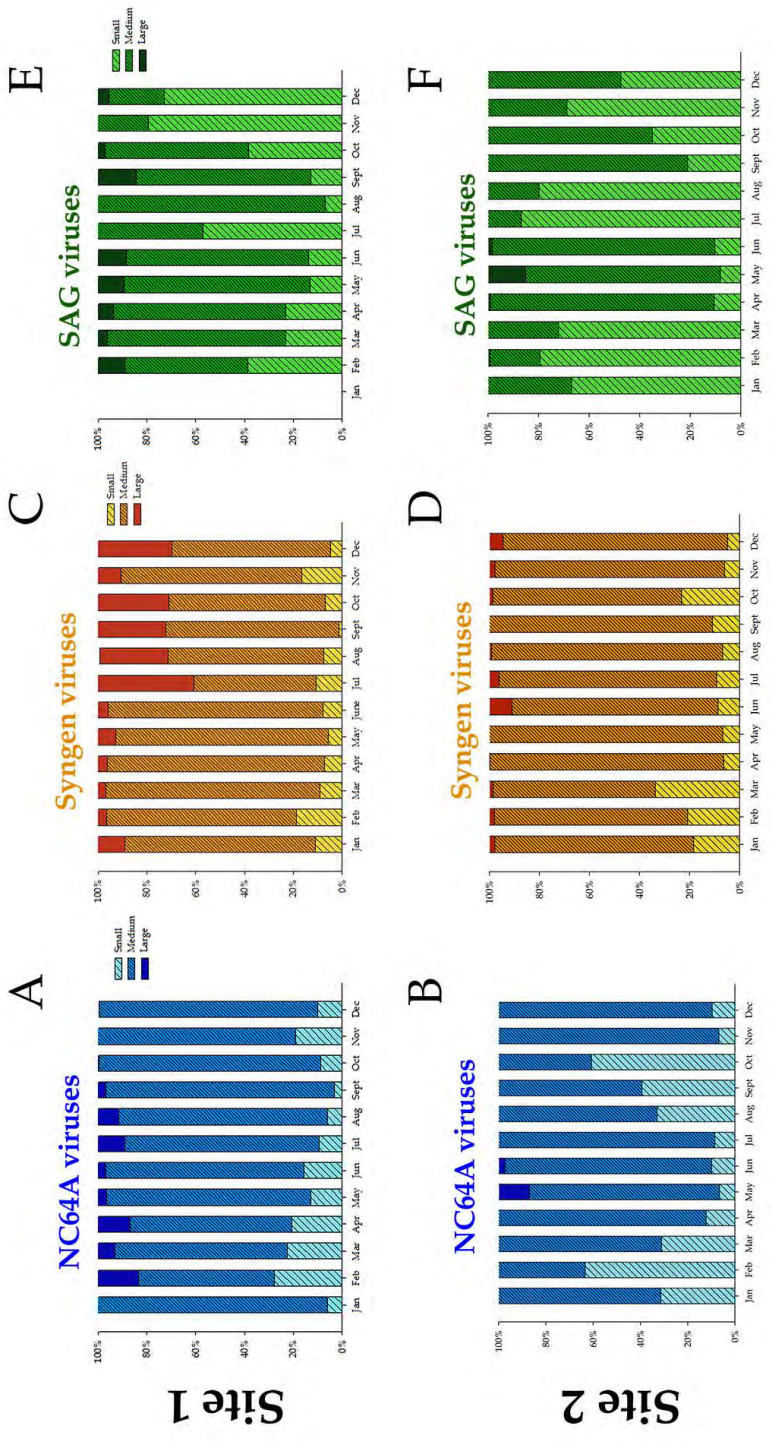
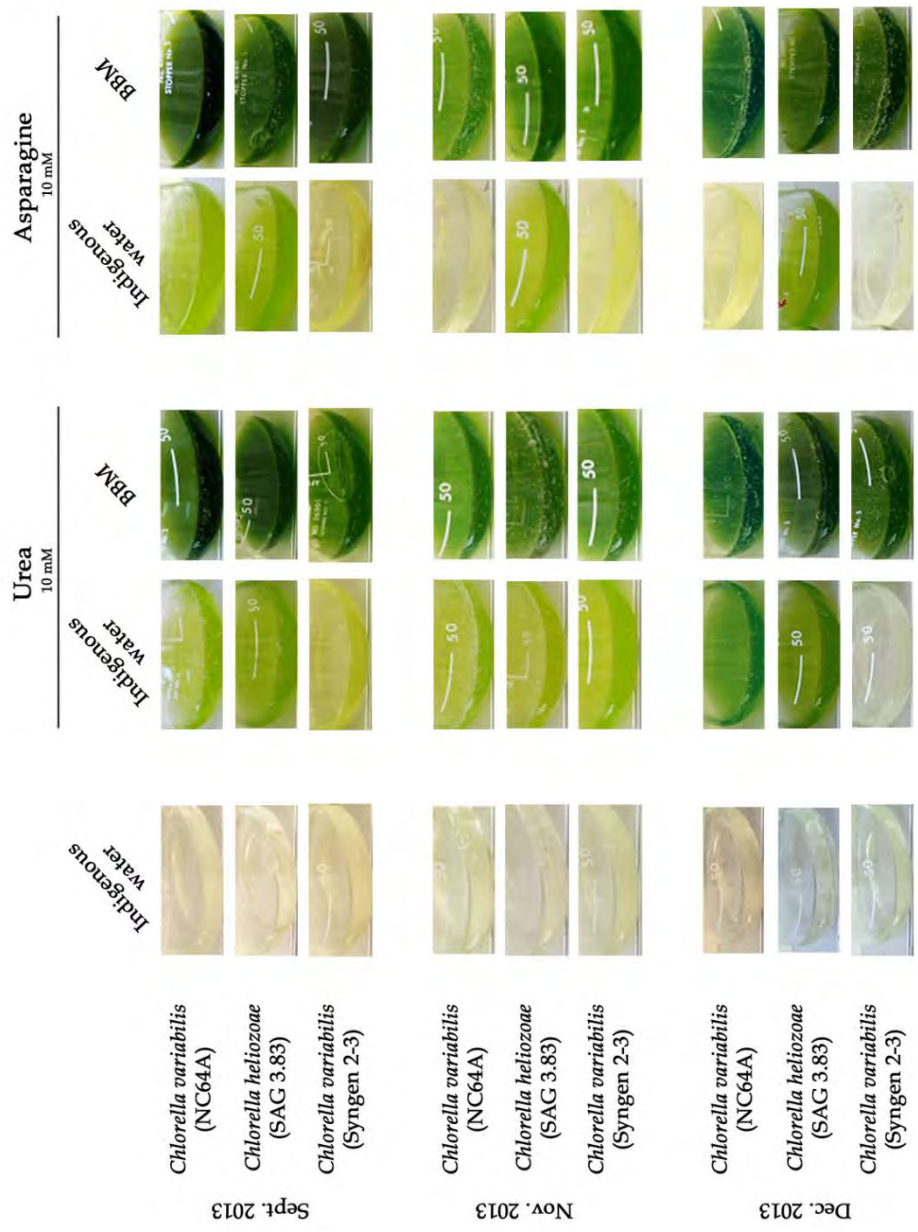
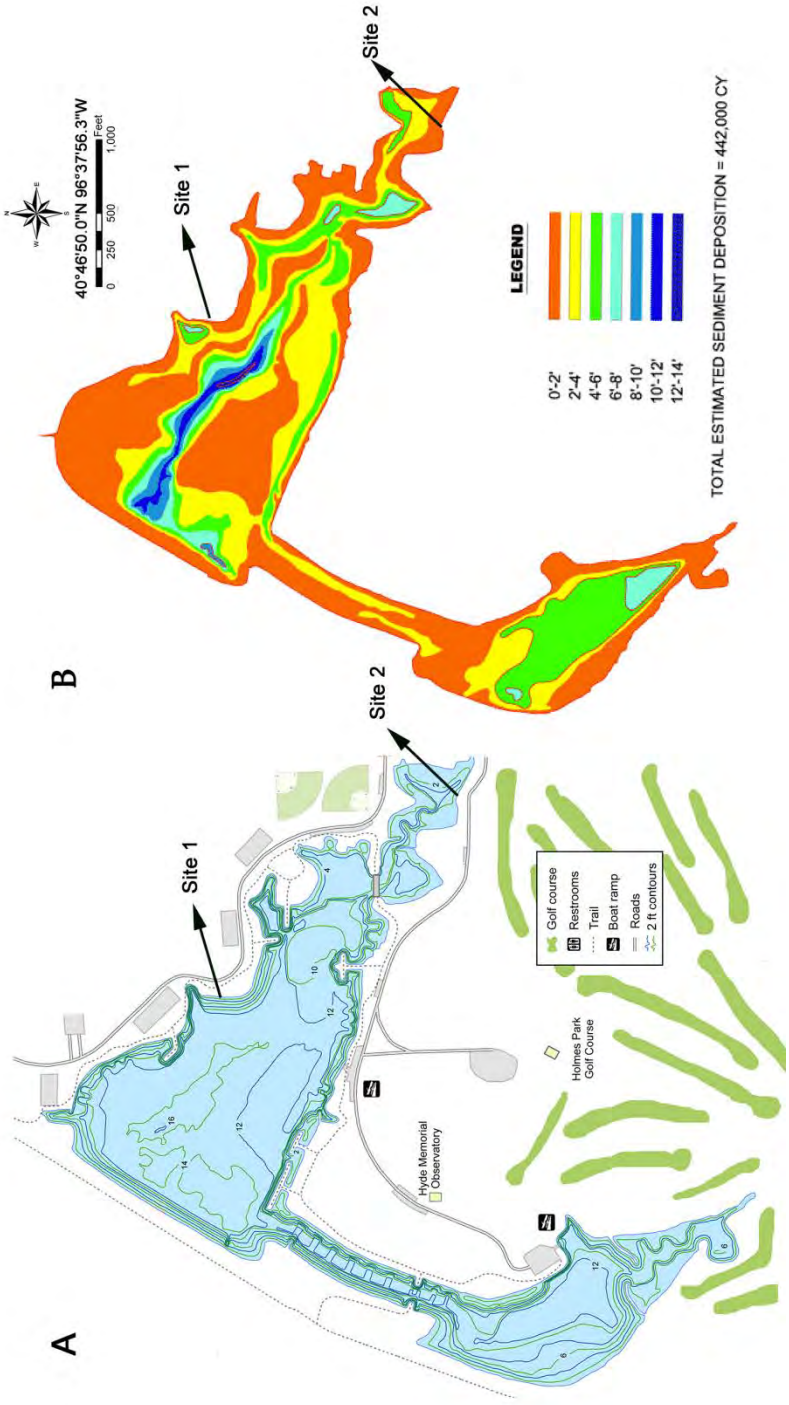

A
B
C
D
E
F
Site 1
Site 2
NC64A viruses
Syngen viruses
SAG viruses
Small
Medium
Large
Jan
Feb
Mar
Apr
May
June
Jul
Aug
Sept
Oct
Nov
Dec
100%
80%
60%
40%
20%
0%

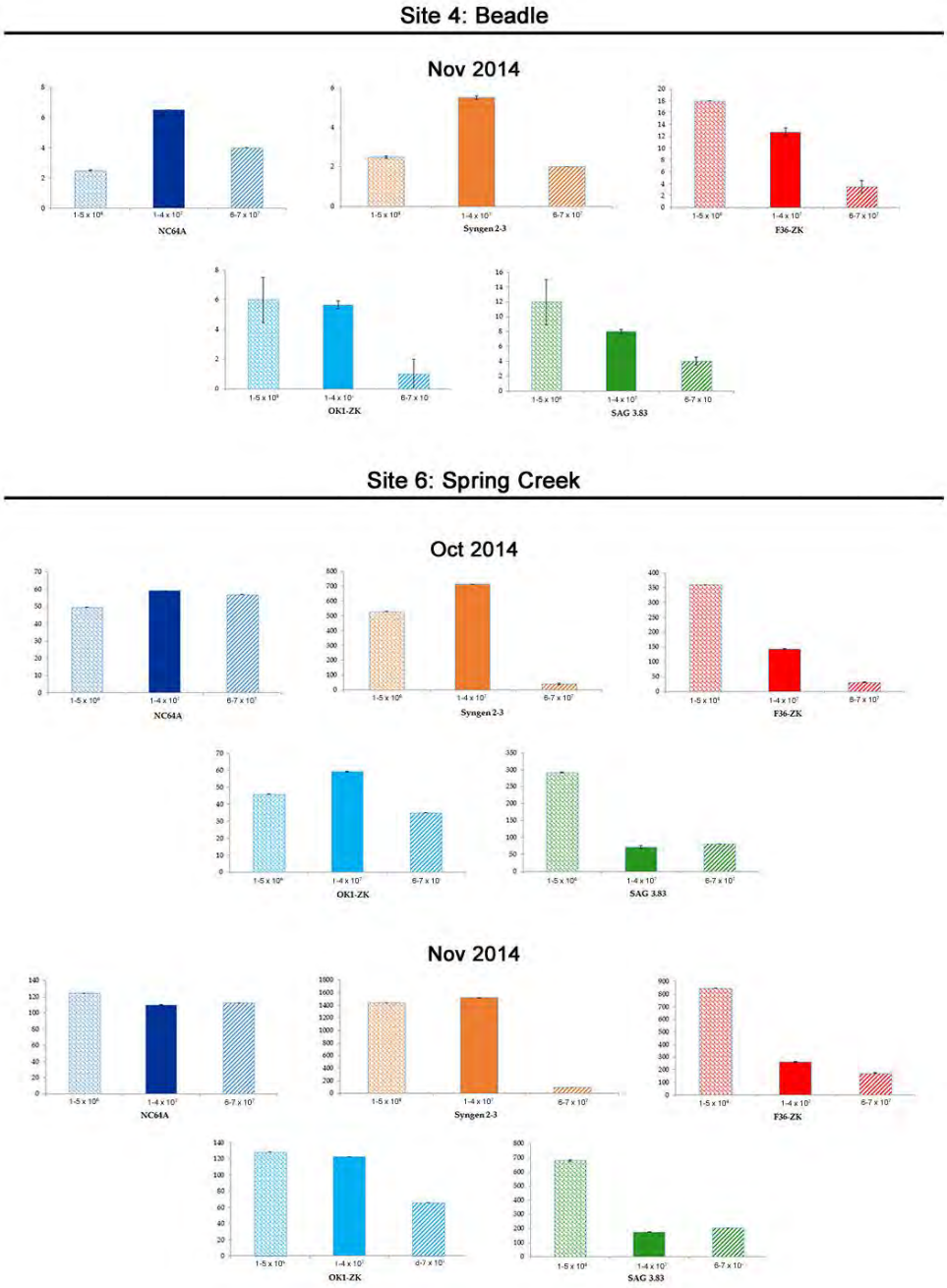
Figure 6.


Asparagine
10 mM
Urea
10 mM
BBM
Indigenous water
Chlorella variabilis (NC64A)
Chlorella heliozoae (SAG 3.83)
Chlorella variabilis (Syngen 2-3)
Sept. 2013
Nov. 2013
Dec. 2013

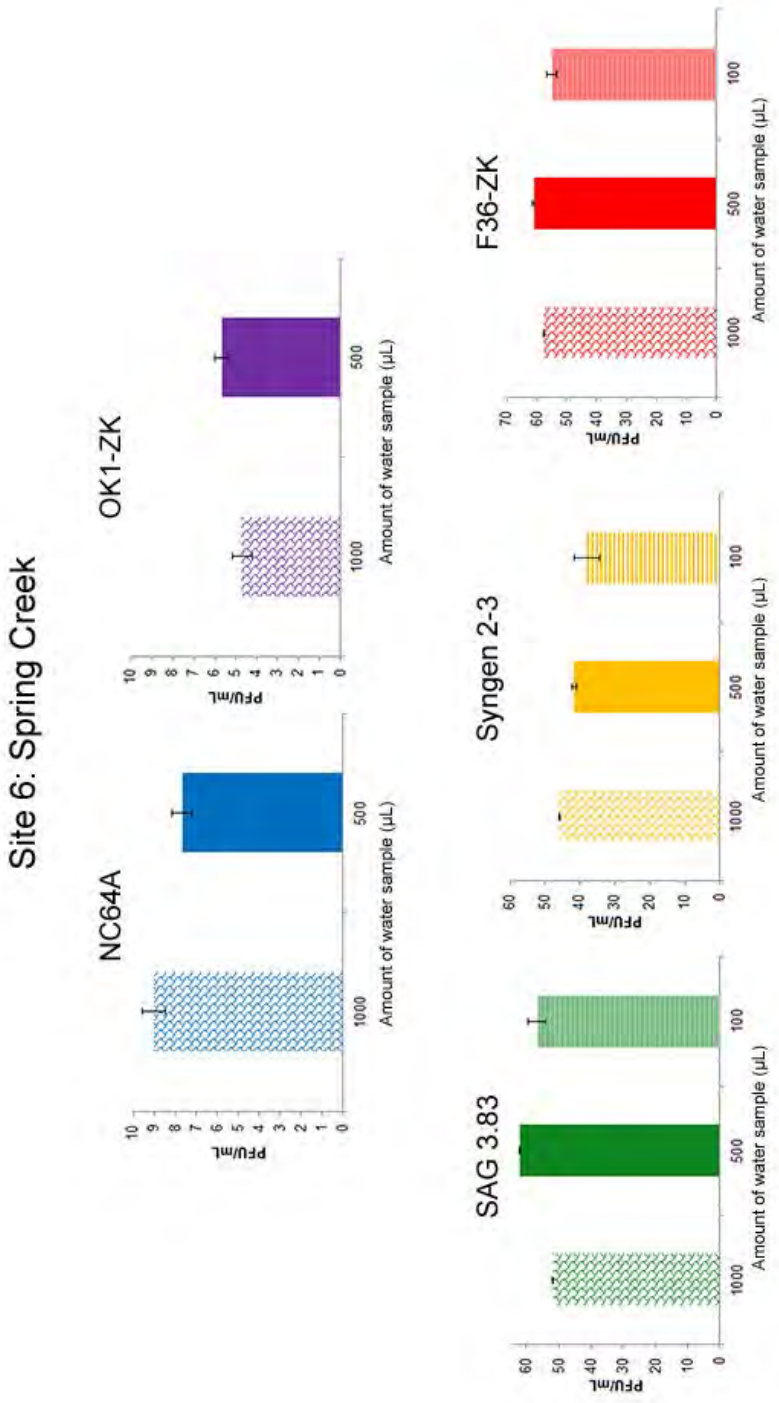
Supplementary Figure S1.



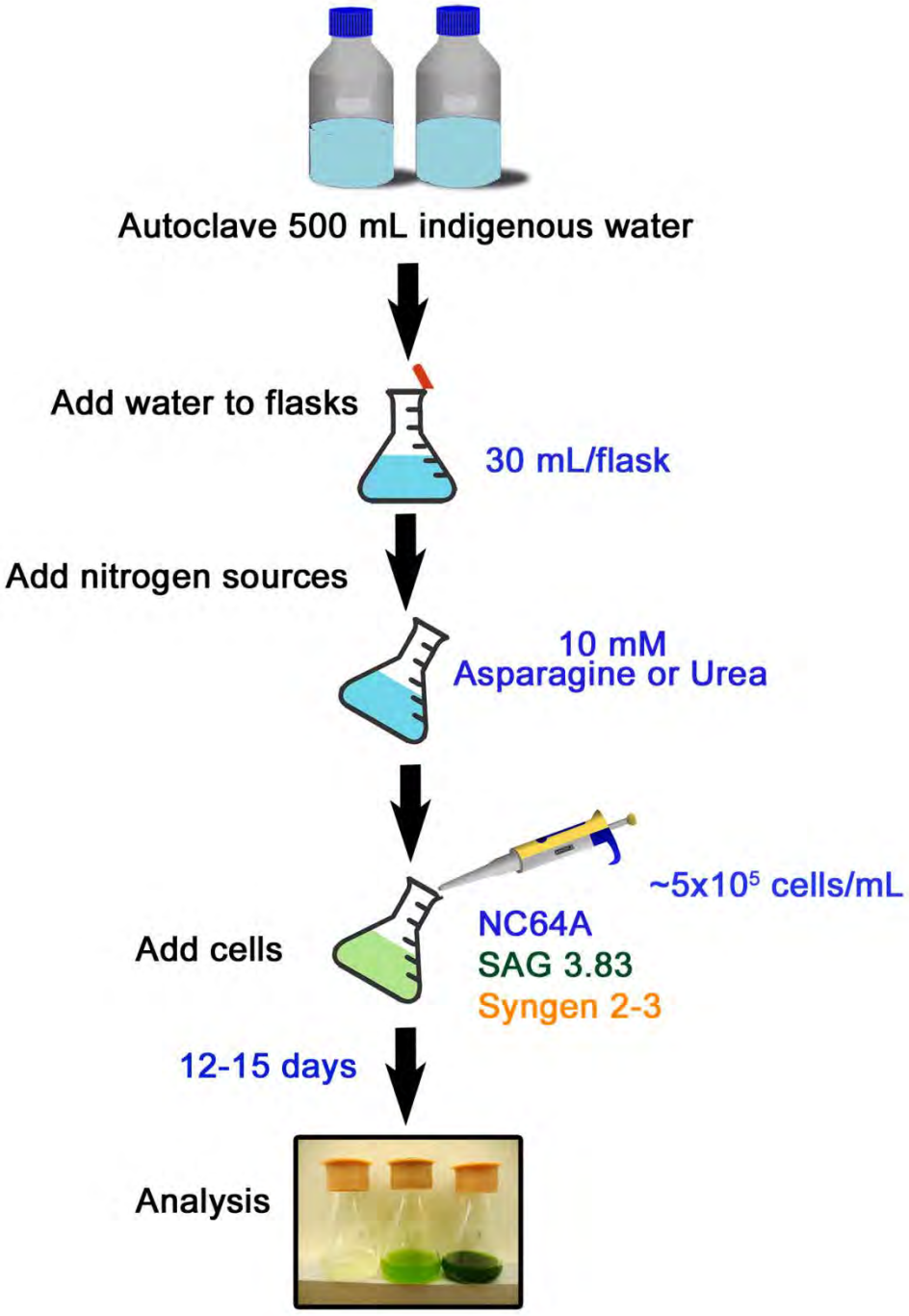
Supplementary Figure S2.



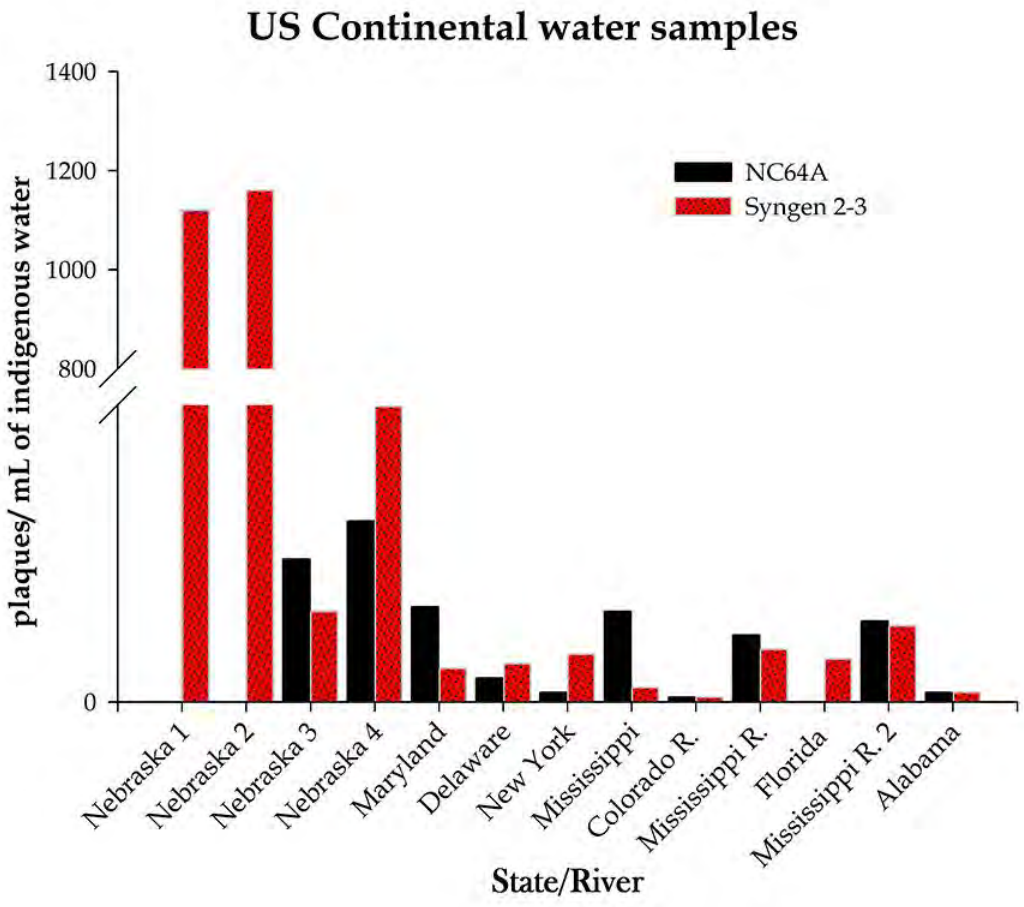
Supplementary Figure S3.



Supplementary Figure S4.

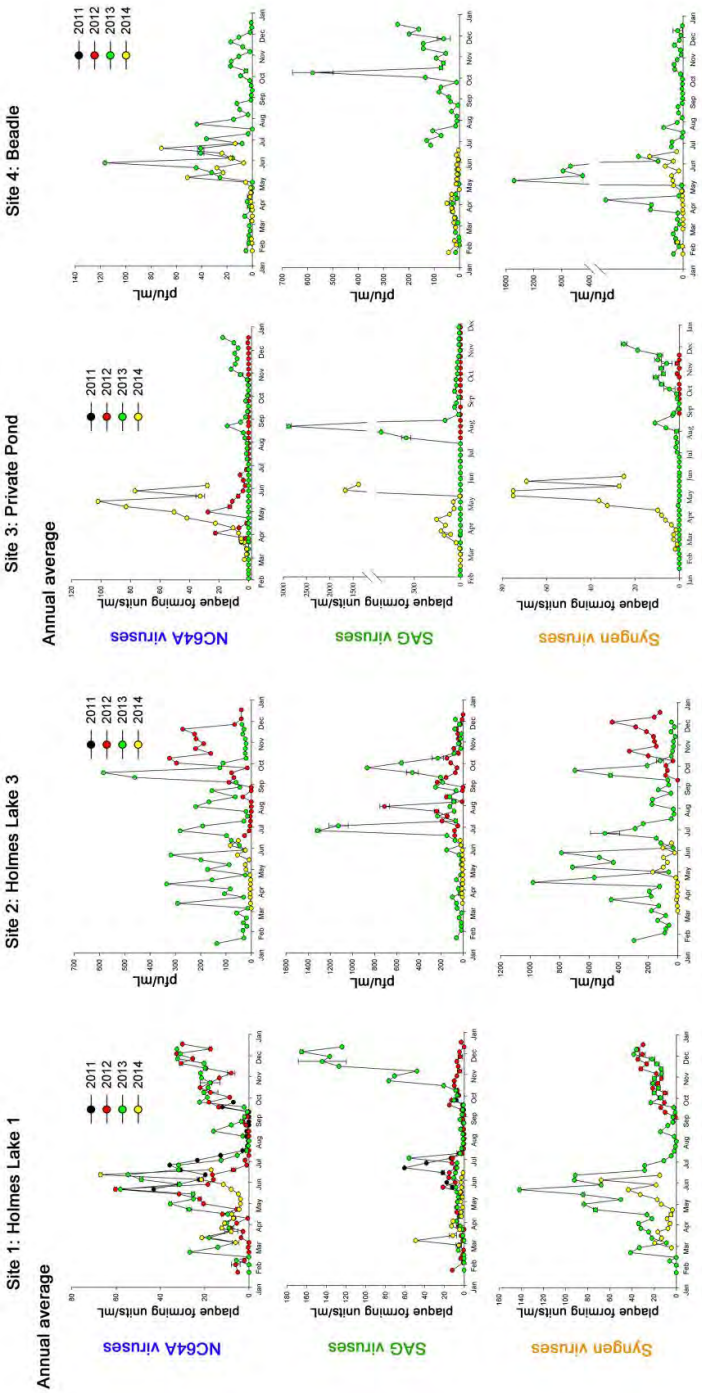


Supplementary Figure S5.

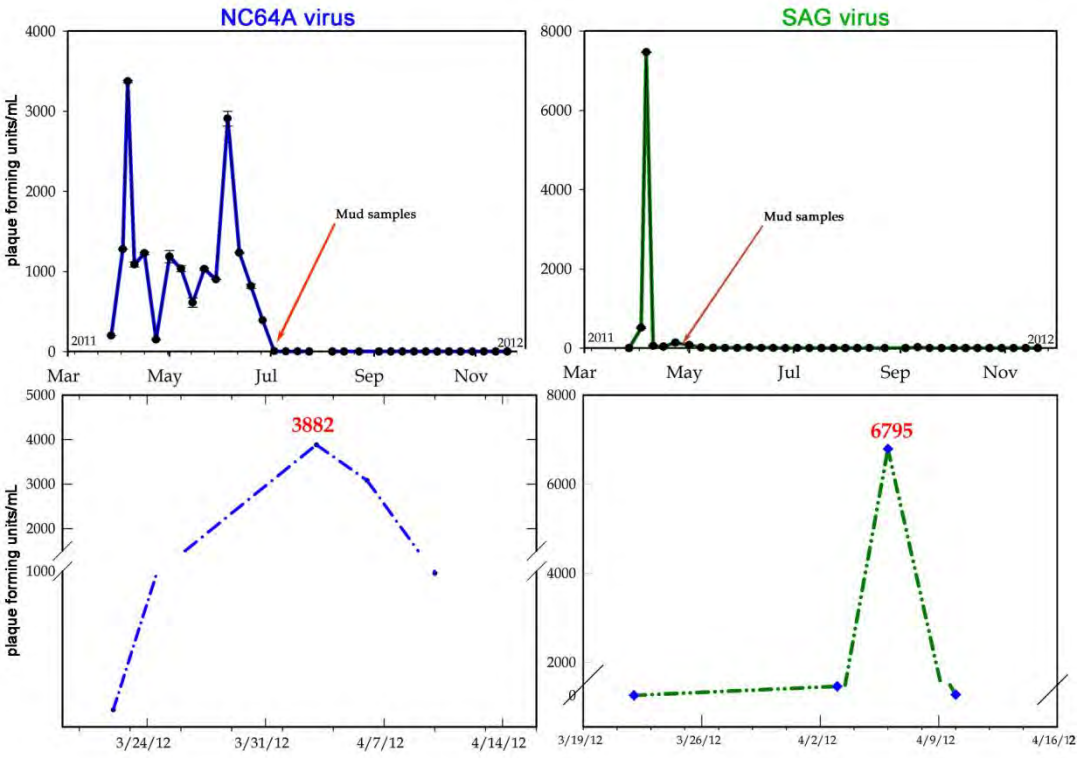


Supplementary Figure S6.

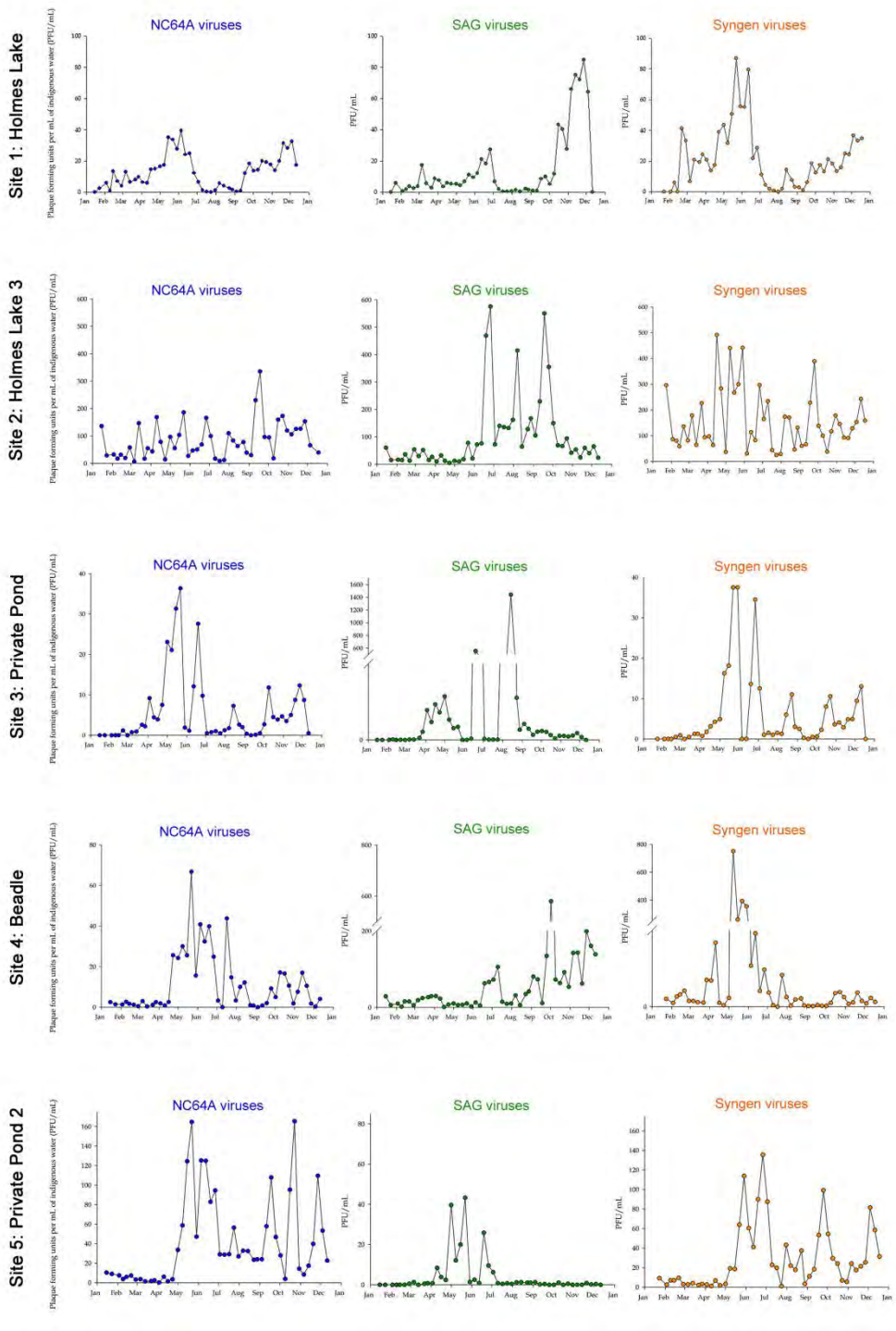
Supplementary Figure S7.



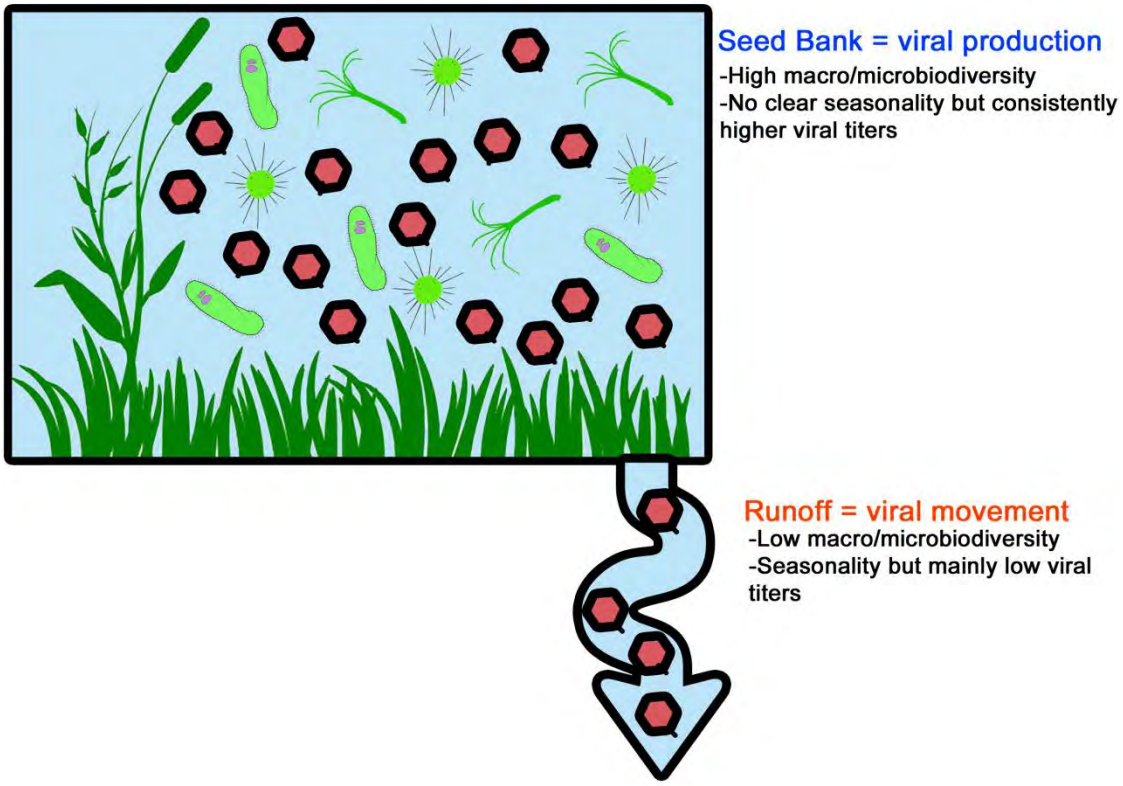
Supplementary Figure S8.



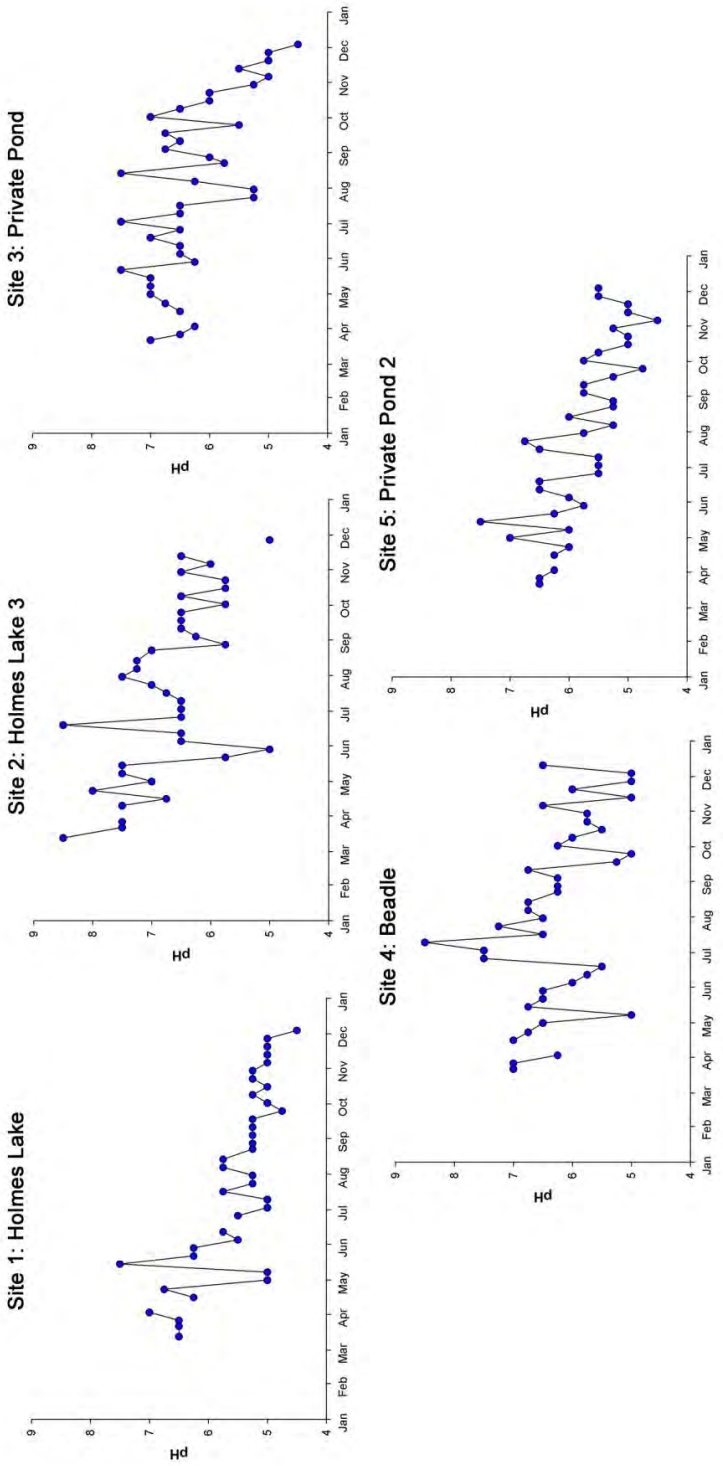
Supplementary Figure S9.



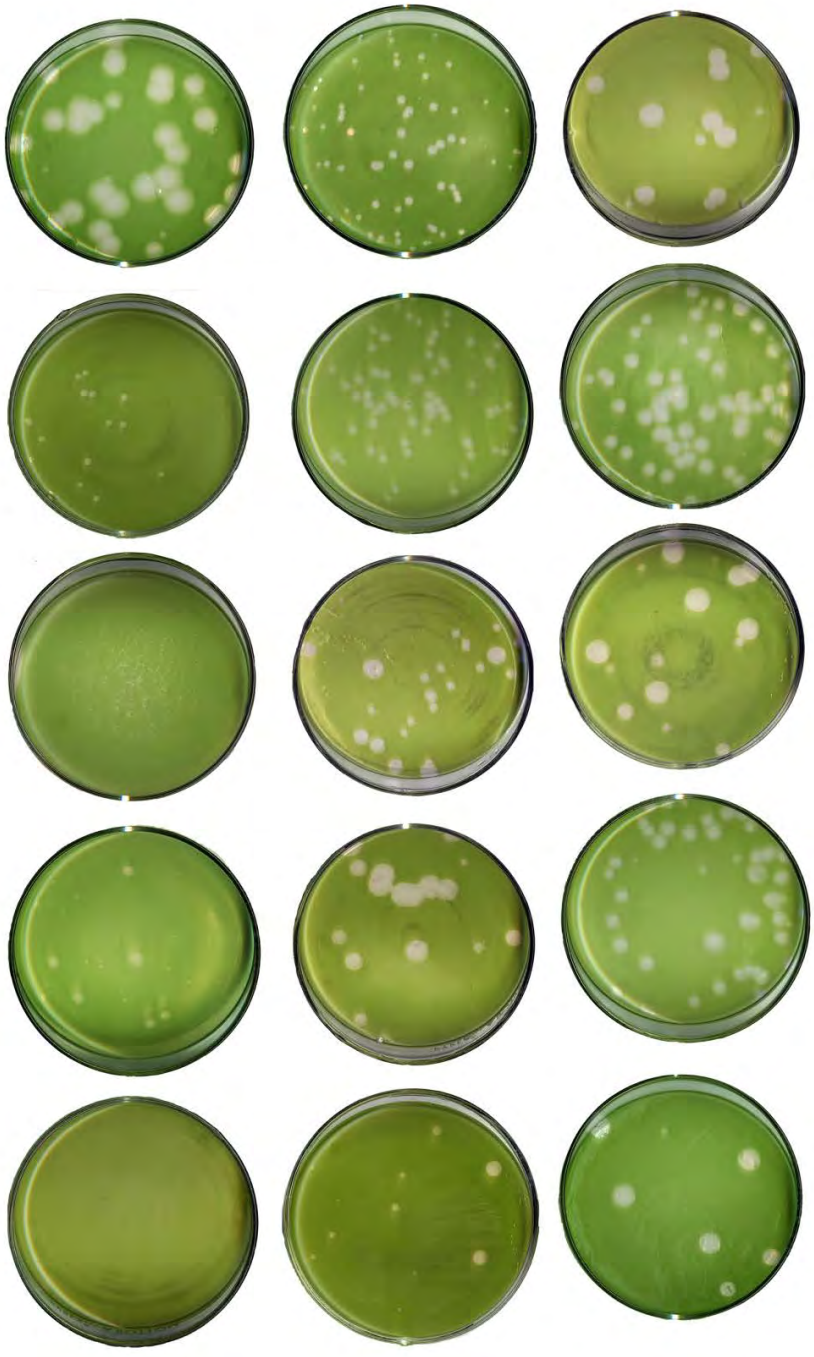
Supplementary Figure S10.



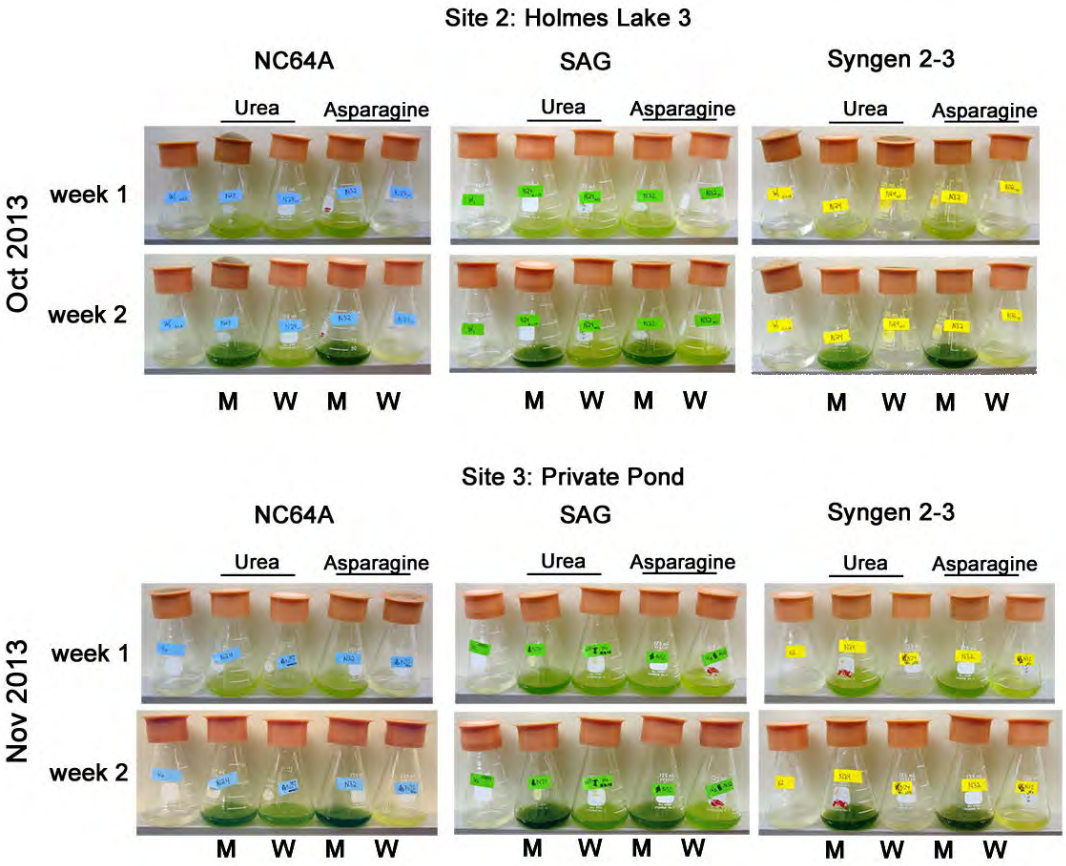
Supplementary Figure S11.



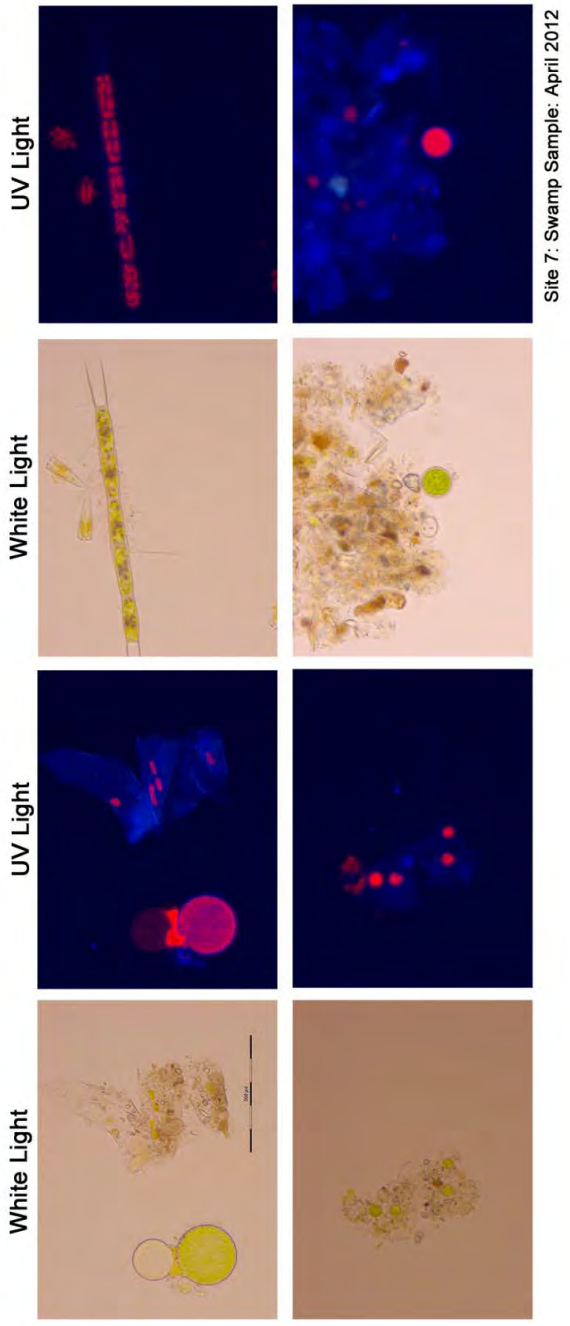
Supplementary Figure S12.



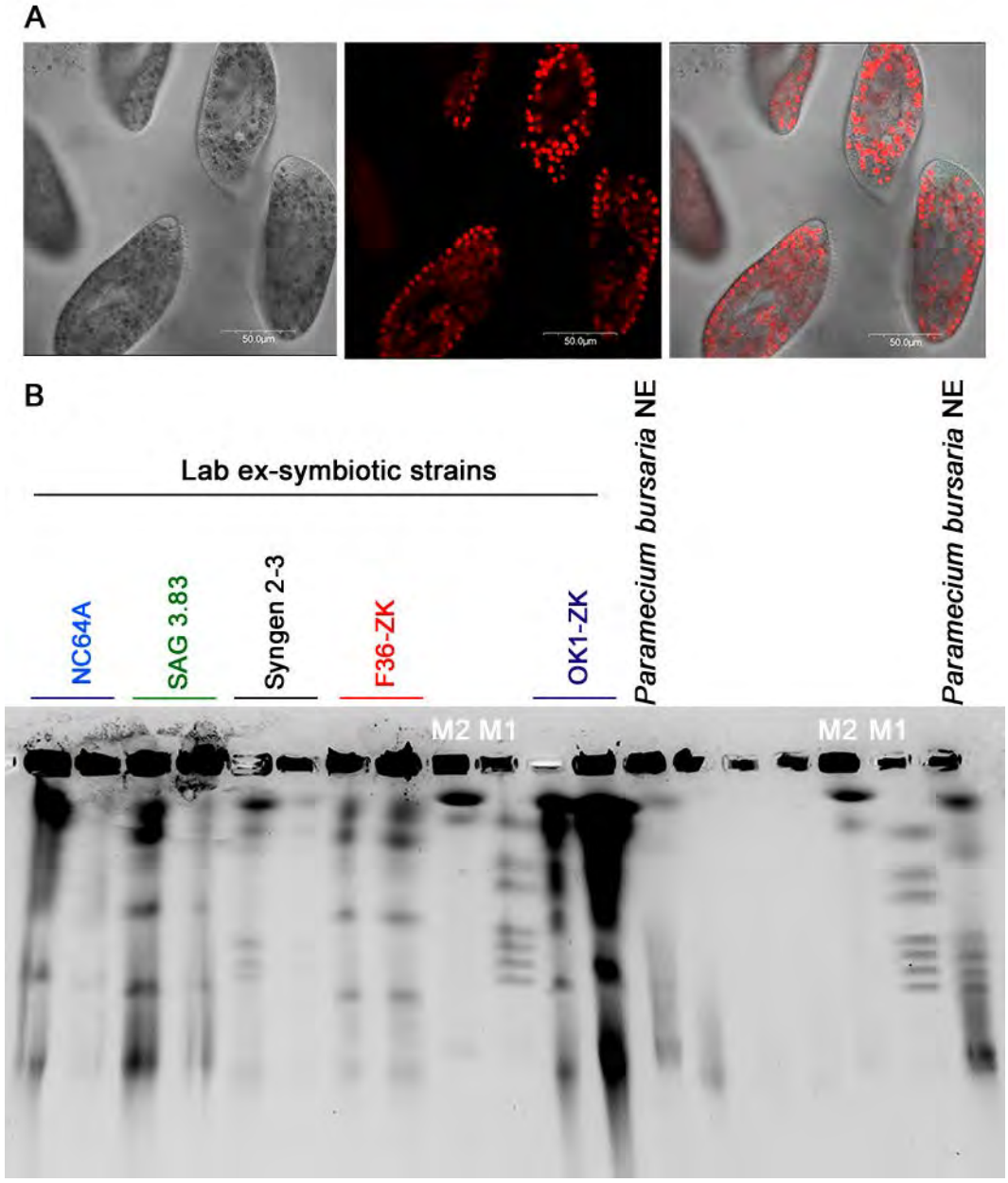
Supplementary Figure S13.



Supplementary Figure S14.



Supplementary Figure S15.



Supplementary Figure S16.

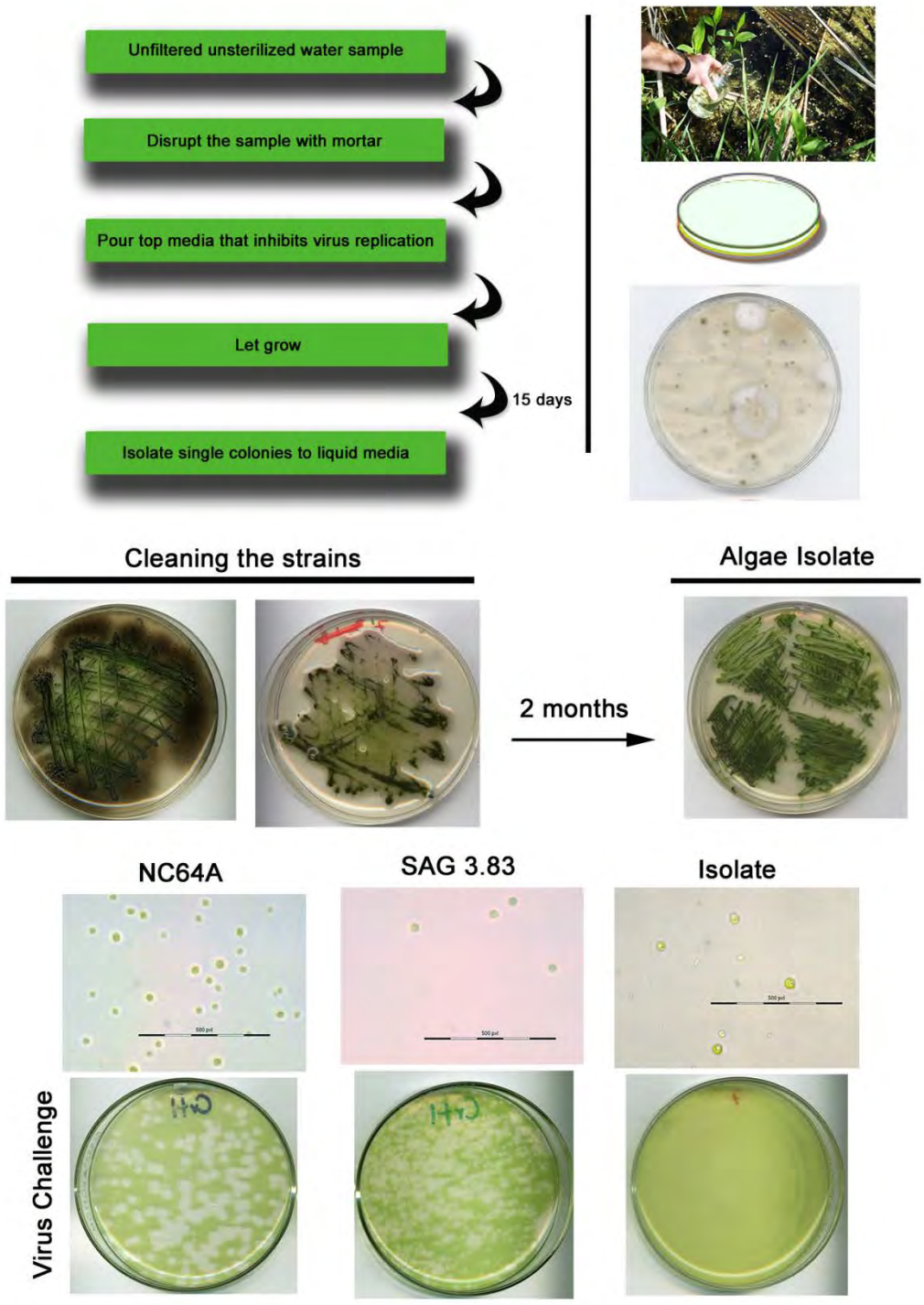


Table 1.

Date collected	Temp (oC)	Chlorophyll a (V)	ORP (mV)	DO (% sat)	pH	C (umhos/cm)	T(NTU)
May 2010	14	0.0284	339	91.4	8.56	566	1.4
July 2010	29	0.0729	209	194.1	9.41	272.6	24
Aug. 2010	29	0.0231	275	105.7	9.23	293.6	16.4
Sept. 2010	21	0.0342	417	74.1	7.69	327.2	11.8

ORD= oxygen reduction potential, DO= Disolved oxygen, C= Specific Conductance, T= Turbidity
DO= Disolved oxygen
C= Specific Conductance
T= Turbidity

CHAPTER III

EXPANSION OF THE *CHLOROVIRUS* GENUS

Characterization of a New *Chlorovirus* Type with Permissive and Non-Permissive Features on Phylogenetically Related Algae Strains

Short title: Characterization of a New Chlorovirus Type

Cristian F. Quispe[1,3], Olivia Sonderman[1,2], Michelle McQuinn[1], Irina Agarkova[1,2], Garry A. Duncan[4], and James L. Van Etten [1,2]

[1] Nebraska Center for Virology, [2] Department of Plant Pathology and [3] School of Biological Science, University of Nebraska-Lincoln, NE 68583-0900. [4] Department of Biology, Nebraska Wesleyan University, Lincoln, Nebraska, USA.

Correspondence: Cristian Quispe. School of Biological Sciences. University of Nebraska-Lincoln. 204 Morrison Center. Lincoln, NE 68583. Fax: 402.472.3323. Phone: (402) 472-5776. Email: quispecristian@gmail.com

Conflict of interest

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this paper.

Author Contributions: C.F.Q. and J.L.V.E. designed research; C.F.Q., O.S., M.M. and I.A. performed the experiments; G.A.D. assembled the viral genome; C.F.Q., I.A. and J.L.V.E. analyzed data; and C.F.Q., O.S. and J.L.V.E. wrote the paper.

Characterization of a New *Chlorovirus* Type with Permissive and Non-Permissive Features on Phylogenetically Related Algae Strains

Abstract

Chloroviruses are large, icosahedral dsDNA viruses that are ubiquitous in freshwater reaching titers as high as thousands of plaque forming units (PFU) per ml of native water. Previously, *Paramecium bursaria chlorella virus* 1 (PBCV-1) was described as the *Chlorovirus* prototype that replicates in two *Chlorella variabilis* algal strains, NC64A and Syngen 2-3. Recently, it was discovered that PBCV-1 could also replicate in the *Chlorella variabilis* OK1-ZK strain. These three strains are ex-symbionts originally isolated from the protozoan *Paramecium bursaria*. As part of a three-year systematic study to monitor chloroviruses in natural aquatic environments in Nebraska, the three strains were used for plaque assays. Surprisingly, the PFUs on Syngen 2-3 lawns were significantly higher than the PFUs on NC64A and OK1-ZK from the same indigenous samples. These unexpected discrepancies led to the discovery of viruses that only infect Syngen 2-3 cells. As a result, a new *Chlorovirus* genus named *Only Syngen* (OSy) viruses, that form plaques only on the ex-symbiotic Syngen 2-3 strain but not on NC64A or OK1-ZK lawns was discovered from native water. The *Only Syngen Nebraska virus* 5 (OSyNE-5) was selected as the prototype virus for the genus. OSyNE-5 resembled PBCV-1 and the other chloroviruses in that it had a dsDNA genome of 323-kb and had a distinct icosahedral shape with a diameter of around 180–190 nm. Interestingly, OSyNE-5 contained two major capsid

proteins that migrated slightly slower than the PBCV-1 homolog. Additionally, gene synteny, nucleotide conservation and phylogenetic affinity were highly conserved among OSyNE-5 and PBCV-1, likely because both viruses replicate in Syngen 2-3 (permissive) cells. Intriguingly, OSyNE-5 was also able to attach and initiate infection in NC64A and OK1-ZK, which resulted in the death of the algae. However, infectious particles were not recovered. Thus, OSy viruses have permissive and non-permissive features in phylogenetically related algal species.

Keywords: *Chlorovirus*, *Chlorella variabilis*, OSyNE-5, non-permissive cells, permissive cells

Introduction

Large dsDNA-containing viruses that infect algae comprise the family *Phycodnaviridae*. They have genomes ranging from 160 to 560 kb that contain up to 600 protein-encoding genes (CDSs) (Van Etten 2003; Van Etten et al. 2010; Wilson et al. 2009). These large viruses are found in aqueous environments throughout the world in both fresh and marine waters. Currently phycodnaviruses are classified into six genera. Members of one genus, *Chlorovirus,* are icosahedral, plaque-forming viruses that replicate in certain ex-symbiotic, unicellular chlorella-like green algae. Chloroviruses are cosmopolitan residents of inland waters with titers as high as thousands of plaque forming units (PFU) per ml of indigenous water (Van Etten et al. 1985a; Van Etten et al. 1985b; Van Etten et al. 2002; Yamada et al. 2006).

Chlorovirus hosts, which are normally symbionts in nature, are often referred to as zoochlorellae (Karakashian 1975; Kodama et al. 2014). They are associated with either the protozoan *Paramecium bursaria*, the coelenterate *Hydra viridis* or the heliozoon *Acanthocystis turfacea* (Van Etten and Dunigan 2012). Zoochlorellae are resistant to viruses in their symbiotic state. Fortunately, some zoochlorellae grow independently of their partners in the laboratory, permitting plaque assay of the viruses and synchronous infection of their hosts, aiding the study of the virus life cycle in detail (Van Etten et al. 1983b). Three such zoochlorellae are *Chlorella* NC64A [renamed *Chlorella variabilis* (Proschold et al.

2011), and its viruses are called NC64A viruses]; *Chlorella* SAG 3.83 (renamed *Chlorella heliozoae,* and its viruses are called SAG viruses); and *Chlorella* Pbi (renamed *Micractinium conductrix,* and its viruses are called Pbi viruses). However, little is known about the natural history of the chloroviruses but we suspect that many more chlorovirus hosts and viruses exist in nature.

Paramecium bursaria chlorella virus 1 (PBCV-1) is the type member of the genus *Chlorovirus.* PBCV-1 infects and forms plaques on two *Chlorella variabilis* strains, NC64A and Syngen 2-3 (Van Etten et al. 1983a). Recently, PBCV-1 was also shown to replicate in *C. variabilis* OK1-ZK cells (Quispe, unpublished results). The three *C. variabilis* strains are endosymbionts of the protozoan *P. bursaria*. At the time the plaque assay was developed in 1983, we assumed that chlorella strains NC64A and Syngen 2-3 were identical, and consequently we focused our studies on PBCV-1 and NC64A for the past 35 years (Van Etten and Dunigan 2012). However, a recent taxonomic study on rDNA from zoochlorellae established that NC64A, Syngen 2-3, and OK1-ZK were similar, but not identical strains (Kamako et al. 2005; Proschold et al. 2011). This report prompted us to look for viruses in native water that would plaque on Syngen 2-3 lawns. Surprisingly, as reported in this manuscript, viruses that formed plaques on Syngen 2-3 were more common in indigenous waters than viruses that formed plaques on NC64A and OK1-ZK at certain times of the year. This observation led to the hypothesis that a yet unknown chlorovirus type might only replicate in

Syngen 2-3 cells but not in the NC64A and OK1-ZK strains. This manuscript describes this new group of chloroviruses, designated *Only Syngen* viruses (OSy), which only replicate in permissive Syngen 2-3 cells. In addition, the OSy viruses can initiate infection in phylogenetically related *C. variabilis* strains (NC64A and OK1-ZK) but are unable to complete virus replication (non-permissive cells).

Results and Discussion

Difference in number of plaques from indigenous water samples on lawns of Syngen 2-3, OK1-ZK and NC64A

Chlorovirus PBCV-1 infects and forms plaques on three *C. variabilis* strains. Indigenous water samples were collected and plaque assayed on NC64A, OK1-ZK and Syngen 2-3 lawns as part of a three-year systematic study of chloroviruses in Nebraska. Unexpectedly, viruses that plaqued on Syngen 2-3 lawns were up to ten times more prevalent than viruses that plaqued on NC64A and OK1-ZK lawns (Fig. 1a and Supp. Fig. 1). Usually, samples collected earlier in the year (January to April) showed the highest differences in numbers; in contrast, the number of PFUs on NC64A and OK1-ZK lawns was very similar in all the indigenous samples tested. This observation led to the prediction that an unknown chlorovirus type replicated in Syngen 2-3 cells but not in NC64A and OK1-ZK cells. That is, Syngen 2-3 might serve as a host for two distinct virus

populations: viruses such as PBCV-1 that replicate in NC64A, OK1-ZK and Syngen 2-3 cells and viruses that only replicate in Syngen 2-3 cells.

Isolation of a new Chlorovirus type

To investigate this unexpected difference in plaque-forming viruses in the indigenous water samples, 314 single plaques were isolated from Syngen 2-3 lawns that showed the greatest difference in plaque number when compared to NC64A and OK1-ZK lawns (Supp. Fig. 2). The plaques were then inoculated on the three *C. variabilis* strains in both liquid and solid cultures (Supp. Fig. 3). Seventy-five percent of the 314 viruses only formed plaques on Syngen 2-3 cells whereas twenty-five percent formed plaques on and lysed all three strains (Fig. 1b). This new *Chlorovirus* genus was named Only Syngen viruses (OSy) because they only formed plaques on Syngen 2-3 cells (Fig. 1c and Supp. Table 1). Using the same procedure, we isolated OSy viruses from other water samples, including one sample collected in Florida, suggesting that the OSy genus is ubiquitous in geographically distant sites within the continental United States (Supp. Table 2). One isolate from Nebraska, called OSyNE-5, was selected as the prototype virus for the group and characterized further.

Morphology of OSyNE-5 virus

Electron microscope analysis indicates that OSyNE-5 has the same icosahedral morphology as chlorovirus PBCV-1 (Fig. 2 and Supp. Fig. 4) with a diameter of

180-190 nm, which is similar to the 190 nm diameter measured along the five-fold axis for PBCV-1 (Yan et al. 2000). Another distinct characteristic of chlorovirus virions is the presence of a single bilayered membrane that is derived from the host (Milrot et al. 2015). The membrane is located underneath the outer capsid shell, and it is required for virus infectivity. For instance, PBCV-1 is sensitive to a short time exposure to chloroform (Skrdla et al. 1984). Thus the effect of chloroform on the specific infectivity of OSyNE-5 was determined (results not shown). Similar to PBCV-1, OSyNE-5 infectivity was rapidly reduced by chloroform exposure, suggesting that chloroform destroyed the membrane integrity of the virion particle.

Analysis of the major capsid protein (MCP) of OSyNE-5

The PBCV-1 virion is composed of a mixture of 148 viral encoded proteins and one host-deived protein (Dunigan et al. 2012). The PBCV-1 MCP (A430L) is a 54 kDa glycoprotein that represents about 40% by weight of the total protein content in the virion and is predicted to be present in approximately 5000 copies per particle (Nandhagopal et al. 2002; Yan et al. 2000). The PBCV-1 MCP is post-translationally modified and has 4 N-glycosylation sites (Nandhagopal et al. 2002; Klose et al., unpublished results). The size of the MCPs in the other chlorovirus types ranges between 51 kDa and 55 kDa in size (DeCastro et al., in press). To determine if OSyNE-5 has comparable proteomic features, the protein compositions of OSyNE-5 and PBCV-1 were compared by SDS-PAGE (Fig. 3).

The OSyNE-5 SDS-PAGE profile resembled the PBCV-1 profile but with some differences (Supp. Fig. 5). For instance, OSyNE-5 appears to have two MCPs that migrated slightly slower than the PBCV-1 MCP with predicted molecular weights of 54 and 55 kDa. Thus OSyNE-5 resembles another chlorovirus, named NY-2A, that also may have two MCPs (DeCastro et al., in press).

General genomic comparison of OSyNE-5 and PBCV-1

The OSyNE-5 genome was sequenced, annotated and compared to PBCV-1 (Supp. Fig. 6). The genome was 323,150 bp with a G+C content of 42%. Thus, OSyNE-5 has a slightly smaller genome and a slightly increased G+C content compared to PBCV-1 (331-kb and 40% G+C content) (Supp. Fig. 7). Interestingly, all of the NC64A viruses have a G+C content of about 40%, whereas the G+C content of the Pbi and SAG viruses are 45% and 49%, respectively (Jeanniard et al. 2013).

Gene prediction algorithms identified 765 open reading frames (ORFs) in the OSyNE-5 genome (Table 5) of which 348 were classified as major CDSs and 417 as minor ORFs. We classified potential CDSs based on the parameters used to resequence and reannotate the PBCV-1 genome (Dunigan et al. 2012). (Dunigan et al. 2012). Half of the OSyNE-5 genome (52% =399 ORFs) were PBCV-1 orthologs with high e-values (Fig. 4c and Table 1). Additionally OSyNE-5 was predicted to contain 14 tRNA genes (Table 2). These features are

comparable to PBCV-1 that has 801 ORFs (416 major CDSs and 386 minor ORFs) and 11 tRNA genes (Fig. 4c) (Dunigan et al. 2012). (Dunigan et al. 2012).

A previous report of 41 sequenced chloroviruses indicated that gene synteny, nucleotide conservation and phylogenetic affinity were highly conserved among viruses infecting the same algal host, with only a few localized rearrangements (Supp. Fig. 8). In contrast, low synteny was observed between chloroviruses that infected different hosts (Jeanniard et al. 2013). At the genome level, OSyNE-5 and PBCV-1 exhibited high synteny despite the fact that PBCV-1 formed plaques on NC64A while OSyNE-5 did not (Fig. 4 a,b). We speculate that the ability of both viruses to replicate in Syngen 2-3 cells could explain their high genomic colinearity.

Phylogenetic analysis

Phylogenetic relationships between OSyNE-5, other chlorovirus and *Ostreococcus* virus (outgroup) genomes were determined using previous analysis of the concatenated alignment of core chlorovirus genes (Jeanniard et al. 2013). The resulting maximum likelihood (ML) phylogenetic tree is presented in Fig. 5. While most of the newly sequenced OSyNE-5 genome is a close relative of previously sequenced NC64A viruses including PBCV-1, the isolated phylogenetic position of virus OSyNE-5 within the NC64A virus clade makes it the first representative of this new OSy genus of chloroviruses. This new virus

genus resides within the two separate phylogenetic sub-groups of NC64A viruses –one contains PBCV-1 and the other NY-2A. Both sub-groups share almost perfect gene colinearity since they replicate in the same host. Similarly, OSy and NC64A viruses replicate on Syngen 2-3 cells. Thus, they are clustered together likely because viruses infecting the same algal host always cluster in monophyletic clades (Jeanniard et al. 2013). Consequently, the 29 core proteins identified in the OSyNE-5 genome to create the phylogeny share a high (average of 89%) amino acid identity with their PBCV-1 orthologs (Table 3). In comparison, the protein sequence identity between clades of chlorovirus genus ranged from 63.1% (NC64A vs. Pbi viruses) to 70.6% (Pbi vs. SAG viruses) (Jeanniard et al. 2013). Thus, the molecular phylogenetic analysis indicates a high phylogenetic affinity between the OSy and the NC64A viruses.

OSyNE-5 virus attaches to non-permissive cells NC64A and OK1-ZK

One of the PBCV-1 icosahedron vertices has a spike structure; in addition, several fibers extending from the capsid are believed to play a role in virus attachment to the host (Cherrier et al. 2009; Van Etten et al. 1991; Zhang et al. 2011). Attachment of PBCV-1 and the other chloroviruses is host specific; thus, virus attachment is the major factor in limiting the host range of the chloroviruses (Van Etten and Dunigan 2012). In permissive cells, attachment leads to host cell wall degradation at the point of attachment by a virus-associated enzyme(s); the viral internal membrane presumably fuses with the host membrane, facilitating

entry of the viral DNA and virion-associated proteins into the cell (Thiel et al. 2010). An empty capsid remains attached to the algal surface (Meints et al. 1984) while virus infection causes rapid depolarization of the host plasma membrane that leads to the inhibition of secondary active transporters, consequently altering cellular solute uptake (Agarkova et al. 2008). During a successful viral infection DNA dyes such as SYBR® gold rapidly diffuse inside the infected cell and cause DNA to fluoresce yellow after UV exposure. The strong genome synteny and high phylogenetic affinity between OSyNE-5 and PBCV-1 led to speculation about possible interactions of OSyNE-5 on PBCV-1's host range. Thus, we conducted comparative infection studies between OSyNE-5 and PBCV-1 on NC64A, OK1-ZK and Syngen 2-3 cells, initially focusing on virus attachment.

Fluorescent microscopy and flow cytometry analyses using the SYBR® Gold stain were performed to determine if OSyNE-5 attaches to Syngen 2-3, OK1-ZK and NC64A cells. PBCV-1 and an SAG virus (ATCV-1) served as positive and negative controls, respectively (Supp. Fig. 9). Purified OSyNE-5, PBCV-1 and ATCV-1 viruses and actively growing NC64A, Syngen 2-3 and OK1-ZK (1×10^6 cells/ml) cells were used. Cultures were infected at MOI of 10, and 1 hr pi cells were mixed with the SYBR® Gold stain and analyzed by fluorescent microscopy and flow cytometry.

The fluorescent microscopy analysis showed that the three chlorella strains, both alone or after mixing with the non-host ATCV-1 virus, were negative for DNA staining. Syngen 2-3 cells infected with either OSyNE-5 or PBCV-1 showed rapid SYBR® Gold uptake and consequently positive DNA fluorescent staining at the point of viral infection. Surprisingly, *C. variabilis* NC64A and OK1-ZK cells were also positive for DNA staining after OSyNE-5 attachment (Fig. 6a and Supp. Fig. 10).

To corroborate and quantify our results, flow cytometry analysis was performed following a similar procedure. Samples of infected, uninfected and control cells were run at 1 x 10^6 cells/ml and approximately 1 x 10^4 cell events were collected per sample. Similar to the fluorescent microscopy analysis, OSyNE-5 significantly increased the population of cells with higher SYBR® Gold fluorescent intensity compared to uninfected cells (Fig. 6b). The increase in the OSyNE-5-induced fluorescent was similar to that observed in PBCV-1 on Syngen 2-3, NC64A and OK1-ZK. Background staining accounted for only 10% of the total cell population across the strains. Consequently, OSyNE-5 attached and initiated infection not only in Syngen 2-3, but also on NC64A and OK1-ZK cells. Additionally, we determined that OSyNE-5 did not attach or initiate infection on other *Chlorella* ex-symbiotic strains, e.g., SAG 3.83, Pbi, and F36-ZK (data not shown).

Non-permissive cells pre-challenged with OSyNE-5 virus avoided secondary PBCV-1 infection

Chlorovirus infection in permissive cells prevents infection by a second chlorovirus (Greiner et al. 2009). To determine if OSyNE-5 infection on non-permissive cells inhibited subsequent PBCV-1 infection, actively growing Syngen 2-3, NC64A and OK1-ZK cells at 1 x 10^7 cells/ml were inoculated with either OSyNE-5, PBCV-1 or ATCV-1 virus at an MOI of 10 for 30 min (Supp. Fig. 12). Cells were pelleted by centrifugation, culture supernatants discarded, and the cells were washed and resuspended two times in MBBM. Cells were then challenged with PBCV-1 at an MOI of 0.01 and lysates plaque assayed on NC64A cells 96 hrs post infection (pi). First, we established that OSyNE-5 infection on permissive (Syngen 2-3) cells prevented subsequent PBCV-1 infection. Interestingly, similar to the scenario on permissive cells, OSyNE-5 infection on non-permissive (NC64A or OK1-ZK) cells also prevented the secondary PBCV-1 infection. Thus, OSyNE-5 prevents subsequent PBCV-1 infection on permissive and non-permissive cells (Fig. 7). Control experiments with ATCV-1 or PBCV-1 primary infection followed by the secondary PBCV-1 infection produced the expected results.

OSyNE-5 attachment kills non-permissive cells

The preceding experiments established that OSyNE-5 interaction with non-permissive cells (NC64A and OK1-ZK) results in initiation of infection but no virus replication. This leads to the question: what is the fate of non-permissive cells after exposure to OSyNE-5? Thus, we inoculated Syngen 2-3, NC64A and OK1-

ZK cells ($5x10^5$ cells/ml) with OSyNE-5 at low (0.01) and high (20) MOIs (Supp. Fig. 11). Ten minutes pi we transferred cells to 25 ml of MBBM. Uninfected cells were used as controls. At seven days pi, cell viability was evaluated by visualizing cell culture growth. OSyNE-5 completely lysed the cell cultures at both low and high MOIs in permissive cells (Syngen 2-3). Likewise, there was no cell growth in non-permissive cells (NC64A and OK1-ZK) inoculated at high MOI with OSyNE-5, suggesting that all cells were killed after infection at high MOI. In contrast, cell growth similar to uninfected controls was observed in non-permissive cells (NC64A and OK1-ZK) infected with OSyNE-5 at low MOI (Fig. 8). Consequently, the non-permissive cells were killed after addition of OSyNE-5 at high MOI even though no virus replication occurred.

OSyNE-5 infection leads to host nuclear DNA degradation in permissive and non-permissive cells

C. variabilis NC64A genomic DNA is distributed into 13 chromosomes that range in size from 1.1 to 6.5 Mb (Blanc et al. 2010). NC64A nuclear DNA begins to be degraded to 150 to 200 kb segments by PBCV-1 encoded and virion packaged DNA restriction endonucleases within minutes after PBCV-1 infection (Agarkova et al. 2006). In contrast, the infecting 331-kb PBCV-1 DNA, which is methylated in the restriction sites, remains intact during infection. The observation that OSyNE-5 initiates infection in NC64A and OK-ZK-1 cells without completing its replication cycle prompted us to examine NC64A, OK1-ZK and Syngen 2-3

chromosomal DNA integrity by pulsed field gel electrophoresis (PFGE) after OSyNE-5 inoculation. Actively growing permissive and non-permissive cultures were infected with OSyNE-5 at an MOI of 10; PBCV-1 was used as a control. We observed that DNA degradation patterns in the three *C. variabilis* strains following OSyNE-5 inoculation were similar to those observed with PBCV-1, although the kinetics were slightly slower (Fig. 9). Thus, even though OSyNE-5 cannot complete its replication cycle in NC64A or OK1-ZK cells, host DNA degradation occurs, suggesting that at least some early and/or early/late transcripts may be synthesized (Supp. Fig. 13). It was difficult to conclusively establish if OSyNE-5 host DNA degradation was followed by viral DNA accumulation in later time points.

Genomic information present only in the OSyNE-5 genome

The preceding experiments indicate that OSyNE-5 can initiate replication in NC64A and OK1-ZK cells but is unable to complete its replication. As noted in Fig. 4, there are similarities in the synteny between PBCV-1 and OSyNE-5 genomes. However, one or more gene differences between the two viruses must explain the discrepancies in their ability to infect and complete viral replication in NC64A or OK1-ZK cells. The viruses share 399 ORFs (Table 1). They are scattered in the genome and include 305 major CDSs and 94 minor ORFs. There are 366 ORFs present only in the OSyNE-5 genome, and most of them did not have predicted functions (Table 4). They include 43 major CDSs (protein length

between 150-877 aa) and 323 minor ORFs. In contrast, 403 ORFs are present only in the PBCV-1 genome, and they include 97 major CDSs and 306 minor ORFs (Fig. 4c). Thus, while the majority of major CDSs are shared between both viruses, minor ORFs are divergent and probably important for the differential host permissibility of both viruses.

Additionally, we closely compared three regions that were present exclusively in the OSyNE-5 genome (labeled as a', b' and c' on Fig. 4a) to search for specific genetic differences. An inverted stretch of 20-kb from 71- to 91-kb (group a') within the OSyNE-5 genome was absent in the PBCV-1 genome. Additionally, two regions (groups b' and c') include the stretch between 125- and 130-kb and 310- to 315-kb respectively. Blast hits for these three regions are summarized in Table 6. Most hits have low protein identity (<61% averaged) after searching using the blastp algorithm. These sequences were completely absent in most NC64A viruses but present in some SAG or Pbi genome sequences.

The left most region between 0- to 20-kb in the PBCV-1 genome did not resemble any sequences in OSyNE-5 (Fig. 4). This last genomic difference, although interesting, may not explain the difference in host permissibility between the two viruses because a NC64A virus KS1B isolated in Kansas, USA contained a 35-kb deletion in the left end of the genome when compared to PBCV-1 (Jeanniard et al. 2013). Thus, a 35-kb section in the left end of the PBCV-1

genome that encompasses 29 CDSs is dispensable in a natural environment. In addition, spontaneous antigenic variants of PBCV-1 containing 27- to 37-kb deletions (12% of the PBCV-1 genome) in the left end of the 331-kb genome replicate in *C. variabilis* NC64A and Syngen 2-3 cells in laboratory conditions, albeit not as successfully as the wild type PBCV-1 (Landstein et al. 1995; Quispe, unpublished results).

Conclusions

This manuscript describes the identification of OSy viruses, a new group of chloroviruses that infects *C. variabilis* Syngen 2-3. These viruses are most closely related at the genomic and phylogenetic level to the NC64A viruses. Interestingly, the OSy viruses can initiate infection, including carrying out many of the early infection events in *C. variabilis* (NC64A and OK1-ZK) cells, but they are unable to complete virus replication (non-permissive cells). In contrast, the NC64A viruses are able to infect and complete replication in three permissive *C. variabilis* strains (NC64A, Syngen 2-3, and OK1-ZK). All of our previous studies on the isolation of chlorella cells that were resistant to the chloroviruses involved the loss of the ability of the viruses to attach to the cells, presumably due to a change in the receptor. However, the finding in this report that OSy viruses can initiate infection in NC64A and OK1-ZK cells and not complete replication opens up a new area of investigation. They are blocked at later stage(s) of infection and we are actively trying to determine where the block is located.

Materials and methods

Cell cultures and virus purification

C. variabilis NC64A and Syngen 2-3 were maintained as slant stocks at 4°C. *C. variabilis* OK1-ZK (NIES-2541) was obtained from the Japanese Culture Collection of the National Institute for Environmental Studies (http://www.nies.go.jp/index-e.html). NC64A, Syngen 2-3 and *OK1-ZK* strains were grown on Modified Bold's Basal Medium with 1% thiamine (V/V) (Complete-MBBM) (Bischoff, H. & Bold, H. 1963; Van Etten et al. 1983a). All experiments were performed with cells growing at early log phase (4 - 7 x 10^6 cells/ml). Cell cultures were shaken (200 rpm) at 26^{o}C under continuous light. Procedures for producing and purifying chloroviruses have been described previously (Agarkova, et al. 2006; Dunigan et al. 2012; Van Etten et al. 1983b).

Plaque assays

Water samples were analyzed by plaque assay on the three *Chlorella variabilis* strains. Plaque assays were performed as previously described (Van Etten et al. 1983a) with minor modifications. The strains were grown to early log phase (4 - 7 x 10^6 cells/ml) and concentrated tenfold (4 - 7 x 10^7 cells/ml) by centrifugation for the plaque assays. Three ml of MBBM top agar (7 g/L agar) were mixed with 300 ul of concentrated actively growing cells and the water sample. Adequate amounts of 0.45 um filtered water samples were plated to produce 25 - 120 plaques/plate when possible. The samples were poured over solidified MBBM-

containing agar (15 g/L). Plates were then incubated for several days in constant light at 26^{o}C and plaque averages were determined from four plates per sample/strain.

Isolation of OSy viruses

Plaque assays of indigenous water samples with significantly higher numbers of plaques in Syngen 2-3 compared to NC64A and OK1-ZK were selected. A total of 314 single plaques were isolated from Syngen 2-3 lawns and transferred to liquid cultures of Syngen 2-3, OK1-ZK and NC64A cells to amplify the viruses. After incubating in MBBM at 26^{o}C in continuous light for seven days, tubes were centrifuged at 5000 rpm for 5 min to pellet fragments and whole cells. A clear tube indicated lysis whereas a green pellet of intact algae cells suggested no infection. Viruses that lysed Syngen 2-3 cells without lysing NC64A or OK1-ZK were selected. These viruses were diluted and plaque assayed on the three strains. Viruses that formed plaques only on Syngen 2-3 lawns were selected and re-plaqued at least two times then amplified in liquid culture for final virus purification. Lysate and purified viruses were stored at 4°C in glass vials.

Transmission electron microscopy

For electron microscopic studies, freshly purified virus particles were added to a strip of parafilm and transferred to 400-mesh copper grids (Electron Microscopy Sciences) supported by carbon-coated Formvar film. The grid was then

submerged in buffer suspension for 3 min and air-dried for 60 sec. Negative staining solution (2% w/v aqueous phosphate tungsten acid, pH 7.2) (1:1 ratio) was added for 5 min and air-dried for 1 h at room temperature. Virus particles were visualized uisng a Hitachi H7500 transmission electron microscope in the Morrison Microscopy Core Research Facility at the University of Nebraska-Lincoln.

SDS-PAGE gel analysis

Virion proteins were solubilized and separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE profiles of OSyNE-5 and PBCV-1 were performed using 15 µg of protein extracted from freshly purified particles and run on a 4-20% Tris-Glycine PAGE® Gold Precast Protein Gel before silver staining.

Chloroform-isoamyl alcohol DNA isolation

DNA was isolated as previously described (Doyle 1987) with modifications. Five hundred µl of freshly purified virus particles ($1x10^{10}$ PFU/ml) were mixed with 9 µl DNAse I (Invitrogen 153 U/µl) and incubated at room temperature for 60 min to remove external DNA molecules. Then, 20 µl of 500 mM EDTA (Sigma, pH 8.0), 20 µl of proteinase K (20 mg/ml), 4 µl of 30% Na sarcosyl and 2 µl of calcium acetate (1M, Sigma) were added to the sample and vortexed briefly. Samples were incubated at 65^{o}C for 30 min. Three hundred µl of buffer-saturated phenol

and 300 µl of chloroform-isoamyl alcohol (24:1) were added to the sample and gently mixed by inversion. Tubes were centrifuged at maximum speed for 5 min at 4^{o}C, and the upper aqueous layer was transferred to a fresh tube. Six hundred µl of chloroform-isoamyl alcohol (24:1) were added to the tube and mixed by inversion before centrifuging at maximum speed for 5 min at 4^{o}C. The upper aqueous layer was removed and exposed to another 600 µl of chloroform-isoamyl alcohol. DNA was precipitated from the aqueous layer by adding 66 µl of 3 M sodium acetate and 1350 µl of cold 100% ethanol, mixed and held at -20^{o}C for 3 h. The tubes were centrifuged at maximum speed for 15 min at 4^{o}C to pellet the DNA. Pellets were washed once with 1 ml of cold 70% ethanol and centrifuged for 5 min at 4^{o}C. Supernatant was removed and tubes were dried in a vacuum desiccator. Finally, 300 µl of TE buffer (100 mM Tris, 10 mM EDTA, pH 7.4) were added to the pellet and incubated overnight at room temperature to re-suspend the DNA. DNA was evaluated for quantity and quality by measuring absorbance at 260 and 280 nm with a Thermo Scientific NanoDrop 2000 spectrophotometer. DNA was stored at 4^{o}C.

Pulsed-field gel electrophoresis (PFGE)

The genome size of the OSyNE-5 was estimated by PFGE. Studies were carried out according to the procedure of Agarkova et al. 2006 with some adjustments. An equal volume of 2% low-melting-point agarose (Bio-Rad) in suspension buffer (SB) (25 mM Tris, pH 7.5, 20 mM EDTA) at 45°C and freshly purified 1 x 10^{7} virus particles were

poured into plug molds (Bio-Rad, Hercules, CA), and placed for cooling at 4°C for 20 min. After solidification, agarose blocks were incubated in 2 ml digestion buffer (DB) (250 mM EDTA, pH 9.5, 1% *N*-lauroylsarcosine and 1 mg/ml proteinase K) for 24 h. Samples were rinsed two times for 30 min with DB (without proteinase K) and cut into small pieces to fit into gel wells. Agarose blocks inside the wells were sealed with 1% low-melting-point agarose at 45°C in electrophoresis buffer. Intact viral DNAs were separated in a CHEF-DR II (Bio-Rad) unit in 1% 0.5X TBE (45 mM Tris-base, 45 mM boric acid, 1 mM EDTA, pH 8.0) agarose gel. Electrophoresis conditions included a pulse time ramped from 40 to 80 sec for 24 h at 200 V. *Saccharomyces cerevisiae* chromosomes (225 to 1,900 kb) (New England BioLabs, Beverly, MA) were used as DNA size markers. Gels were stained with 0.5 µg/ml ethidium bromide for 30 min and destained in water for 1 h. Images were taken with a ChemiDoc EQ system (Bio-Rad). PFGE studies to evaluate host DNA degradation on Chlorella cells was described previously (Agarkova et al., 2006).

Confocal Fluorescent microscopy

An Axio Imager A1 confocal fluorescent microscope was used with 20x and 40x objectives. SYBR® Gold Stain was visualized by excitation with a mercury vapor short arc lamp HBO®.

NC64A, Syngen 2-3 and OK1-ZK cells were grown to early log phase (4 - 7 x 10^6 cells/ml) and concentrated ten fold. 300 µl of cells were mixed with 30 µl of SYBR® Gold Stain (10X) following manufacturer's instructions. Cells were aliquoted in 30 µl and infected with 10 µl of purified virus at an MOI of 20. After

30 min, cells were visualized under the microscope. Pictures were taken with a ProgRes® C14plus camera.

Genomic library preparation and sequencing

Genomic libraries were prepared using the Nextera XT library prep kit and 1 ng of genomic DNA. DNA was fragmented and tagged with sequencing adapters in a single step reaction enabling dual-indexed sequencing of pooled libraries. Libraries were multiplexed, pooled and denatured following manufacturer's protocol. The sequencing was done in one lane of 100 bp paired ends run on illumina HiSeq 2500 using illumina TrueSeq Rapid PE Cluster Kit and TrueSeq Rapid SBS Kit. The Illumina Sequence Analysis Viewer monitored the quality scores of the run.

Sequence assembly, gene prediction and annotation

Assembly was performed using the CLC Genomics Workbench version 8.5. Reads were trimmed and aligned to the PBCV-1 reference genome using guided assembly. Duplicated reads were removed. ORF prediction and annotation were performed and their sequences extracted from the genome. Sequences were compared using the blastx algorithm against the annotated proteins of the PBCV-1 genome (max e-value 1e-20). Annotation for the OSYNE-5 gene set were performed by taking both the best PBCV-1, and the best Swissprot hit, and the

function of the single best BLAST hit was assigned to the ORF. Transfer RNAs were predicted using the tRNAscan-SE 1.21 software (Schattner et al. 2005).

Phylogenetic analysis

The evolutionary history was inferred by using the Maximum Likelihood (ML) method based on the JTT matrix-based model (Jones et al. 1992). The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 47 core-viral concatenated amino acid sequences. Forty-three are chlorovirus sequences and four are ostreococcus viral sequences (out group). All positions containing gaps and missing data were eliminated. There were a total of 7762 positions in the final dataset. Evolutionary analyses were conducted in MEGA 6.0 software (Tamura et al. 2013).

Flow cytometry analysis

Viral attachment was analyzed by flow cytometry in the Morrison Core Research Facility at the University of Nebraska-Lincoln using BD FACSCalibur for acquisition and FlowJo software for data analysis. Samples were run at 1 x 10^6 cells/ml, and approximately 1 x 10^4 cell events were collected per sample.

SYBR® Gold Stain was diluted in Tris buffer (50 mM Tris, 10 mM EDTA, pH 7.4) and mixed with uninfected and infected cells following manufacturer's instructions. Chorella cells were gated based on light scatter properties and analyzed in the FL-1 channel for SYBR® Gold stain intensity (488 nm excitation, 530/30 nm emission). Histograms are representative of 3 independent experiments.

Acknowledgements

Funding for this work was partially provided by the NSF-EPSCoR grant EPS-1004094 (JVE), the COBRE program of the National Center for Research Resources Grant P20-RR15535 (JVE), the National Center for Research Resources grant 5P20RR016469 (GAD), and the National Institute for General Medical Science grant 8P20GM103427 (GAD). OS was supported by UNL UCARE and ARD scholarships, and MM was a UNL volunteer in the lab. We thank Danielle Shae and Han Chen in the Morrison Core Research Facility and David D. Dunigan at UNL for technical assistance. Additionally, we would like to acknowledge the UNMC High-Throughput DNA Sequencing and Genotyping Core Facility for sequencing the OSy genomes.

References

Agarkova I, Dunigan D, Gurnon J, Greiner T, Barres J, Thiel G, Van Etten JL. 2008. Chlorovirus-mediated membrane depolarization of chlorella alters secondary active transport of solutes. Journal of Virology 82(24):12181-90.

Agarkova IV, Dunigan DD, Van Etten JL. 2006. Virion-associated restriction endonucleases of chloroviruses. Journal of Virology 80(16):8114-23.

Bischoff, H. W. & Bold, H. C. 1963. Phycological studies IV. some soil algae from enchanted rock and related algal species. No. 6318 ed. Austin, Texas.: University of Texas Publication.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, et al. 2010. The chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. The Plant Cell 22(9):2943-55.

Cherrier MV, Kostyuchenko VA, Xiao C, Bowman VD, Battisti AJ, Yan X, Chipman PR, Baker TS, Van Etten JL, Rossmann MG. 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. Proceedings of the National Academy of Sciences of the United States of America 106(27):11085-9.

Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-5.

Dunigan DD, Cerny RL, Bauman AT, Roach JC, Lane LC, Agarkova IV, Wulser K, Yanai-Balser GM, Gurnon JR, Vitek JC, et al. 2012. Paramecium bursaria chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. Journal of Virology 86(16):8821-34.

Greiner T, Frohns F, Kang M, Van Etten JL, Kasmann A, Moroni A, Hertel B, Thiel G. 2009. Chlorella viruses prevent multiple infections by depolarizing the host membrane. The Journal of General Virology 90(Pt 8):2033-9.

Jeanniard A, Dunigan DD, Gurnon JR, Agarkova IV, Kang M, Vitek J, Duncan G, McClung OW, Larsen M, Claverie JM, et al. 2013. Towards defining the chloroviruses: A genomic journey through a genus of large DNA viruses. BMC Genomics 14:158.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences : CABIOS 8(3):275-82.

Kamako S, Hoshina R, Ueno S, Imamura N. 2005. Establishment of axenic endosymbiotic strains of japanese paramecium bursaria and the utilization of

carbohydrate and nitrogen compounds by the isolated algae. Eur J Protistol 41(3):193-202.

Karakashian MW. 1975. Symbiosis in paramecium bursaria. Symp Soc Exp Biol (29)(29):145-73.

Kodama Y, Suzuki H, Dohra H, Sugii M, Kitazume T, Yamaguchi K, Shigenobu S, Fujishima M. 2014. Comparison of gene expression of paramecium bursaria with and without chlorella variabilis symbionts. BMC Genomics 15:183,2164-15-183.

Landstein D, Burbank DE, Nietfeldt JW, Van Etten JL. 1995. Large deletions in antigenic variants of the chlorella virus PBCV-1. Virology 214(2):413-20.

Meints RH, Lee K, Burbank DE, Van Etten JL. 1984. Infection of a chlorella-like alga with the virus, PBCV-1: Ultrastructural studies. Virology 138(2):341-6.

Milrot E, Mutsafi Y, Fridmann-Sirkis Y, Shimoni E, Rechav K, Gurnon JR, Van Etten JL, Minsky A. 2015. Virus?host interactions: Insights from the replication cycle of the large paramecium bursaria chlorella virus. Cell Microbiol :n/a,n/a.

Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL, Rossmann MG. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proceedings of the National Academy of Sciences of the United States of America 99(23):14758-63.

Proschold T, Darienko T, Silva PC, Reisser W, Krienitz L. 2011. The systematics of zoochlorella revisited employing an integrative approach. Environ Microbiol 13(2):350-64.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research 33(suppl 2):W686-9.

Skrdla MP, Burbank DE, Xia Y, Meints RH, Van Etten JL. 1984. Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. Virology 135(2):308-15.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725-9.

Thiel G, Moroni A, Dunigan D, Van Etten JL. 2010. Initial events associated with virus PBCV-1 infection of chlorella NC64A. Progress in Botany Fortschritte Der Botanik 71(3):169-83.

Van Etten JL, Burbank DE, Kuczmarski D, Meints RH. 1983a. Virus infection of culturable chlorella-like algae and dlevelopment of a plaque assay. Science 219(4587):994-6.

Van Etten JL. 2003. Unusual life style of giant chlorella viruses. Annual Review of Genetics 37:153-95.

Van Etten JL and Dunigan DD. 2012. Chloroviruses: Not your everyday plant virus. Trends in Plant Science 17(1):1-8.

Van Etten JL, Lane LC, Dunigan DD. 2010. DNA viruses: The really big ones (giruses). Annual Review of Microbiology 64:83-99.

Van Etten JL, Lane LC, Meints RH. 1991. Viruses and viruslike particles of eukaryotic algae. Microbiological Reviews 55(4):586-620.

Van Etten JL, Burbank DE, Schuster AM, Meints RH. 1985a. Lytic viruses infecting a chlorella-like alga. Virology 140(1):135-43.

Van Etten JL, Van Etten CH, Johnson JK, Burbank DE. 1985b. A survey for viruses from fresh water that infect a eucaryotic chlorella-like green alga. Applied and Environmental Microbiology 49(5):1326-8.

Van Etten JL, Burbank DE, Xia Y, Meints RH. 1983b. Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. Virology 126(1):117-25.

Van Etten JL, Graves MV, Muller DG, Boland W, Delaroque N. 2002. Phycodnaviridae--large DNA algal viruses. Archives of Virology 147(8):1479-516.

Wilson WH, Van Etten JL, Allen MJ. 2009. The phycodnaviridae: The story of how tiny giants rule the world. Current Topics in Microbiology and Immunology 328:1-42.

Yamada T, Onimatsu H, Van Etten JL. 2006. Chlorella viruses. Advances in Virus Research 66:293-336.

Yan X, Olson NH, Van Etten JL, Bergoin M, Rossmann MG, Baker TS. 2000. Structure and assembly of large lipid-containing dsDNA viruses. Nature Structural Biology 7(2):101-3.

Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Van Etten JL, Rossmann MG. 2011. Three-dimensional structure and function of the paramecium bursaria chlorella virus capsid. Proceedings of the National Academy of Sciences of the United States of America 108(36):14837-42.

Figure Legends

Figure 1. (a) Three independent inland water samples collected from different sites in Lincoln, Nebraska that show significantly higher plaque numbers on *C. variabilis* Syngen 2-3 lawns compared to *C. variabilis* NC64A and OK1-ZK lawns. (b) Distribution of isolates between NC64A and OSy viruses. Three hundred and fourteen plaques were isolated; 75% belong to the OSy and 25% to the NC64A genus highlighting the ubiquity and abundance of OSy viruses in Lincoln inland waters. (c) Plaque assay of OSyNE-5 and PBCV-1 viruses on NC64A, Syngen 2-3 and OK1-ZK cells. OSyNE-5 only forms plaques on Syngen 2-3 cells.

Figure 2. Electron micrographs of OSyNE-5 and PBCV-1 after negative staining of purified viral particles. Pictures reveal similarities in shape (icosahedral morphology) and size (140-190 nm) between OSyNE-5 and PBCV-1.

Figure 3. SDS-PAGE profile of the virion protein compositions of OSyNE-5 and PBCV-1 purified particles. The OSyNE-5 SDS-PAGE profile resembles the PBCV-1 profile but with slight differences. OSyNE-5 appears to have two major capsid proteins (MCPs) that migrated slightly slower than the single PBCV-1 MCP. The PBCV-1 MCP (A430L) at 54 kDa is indicated.

Figure 4. Genome comparison of chlorovirus OSyNE-5 and the prototype PBCV-1 as reference. (a) Progressive Mauve alignment on default settings for the genomes of OSyNE-5 (top) and PBCV-1 (bottom). The degree of DNA sequence similarity is indicated by the height of each colored block. Homologous regions are connected by lines between genomes, and blocks below the center line in OSyNE-5 indicate regions with inverse orientation in comparison to PBCV-1. White spaces within blocks represent small-localized areas of the genome sequences that were not aligned. The largest of these is the stretch from around 71- kb to 91-kb (a') within the inverted OSyNE-5 section, which is present in OSyNE-5 but not in the PBCV-1 genome. Blocks below the central line represent sequences that are inverted in comparison to the PBCV-1 arrangement. (b) Dot plot alignments of the genome sequences of OSyNE-5 (vertical) against PBCV-1 (horizontal). Reverse slopes (red points) represent sequences that are

inverted between the two genomes. (c) Venn diagram illustrating the comparisons of OSyNE-5 and PBCV-1 genomes. Number of major CDSs (red) and minor ORFs (blue) are displayed for each genome. While the majority of major CDSs are shared between both viruses, minor ORFs are divergent and likely important for the differential host permissibility of both viruses.

Figure 5. Phylogenetic tree shows the evolutionary relationships between 47 viral concatenated amino acid sequences (7762 gap-free sites). The Maximum Likelihood (ML) tree was constructed using the MEGA 6.0 software (www.megasoftware.net) with the ML algorithm and default settings. Bar: 0.2 substitutions per amino acid site. The new OSy *Chlorovirus* genus is indicated in red and resides within the two separate phylogenetic sub-groups of NC64A viruses –one contains PBCV-1 and the other NY-2A. Both sub-groups share almost perfect gene colinearity since they replicate in the same host. Branch support was estimated from 1000 bootstrap replicates. Four Ostreococcus virus sequences serve as an outgroup to root the tree.

Figure 6. Attachment analysis of infected and uninfected *C. variabilis* cells with three chlorovirus types 1h post infection (pi) and stained with the DNA dye SYBR® Gold. (a) Fluorescent microscopy observations. Cells are stained by the fluorescent DNA dye if the virus attaches and initiates infection. Chlorophyll stains red and DNA yellow under UV light. (b) Infected and uninfected cells mixed with OSyNE-5 and PBCV-1 (1h pi). The histogram depicts the level of SYBR® Gold stain intensity on infected and uninfected cells. SYBR® Gold intensity increases when chlorella cells are infected by the respective chloroviruses. The gating was set up on an uninfected control sample stained with SYBR® Gold and applied to the infected samples. Negative control cells are plotted in green and black. Experimental cells mixed with the DNA dye (SYBR® Gold) 1h pi are plotted in red. Samples were run at 1 x 10^6 cells/ml, and approximately 1 x 10^4 cell events were collected per sample.

Figure 7. OSyNE-5 inhibits PBCV-1 replication on permissive and non-permissive cells. Number of PBCV-1 plaques on NC64A lawns after 96 hrs (y-axis) following pre-challenge infections (MOI =10) with ATCV-1, PBCV-1, or OSyNE-5 respectively on

NC64A, Syngen 2-3 and OK1-ZK cells followed by a second challenge with PBCV-1 (x-axis). The lack of plaques for all zoochlorella strains after primary OSyNE-5 infection indicates that OSyNE-5 attachment inhibits secondary infection by PBCV-1. ATCV-1 and PBCV-1 serve as controls.

Figure 8. Viability test of NC64A, OK1-ZK, and Syngen 2-3 cells upon infection with OSyNE-5 at high and low MOI (20 and 0.01 respectively). A viable culture of OK1-ZK and NC64A cells following OSyNE-5 infection at low MOI establishes that viral attachment but not replication happens in non-permissive cells. In contrast, OSyNE-5 infection at high MOI in non-permissive cells triggers cell death after viral attachment. OSyNE-5 infection on permissive cells (Syngen 2-3) at high and low MOI serves as control. Pictures were taken 7 days pi.

Figure 9. PFGE kinetics of DNA degradation of NC64A and Syngen 2-3 cells upon infection with OSyNE-5 virus. Although OSyNE-5 cannot complete its replication cycle, degradation of NC64A host DNA occurs. Additionally, OSyNE-5 appears to degrade host chromosomal DNA at a slower rate than PBCV-1 in both strains.

Table Legends

Table 1. Predicted ORFs in the OSyNE-5 genome that are close orthologs to the annotated ORFs in the PBCV-1 genome.

Table 2. Fourteen tRNAs predicted in the OSyNE-5 genome.

Table 3. Twenty-nine identified core proteins from the OSyNE-5 virus used for the phylogenetic analysis.

Table 4. Predicted ORFs exclusively present in the OSyNE-5 genome.

Table 5. OSyNE-5 genes and gene annotations.

Table 6. Blast results for the three regions that are present exclusively in the OSyNE-5 genome (labeled as a', b' and c' on Fig. 4a).

Supplementary Figure Legends

Supplementary Figure 1. Significantly higher plaque numbers on Syngen 2-3 lawn compared to NC64A and OK1-ZK lawns observed after plating 1 ml of indigenous water collected from Lincoln, Nebraska.

Supplementary Figure 2. Experimental flow chart of OSy viruses isolation from indigenous water. Single plaques were selected from Syngen 2-3 lawns showing higher plaque numbers when compared to NC64A and OK1-ZK. Virus was amplified in Syngen 2-3 cells for seven days and challenged against the three strains. Virus that exclusively lysed Syngen 2-3 cells (showing a clear tube) without lysing NC64A cells or OK1-ZK (showing a green tube) were selected for plaque assay.

Supplementary Figure 3. OSyNE isolates challenged against Syngen 2-3 and NC64A strains at MOI=0.01. NC64A cells persist when incubated with OSyNE viral lysates that would lyse Syngen 2-3 cells.

Supplementary Figure 4. Electron micrographs of additional OSyNE viruses at various resolutions.

Supplementary Figure 5.Virion proteins of PBCV-1 that replicated on NC64A cells (blue), PBCV-1 that replicated on Syngen 2-3 cells (black), and other OSy viruses (Florida isolate OSyF-3, Nebraska isolate OSyNE-M2, and the prototype OSyNE-5).

Supplementary Figure 6. Restriction enzyme analysis of OSyNE viral DNAs compared to PBCV-1 and ATCV-1 DNAs. Prototype OSyNE-5 highlighted in red by an asterisk.

Supplementary Figure 7. Pulsed field gel electrophoresis of genomic DNAs from PBCV-1 and OSyNE-5.

Supplementary Figure 8. Dot plot alignment of NC64A, SAG, Pbi, and OSy viruses. Each dot represents a nucleotide match between the two sequenced genomes in the same orientation or in reverse orientation.

Supplementary Figure 9. Experimental flow chart for OSy viruses attachment analysis. Preparation of uninfected and infected *C. variabilis* cells (1h pi). stained with the DNA dye SYBR® Gold for fluorescent microscopy and flow cytometry analysis.

Supplementary Figure 10. Fluorescent microscopy observations of uninfected and infected NC64A cells stained with the DNA dye SYBR® Gold. NC64A cells mixed with OSy viruses or PBCV-1 show similar increase staining intensity of the fluorescent DNA dye (1 hr pi) indicating that OSy virus attach and initiate infection in permissive and non-permissive cells.

Supplementary Figure 11. Schematic of viability test of NC64A, OK1-ZK, and Syngen 2-3 cells upon infection with OSyNE-5 at high and low MOI (20 and 0.01 respectively).

Supplementary Figure 12. Experimental flow chart to test if OSyNE-5 inhibits PBCV-1 replication on permissive and non-permissive cells.

Supplementary Figure 13. Schematic model for OSyNE-5 infection on non-permissive cells. When OSyNE-5 infection starts on NC64A cells, host DNA degradation occurs suggesting that at least some early and/or early/late transcripts may be synthetized; nevertheless, the molecular mechanisms remain elusive.

Supplementary Table Legends

Supplementary Table 1. Summary of OSy virus isolates from different sites within the USA.

Supplementary Table 2. List of sequenced OSy isolates from different sites within the USA.

Figure 1

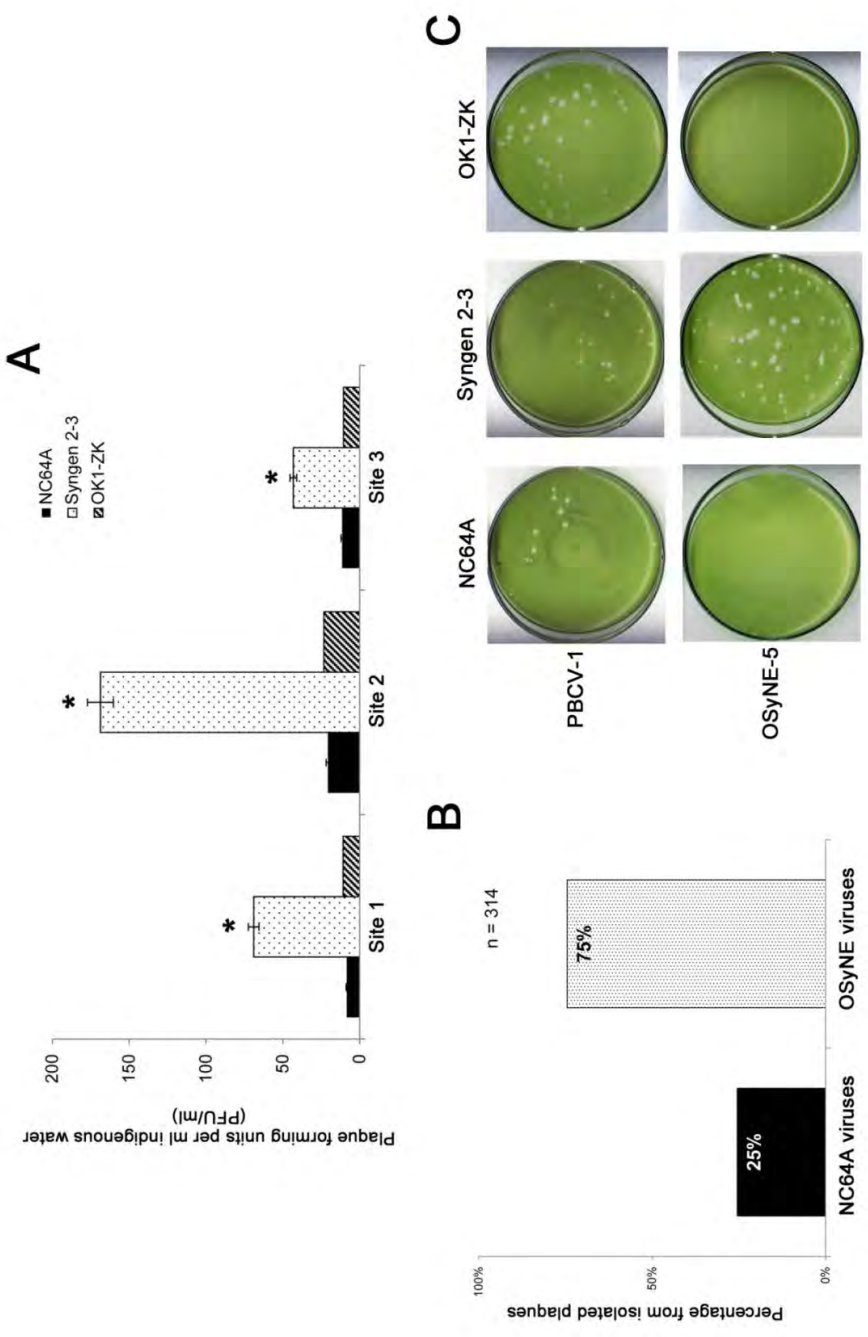

A
NC64A
Syngen 2-3
OK1-ZK
Plaque forming units per ml indigenous water (PFU/ml)
200
150
100
50
0
Site 1
Site 2
Site 3
*
*
*
B
n = 314
75%
25%
NC64A viruses
OSyNE viruses
Percentage from isolated plaques
100%
50%
0%
C
NC64A
Syngen 2-3
OK1-ZK
PBCV-1
OSyNE-5

Figure 2

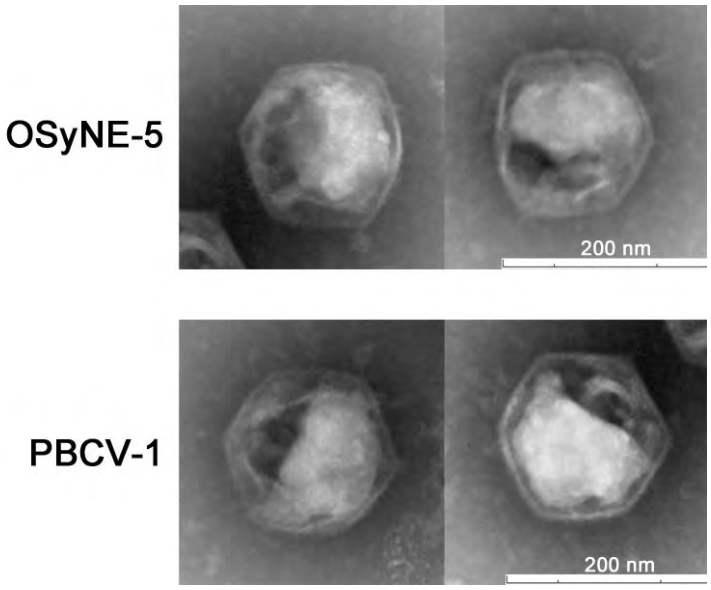


Figure 3

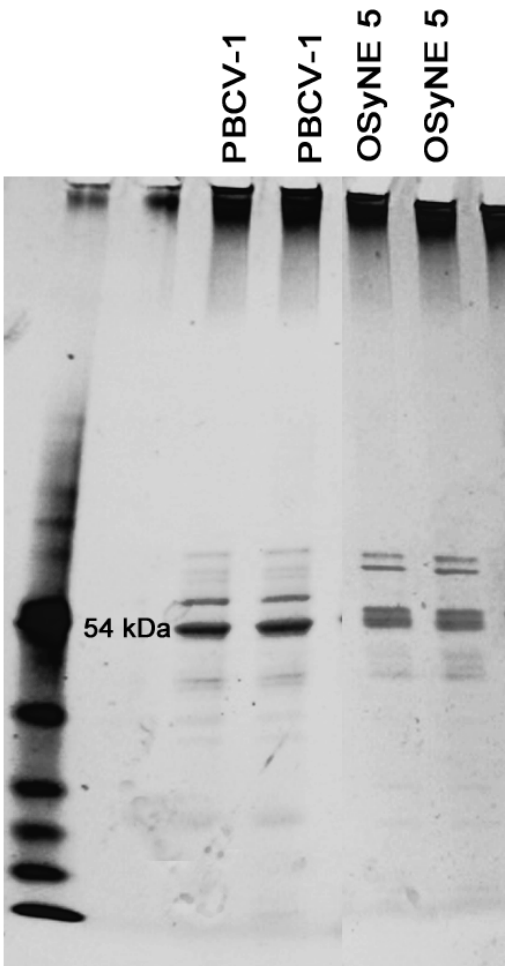


Figure 4

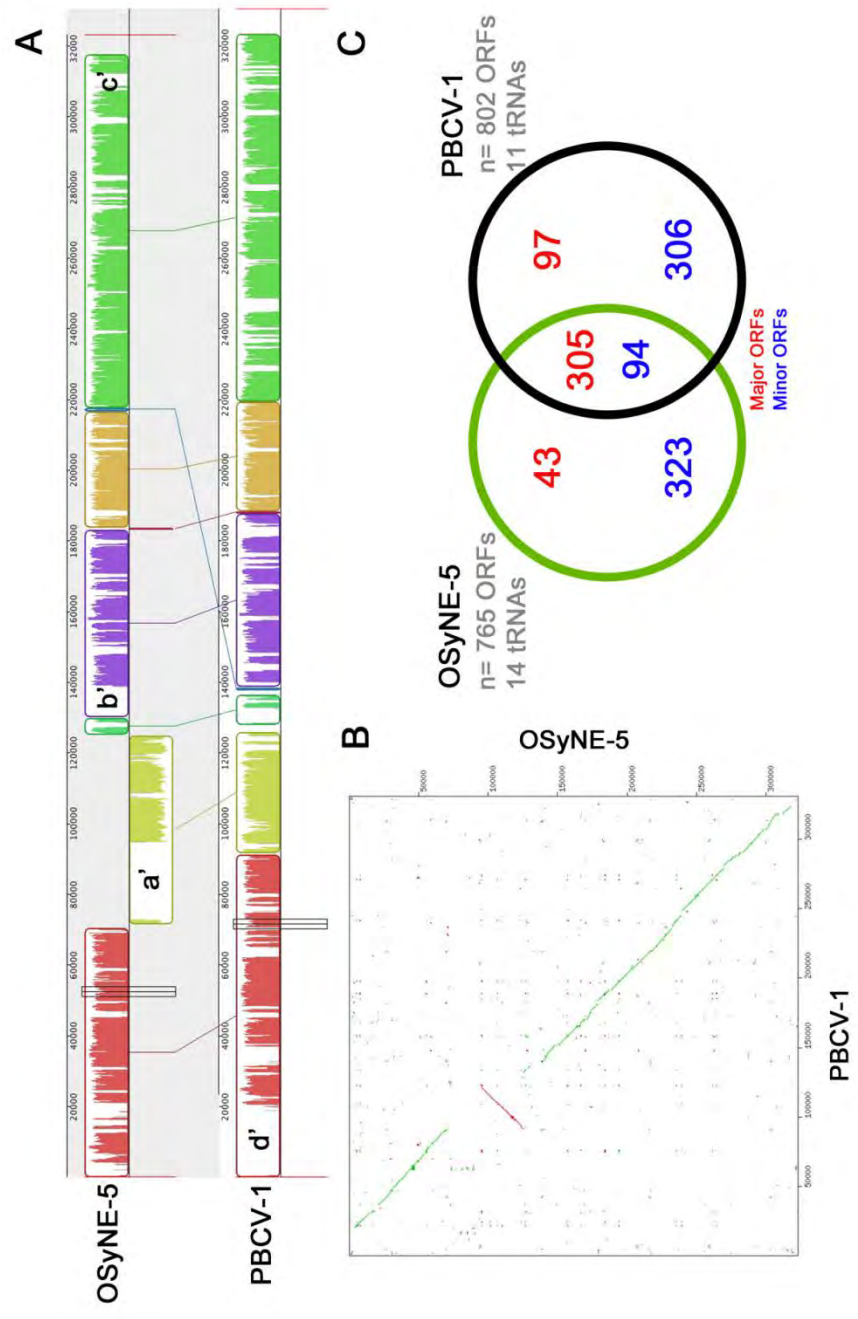

A
OSyNE-5
PBCV-1
a'
b'
c'
d'
B
OSyNE-5
PBCV-1
C
PBCV-1
n= 802 ORFs
11 tRNAs
OSyNE-5
n= 765 ORFs
14 tRNAs
97
306
305
94
43
323
Major ORFs
Minor ORFs

Figure 5

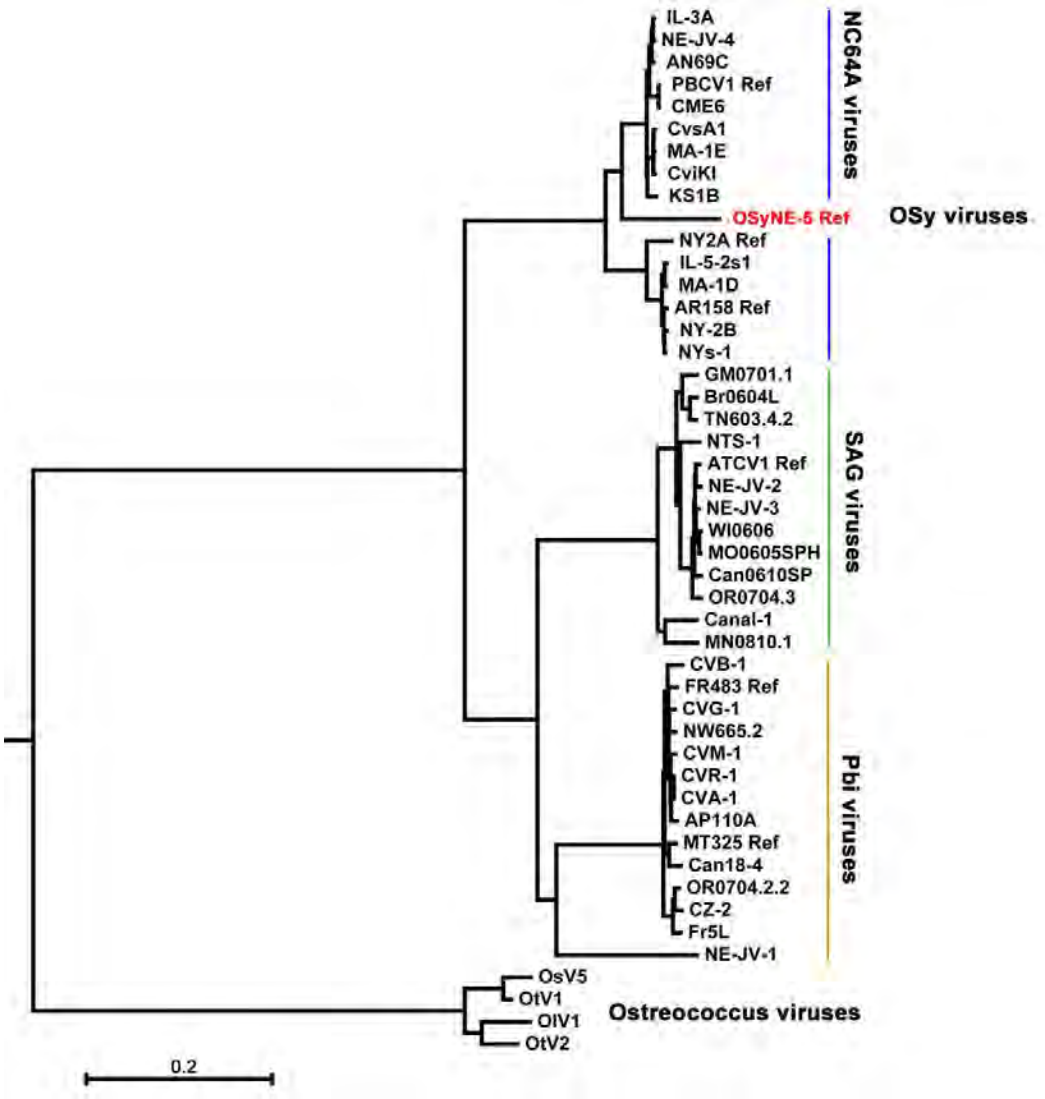

IL-3A
NE-JV-4
AN69C
PBCV1 Ref
CME6
CvsA1
MA-1E
CviKI
KS1B
OSyNE-5 Ref
NY2A Ref
IL-5-2s1
MA-1D
AR158 Ref
NY-2B
NYs-1
GM0701.1
Br0604L
TN603.4.2
NTS-1
ATCV1 Ref
NE-JV-2
NE-JV-3
WI0606
MO0605SPH
Can0610SP
OR0704.3
Canal-1
MN0810.1
CVB-1
FR483 Ref
CVG-1
NW665.2
CVM-1
CVR-1
CVA-1
AP110A
MT325 Ref
Can18-4
OR0704.2.2
CZ-2
Fr5L
NE-JV-1
OsV5
OtV1
OlV1
OtV2
0.2
NC64A viruses
OSy viruses
SAG viruses
Pbi viruses
Ostreococcus viruses

Figure 6

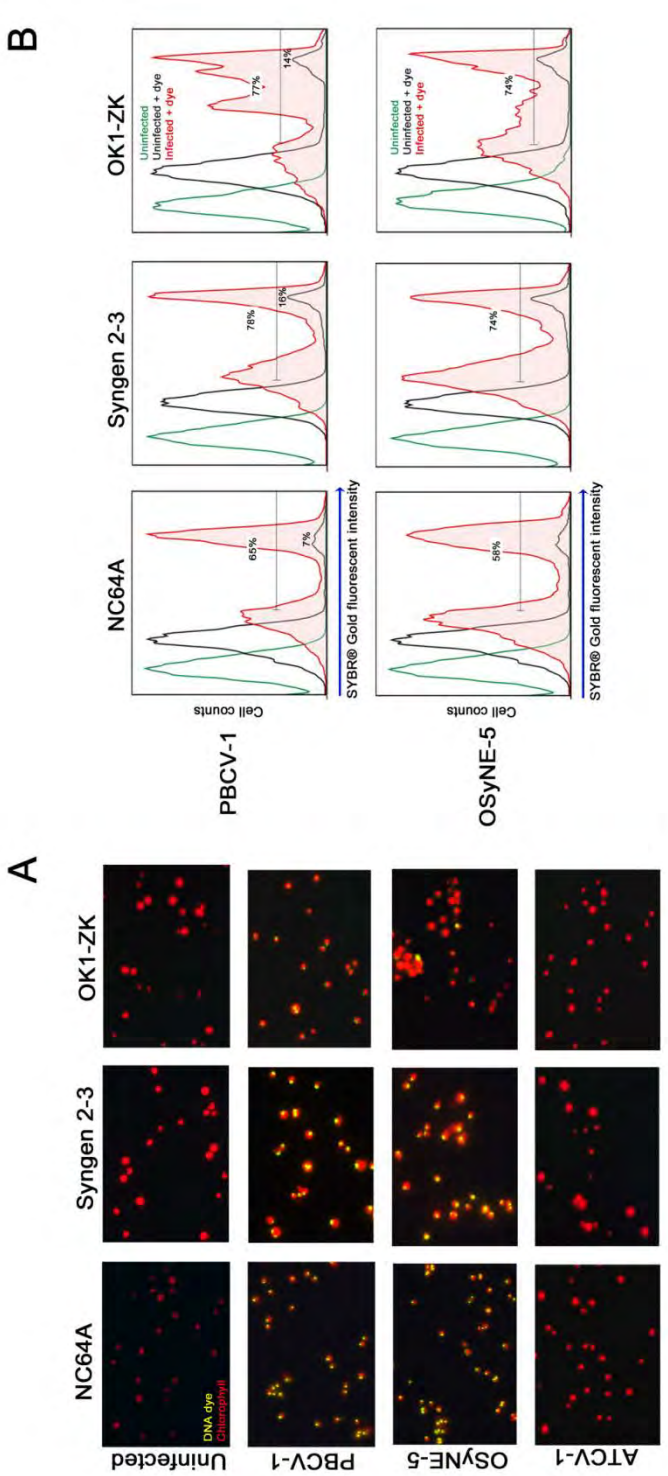

A
NC64A
Syngen 2-3
OK1-ZK
Uninfected
PBCV-1
OSyNE-5
ATCV-1
DNA dye
Chlorophyll
B
NC64A
Syngen 2-3
OK1-ZK
PBCV-1
OSyNE-5
Cell counts
SYBR® Gold fluorescent intensity
Uninfected
Uninfected + dye
Infected + dye
65%
7%
78%
16%
77%
14%
56%
74%
74%

Figure 7

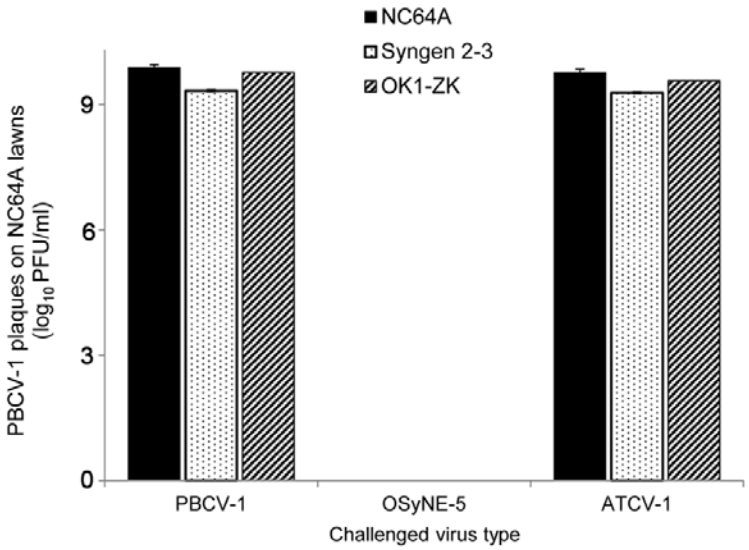

NC64A
Syngen 2-3
OK1-ZK
PBCV-1 plaques on NC64A lawns
(log10 PFU/ml)
0
3
6
9
PBCV-1
OSyNE-5
ATCV-1
Challenged virus type

Figure 8

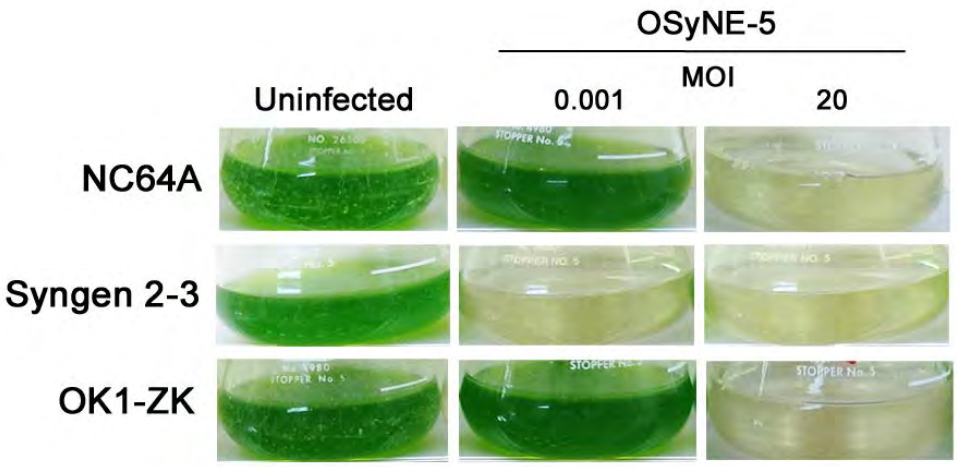


Figure 9

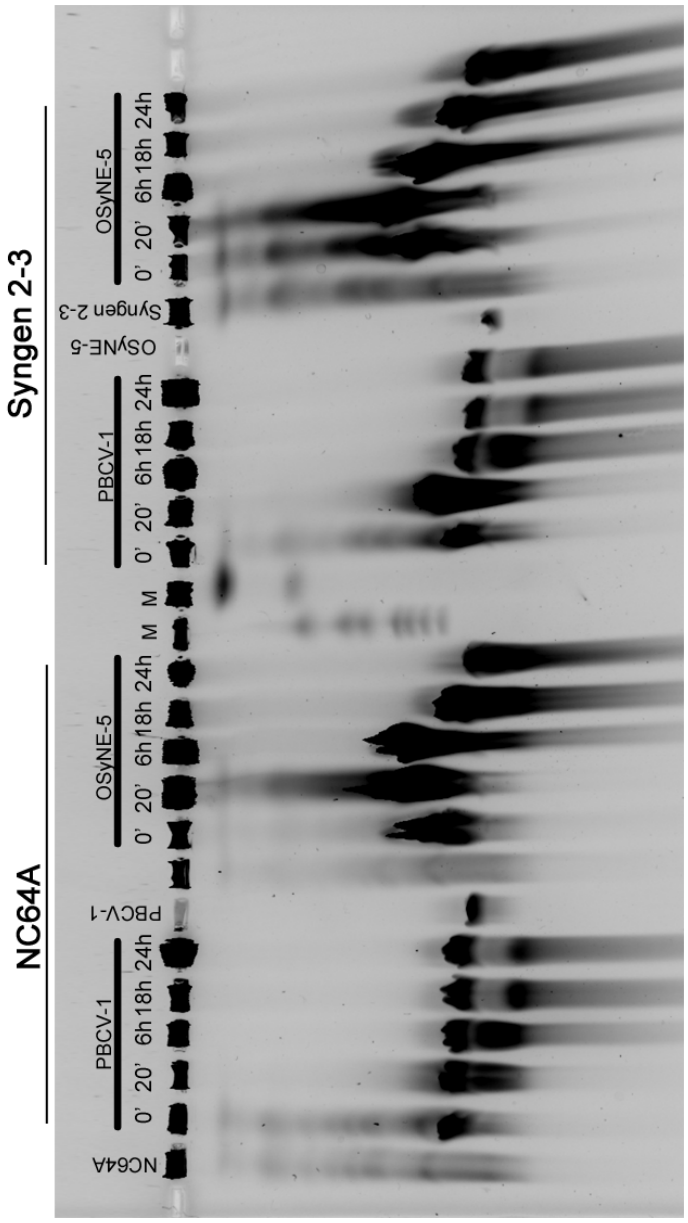

NC64A
Syngen 2-3
NC64A
PBCV-1
0' 20' 6h 18h 24h
PBCV-1
OSyNE-5
0' 20' 6h 18h 24h
OSyNE-5
M
PBCV-1
0' 20' 6h 18h 24h
OSyNE-5
0' 20' 6h 18h 24h
OSyNE-5
Syngen 2-3

Table 1

ORF ID	Start	End	AA length	PBCV-1 best match [e-value]	Swissprot best match (<1e-5)
OSYNE5|428R	179322	180266	315	A007/008L [hypothetical protein] [4e-32]	Q8Q0U0.1 [Putative ankyrin repeat protein MM_0045] [4e-29]
OSYNE5|193R	84424	88875	1484	A025/027/029L [hypothetical protein] [1e-10]	N/A
OSYNE5|010R	4153	5079	309	A034R [Protein kinase] [0.0]	N/A
OSYNE5|015L	6885	7199	105	A037L [hypothetical protein] [4e-52]	N/A
OSYNE5|017R	7248	7727	160	a038R [hypothetical protein] [7e-13]	N/A
OSYNE5|016L	7239	7685	149	A039L [hypothetical protein] [2e-66]	O49484.1 [SKP1-like protein 11] [1e-31]
OSYNE5|018R	7770	9017	416	A041R [hypothetical protein] [0.0]	Q4JV51.1 [Translation initiation factor IF-2] [1e-09]
OSYNE5|019L	7863	8135	91	a042L [hypothetical protein] [6e-16]	N/A
OSYNE5|160R	67402	67629	76	a043R [hypothetical protein] [3e-16]	N/A
OSYNE5|021L	9014	10849	612	A044L [hypothetical protein] [0.0]	Q5UR45.1 [Putative AAA family ATPase L572] [2e-19]
OSYNE5|025R	10917	11294	126	A048R [hypothetical protein] [8e-63]	N/A
OSYNE5|026L	11291	11950	220	A049L [hypothetical protein] [6e-120]	O07592.1 [Putative glycerophosphoryl diester phosphodiesterase] [6e-27]
OSYNE5|027L	11967	12392	142	A050L [Pyrimidine dimer-specific glycosylase] [2e-84]	P04418.1 [Endonuclease V] [7e-27]
OSYNE5|030R	13557	14645	363	A053R [hypothetical protein] [0.0]	P52643.1 [D-lactate dehydrogenase] [9e-85]
OSYNE5|031L	13673	13996	108	a054L [hypothetical protein] [1e-56]	N/A
OSYNE5|032L	14162	14407	82	a056L [hypothetical protein] [1e-39]	N/A
OSYNE5|037R	17004	17711	236	A057aR [hypothetical protein] [1e-23]	N/A
OSYNE5|049R	21733	21936	68	A067R [hypothetical protein] [5e-32]	N/A
OSYNE5|050R	21969	23033	355	A071R [hypothetical protein] [0.0]	N/A
OSYNE5|051L	22290	22592	101	a073L [hypothetical protein] [5e-22]	N/A
OSYNE5|053L	22626	22886	87	a074L [hypothetical protein] [1e-13]	N/A
OSYNE5|055L	23043	23882	280	A075L [hypothetical protein] [0.0]	N/A
OSYNE5|058L	24499	24798	100	A076L [hypothetical protein] [7e-35]	N/A
OSYNE5|059L	24767	25068	100	A077L [hypothetical protein] [4e-54]	N/A
OSYNE5|063L	26798	27025	76	a078cL [hypothetical protein] [2e-10]	N/A
OSYNE5|060R	25164	26060	299	A078R (N-carbamoylputrescine amidohydrolase) [0.0]	Q3HVN1.1 [N-carbamoylputrescine amidase] [2e-94]
OSYNE5|064R	26831	27559	243	A079R [hypothetical protein] [2e-160]	N/A
OSYNE5|729L	303526	303849	108	a080L [hypothetical protein] [1e-19]	N/A
OSYNE5|067L	27563	28132	190	A081L [hypothetical protein] [2e-108]	N/A
OSYNE5|068L	28205	28735	177	A084L [hypothetical protein] [4e-74]	N/A
OSYNE5|069R	28881	29561	227	A085R [Prolyl 4-hydroxylase] [3e-133]	Q5UP57.1 [Putative prolyl 4-hydroxylase] [3e-23]
OSYNE5|073R	30634	31122	163	A088R [hypothetical protein] [4e-35]	N/A
OSYNE5|071R	29610	30524	305	A086R [hypothetical protein] [9e-138]	N/A
OSYNE5|075L	31204	32505	434	A092/093L [hypothetical protein] [0.0]	N/A
OSYNE5|076L	32543	33643	367	A094L [beta-1-3-glucanase] [0.0]	P23903.1 [Glucan endo-13-beta-glucosidase A1] [2e-28]
OSYNE5|077R	32892	33116	75	a095R [hypothetical protein] [9e-43]	N/A
OSYNE5|082R	33797	35488	564	A098R [Hyaluronan synthase] [0.0]	O06650.2 [Hyaluronan synthase 3] [3e-65]
OSYNE5|084R	35643	37430	596	A100R [Glutamine:fructose-6-phosphate amidotransferase] [0.	Q7WE36.3 [Glutamine--fructose-6-phosphate aminotransferase] [0.0]
OSYNE5|085L	35781	36044	88	a101L [hypothetical protein] [6e-27]	N/A
OSYNE5|087R	37599	38591	331	A103R [mRNA guanylyltransferase] [0.0]	Q84424.1 [GTP--RNA guanylyltransferase] [0.0]
OSYNE5|088L	38079	38468	130	a104L [hypothetical protein] [3e-22]	N/A
OSYNE5|090L	38578	39450	291	A105L [hypothetical protein] [3e-168]	Q6DCJ1.2 [Ubiquitin carboxyl-terminal hydrolase 22-B] [1e-06]
OSYNE5|092L	39485	40357	291	A107L [hypothetical protein] [0.0]	P61998.1 [Transcription initiation factor IIB] [4e-09]
OSYNE5|093R	40181	40369	63	A108aR [hypothetical protein] [2e-26]	N/A
OSYNE5|094L	40459	40944	162	A108bL [hypothetical protein] [8e-110]	N/A
OSYNE5|096R	41069	43651	861	A111/114R [hypothetical protein] [0.0]	N/A
OSYNE5|098L	41540	42064	175	a113L [hypothetical protein] [2e-35]	N/A
OSYNE5|99R	42369	42614	82	a116R [hypothetical protein] [1e-21]	N/A
OSYNE5|101R	43699	44739	347	A118R [GDP-D-mannose dehydratase] [0.0]	Q9JRN5.1 [GDP-mannose 46-dehydratase] [1e-142]
OSYNE5|102R	44759	45073	105	A121R [hypothetical protein] [1e-67]	N/A
OSYNE5|178R	73060	75918	953	A122/123R [hypothetical protein] [1e-12]	N/A
OSYNE5|104R	47015	47935	307	A122/123R [hypothetical protein] [1e-156]	Q37893.1 [Pre-neck appendage protein] [1e-11]
OSYNE5|204R	88972	93492	1507	A122/123R [hypothetical protein] [3e-11]	N/A
OSYNE5|106L	47937	48479	181	A125L [hypothetical protein] [5e-132]	P49373.1 [Transcription elongation factor S-II] [3e-14]
OSYNE5|108R	48513	49229	239	A127R [hypothetical protein] [5e-153]	N/A
OSYNE5|116R	50475	50792	106	A130R [hypothetical protein] [4e-47]	N/A
OSYNE5|117L	50785	51192	136	A131L [hypothetical protein] [7e-78]	N/A
OSYNE5|118L	51333	51851	173	A134L [hypothetical protein] [8e-94]	N/A
OSYNE5|119L	51824	52147	108	A135L [hypothetical protein] [2e-25]	N/A
OSYNE5|120R	51897	52337	147	A136R [hypothetical protein] [1e-60]	N/A
OSYNE5|122R	52396	52619	74	A137R [hypothetical protein] [2e-24]	N/A
OSYNE5|123R	52665	53471	269	A138R [hypothetical protein] [2e-87]	N/A
OSYNE5|125L	53468	53779	104	A139L [hypothetical protein] [3e-47]	N/A
OSYNE5|127L	56426	56866	147	A150L [hypothetical protein] [1e-60]	N/A
OSYNE5|128R	56962	58356	465	A153R [hypothetical protein] [0.0]	Q5UQ46.1 [Putative ATP-dependent RNA helicase L396] [1e-55]
OSYNE5|110R	49301	50455	385	A154L [hypothetical protein] [1e-174]	N/A
OSYNE5|114L	50035	50221	62	a155R [hypothetical protein] [2e-16]	N/A
OSYNE5|112R	49386	49595	70	a156L [hypothetical protein] [2e-06]	N/A
OSYNE5|135L	59252	59590	113	A157L [hypothetical protein] [8e-63]	N/A
OSYNE5|137L	59638	59958	107	A158L [hypothetical protein] [3e-36]	N/A
OSYNE5|139R	59966	60295	110	A159R [hypothetical protein] [5e-24]	N/A
OSYNE5|141R	60114	60551	146	A161R [hypothetical protein] [5e-26]	N/A
OSYNE5|142L	60552	61772	407	A162L [hypothetical protein] [0.0]	N/A
OSYNE5|143R	61813	63159	449	A163R [hypothetical protein] [0.0]	N/A
OSYNE5|145L	62745	63023	93	a164L [hypothetical protein] [3e-30]	N/A
OSYNE5|148L	63755	64216	154	A165aL [hypothetical protein] [8e-84]	N/A
OSYNE5|147L	63386	63724	113	A165L [hypothetical protein] [6e-67]	N/A
OSYNE5|149R	64279	65085	269	A166R [hypothetical protein] [9e-179]	Q5UQV1.1 [Uncharacterized protein R354] [1e-12]
OSYNE5|150L	64483	64935	151	a167L [hypothetical protein] [2e-24]	N/A
OSYNE5|153R	65124	65621	166	A168R [hypothetical protein] [1e-106]	N/A
OSYNE5|155R	65622	66704	361	A169R [Aspartate transcarbamylase] [0.0]	Q43087.1 [Aspartate carbamoyltransferase 2 chloroplastic] [5e-98]
OSYNE5|157L	65819	66148	110	a170L [hypothetical protein] [5e-50]	N/A
OSYNE5|161R	67486	67990	175	A171R [hypothetical protein] [4e-111]	N/A
OSYNE5|163L	67993	68820	276	A173L [hypothetical protein] [0.0]	Q91F63.1 [Probable lipid hydrolase 463L] [3e-18]
OSYNE5|165L	69171	69473	101	A176L [hypothetical protein] [2e-18]	N/A
OSYNE5|164L	68941	69150	70	A176L [hypothetical protein] [3e-41]	N/A
OSYNE5|166R	69979	70716	246	A177R [hypothetical protein] [1e-154]	N/A
OSYNE5|168L	70279	70497	73	a178L [hypothetical protein] [6e-24]	N/A
OSYNE5|369L	154459	154791	111	A180R [hypothetical protein] [1e-50]	O34693.1 [Uncharacterized protein YloA] [1e-14]

OSYNE5|291R	124022	124792	257	a183L [hypothetical protein] [8e-09]	N/A
OSYNE5|287L	120694	122952	753	A185R [hypothetical protein] [0.0]	P30321.2 [DNA polymerase] [0.0]
OSYNE5|285L	120113	120559	149	A188aR [hypothetical protein] [2e-100]	P30321.2 [DNA polymerase] [1e-89]
OSYNE5|286R	120546	120794	83	a188L [hypothetical protein] [3e-46]	N/A
OSYNE5|279L	116159	120070	1304	A189/192R [hypothetical protein] [0.0]	N/A
OSYNE5|284R	119784	120026	81	a190L [hypothetical protein] [6e-28]	N/A
OSYNE5|278R	115350	116156	269	A193L [hypothetical protein] [0.0]	Q84513.1 [Probable DNA polymerase sliding clamp 1] [0.0]
OSYNE5|277R	114887	115345	153	A196L [hypothetical protein] [1e-102]	N/A
OSYNE5|276L	114877	115152	92	a197R [hypothetical protein] [4e-47]	N/A
OSYNE5|275L	114537	114839	101	A199R [hypothetical protein] [2e-54]	N/A
OSYNE5|274L	114079	114435	119	A200R [hypothetical protein] [6e-82]	N/A
OSYNE5|273R	113772	114056	95	A201L [hypothetical protein] [1e-42]	N/A
OSYNE5|272R	113412	113753	114	A202L [hypothetical protein] [8e-77]	N/A
OSYNE5|270L	112702	113346	215	A203R [hypothetical protein] [7e-147]	N/A
OSYNE5|268L	112033	112656	206	A205R [hypothetical protein] [2e-108]	N/A
OSYNE5|269R	112237	112449	71	a206L [hypothetical protein] [7e-16]	N/A
OSYNE5|263L	110852	111970	373	A207R [Arginine/Ornithine decarboxylase] [0.0]	P27117.1 [Ornithine decarboxylase] [2e-87]
OSYNE5|305L	130414	131586	391	A208R [hypothetical protein] [1e-11]	N/A
OSYNE5|260L	109852	110724	291	A208R [hypothetical protein] [6e-80]	N/A
OSYNE5|261L	109886	110318	151	a211R [hypothetical protein] [1e-06]	N/A
OSYNE5|258R	109209	109655	149	A213L [hypothetical protein] [8e-102]	N/A
OSYNE5|256R	108749	109168	140	A214L [hypothetical protein] [2e-81]	N/A
OSYNE5|252R	106424	107359	312	A215L [Alkaline alginate lyase vAL-1] [0.0]	N/A
OSYNE5|253L	106549	106770	74	a216R [hypothetical protein] [6e-19]	N/A
OSYNE5|250R	105250	106404	385	A217L [hypothetical protein] [0.0]	N/A
OSYNE5|246L	103245	105140	632	A219/222/226R [hypothetical protein] [0.0]	Q9U720.1 [Cellulose synthase catalytic subunit A (UDP-forming)] [1e-06]
OSYNE5|248L	104210	104530	107	a223R [hypothetical protein] [2e-21]	N/A
OSYNE5|245R	103239	103892	218	a225L [hypothetical protein] [2e-35]	N/A
OSYNE5|243R	102823	103236	138	A227L [hypothetical protein] [2e-92]	N/A
OSYNE5|244L	102957	103178	74	a228R [hypothetical protein] [9e-40]	N/A
OSYNE5|242R	102568	102801	78	A229L [hypothetical protein] [1e-47]	N/A
OSYNE5|239L	101953	102543	197	A230R [hypothetical protein] [1e-129]	N/A
OSYNE5|235R	101800	101939	380	A231L [hypothetical protein] [0.0]	N/A
OSYNE5|236L	10 :6	101159	78	a232R [hypothetical protein] [8e-22]	N/A
OSYNE5|234L	100427	100750	108	A233R [hypothetical protein] [1e-58]	N/A
OSYNE5|233R	100104	100430	109	A234L [hypothetical protein] [6e-52]	N/A
OSYNE5|232R	99814	100116	101	a236L [hypothetical protein] [1e-37]	N/A
OSYNE5|230L	98481	100031	517	A237R [Homospermidine synthase] [0.0]	Q98H64.1 [Homospermidine synthase] [3e-63]
OSYNE5|229R	98039	98476	146	A239L [hypothetical protein] [2e-73]	N/A
OSYNE5|223L	95698	97875	726	A241R [hypothetical protein] [0.0]	P47047.1 [ATP-dependent RNA helicase DOB1] [3e-98]
OSYNE5|220L	94786	95670	295	A246aR [hypothetical protein] [3e-114]	Q5KSL6.1 [Diacylglycerol kinase kappa (kinase kappa) [3e-12]
OSYNE5|175L	71978	72844	289	A248R [Protein kinase] [1e-158]	Q96NX5.3 [Calcium/calmodulin-dependent protein kinase type 1G] [1e-27]
OSYNE5|176R	71994	72422	143	a249L [hypothetical protein] [9e-38]	N/A
OSYNE5|172L	71671	71955	95	A250R [Potassium ion channel protein (Kcv)] [1e-55]	Q84568.1 [Potassium channel protein kcv] [2e-52]
OSYNE5|292R	125295	125750	152	A253R [hypothetical protein] [2e-79]	N/A
OSYNE5|293R	125781	127298	506	A260R [chitenase] [0.0]	P32470.2 [Chitinase 1 Flags: Precursor] [5e-56]
OSYNE5|300L	128684	129334	217	A262/263L [hypothetical protein] [4e-119]	N/A
OSYNE5|301L	129355	130104	250	A265L [hypothetical protein] [4e-169]	N/A
OSYNE5|389R	164860	165702	281	A267L [hypothetical protein] [9e-45]	Q5UQL9.1 [Uncharacterized protein R423] [9e-15]
OSYNE5|525L	216782	217606	275	A271L [hypothetical protein] [1e-149]	Q55EQ3.2 [Uncharacterized abhydrolase protein DDB_G0269086] [2e-08]
OSYNE5|526R	217244	217546	101	a272aR [hypothetical protein] [9e-14]	N/A
OSYNE5|304L	130118	130354	79	A273L [hypothetical protein] [2e-15]	N/A
OSYNE5|330R	139082	139825	248	A275R [hypothetical protein] [4e-159]	N/A
OSYNE5|331L	139810	140649	280	A277L [Protein kinase] [4e-159]	Q5B4Z3.2 [Serine/threonine-protein kinase sepH] [3e-19]
OSYNE5|333L	140725	142026	434	A278L [Protein kinase] [0.0]	N/A
OSYNE5|332R	140714	141094	127	a281R [hypothetical protein] [1e-38]	N/A
OSYNE5|335L	142527	143525	333	A284L [Amindase] [8e-178]	P54966.1 [Uncharacterized protein A284L] [1e-174]
OSYNE5|337R	143433	144539	369	A286R [hypothetical protein] [0.0]	N/A
OSYNE5|297L	127864	128619	252	A287R [hypothetical protein] [3e-101]	N/A
OSYNE5|639R	262485	263297	271	A287R [hypothetical protein] [3e-101]	Q36580.1 [Probable intron-encoded endonuclease 1] [2e-08]
OSYNE5|640L	262505	262699	65	a288L [hypothetical protein] [3e-18]	N/A
OSYNE5|341L	144593	145447	285	A289L [Protein kinase] [2e-160]	A8WYE4.1 [Serine/threonine-protein kinase par-1] [1e-27]
OSYNE5|342R	144747	145121	125	a290R [hypothetical protein] [2e-50]	N/A
OSYNE5|344R	145238	145480	81	a291R [hypothetical protein] [3e-30]	N/A
OSYNE5|345L	145533	146546	338	A292L [Chitosanase] [0.0]	Q07921.1 [Chitosanase Flags: Precursor] [3e-14]
OSYNE5|347R	145921	146121	67	a293R [hypothetical protein] [2e-30]	N/A
OSYNE5|346R	145615	145908	98	a293R [hypothetical protein] [2e-34]	N/A
OSYNE5|349L	146558	147496	313	A295L [Fucose synthetase] [0.0]	Q9LMU0.1 [Putative GDP-L-fucose synthase 2] [1e-124]
OSYNE5|352R	147546	148025	160	A296R [hypothetical protein] [2e-46]	N/A
OSYNE5|358L	150151	150675	175	A297L [hypothetical protein] [6e-101]	N/A
OSYNE5|360L	151280	151954	225	A298L [hypothetical protein] [1e-143]	N/A
OSYNE5|361R	151752	151973	74	a299R [hypothetical protein] [3e-40]	N/A
OSYNE5|362L	151975	152709	245	A301L [hypothetical protein] [9e-105]	N/A
OSYNE5|364R	152764	153000	79	A304R [hypothetical protein] [1e-34]	N/A
OSYNE5|365L	153041	153655	205	A305L [Protein phosphatase] [6e-127]	Q9VVW5.2 [phosphatase] [8e-15]
OSYNE5|367L	153680	153940	87	A306L [hypothetical protein] [1e-44]	N/A
OSYNE5|366R	153664	153945	94	a307R [hypothetical protein] [5e-27]	N/A
OSYNE5|368L	153977	154330	118	A308L [hypothetical protein] [4e-58]	N/A
OSYNE5|370L	154854	155366	171	A310L [hypothetical protein] [2e-110]	N/A
OSYNE5|371L	155434	156150	239	A312L [hypothetical protein] [2e-167]	N/A
OSYNE5|372L	156358	156573	72	A313L [hypothetical protein] [3e-30]	N/A
OSYNE5|374R	156853	156895	81	A314R [hypothetical protein] [1e-44]	N/A
OSYNE5|501L	207941	208759	273	A315L [hypothetical protein] [1e-22]	N/A
OSYNE5|133L	58359	59171	271	A315L [hypothetical protein] [2e-87]	P13329.1 [Probable mobile endonuclease B] [7e-08]
OSYNE5|404L	169639	170493	285	A315L [hypothetical protein] [3e-79]	Q5UPT6.1 [Uncharacterized HNH endonuclease L245] [5e-08]
OSYNE5|497R	205745	206491	249	A315L [hypothetical protein] [4e-07]	N/A
OSYNE5|375L	157432	157713	94	a317L [hypothetical protein] [9e-30]	N/A
OSYNE5|376R	158575	158943	122	A320R [hypothetical protein] [3e-59]	N/A
OSYNE5|377R	158960	159339	120	A321R [hypothetical protein] [2e-73]	N/A
OSYNE5|378L	159450	159980	177	A322L [hypothetical protein] [3e-97]	N/A
OSYNE5|379R	159760	159975	72	a323R [hypothetical protein] [1e-28]	N/A

OSYNE5|380L	160032	161366	445	A324L [hypothetical protein] [0.0]	Q5UQN9.1 [Uncharacterized protein R449] [8e-09]
OSYNE5|382L	161447	162085	213	A326L [hypothetical protein] [3e-144]	N/A
OSYNE5|386R	162183	162434	84	a327R [hypothetical protein] [2e-22]	N/A
OSYNE5|385L	162116	163183	356	A328L [hypothetical protein] [0.0]	N/A
OSYNE5|326R	137050	138066	339	A328L [hypothetical protein] [1e-39]	N/A
OSYNE5|394R	166123	166359	79	a329cR [hypothetical protein] [8e-31]	N/A
OSYNE5|388R	163273	163563	97	A329R [hypothetical protein] [2e-36]	N/A
OSYNE5|761R	317885	320122	746	A330R [hypothetical protein] [1e-38]	P16157.3 [Ankyrin-1] [5e-48]
OSYNE5|395L	166372	167553	394	A333L [hypothetical protein] [0.0]	C3PH19.1 [Translation initiation factor IF-2] [8e-07]
OSYNE5|397R	166695	167033	113	a335R [hypothetical protein] [4e-07]	N/A
OSYNE5|398R	167224	167481	86	a336R [hypothetical protein] [5e-29]	N/A
OSYNE5|402L	168532	169083	184	A337L [hypothetical protein] [3e-54]	N/A
OSYNE5|399L	167594	168394	267	A337L [hypothetical protein] [3e-92]	N/A
OSYNE5|403L	169145	169552	135	A341L [hypothetical protein] [4e-78]	N/A
OSYNE5|406L	170573	172287	565	A342L [hypothetical protein] [0.0]	N/A
OSYNE5|408R	170959	171219	87	a343R [hypothetical protein] [1e-30]	N/A
OSYNE5|409L	171256	171762	169	a345L [hypothetical protein] [2e-29]	N/A
OSYNE5|738L	307468	308046	193	A348R [hypothetical protein] [2e-31]	N/A
OSYNE5|410L	172365	172742	126	A349L [hypothetical protein] [1e-73]	N/A
OSYNE5|411L	172705	172911	69	A349L [hypothetical protein] [2e-17]	N/A
OSYNE5|412R	172787	173155	123	A350R [hypothetical protein] [1e-79]	N/A
OSYNE5|415L	173269	173892	208	A352L [hypothetical protein] [1e-133]	Q5UQF7.1 [Uncharacterized protein R489 Flags: Precursor] [5e-07]
OSYNE5|417R	173586	173792	69	a353R [hypothetical protein] [1e-05]	N/A
OSYNE5|418L	173956	174969	338	A357L [hypothetical protein] [4e-171]	N/A
OSYNE5|419R	174317	174937	207	a358R [hypothetical protein] [4e-08]	N/A
OSYNE5|420L	174612	174842	77	a359L [hypothetical protein] [2e-16]	N/A
OSYNE5|421R	175039	178653	1205	A363R [hypothetical protein] [0.0]	N/A
OSYNE5|423L	176778	176981	68	a364L [hypothetical protein] [3e-41]	N/A
OSYNE5|427R	178738	179202	155	A373R [hypothetical protein] [2e-55]	N/A
OSYNE5|431L	180380	181105	242	A378L [hypothetical protein] [2e-112]	N/A
OSYNE5|433L	181129	181758	210	A379L [hypothetical protein] [3e-139]	N/A
OSYNE5|434R	181931	183391	487	A383R [Capsid protein] [0.0]	P30328.3 [Major capsid protein] [6e-39]
OSYNE5|437R	183415	183600	62	A384bL [hypothetical protein] [1e-27]	N/A
OSYNE5|438L	183683	185533	617	A384dL [Capsid protein] [0.0]	Q4JV51.1 [Translation initiation factor IF-2] [4e-06]
OSYNE5|439L	183706	183906	67	a385L [hypothetical protein] [4e-07]	N/A
OSYNE5|444R	185241	185540	100	a391R [hypothetical protein] [8e-32]	N/A
OSYNE5|445R	185624	186388	255	A392R [hypothetical protein] [2e-176]	Q196X2.1 [Uncharacterized protein 068R] [1e-42]
OSYNE5|447R	186645	187049	135	A394R [hypothetical protein] [3e-61]	N/A
OSYNE5|449R	188703	188954	84	A395R [hypothetical protein] [4e-47]	N/A
OSYNE5|450L	189107	189565	153	A396L [hypothetical protein] [1e-89]	N/A
OSYNE5|452L	189620	189976	119	A398L [hypothetical protein] [3e-76]	N/A
OSYNE5|453R	190049	190633	195	A399R [hypothetical protein] [7e-115]	N/A
OSYNE5|456R	190666	191019	118	A400R [hypothetical protein] [1e-79]	N/A
OSYNE5|458R	191057	191911	285	A401R [hypothetical protein] [0.0]	N/A
OSYNE5|460R	191790	192632	281	A402R [hypothetical protein] [1e-150]	N/A
OSYNE5|461R	192751	193032	94	A403R [hypothetical protein] [1e-64]	N/A
OSYNE5|462R	193063	193650	196	A404R [hypothetical protein] [1e-28]	N/A
OSYNE5|468L	195211	195411	67	a406L [hypothetical protein] [2e-29]	N/A
OSYNE5|469L	195448	196080	211	A407L [hypothetical protein] [5e-130]	N/A
OSYNE5|470L	196119	196886	256	A408L [hypothetical protein] [1e-145]	N/A
OSYNE5|471R	196423	196908	162	a409R [hypothetical protein] [5e-41]	N/A
OSYNE5|472L	196892	197221	110	A410L [hypothetical protein] [1e-62]	N/A
OSYNE5|473R	197309	197815	169	A411R [hypothetical protein] [3e-77]	N/A
OSYNE5|474R	197824	198390	189	A412R [hypothetical protein] [7e-125]	N/A
OSYNE5|476L	198391	199095	235	A413L [hypothetical protein] [9e-109]	N/A
OSYNE5|478R	199174	199392	73	A414R [hypothetical protein] [2e-39]	N/A
OSYNE5|479R	199468	200031	188	A416R [hypothetical protein] [5e-119]	Q197D1.1 [Putative kinase protein 029R] [6e-22]
OSYNE5|480L	200007	201296	430	A417L [hypothetical protein] [0.0]	A6UWR5.1 [Replication factor C large subunit large subunit] [6e-06]
OSYNE5|482R	201097	201333	79	a419R [hypothetical protein] [1e-32]	N/A
OSYNE5|483L	201328	201540	71	A420L [hypothetical protein] [3e-41]	N/A
OSYNE5|484R	201585	201884	100	A421R [hypothetical protein] [1e-46]	N/A
OSYNE5|486R	201908	202102	65	A422aR [hypothetical protein] [2e-35]	N/A
OSYNE5|487R	202113	202601	163	A423R [hypothetical protein] [5e-65]	N/A
OSYNE5|491R	203047	203391	115	A426R [hypothetical protein] [2e-58]	N/A
OSYNE5|492L	203388	203753	122	A427L [hypothetical protein] [2e-48]	P0A616.2 [Thioredoxin AltName: MPT46] [1e-06]
OSYNE5|494L	203803	204207	135	A428L [hypothetical protein] [3e-28]	P27951.1 [IgA FC receptor] [7e-07]
OSYNE5|495L	204233	205594	454	A429L [hypothetical protein] [0.0]	Q5ZIJ9.1 [E3 ubiquitin-protein ligase MIB2] [4e-06]
OSYNE5|499L	206548	207861	438	A430L [Major capsid protein] [0.0]	P30328.3 [Major capsid protein AltName: VP54] [0.0]
OSYNE5|216R	93538	94773	412	A430L [Major capsid protein] [1e-97]	P30328.3 [Major capsid protein AltName: VP54] [1e-94]
OSYNE5|503R	208877	209323	149	A432R [hypothetical protein] [6e-80]	N/A
OSYNE5|504R	209086	209304	73	a433R [hypothetical protein] [2e-13]	N/A
OSYNE5|505L	209327	209518	64	A436L [hypothetical protein] [3e-30]	N/A
OSYNE5|506L	209549	209875	109	A437L [hypothetical protein] [1e-63]	P15250.1 [Chromosomal protein MC1b] [5e-06]
OSYNE5|507L	209904	210140	79	A438L [Glutaredoxin] [8e-51]	Q1RHJ0.1 [Glutaredoxin-1] [6e-09]
OSYNE5|508R	210163	210501	113	A439R [hypothetical protein] [1e-69]	N/A
OSYNE5|509L	210658	211071	138	A441L [hypothetical protein] [1e-93]	N/A
OSYNE5|510R	210706	211056	117	a442R [hypothetical protein] [3e-66]	N/A
OSYNE5|511R	211212	212138	309	A443R [hypothetical protein] [0.0]	N/A
OSYNE5|515L	212870	213184	105	A444L [hypothetical protein] [1e-54]	N/A
OSYNE5|517L	213250	214638	463	A445L [hypothetical protein] [0.0]	Q98496.1 [Uncharacterized protein A445L] [0.0]
OSYNE5|519R	213554	214123	190	a446R [hypothetical protein] [6e-59]	N/A
OSYNE5|521L	214699	215019	107	A448L [Protein disulphide isomerase] [9e-70]	P52588.1 [Protein disulfide-isomerase Flags: Precursor] [3e-08]
OSYNE5|522R	214998	215753	252	A449R [hypothetical protein] [9e-126]	N/A
OSYNE5|524R	215987	216745	253	A450R [hypothetical protein] [1e-178]	N/A
OSYNE5|527L	217722	218135	138	A452L [hypothetical protein] [1e-28]	N/A
OSYNE5|528L	218303	219169	289	A454L [hypothetical protein] [0.0]	N/A
OSYNE5|529L	219206	221161	654	A456L [hypothetical protein] [0.0]	N/A
OSYNE5|530R	219252	220103	284	a459R [hypothetical protein] [7e-57]	N/A
OSYNE5|531R	220313	220567	85	a460R [hypothetical protein] [4e-39]	N/A
OSYNE5|535R	221250	221480	77	A461R [hypothetical protein] [6e-25]	N/A
OSYNE5|534L	221133	221912	260	a463L [hypothetical protein] [2e-70]	N/A

OSYNE5|536R	221514	222326	271	A464R [Rnase III] [0.0]	Q98514.1 [Putative protein A464R] [2e-179]
OSYNE5|537R	222364	222720	119	A465R [hypothetical protein] [1e-78]	Q5UQV6.1 [Probable FAD-linked sulfhydryl oxidase R368] [9e-16]
OSYNE5|538Lc	222745	223683	313	A467L [hypothetical protein] [0.0]	N/A
OSYNE5|540R	223832	225157	442	A468R [hypothetical protein] [0.0]	N/A
OSYNE5|541R	225206	225796	197	A470R [hypothetical protein] [2e-127]	N/A
OSYNE5|544R	225848	226369	174	A471R [hypothetical protein] [3e-115]	Q5UQ75.1 [Uncharacterized protein L507] [7e-29]
OSYNE5|545R	226507	227481	325	A476R [hypothetical protein] [0.0]	P50650.1 [Ribonucleoside-diphosphate reductase small chain] [1e-135]
OSYNE5|546L	226639	226929	97	a477L [hypothetical protein] [5e-31]	N/A
OSYNE5|170R	70804	71676	291	A478L [hypothetical protein] [8e-112]	Q5UQL9.1 [Uncharacterized protein R423] [1e-40]
OSYNE5|550L	228382	228657	92	A480L [hypothetical protein] [5e-42]	N/A
OSYNE5|552L	228685	229368	228	A481L [hypothetical protein] [3e-126]	N/A
OSYNE5|555R	229447	230091	215	A482R [hypothetical protein] [3e-128]	N/A
OSYNE5|557L	230086	230553	156	A484L [hypothetical protein] [7e-98]	N/A
OSYNE5|559R	230636	231079	148	A485R [hypothetical protein] [7e-96]	N/A
OSYNE5|561R	231403	232362	320	A488R [hypothetical protein] [0.0]	Q5UQL4.1 [Uncharacterized protein L417] [9e-11]
OSYNE5|562R	232148	232378	77	a489R [hypothetical protein] [2e-12]	N/A
OSYNE5|549L	227462	228346	295	A490L [hypothetical protein] [2e-149]	Q5UQL9.1 [Uncharacterized protein R423] [9e-33]
OSYNE5|563R	232412	232642	77	A491R [hypothetical protein] [1e-42]	N/A
OSYNE5|564L	232639	233184	182	A492L [hypothetical protein] [2e-93]	N/A
OSYNE5|565R	233226	234320	365	A494R [hypothetical protein] [0.0]	Q98544.1 [Putative transcription factor A494R] [0.0]
OSYNE5|255L	107812	106612	267	A495R [hypothetical protein] [5e-17]	N/A
OSYNE5|566R	234374	234814	147	A497R [hypothetical protein] [1e-83]	Q9T1Q1.1 [Putative protein p47] [4e-08]
OSYNE5|568L	234864	235916	351	A500L [hypothetical protein] [7e-73]	Q8DQN5.1 [Zinc metalloprotease ZmpB Flags: Precursor] [6e-08]
OSYNE5|572L	235950	236237	96	A502L [hypothetical protein] [1e-59]	N/A
OSYNE5|574L	236283	237122	280	A503L [hypothetical protein] [0.0]	N/A
OSYNE5|577L	237201	238700	500	A505L [hypothetical protein] [0.0]	N/A
OSYNE5|579R	237957	238145	63	a507R [hypothetical protein] [2e-18]	N/A
OSYNE5|578R	237585	237797	71	a507R [hypothetical protein] [6e-27]	N/A
OSYNE5|580R	238231	238536	102	a508R [hypothetical protein] [1e-21]	N/A
OSYNE5|585L	239469	239717	83	A519L [hypothetical protein] [2e-49]	N/A
OSYNE5|586L	239722	240021	100	A520L [hypothetical protein] [2e-56]	N/A
OSYNE5|588L	240616	241233	206	A521aL [hypothetical protein] [4e-133]	O55742.1 [Uncharacterized protein 136R] [9e-09]
OSYNE5|587L	240039	240587	183	A521L [hypothetical protein] [2e-69]	N/A
OSYNE5|589R	240911	241123	71	a522R [hypothetical protein] [7e-35]	N/A
OSYNE5|590R	241288	241821	178	A523R [hypothetical protein] [1e-118]	N/A
OSYNE5|591L	241518	241718	67	a524L [hypothetical protein] [2e-37]	N/A
OSYNE5|592R	241864	242304	147	A526R [hypothetical protein] [2e-77]	N/A
OSYNE5|594R	242282	242596	105	A527R [hypothetical protein] [4e-56]	N/A
OSYNE5|595R	242711	242983	91	a528R [hypothetical protein] [1e-07]	N/A
OSYNE5|596L	242928	243146	73	a529L [hypothetical protein] [1e-39]	N/A
OSYNE5|597R	242947	243999	351	A530R [hypothetical protein] [0.0]	P36216.1 [Cytosine-specific methyltransferase CviJI] [8e-98]
OSYNE5|599L	243996	244190	65	A531L [hypothetical protein] [2e-32]	N/A
OSYNE5|600L	244222	244461	80	A532L [hypothetical protein] [1e-50]	N/A
OSYNE5|601R	244740	246335	532	A533R [hypothetical protein] [0.0]	N/A
OSYNE5|603L	246337	246561	75	A535L [hypothetical protein] [3e-39]	N/A
OSYNE5|604L	246627	246881	85	A536L [hypothetical protein] [7e-22]	N/A
OSYNE5|605L	246886	247728	281	A537L [hypothetical protein] [4e-140]	N/A
OSYNE5|607R	247722	248330	203	A539R [hypothetical protein] [3e-108]	N/A
OSYNE5|609L	248346	251747	1134	A540L [hypothetical protein] [0.0]	N/A
OSYNE5|616R	251868	252764	299	A544R [ATP-dependent DNA ligase] [0.0]	P44121.2 [DNA ligase] [3e-11]
OSYNE5|617L	252239	252460	74	a545L [hypothetical protein] [4e-45]	N/A
OSYNE5|619L	252746	253975	410	A546L [hypothetical protein] [0.0]	N/A
OSYNE5|621L	253962	255332	457	A548L [hypothetical protein] [0.0]	Q91ZW3.1 [SWI/SNF actin-dependent regulator of chromatin] [4e-35]
OSYNE5|624R	254868	255125	86	a550R [hypothetical protein] [1e-40]	N/A
OSYNE5|626L	255424	255861	146	A551L [dUTP pyrophosphatase] [1e-77]	O41033.1 [Deoxyuridine 5-triphosphate nucleotidohydrolase] [1e-74]
OSYNE5|627R	255988	256941	318	A552R [hypothetical protein] [0.0]	N/A
OSYNE5|629L	256956	258452	499	A554/556/557L [hypothetical protein] [0.0]	Q82VP4.1 [tRNA(Ile)-lysidine synthase] [2e-18]
OSYNE5|631R	257334	257816	161	a555R [hypothetical protein] [3e-41]	N/A
OSYNE5|632L	258552	259754	401	A558L [Capsid protein] [0.0]	P30328.3 [Major capsid protein AltName: VP54] [5e-77]
OSYNE5|635L	259867	260538	224	A559L [hypothetical protein] [5e-101]	N/A
OSYNE5|636R	259889	260365	159	a560R [hypothetical protein] [3e-32]	N/A
OSYNE5|644R	264703	265260	186	A565R [hypothetical protein] [5e-112]	N/A
OSYNE5|645L	265273	265698	142	A567L [hypothetical protein] [1e-46]	N/A
OSYNE5|646R	265989	266243	85	a569R [hypothetical protein] [8e-25]	N/A
OSYNE5|649L	266389	266790	134	A570L [hypothetical protein] [4e-82]	N/A
OSYNE5|651R	266847	267197	117	A571R [hypothetical protein] [2e-73]	N/A
OSYNE5|652R	267211	267756	182	A572R [hypothetical protein] [1e-119]	N/A
OSYNE5|654L	267762	268508	249	A574L [hypothetical protein] [1e-164]	O41056.1 [Probable DNA polymerase sliding clamp 2] [2e-161]
OSYNE5|656L	268583	269089	169	A575L [hypothetical protein] [8e-115]	N/A
OSYNE5|658L	269196	269603	136	A577L [hypothetical protein] [3e-78]	N/A
OSYNE5|659L	269619	272804	1062	A583L [DNA topoisomerase II] [0.0]	P08096.2 [DNA topoisomerase 2] [0.0]
OSYNE5|660R	269980	270192	71	a584R [hypothetical protein] [5e-34]	N/A
OSYNE5|661R	270033	270641	203	a585R [hypothetical protein] [3e-19]	N/A
OSYNE5|663R	271350	271556	69	a587R [hypothetical protein] [1e-21]	N/A
OSYNE5|666L	272934	274043	370	A590L [hypothetical protein] [2e-66]	N/A
OSYNE5|671R	276121	276549	143	A596R [hypothetical protein] [1e-87]	Q9VWA2.1 [Probable deoxycytidylate deaminase] [6e-26]
OSYNE5|673L	277197	278351	385	A598L [hypothetical protein] [0.0]	P54772.1 [Histidine decarboxylase] [2e-61]
OSYNE5|674R	277503	277688	62	a599R [hypothetical protein] [6e-22]	N/A
OSYNE5|676R	277947	278273	109	a600aR [hypothetical protein] [1e-09]	N/A
OSYNE5|678R	279399	279696	100	A601R [hypothetical protein] [9e-47]	N/A
OSYNE5|680L	279954	280376	141	A602L [hypothetical protein] [5e-78]	N/A
OSYNE5|681R	280481	280795	105	a603R [hypothetical protein] [3e-60]	N/A
OSYNE5|682L	280983	281474	164	A604L [hypothetical protein] [2e-70]	N/A
OSYNE5|684L	281485	281961	159	A605L [hypothetical protein] [3e-80]	N/A
OSYNE5|688R	283412	284584	391	A607R [hypothetical protein] [0.0]	Q02357.2 [Ankyrin-1 AltName: Erythrocyte ankyrin] [9e-10]
OSYNE5|689L	284593	285762	390	A609L [UDP-glucose dehydrogenase] [0.0]	O33952.1 [UDP-glucose 6-dehydrogenase dehydrogenase] [1e-144]
OSYNE5|690L	285844	286203	120	A612L [Histone H3K27 methylase] [1e-75]	Q9Y7Q6.1 [SET domain-containing protein 7] [5e-07]
OSYNE5|691L	286257	287969	571	A614L [Protein kinase] [0.0]	N/A
OSYNE5|694R	288040	289011	324	A617R [hypothetical protein] [0.0]	Q5UQJ6.1 [Putative serine/threonine-protein kinase R400] [1e-11]
OSYNE5|695L	289028	289436	137	A618L [hypothetical protein] [6e-65]	N/A
OSYNE5|696L	289453	290139	229	A619L [hypothetical protein] [3e-90]	N/A

OSYNE5|699L	290179	290430	84	A620L [hypothetical protein] [1e-52]	N/A
OSYNE5|701L	290450	290806	119	A621L [hypothetical protein] [3e-68]	N/A
OSYNE5|702L	290864	292441	526	A622L [Capsid protein] [0.0]	A7U6E9.1 [Major capsid protein] [9e-67]
OSYNE5|703L	292601	292792	64	A623aL [hypothetical protein] [1e-07]	N/A
OSYNE5|704R	292728	293096	123	A624R [hypothetical protein] [1e-65]	N/A
OSYNE5|706R	293115	294431	439	A627R [hypothetical protein] [0.0]	N/A
OSYNE5|709L	294450	294773	108	A628L [hypothetical protein] [9e-24]	N/A
OSYNE5|711R	294936	297239	768	A629R [hypothetical protein] [0.0]	Q03604.1 [Ribonucleoside-diphosphate reductase large subunit] [0.0]
OSYNE5|712L	296075	296302	76	a632L [hypothetical protein] [8e-37]	N/A
OSYNE5|715R	297277	297642	122	A633R [hypothetical protein] [3e-78]	N/A
OSYNE5|718L	299771	300085	105	a634aL [hypothetical protein] [1e-08]	N/A
OSYNE5|716L	297643	298062	140	A634L [hypothetical protein] [7e-84]	N/A
OSYNE5|719R	299913	300176	88	A635R [hypothetical protein] [3e-44]	N/A
OSYNE5|720R	300234	300671	146	A637R [hypothetical protein] [1e-63]	N/A
OSYNE5|721R	300765	302081	439	A643R [hypothetical protein] [0.0]	N/A
OSYNE5|724R	302120	302635	172	A644R [hypothetical protein] [3e-116]	Q5UQL1.1 [Uncharacterized protein R409] [3e-07]
OSYNE5|726R	302729	303106	126	A645R [hypothetical protein] [2e-71]	N/A
OSYNE5|328R	138196	138774	193	A647R [hypothetical protein] [8e-63]	N/A
OSYNE5|728R	303394	304167	258	A649R [hypothetical protein] [2e-164]	N/A
OSYNE5|132R	58349	58633	95	a650cR [hypothetical protein] [1e-08]	N/A
OSYNE5|065L	27303	27563	87	a650L [hypothetical protein] [2e-10]	N/A
OSYNE5|730L	304174	304767	198	A654L [hypothetical protein] [2e-127]	N/A
OSYNE5|731L	304263	304577	105	A655L [hypothetical protein] [1e-32]	N/A
OSYNE5|732L	304835	305482	216	A656L [hypothetical protein] [6e-68]	N/A
OSYNE5|733L	305669	306241	191	A659L [hypothetical protein] [2e-96]	N/A
OSYNE5|734R	305854	306162	103	a660R [hypothetical protein] [1e-22]	N/A
OSYNE5|735L	306265	306780	172	A662L [hypothetical protein] [8e-96]	Q54FR4.1 [PXMP2/4 family protein 4] [2e-08]
OSYNE5|736L	306866	307327	154	A664L [hypothetical protein] [2e-70]	N/A
OSYNE5|739L	308181	308648	156	A665L [hypothetical protein] [7e-88]	N/A
OSYNE5|748L	312215	315056	948	A666L [hypothetical protein] [0.0]	O94489.1 [Elongation factor 3] [0.0]
OSYNE5|749R	312665	312994	110	a667R [hypothetical protein] [5e-44]	N/A
OSYNE5|750R	313164	313382	73	a669R [hypothetical protein] [3e-42]	N/A
OSYNE5|753R	314546	315073	176	a670R [hypothetical protein] [7e-87]	N/A
OSYNE5|754R	315104	315751	216	A672R [hypothetical protein] [2e-110]	Q8GXE6.2 [Potassium channel] [4e-18]
OSYNE5|756R	315987	316637	217	A674R [Thymidylate synthase X] [1e-153]	O41156.1 [Probable thymidylate synthase] [1e-150]
OSYNE5|758R	316676	317788	371	A676R [hypothetical protein] [0.0]	N/A
OSYNE5|003L	1898	2083	62	a690R [hypothetical protein] [2e-16]	N/A

Table 2

tRNAs	start	end	nt
OSYNE5.tRNA1-LeuTAA	163841	163924	502
OSYNE5.tRNA2-LeuTAA	164067	164150	502
OSYNE5.tRNA3-AsnGTT	164259	164333	796
OSYNE5.tRNA4-GlyTCC	164336	164407	668
OSYNE5.tRNA5-AsnGTT	164430	164504	796
OSYNE5.tRNA6-LysCTT	164507	164580	765
OSYNE5.tRNA7-ArgTCT	164607	164679	602
OSYNE5.tRNA8-ArgTCT	164706	164778	602
OSYNE5.tRNA9-TyrGTA	165777	165862	584
OSYNE5.tRNA10-AspGTC	165889	165961	684
OSYNE5.tRNA11-LeuCAA	165962	166043	424
OSYNE5.tRNA12-ThrTGT	166046	166118	717
OSYNE5.tRNA13-LeuTAA	322020	322103	502
OSYNE5.tRNA14-LeuTAA	322211	322294	502

Table 3

DSyNE-5 ORF ID	PBCV-1 best match [e-value]	AA DSyNE-5	AA PBCV-1	% Identity	NCBI ID	PHI-BLAST	Swissprot best match (<1e-5)2
069R	A065R Prolyl 4-hydroxylase [3e-133]	227	242	175/230(76%)	NP_048433	Prolyl 4-hydroxylase	Q5UP57.1 Putative prolyl 4-hydroxylase [3e-23]
087R	A103R mRNA guanylyltransferase [0.0]	331	330	273/330(83%)	NP_048451	mRNA guanylyltransferase	Q84424.1 mRNA-capping enzyme [0.0]
090L	A105L hypothetical protein [3e-168]	291	284	218/285(76%)	NP_048453	Ubiquitin carboxyl-terminal hydrolase	A105L hypothetical protein [3e-168]
106L	A125L hypothetical protein [5e-132]	181	180	171/180(95%)	NP_048472	Transcription elongation factor	P49373.1 Transcription elongation factor S-II [3e-14]
128R	A153R hypothetical protein [0.0]	465	459	377/459(82%)	NP_048501	DNA or RNA helicase of superfamily II	Q5UQ46.1 Putative ATP-dependent RNA helicase L396 [1e-55]
149R	A176R hypothetical protein [9e-179]	269	268	232/268(87%)	NP_048514.2	Restriction endonuclease	Q5UQV1.1 Uncharacterized protein R354 [1e-12]
163L	A173L hypothetical protein [0.0]	276	288	245/278(88%)	NP_048521.1	Patatin-like phospholipase family	Q91F63.1 Probable lipid hydrolase 463L [3e-18]
278R	A193L hypothetical protein [0.0]	269	262	246/262(94%)	NP_048540	Proliferating cell nuclear antigen	Q84513.1 Probable DNA polymerase sliding clamp 1 [0.0]
301L	A265L hypothetical protein [4e-169]	250	249	219/249(88%)	NP_048819.2	Glycosyl hydrolase	N/A
380L	A324L hypothetical protein [0.0]	445	453	391/453(86%)	NP_048660	Transposase	Q5UQN9.1 Uncharacterized protein R449 [8e-09]
408L	A342L hypothetical protein [0.0]	565	576	519/558(93%)	NP_048699	2',3'-cyclic nucleotide 2'-phosphodiesterase	N/A
415L	A352L hypothetical protein [1e-133]	208	207	193/207(93%)	NP_048709	Vacuolar protein sorting-associated protein 72 homolog	Q5UQF7.1 Uncharacterized protein R469 Flags: Precursor [5e-07]
445R	A392R hypothetical protein [2e-176]	255	259	238/254(94%)	NP_048749.2	A32 virion packaging ATPase	Q196X2.1 Uncharacterized protein D88R [1e-42]
460R	A402R hypothetical protein [1e-150]	281	227	199/227(88%)	NP_048759	Leucyl-tRNA synthetase	N/A
469L	A407L hypothetical protein [5e-130]	211	210	195/210(93%)	NP_048764	ABC transporter	N/A
517L	A445L hypothetical protein [0.0]	463	462	427/462(92%)	NP_048802.2	ABC transporter	Q98496.1 Uncharacterized protein A445L [0.0]
536R	A464R Rnase III [0.0]	271	275	237/251(91%)	NP_048820	RNAse III	Q98514.1 Rnase III
538Lc	A467L hypothetical protein [0.0]	313	312	275/313(88%)	NP_048823.1	Ubiquitin--protein ligase molybdopterin-converting factor	N/A
540R	A468R hypothetical protein [0.0]	442	443	379/443(86%)	NP_048824.1	VV D5 family helicase	N/A
545R	A476R hypothetical protein [0.0]	325	324	303/324(94%)	NP_048832.1	Ribonucleotide reductase	P50650.1 Ribonucleoside-diphosphate reductase small chain [1e-135]
555R	A482R hypothetical protein [3e-128]	215	215	173/216(80%)	NP_048838.1	Viral transcription factor 2	N/A
561R	A488R hypothetical protein [0.0]	320	317	304/319(95%)	NP_048844.1	alanyl-tRNA synthetase	Q5UQL4.1 Uncharacterized protein L417 [9e-11]
565R	A494R hypothetical protein [0.0]	365	360	330/364(91%)	NP_048850.1	NCLDV-specific basal transcription factor A2L	Q98544.1 Putative transcription factor A494R [0.0]
621L	A548L hypothetical protein [0.0]	457	495	410/458(90%)	NP_048904.2	DNA/RNA helicase protein	Q91ZW3.1 SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 [4e-35]
626L	A551L dUTP pyrophosphatase [1e-77]	146	141	112/131(85%)	NP_048907.1	Uncharacterized	O41033.1 Deoxyuridine 5-triphosphate nucleotidohydrolase [1e-74]
627R	A552R hypothetical protein [0.0]	318	317	275/317(87%)	NP_048908.3	Transcription factor	N/A
711R	A629R hypothetical protein [0.0]	766	771	724/772(94%)	NP_048985.1	Ribonucleoside-diphosphate reductase large chain	Q03604.1 Ribonucleoside-diphosphate reductase large subunit [0.0]
724R	A644R hypothetical protein [3e-116]	172	173	155/170(91%)	NP_049000.2	Uncharacterized	Q5UQL1.1 Uncharacterized protein R409 [3e-07]
756R	A674R Thymidylate synthase X [1e-153]	217	216	196/216(91%)	NP_049030.1	Thymidylate synthase X	O41156.1 Probable thymidylate synthase [1e-150]

Table 4

ORF ID	Start	End	AA length	Swissprot best match (<1e-5)
OSYNE5|001R	827	1018	64	N/A
OSYNE5|002R	1786	1971	62	N/A
OSYNE5|004L	2230	2475	82	N/A
OSYNE5|005R	2314	2511	66	N/A
OSYNE5|006R	2923	4077	385	B4F777.1 [High mobility group nucleosome-binding protein 5] [2e-14]
OSYNE5|007R	2996	3373	126	Q7UZZ9.1 [Translation initiation factor IF-2] [6e-10]
OSYNE5|008L	3002	4087	362	P16884.4 [Neurofilament heavy polypeptide] [2e-09]
OSYNE5|09R	3530	4189	220	N/A
OSYNE5|011L	4191	4445	85	N/A
OSYNE5|012R	5592	5798	69	N/A
OSYNE5|013R	6034	6234	67	N/A
OSYNE5|014L	6258	6488	77	N/A
OSYNE5|020L	8141	8695	185	N/A
OSYNE5|022R	9122	9316	65	N/A
OSYNE5|023R	9213	9521	103	N/A
OSYNE5|024R	9776	9979	68	N/A
OSYNE5|028L	12612	13421	270	Q9X1E3.1 [Probable glycerol uptake facilitator protein] [3e-35]
OSYNE5|029R	12981	13202	74	N/A
OSYNE5|033L	14833	15087	85	N/A
OSYNE5|034R	14969	15283	105	N/A
OSYNE5|035L	15097	16152	352	N/A
OSYNE5|036R	15545	15805	87	N/A
OSYNE5|038R	17744	18487	248	N/A
OSYNE5|039L	17776	18030	85	N/A
OSYNE5|040L	18088	18435	116	N/A
OSYNE5|041L	18584	19210	209	N/A
OSYNE5|042R	18860	19123	88	N/A
OSYNE5|043R	18978	19181	68	N/A
OSYNE5|044R	19285	20196	304	N/A
OSYNE5|045L	19991	20242	84	N/A
OSYNE5|046L	20193	20375	61	N/A
OSYNE5|047R	20221	21012	264	N/A
OSYNE5|048R	20576	20851	92	N/A
OSYNE5|052R	22516	22743	76	N/A
OSYNE5|054L	22792	23019	76	N/A
OSYNE5|056R	24116	24430	105	N/A
OSYNE5|057L	24166	24435	90	N/A
OSYNE5|061R	25444	25686	81	N/A
OSYNE5|062L	26384	26599	72	N/A
OSYNE5|066R	27546	27776	77	N/A
OSYNE5|070L	28994	29176	61	N/A
OSYNE5|072L	30209	30412	68	N/A
OSYNE5|074R	30893	31147	85	N/A
OSYNE5|078R	32902	33180	93	N/A
OSYNE5|079L	33097	33390	98	N/A
OSYNE5|080R	33214	33462	83	N/A
OSYNE5|081L	33712	33915	68	N/A
OSYNE5|083R	34191	34394	68	N/A
OSYNE5|086L	37247	37633	129	N/A
OSYNE5|089R	38350	38535	62	N/A
OSYNE5|091L	39024	39239	72	N/A
OSYNE5|095L	41030	41221	64	N/A
OSYNE5|097R	41358	41660	101	N/A
OSYNE5|100L	42643	42861	73	N/A
OSYNE5|103L	45152	45340	63	N/A
OSYNE5|105L	47617	47868	84	N/A
OSYNE5|107R	48073	48282	70	N/A
OSYNE5|109R	49045	49245	67	N/A
OSYNE5|111L	49343	49933	197	N/A

OSYNE5|113L	49867	50118	84	N/A
OSYNE5|115L	50225	50674	150	N/A
OSYNE5|121L	52065	52349	95	N/A
OSYNE5|124L	53204	53401	66	N/A
OSYNE5|126L	54577	54831	85	N/A
OSYNE5|129R	57231	57413	61	N/A
OSYNE5|130L	57531	57893	121	N/A
OSYNE5|131L	57667	57951	95	N/A
OSYNE5|134R	59178	59366	63	N/A
OSYNE5|136R	59263	59457	65	N/A
OSYNE5|138R	59780	59965	62	N/A
OSYNE5|140L	60068	60262	65	N/A
OSYNE5|144L	62103	62297	65	N/A
OSYNE5|146R	63360	63614	85	N/A
OSYNE5|151L	64808	65041	78	N/A
OSYNE5|152L	65117	65410	98	N/A
OSYNE5|154L	65524	65706	61	N/A
OSYNE5|156R	65783	66013	77	N/A
OSYNE5|158L	66078	66284	69	N/A
OSYNE5|159L	67216	67452	79	N/A
OSYNE5|162L	67654	67851	66	N/A
OSYNE5|167L	70269	70460	64	N/A
OSYNE5|169L	70736	71014	93	N/A
OSYNE5|171L	71121	71384	88	N/A
OSYNE5|173R	71687	71908	74	N/A
OSYNE5|174R	71912	72316	135	N/A
OSYNE5|177L	72364	72576	71	N/A
OSYNE5|179L	73125	73391	89	N/A
OSYNE5|180L	73401	73586	62	N/A
OSYNE5|181L	73596	73793	66	N/A
OSYNE5|182L	74028	74213	62	N/A
OSYNE5|183L	74541	74795	85	N/A
OSYNE5|184L	74976	75185	70	N/A
OSYNE5|185L	76901	77083	61	N/A
OSYNE5|186L	77010	77255	82	N/A
OSYNE5|187L	77336	77587	84	N/A
OSYNE5|188R	77995	78276	94	N/A
OSYNE5|189L	79316	79501	62	N/A
OSYNE5|190L	80013	80723	237	N/A
OSYNE5|191L	80970	81218	83	N/A
OSYNE5|192L	82306	82788	161	N/A
OSYNE5|194L	84456	84701	82	N/A
OSYNE5|195L	84903	85178	92	N/A
OSYNE5|196L	85179	85418	80	N/A
OSYNE5|197L	85425	85676	84	N/A
OSYNE5|198R	86843	87124	94	N/A
OSYNE5|199L	87045	87242	66	N/A
OSYNE5|200R	87162	87353	64	N/A
OSYNE5|201R	87381	87599	73	N/A
OSYNE5|202R	87783	87977	65	N/A
OSYNE5|203L	88491	88706	72	N/A
OSYNE5|205L	89184	89468	95	N/A
OSYNE5|206L	89673	89855	61	N/A
OSYNE5|207L	89862	90434	191	N/A
OSYNE5|208L	90453	90638	62	N/A
OSYNE5|209L	90705	90998	98	N/A
OSYNE5|210L	91035	91217	61	N/A
OSYNE5|211L	91174	91398	75	N/A
OSYNE5|212R	91470	91844	125	N/A
OSYNE5|213L	91608	91904	99	N/A
OSYNE5|214L	92703	93053	117	N/A

OSYNE5|215L	93090	93344	85	N/A
OSYNE5|217L	94033	94350	106	N/A
OSYNE5|218R	94349	94543	65	N/A
OSYNE5|219R	94774	95016	81	N/A
OSYNE5|221R	94838	95134	99	N/A
OSYNE5|222L	94947	95144	66	N/A
OSYNE5|224R	95737	95982	82	N/A
OSYNE5|225R	95789	96187	133	N/A
OSYNE5|226R	96275	96535	87	N/A
OSYNE5|227R	96800	96988	63	N/A
OSYNE5|228R	97091	97300	70	N/A
OSYNE5|231L	98536	98799	88	N/A
OSYNE5|237L	101366	101710	115	N/A
OSYNE5|238L	101428	101637	70	N/A
OSYNE5|240R	102071	102253	61	N/A
OSYNE5|241R	102254	102523	90	N/A
OSYNE5|247R	103261	103473	71	N/A
OSYNE5|249R	104475	104852	126	N/A
OSYNE5|251L	105736	106254	173	N/A
OSYNE5|254L	106727	107017	97	N/A
OSYNE5|257L	108752	109024	91	N/A
OSYNE5|259L	109418	109642	75	N/A
OSYNE5|262R	110300	110956	219	N/A
OSYNE5|264R	110914	111099	62	N/A
OSYNE5|265L	111229	111486	86	N/A
OSYNE5|266R	111317	111619	101	N/A
OSYNE5|267R	111570	111839	90	N/A
OSYNE5|271R	112748	112936	63	N/A
OSYNE5|280R	116728	116961	78	N/A
OSYNE5|281R	117621	118028	136	N/A
OSYNE5|282R	117979	118182	68	N/A
OSYNE5|283R	118219	118434	72	N/A
OSYNE5|288R	120986	121402	139	N/A
OSYNE5|289R	122548	122988	147	N/A
OSYNE5|290R	123028	123255	76	N/A
OSYNE5|294R	125969	126232	88	N/A
OSYNE5|295R	126958	127230	91	N/A
OSYNE5|296R	127332	127880	183	P26840.1 [Probable macrolide acetyltransferase] [3e-38]
OSYNE5|298R	128041	128325	95	N/A
OSYNE5|299L	128647	128913	89	N/A
OSYNE5|302R	129554	129808	85	N/A
OSYNE5|303R	129877	130062	62	N/A
OSYNE5|306R	130484	130882	133	N/A
OSYNE5|307R	130789	131064	92	N/A
OSYNE5|308R	131009	131236	76	N/A
OSYNE5|309R	131727	132860	378	P52284.1 [Adenine-specific methyltransferase CviRI] [0.0]
OSYNE5|310L	131759	132130	124	N/A
OSYNE5|311L	132227	132547	107	N/A
OSYNE5|312L	132432	132716	95	N/A
OSYNE5|313R	132893	134608	572	N/A
OSYNE5|314L	133099	133317	73	N/A
OSYNE5|315R	133146	133355	70	N/A
OSYNE5|316R	133933	134169	79	N/A
OSYNE5|317L	134095	134304	70	N/A
OSYNE5|318L	134332	134547	72	N/A
OSYNE5|319R	134630	135745	372	P10835.1 [Adenine-specific methyltransferase CviBIII] [3e-65]
OSYNE5|320L	134740	135027	96	N/A
OSYNE5|321L	135055	135630	192	N/A
OSYNE5|322L	135722	135934	71	N/A
OSYNE5|323R	135776	136954	393	P52284.1 [Adenine-specific methyltransferase CviRI] [5e-151]
OSYNE5|324L	136312	136677	122	N/A

OSYNE5|325L	136358	136663	102	N/A
OSYNE5|327R	137903	138154	84	N/A
OSYNE5|329L	138629	138817	63	N/A
OSYNE5|334L	141855	142043	63	N/A
OSYNE5|336L	143411	143596	62	N/A
OSYNE5|338L	143777	143959	61	N/A
OSYNE5|339L	144084	144518	145	N/A
OSYNE5|340R	144517	144717	67	N/A
OSYNE5|343L	144793	144984	64	N/A
OSYNE5|348R	146289	146486	66	N/A
OSYNE5|350R	146583	146930	116	N/A
OSYNE5|351L	147527	147772	82	N/A
OSYNE5|353L	148060	149010	317	N/A
OSYNE5|354R	148110	148346	79	N/A
OSYNE5|355L	148866	149087	74	N/A
OSYNE5|356L	149271	149459	63	N/A
OSYNE5|357R	149856	150113	86	N/A
OSYNE5|359R	150795	151256	154	N/A
OSYNE5|363L	152088	152291	68	N/A
OSYNE5|373L	156647	156973	109	N/A
OSYNE5|381R	160599	161282	228	N/A
OSYNE5|383R	161525	161707	61	N/A
OSYNE5|384R	161715	161924	70	N/A
OSYNE5|387L	163254	163577	108	N/A
OSYNE5|390L	165462	165680	73	N/A
OSYNE5|391L	165908	166210	101	N/A
OSYNE5|392L	165943	166131	63	N/A
OSYNE5|393R	165971	166195	75	N/A
OSYNE5|396R	166432	166875	148	N/A
OSYNE5|400R	167720	167944	75	N/A
OSYNE5|401L	168235	168471	79	N/A
OSYNE5|405L	170390	170572	61	N/A
OSYNE5|407R	170579	170809	77	N/A
OSYNE5|413L	172847	173104	86	N/A
OSYNE5|414R	173239	173448	70	N/A
OSYNE5|416L	173532	173735	68	N/A
OSYNE5|422L	175996	176253	86	N/A
OSYNE5|424R	177671	177859	63	N/A
OSYNE5|425L	178005	178277	91	N/A
OSYNE5|426L	178325	178576	84	N/A
OSYNE5|429L	179656	179943	96	N/A
OSYNE5|430L	179720	180004	95	N/A
OSYNE5|432R	180751	180957	69	N/A
OSYNE5|435L	182383	182613	77	N/A
OSYNE5|436R	183111	183296	62	N/A
OSYNE5|440R	184408	184740	111	N/A
OSYNE5|441R	184755	184991	79	N/A
OSYNE5|442R	184955	185428	158	N/A
OSYNE5|443R	184963	185208	82	N/A
OSYNE5|446R	186424	186633	70	N/A
OSYNE5|448L	187135	188619	495	N/A
OSYNE5|451R	189363	189551	63	N/A
OSYNE5|454L	190336	190722	129	N/A
OSYNE5|455L	190614	190805	64	N/A
OSYNE5|457L	190996	191193	66	N/A
OSYNE5|459L	191576	192049	158	N/A
OSYNE5|463R	193280	193627	116	N/A
OSYNE5|464L	193307	193555	83	N/A
OSYNE5|465R	193776	193967	64	N/A
OSYNE5|466L	194686	194886	67	N/A
OSYNE5|467L	194965	195153	63	N/A

OSYNE5|475L	198198	198407	70	N/A
OSYNE5|477R	198533	198925	131	N/A
OSYNE5|481R	200125	200370	82	N/A
OSYNE5|485R	201865	202083	73	N/A
OSYNE5|488R	202608	203012	135	N/A
OSYNE5|489R	202756	203028	91	N/A
OSYNE5|490L	202794	203036	81	N/A
OSYNE5|493R	203397	203621	75	N/A
OSYNE5|496L	204793	204993	67	N/A
OSYNE5|498L	206030	206293	88	N/A
OSYNE5|500R	207020	207241	74	N/A
OSYNE5|502R	208202	208429	76	N/A
OSYNE5|512L	212153	212815	221	N/A
OSYNE5|513R	212394	212777	128	N/A
OSYNE5|514R	212600	212794	65	N/A
OSYNE5|516R	212964	213203	80	N/A
OSYNE5|518R	213284	213508	75	N/A
OSYNE5|520R	214675	214950	92	N/A
OSYNE5|523R	215302	215577	92	N/A
OSYNE5|532R	220500	220688	63	N/A
OSYNE5|533R	220977	221168	64	N/A
OSYNE5|539L	223245	223451	69	N/A
OSYNE5|542L	225283	225486	68	N/A
OSYNE5|543L	225523	225705	61	N/A
OSYNE5|547R	226820	227155	112	N/A
OSYNE5|548L	227118	227372	85	N/A
OSYNE5|551R	228623	228895	91	N/A
OSYNE5|553R	228883	229128	82	N/A
OSYNE5|554L	229326	229580	85	N/A
OSYNE5|556L	229648	229971	108	N/A
OSYNE5|558R	230189	230395	69	N/A
OSYNE5|560R	231093	231359	89	N/A
OSYNE5|567L	234429	234647	73	N/A
OSYNE5|569R	234982	235203	74	N/A
OSYNE5|570L	235229	235573	115	N/A
OSYNE5|571R	235897	236427	177	N/A
OSYNE5|573L	235982	236167	62	N/A
OSYNE5|575R	236536	236727	64	N/A
OSYNE5|576L	237059	237301	81	N/A
OSYNE5|581R	238884	239447	188	N/A
OSYNE5|582L	238893	239093	67	N/A
OSYNE5|583R	238972	239160	63	N/A
OSYNE5|584L	239133	239363	77	N/A
OSYNE5|593L	242245	242439	65	N/A
OSYNE5|598L	243595	243819	75	N/A
OSYNE5|602R	246193	246546	118	N/A
OSYNE5|606R	247526	247711	62	N/A
OSYNE5|608L	247725	248012	96	N/A
OSYNE5|610R	249160	249354	65	N/A
OSYNE5|611R	250321	250527	69	N/A
OSYNE5|612L	250588	250770	61	N/A
OSYNE5|613R	250730	250921	64	N/A
OSYNE5|614R	250969	251217	83	N/A
OSYNE5|615R	251497	251688	64	N/A
OSYNE5|618L	252417	252767	117	N/A
OSYNE5|620R	253894	254148	85	N/A
OSYNE5|622L	254516	254731	72	N/A
OSYNE5|623R	254560	254751	64	N/A
OSYNE5|625R	255019	255306	96	N/A
OSYNE5|628L	256826	257011	62	N/A
OSYNE5|630R	256960	257361	134	N/A

OSYNE5|633R	258568	258804	79	N/A
OSYNE5|634R	258904	259089	62	N/A
OSYNE5|637R	261824	262051	76	N/A
OSYNE5|638R	262271	262591	107	N/A
OSYNE5|641L	262868	263158	97	N/A
OSYNE5|642R	263267	263539	91	N/A
OSYNE5|643R	264253	264528	92	N/A
OSYNE5|647R	266262	266513	84	N/A
OSYNE5|648R	266377	266625	83	N/A
OSYNE5|650L	266559	266783	75	N/A
OSYNE5|653L	267517	267765	83	N/A
OSYNE5|655R	267877	268095	73	N/A
OSYNE5|657R	269070	269273	68	N/A
OSYNE5|662R	271074	271262	63	N/A
OSYNE5|664R	271519	271719	67	N/A
OSYNE5|665R	272266	272499	78	N/A
OSYNE5|667R	273788	274006	73	N/A
OSYNE5|668R	274259	274906	216	N/A
OSYNE5|669L	275353	275538	62	N/A
OSYNE5|670L	275459	275644	62	N/A
OSYNE5|672R	276636	277202	189	N/A
OSYNE5|675L	277855	278052	66	N/A
OSYNE5|677R	278428	279345	306	Q3SYV9.1 [ADP-ribosylhydrolase 3] [6e-18]
OSYNE5|679L	279688	279939	84	P51423.2 [Ubiquitin-60S ribosomal protein L40] [3e-46]
OSYNE5|683L	281224	281481	86	N/A
OSYNE5|685R	281990	282193	68	N/A
OSYNE5|686R	282021	283382	454	N/A
OSYNE5|687L	282682	282915	78	N/A
OSYNE5|692L	286579	286782	68	N/A
OSYNE5|693R	287568	287798	77	N/A
OSYNE5|697R	289639	290061	141	N/A
OSYNE5|698R	289787	290182	132	N/A
OSYNE5|700R	290222	290443	74	N/A
OSYNE5|705L	292997	293476	160	N/A
OSYNE5|707L	293325	293645	107	N/A
OSYNE5|708L	293765	293977	71	N/A
OSYNE5|710L	294932	295195	88	N/A
OSYNE5|713L	296238	296555	106	N/A
OSYNE5|714L	296609	296947	113	N/A
OSYNE5|717R	298128	299750	541	Q1DDB7.1 [CTP synthase] [0.0]
OSYNE5|722L	301256	301492	79	N/A
OSYNE5|723L	301709	301927	73	N/A
OSYNE5|725L	302395	302646	84	N/A
OSYNE5|727L	303041	303232	64	N/A
OSYNE5|737R	307363	307623	87	N/A
OSYNE5|740L	308683	311295	871	Q9M2L4.1 [Calcium-transporting ATPase plasma membrane] [2e-170]
OSYNE5|741R	309059	309508	150	N/A
OSYNE5|742R	309566	310048	161	N/A
OSYNE5|743R	310103	310603	167	N/A
OSYNE5|744R	310985	311173	63	N/A
OSYNE5|745R	311246	311548	101	N/A
OSYNE5|746R	311455	311898	148	N/A
OSYNE5|747L	311906	312169	88	N/A
OSYNE5|751L	313651	313836	62	N/A
OSYNE5|752R	313826	314170	115	N/A
OSYNE5|755L	315954	316145	64	N/A
OSYNE5|757R	316258	316467	70	N/A
OSYNE5|759L	316884	317096	71	N/A
OSYNE5|760L	317337	317618	94	N/A
OSYNE5|762L	318046	318396	117	N/A
OSYNE5|763L	319685	319981	99	N/A

OSYNE5|764L	319711	319953	81	N/A
OSYNE5|765R	320171	321349	393	N/A

Table 5

ORF ID	Start	End	AA length	PBCV-1 best match [e-value]	Swissprot best match (<1e-5)
OSYNE5|001R	827	1018	64	N/A	N/A
OSYNE5|002R	1786	1971	62	N/A	N/A
OSYNE5|003L	1898	2083	62	a690R [hypothetical protein] [2e-16]	N/A
OSYNE5|004L	2230	2475	82	N/A	N/A
OSYNE5|005R	2314	2511	66	N/A	N/A
OSYNE5|006R	2923	4077	385	N/A	B4F777.1 [High mobility group nucleosome-binding protein 5] [2e-14]
OSYNE5|007R	2996	3373	126	N/A	Q7UZZ9.1 [Translation initiation factor IF-2] [6e-10]
OSYNE5|008L	3002	4087	362	N/A	P16884.4 [Neurofilament heavy polypeptide] [2e-09]
OSYNE5|009R	3530	4189	220	N/A	N/A
OSYNE5|010R	4153	5079	309	A034R [Protein kinase] [0.0]	N/A
OSYNE5|011L	4191	4445	85	N/A	N/A
OSYNE5|012R	5592	5798	69	N/A	N/A
OSYNE5|013R	6034	6234	67	N/A	N/A
OSYNE5|014L	6258	6488	77	N/A	N/A
OSYNE5|015L	6885	7199	105	A037L [hypothetical protein] [4e-52]	N/A
OSYNE5|016L	7239	7685	149	A039L [hypothetical protein] [2e-66]	Q49484.1 [SKP1-like protein 11] [1e-31]
OSYNE5|017R	7248	7727	160	a038R [hypothetical protein] [7e-13]	N/A
OSYNE5|018R	7770	9017	416	A041R [hypothetical protein] [0.0]	Q4JV51.1 [Translation initiation factor IF-2] [1e-09]
OSYNE5|019L	7863	8135	91	a042L [hypothetical protein] [6e-16]	N/A
OSYNE5|020L	8141	8695	185	N/A	N/A
OSYNE5|021L	9014	10849	612	A044L [hypothetical protein] [0.0]	Q5UR45.1 [Putative AAA family ATPase L572] [2e-19]
OSYNE5|022R	9122	9316	65	N/A	N/A
OSYNE5|023R	9213	9521	103	N/A	N/A
OSYNE5|024R	9776	9979	68	N/A	N/A
OSYNE5|025R	10917	11294	126	A048R [hypothetical protein] [8e-63]	N/A
OSYNE5|026L	11291	11950	220	A049L [hypothetical protein] [6e-120]	O07592.1 [Putative glycerophosphoryl diester phosphodiesterase] [6e-27]
OSYNE5|027L	11967	12392	142	A050L [Pyrimidine dimer-specific glycosylase] [2e-84]	P04418.1 [Endonuclease V] [7e-27]
OSYNE5|028L	12612	13421	270	N/A	Q9X1E3.1 [Probable glycerol uptake facilitator protein] [3e-35]
OSYNE5|029R	12981	13202	74	N/A	N/A
OSYNE5|030R	13557	14645	363	A053R [hypothetical protein] [0.0]	P52643.1 [D-lactate dehydrogenase] [9e-85]
OSYNE5|031L	13673	13996	108	a054L [hypothetical protein] [1e-56]	N/A
OSYNE5|032L	14162	14407	82	a056L [hypothetical protein] [1e-39]	N/A
OSYNE5|033L	14833	15087	85	N/A	N/A
OSYNE5|034R	14969	15283	105	N/A	N/A
OSYNE5|035L	15097	16152	352	N/A	N/A
OSYNE5|036R	15545	15805	87	N/A	N/A
OSYNE5|037R	17004	17711	236	A057aR [hypothetical protein] [1e-23]	N/A
OSYNE5|038R	17744	18487	248	N/A	N/A
OSYNE5|039L	17776	18030	85	N/A	N/A
OSYNE5|040L	18088	18435	116	N/A	N/A
OSYNE5|041L	18584	19210	209	N/A	N/A
OSYNE5|042R	18860	19123	88	N/A	N/A
OSYNE5|043R	18978	19181	68	N/A	N/A
OSYNE5|044R	19285	20196	304	N/A	N/A
OSYNE5|045L	19991	20242	84	N/A	N/A
OSYNE5|046L	20193	20375	61	N/A	N/A
OSYNE5|047R	20221	21012	264	N/A	N/A
OSYNE5|048R	20576	20851	92	N/A	N/A
OSYNE5|049R	21733	21936	68	A067R [hypothetical protein] [5e-32]	N/A
OSYNE5|050R	21969	23033	355	A071R [hypothetical protein] [0.0]	N/A
OSYNE5|051L	22290	22592	101	a073L [hypothetical protein] [5e-22]	N/A
OSYNE5|052R	22516	22743	76	N/A	N/A
OSYNE5|053L	22626	22886	87	a074L [hypothetical protein] [1e-13]	N/A
OSYNE5|054L	22792	23019	76	N/A	N/A
OSYNE5|055L	23043	23882	280	A075L [hypothetical protein] [0.0]	N/A
OSYNE5|056R	24116	24430	105	N/A	N/A
OSYNE5|057L	24166	24435	90	N/A	N/A
OSYNE5|058L	24499	24798	100	A076L [hypothetical protein] [7e-35]	N/A
OSYNE5|059L	24767	25066	100	A077L [hypothetical protein] [4e-54]	N/A
OSYNE5|060R	25164	26060	299	A078R [N-carbamoylputrescine amidohydrolase] [0.0]	Q3HVN1.1 [N-carbamoylputrescine amidase] [2e-94]
OSYNE5|061R	25444	25686	81	N/A	N/A
OSYNE5|062L	26384	26599	72	N/A	N/A
OSYNE5|063L	26796	27025	76	a078cL [hypothetical protein] [2e-10]	N/A
OSYNE5|064R	26831	27559	243	A079R [hypothetical protein] [2e-160]	N/A
OSYNE5|065L	27303	27563	87	a650L [hypothetical protein] [2e-10]	N/A
OSYNE5|066R	27546	27776	77	N/A	N/A
OSYNE5|067L	27563	28132	190	A081L [hypothetical protein] [2e-108]	N/A
OSYNE5|068L	28205	28735	177	A084L [hypothetical protein] [4e-74]	N/A
OSYNE5|069R	28881	29561	227	A085R [Prolyl 4-hydroxylase] [3e-133]	Q5UP57.1 [Putative prolyl 4-hydroxylase] [3e-23]
OSYNE5|070L	28994	29176	61	N/A	N/A
OSYNE5|071R	29610	30524	305	A088R [hypothetical protein] [9e-138]	N/A
OSYNE5|072L	30209	30412	68	N/A	N/A
OSYNE5|073R	30634	31122	163	A088R [hypothetical protein] [4e-35]	N/A
OSYNE5|074R	30893	31147	85	N/A	N/A
OSYNE5|075L	31204	32505	434	A092/093L [hypothetical protein] [0.0]	N/A
OSYNE5|076L	32543	33643	367	A094L [beta-1-3-glucanase] [0.0]	P23903.1 [Glucan endo-13-beta-glucosidase A1] [2e-26]
OSYNE5|077R	32892	33118	75	a095R [hypothetical protein] [9e-43]	N/A
OSYNE5|078R	32902	33180	93	N/A	N/A
OSYNE5|079L	33097	33390	98	N/A	N/A
OSYNE5|080R	33214	33462	83	N/A	N/A
OSYNE5|081L	33712	33915	68	N/A	N/A
OSYNE5|082R	33797	35488	564	A098R [Hyaluronan synthase] [0.0]	O08650.2 [Hyaluronan synthase 3] [3e-85]
OSYNE5|083R	34191	34394	68	N/A	N/A
OSYNE5|084R	35643	37430	595	A100R [Glutamine:fructose-6-phosphate amidotransferase] [0.0]	Q7WE36.3 [Glutamine--fructose-6-phosphate aminotransferase] [0.0]
OSYNE5|085L	35781	36044	88	a101L [hypothetical protein] [6e-27]	N/A
OSYNE5|086L	37247	37633	129	N/A	N/A
OSYNE5|087R	37599	38591	331	A103R [mRNA guanylyltransferase] [0.0]	Q84424.1 [GTP--RNA guanylyltransferase] [0.0]
OSYNE5|088L	38079	38468	130	a104L [hypothetical protein] [3e-22]	N/A
OSYNE5|089R	38350	38535	62	N/A	N/A
OSYNE5|090L	38578	39450	291	A105L [hypothetical protein] [3e-168]	Q6DCJ1.2 [Ubiquitin carboxyl-terminal hydrolase 22-B] [1e-06]
OSYNE5|091L	39024	39239	72	N/A	N/A
OSYNE5|092L	39485	40357	291	A107L [hypothetical protein] [0.0]	P61998.1 [Transcription initiation factor IIB] [4e-09]

OSYNE5|093R	40181	40369	63	A108aR [hypothetical protein] [2e-26]	N/A
OSYNE5|094L	40459	40944	162	A108bL [hypothetical protein] [8e-110]	N/A
OSYNE5|095L	41030	41221	64	N/A	N/A
OSYNE5|096R	41069	43651	861	A111/114R [hypothetical protein] [0.0]	N/A
OSYNE5|097R	41358	41660	101	N/A	N/A
OSYNE5|098L	41540	42064	175	a113L [hypothetical protein] [2e-35]	N/A
OSYNE5|099R	42369	42614	82	a116R [hypothetical protein] [1e-21]	N/A
OSYNE5|100L	42643	42861	73	N/A	N/A
OSYNE5|101R	43699	44739	347	A118R [GDP-D-mannose dehydratase] [0.0]	Q9JRN5.1 [GDP-mannose 46-dehydratase] [1e-142]
OSYNE5|102R	44759	45073	105	A121R [hypothetical protein] [1e-67]	N/A
OSYNE5|103L	45152	45340	63	N/A	N/A
OSYNE5|104R	47015	47935	307	A122/123R [hypothetical protein] [1e-156]	Q37893.1 [Pre-neck appendage protein] [1e-11]
OSYNE5|105L	47617	47868	84	N/A	N/A
OSYNE5|106L	47937	48479	181	A125L [hypothetical protein] [5e-132]	P49373.1 [Transcription elongation factor S-II] [3e-14]
OSYNE5|107R	48073	48282	70	N/A	N/A
OSYNE5|108R	48513	49229	239	A127R [hypothetical protein] [5e-153]	N/A
OSYNE5|109R	49045	49245	67	N/A	N/A
OSYNE5|110R	49301	50455	385	A154L [hypothetical protein] [1e-174]	N/A
OSYNE5|111L	49343	49933	197	N/A	N/A
OSYNE5|112R	49386	49595	70	a156L [hypothetical protein] [2e-06]	N/A
OSYNE5|113L	49867	50118	84	N/A	N/A
OSYNE5|114L	50036	50221	62	a155R [hypothetical protein] [2e-16]	N/A
OSYNE5|115L	50225	50674	150	N/A	N/A
OSYNE5|116R	50475	50792	106	A130R [hypothetical protein] [4e-47]	N/A
OSYNE5|117L	50785	51192	136	A131L [hypothetical protein] [7e-78]	N/A
OSYNE5|118L	51333	51851	173	A134L [hypothetical protein] [8e-94]	N/A
OSYNE5|119L	51824	52147	108	A135L [hypothetical protein] [2e-25]	N/A
OSYNE5|120R	51897	52337	147	A136R [hypothetical protein] [1e-60]	N/A
OSYNE5|121L	52065	52349	95	N/A	N/A
OSYNE5|122R	52398	52619	74	A137R [hypothetical protein] [2e-24]	N/A
OSYNE5|123R	52665	53471	269	A138R [hypothetical protein] [2e-87]	N/A
OSYNE5|124L	53204	53401	66	N/A	N/A
OSYNE5|125L	53468	53779	104	A139L [hypothetical protein] [3e-47]	N/A
OSYNE5|126L	54577	54831	85	N/A	N/A
OSYNE5|127L	56426	56866	147	A150L [hypothetical protein] [1e-60]	N/A
OSYNE5|128R	56962	58356	465	A153R [hypothetical protein] [0.0]	Q5UQ46.1 [Putative ATP-dependent RNA helicase L396] [1e-55]
OSYNE5|129R	57231	57413	61	N/A	N/A
OSYNE5|130L	57531	57893	121	N/A	N/A
OSYNE5|131L	57667	57951	95	N/A	N/A
OSYNE5|132R	58349	58633	95	a650cR [hypothetical protein] [1e-08]	N/A
OSYNE5|133L	58359	59171	271	A315L [hypothetical protein] [2e-87]	P13329.1 [Probable mobile endonuclease B] [7e-08]
OSYNE5|134R	59178	59366	63	N/A	N/A
OSYNE5|135L	59252	59590	113	A157L [hypothetical protein] [8e-63]	N/A
OSYNE5|136R	59263	59457	65	N/A	N/A
OSYNE5|137L	59638	59958	107	A158L [hypothetical protein] [3e-38]	N/A
OSYNE5|138R	59780	59965	62	N/A	N/A
OSYNE5|139R	59966	60295	110	A159R [hypothetical protein] [5e-24]	N/A
OSYNE5|140L	60068	60262	65	N/A	N/A
OSYNE5|141R	60114	60551	146	A161R [hypothetical protein] [5e-26]	N/A
OSYNE5|142L	60552	61772	407	A162L [hypothetical protein] [0.0]	N/A
OSYNE5|143R	61813	63159	449	A163R [hypothetical protein] [0.0]	N/A
OSYNE5|144L	62103	62297	65	N/A	N/A
OSYNE5|145L	62745	63023	93	a164L [hypothetical protein] [3e-30]	N/A
OSYNE5|146R	63360	63614	85	N/A	N/A
OSYNE5|147L	63386	63724	113	A165L [hypothetical protein] [6e-67]	N/A
OSYNE5|148L	63755	64216	154	A165aL [hypothetical protein] [8e-84]	N/A
OSYNE5|149R	64279	65085	269	A166R [hypothetical protein] [9e-179]	Q5UQV1.1 [Uncharacterized protein R354] [1e-12]
OSYNE5|150L	64483	64935	151	a167L [hypothetical protein] [2e-24]	N/A
OSYNE5|151L	64808	65041	78	N/A	N/A
OSYNE5|152L	65117	65410	98	N/A	N/A
OSYNE5|153R	65124	65621	166	A168R [hypothetical protein] [1e-106]	N/A
OSYNE5|154L	65524	65706	61	N/A	N/A
OSYNE5|155R	65622	66704	361	A169R [Aspartate transcarbamylase] [0.0]	Q43087.1 [Aspartate carbamoyltransferase 2 chloroplastic] [5e-98]
OSYNE5|156R	65783	66013	77	N/A	N/A
OSYNE5|157L	65819	66148	110	a170L [hypothetical protein] [5e-50]	N/A
OSYNE5|158L	66078	66284	69	N/A	N/A
OSYNE5|159L	67216	67452	79	N/A	N/A
OSYNE5|160R	67402	67629	76	a043R [hypothetical protein] [3e-16]	N/A
OSYNE5|161R	67466	67990	175	A171R [hypothetical protein] [4e-111]	N/A
OSYNE5|162L	67654	67851	66	N/A	N/A
OSYNE5|163L	67993	68820	276	A173L [hypothetical protein] [0.0]	Q91F63.1 [Probable lipid hydrolase 463L] [3e-18]
OSYNE5|164L	68941	69150	70	A176L [hypothetical protein] [3e-41]	N/A
OSYNE5|165L	69171	69473	101	A176L [hypothetical protein] [2e-18]	N/A
OSYNE5|166R	69979	70716	246	A177R [hypothetical protein] [1e-154]	N/A
OSYNE5|167L	70269	70460	64	N/A	N/A
OSYNE5|168L	70279	70497	73	a178L [hypothetical protein] [6e-24]	N/A
OSYNE5|169L	70736	71014	93	N/A	N/A
OSYNE5|170R	70804	71676	291	A478L [hypothetical protein] [8e-112]	Q5UQL9.1 [Uncharacterized protein R423] [1e-40]
OSYNE5|171L	71121	71384	88	N/A	N/A
OSYNE5|172L	71671	71955	95	A250R [Potassium ioin channel protein (Kcv)] [1e-55]	Q84568.1 [Potassium channel protein kcv] [2e-52]
OSYNE5|173R	71687	71908	74	N/A	N/A
OSYNE5|174R	71912	72316	135	N/A	N/A
OSYNE5|175L	71978	72844	289	A248R [Protein kinase] [1e-158]	Q98NX5.3 [Calcium/calmodulin-dependent protein kinase type 1G] [1e-27]
OSYNE5|176R	71994	72422	143	a249L [hypothetical protein] [9e-38]	N/A
OSYNE5|177L	72364	72576	71	N/A	N/A
OSYNE5|178R	73060	75918	953	A122/123R [hypothetical protein] [1e-12]	N/A
OSYNE5|179L	73125	73391	89	N/A	N/A
OSYNE5|180L	73401	73586	62	N/A	N/A
OSYNE5|181L	73596	73793	66	N/A	N/A
OSYNE5|182L	74028	74213	62	N/A	N/A
OSYNE5|183L	74541	74795	85	N/A	N/A
OSYNE5|184L	74976	75185	70	N/A	N/A
OSYNE5|185L	76901	77083	61	N/A	N/A
OSYNE5|186L	77010	77255	82	N/A	N/A

OSYNE5|187L	77336	77587	84	N/A	N/A
OSYNE5|188R	77995	78276	94	N/A	N/A
OSYNE5|189L	79316	79501	62	N/A	N/A
OSYNE5|190L	80013	80723	237	N/A	N/A
OSYNE5|191L	80970	81218	83	N/A	N/A
OSYNE5|192L	82306	82788	161	N/A	N/A
OSYNE5|193R	84424	88875	1484	A025/027/029L [hypothetical protein] [1e-10]	N/A
OSYNE5|194L	84456	84701	82	N/A	N/A
OSYNE5|195L	84903	85178	92	N/A	N/A
OSYNE5|196L	85179	85418	80	N/A	N/A
OSYNE5|197L	85425	85676	84	N/A	N/A
OSYNE5|198R	86843	87124	94	N/A	N/A
OSYNE5|199L	87045	87242	66	N/A	N/A
OSYNE5|200R	87162	87353	64	N/A	N/A
OSYNE5|201R	87381	87599	73	N/A	N/A
OSYNE5|202R	87783	87977	65	N/A	N/A
OSYNE5|203L	88491	88706	72	N/A	N/A
OSYNE5|204R	88972	93492	1507	A122/123R [hypothetical protein] [3e-11]	N/A
OSYNE5|205L	89184	89468	95	N/A	N/A
OSYNE5|206L	89673	89855	61	N/A	N/A
OSYNE5|207L	89862	90434	191	N/A	N/A
OSYNE5|208L	90453	90638	62	N/A	N/A
OSYNE5|209L	90705	90998	98	N/A	N/A
OSYNE5|210L	91035	91217	61	N/A	N/A
OSYNE5|211L	91174	91398	75	N/A	N/A
OSYNE5|212R	91470	91844	125	N/A	N/A
OSYNE5|213L	91608	91904	99	N/A	N/A
OSYNE5|214L	92703	93053	117	N/A	N/A
OSYNE5|215L	93090	93344	85	N/A	N/A
OSYNE5|216R	93538	94773	412	A430L [Major capsid protein] [1e-97]	P30328.3 [Major capsid protein AltName: VP54] [1e-94]
OSYNE5|217L	94033	94350	106	N/A	N/A
OSYNE5|218R	94349	94543	65	N/A	N/A
OSYNE5|219R	94774	95016	81	N/A	N/A
OSYNE5|220L	94786	95670	295	A246aR [hypothetical protein] [3e-114]	Q5KSL6.1 [Diacylglycerol kinase kappa kinase kappa] [3e-12]
OSYNE5|221R	94838	95134	99	N/A	N/A
OSYNE5|222L	94947	95144	66	N/A	N/A
OSYNE5|223L	95698	97875	726	A241R [hypothetical protein] [0.0]	P47047.1 [ATP-dependent RNA helicase DOB1] [3e-93]
OSYNE5|224R	95737	95982	82	N/A	N/A
OSYNE5|225R	95789	96187	133	N/A	N/A
OSYNE5|226R	96275	96535	87	N/A	N/A
OSYNE5|227R	96800	96988	63	N/A	N/A
OSYNE5|228R	97091	97300	70	N/A	N/A
OSYNE5|229R	98039	98476	146	A239L [hypothetical protein] [2e-73]	N/A
OSYNE5|230L	98481	100031	517	A237R [Homospermidine synthase] [0.0]	Q98H64.1 [Hornospermidine synthase] [3e-63]
OSYNE5|231L	98536	98799	88	N/A	N/A
OSYNE5|232R	99814	100116	101	a236L [hypothetical protein] [1e-37]	N/A
OSYNE5|233R	100104	100430	109	A234L [hypothetical protein] [6e-52]	N/A
OSYNE5|234L	100427	100750	108	A233R [hypothetical protein] [1e-58]	N/A
OSYNE5|235R	100800	101939	380	A231L [hypothetical protein] [0.0]	N/A
OSYNE5|236L	100926	101159	78	a232R [hypothetical protein] [8e-22]	N/A
OSYNE5|237L	101366	101710	115	N/A	N/A
OSYNE5|238L	101428	101637	70	N/A	N/A
OSYNE5|239L	101953	102543	197	A230R [hypothetical protein] [1e-129]	N/A
OSYNE5|240R	102071	102253	61	N/A	N/A
OSYNE5|241R	102254	102523	90	N/A	N/A
OSYNE5|242R	102568	102801	78	A229L [hypothetical protein] [1e-47]	N/A
OSYNE5|243R	102823	103236	138	A227L [hypothetical protein] [2e-92]	N/A
OSYNE5|244L	102957	103178	74	a228R [hypothetical protein] [9e-40]	N/A
OSYNE5|245R	103239	103892	218	a225L [hypothetical protein] [2e-35]	N/A
OSYNE5|246L	103245	105140	632	A219/222/226R [hypothetical protein] [0.0]	Q9U720.1 [Cellulose synthase catalytic subunit A (UDP-forming)] [1e-06]
OSYNE5|247R	103261	103473	71	N/A	N/A
OSYNE5|248L	104210	104530	107	a223R [hypothetical protein] [2e-21]	N/A
OSYNE5|249R	104475	104852	126	N/A	N/A
OSYNE5|250R	105250	106404	385	A217L [hypothetical protein] [0.0]	N/A
OSYNE5|251L	105736	106254	173	N/A	N/A
OSYNE5|252R	106424	107359	312	A215L [Alkaline alginate lyase vAL-1] [0.0]	N/A
OSYNE5|253L	106549	106770	74	a216R [hypothetical protein] [6e-19]	N/A
OSYNE5|254L	106727	107017	97	N/A	N/A
OSYNE5|255L	107812	108612	267	A495R [hypothetical protein] [5e-17]	N/A
OSYNE5|256R	108749	109168	140	A214L [hypothetical protein] [2e-81]	N/A
OSYNE5|257L	108752	109024	91	N/A	N/A
OSYNE5|258R	109209	109655	149	A213L [hypothetical protein] [8e-102]	N/A
OSYNE5|259L	109418	109642	75	N/A	N/A
OSYNE5|260L	109852	110724	291	A208R [hypothetical protein] [6e-80]	N/A
OSYNE5|261L	109866	110318	151	a211R [hypothetical protein] [1e-06]	N/A
OSYNE5|262R	110300	110956	219	N/A	N/A
OSYNE5|263L	110852	111970	373	A207R [Arginine/Ornithine decarboxylase] [0.0]	P27117.1 [Ornithine decarboxylase] [2e-87]
OSYNE5|264R	110914	111099	62	N/A	N/A
OSYNE5|265L	111229	111486	86	N/A	N/A
OSYNE5|266R	111317	111619	101	N/A	N/A
OSYNE5|267R	111570	111839	90	N/A	N/A
OSYNE5|268L	112033	112656	208	A205R [hypothetical protein] [2e-108]	N/A
OSYNE5|269R	112237	112449	71	a206L [hypothetical protein] [7e-16]	N/A
OSYNE5|270L	112702	113346	215	A203R [hypothetical protein] [7e-147]	N/A
OSYNE5|271R	112748	112936	63	N/A	N/A
OSYNE5|272R	113412	113753	114	A202L [hypothetical protein] [8e-77]	N/A
OSYNE5|273R	113772	114056	95	A201L [hypothetical protein] [1e-42]	N/A
OSYNE5|274L	114079	114435	119	A200R [hypothetical protein] [6e-82]	N/A
OSYNE5|275L	114537	114839	101	A199R [hypothetical protein] [2e-54]	N/A
OSYNE5|276L	114877	115152	92	a197R [hypothetical protein] [4e-47]	N/A
OSYNE5|277R	114887	115345	153	A196L [hypothetical protein] [1e-102]	N/A
OSYNE5|278R	115350	116156	269	A193L [hypothetical protein] [0.0]	Q84513.1 [Probable DNA polymerase sliding clamp 1] [0.0]
OSYNE5|279L	116159	120070	1304	A189/192R [hypothetical protein] [0.0]	N/A
OSYNE5|280R	116728	116961	78	N/A	N/A

OSYNE5|281R	117621	118028	136	N/A	N/A
OSYNE5|282R	117979	118182	68	N/A	N/A
OSYNE5|283R	118219	118434	72	N/A	N/A
OSYNE5|284R	119784	120026	81	a190L [hypothetical protein] [6e-28]	N/A
OSYNE5|285L	120113	120559	149	A188aR [hypothetical protein] [2e-100]	P30321.2 [DNA polymerase] [1e-89]
OSYNE5|286R	120546	120794	83	a188L [hypothetical protein] [3e-46]	N/A
OSYNE5|287L	120694	122952	753	A185R [hypothetical protein] [0.0]	P30321.2 [DNA polymerase] [0.0]
OSYNE5|288R	120986	121402	139	N/A	N/A
OSYNE5|289R	122548	122988	147	N/A	N/A
OSYNE5|290R	123028	123255	76	N/A	N/A
OSYNE5|291R	124022	124792	257	a183L [hypothetical protein] [8e-09]	N/A
OSYNE5|292R	125295	125750	152	A253R [hypothetical protein] [2e-79]	N/A
OSYNE5|293R	125781	127298	506	A260R [chitinase] [0.0]	P32470.2 [Chitinase 1 Flags: Precursor] [5e-56]
OSYNE5|294R	125969	126232	88	N/A	N/A
OSYNE5|295R	126958	127230	91	N/A	N/A
OSYNE5|296R	127332	127880	183	N/A	P26840.1 [Probable macrolide acetyltransferase] [3e-38]
OSYNE5|297L	127864	128619	252	A287R [hypothetical protein] [3e-101]	N/A
OSYNE5|298R	128041	128325	95	N/A	N/A
OSYNE5|299L	128647	128913	89	N/A	N/A
OSYNE5|300L	128684	129334	217	A262/263L [hypothetical protein] [4e-119]	N/A
OSYNE5|301L	129355	130104	250	A265L [hypothetical protein] [4e-169]	N/A
OSYNE5|302R	129554	129808	85	N/A	N/A
OSYNE5|303R	129877	130082	62	N/A	N/A
OSYNE5|304L	130118	130354	79	A273L [hypothetical protein] [2e-15]	N/A
OSYNE5|305L	130414	131586	391	A208R [hypothetical protein] [1e-11]	N/A
OSYNE5|306R	130484	130882	133	N/A	N/A
OSYNE5|307R	130789	131064	92	N/A	N/A
OSYNE5|308R	131009	131236	76	N/A	N/A
OSYNE5|309R	131727	132860	378	N/A	P52284.1 [Adenine-specific methyltransferase CviRI] [0.0]
OSYNE5|310L	131759	132130	124	N/A	N/A
OSYNE5|311L	132227	132547	107	N/A	N/A
OSYNE5|312L	132432	132716	95	N/A	N/A
OSYNE5|313R	132893	134608	572	N/A	N/A
OSYNE5|314L	133099	133317	73	N/A	N/A
OSYNE5|315R	133146	133355	70	N/A	N/A
OSYNE5|316R	133933	134169	79	N/A	N/A
OSYNE5|317L	134095	134304	70	N/A	N/A
OSYNE5|318L	134332	134547	72	N/A	N/A
OSYNE5|319R	134630	135745	372	N/A	P10835.1 [Adenine-specific methyltransferase CviBIII] [3e-65]
OSYNE5|320L	134740	135027	96	N/A	N/A
OSYNE5|321L	135055	135630	192	N/A	N/A
OSYNE5|322L	135722	135934	71	N/A	N/A
OSYNE5|323R	135776	136954	393	N/A	P52284.1 [Adenine-specific methyltransferase CviRI] [5e-151]
OSYNE5|324L	136312	136677	122	N/A	N/A
OSYNE5|325L	136358	136663	102	N/A	N/A
OSYNE5|326R	137050	138066	339	A328L [hypothetical protein] [1e-39]	N/A
OSYNE5|327R	137903	138154	84	N/A	N/A
OSYNE5|328R	138196	138774	193	A647R [hypothetical protein] [8e-63]	N/A
OSYNE5|329L	138629	138817	63	N/A	N/A
OSYNE5|330R	139082	139825	248	A275R [hypothetical protein] [4e-159]	N/A
OSYNE5|331L	139810	140649	280	A277L [Protein kinase] [4e-159]	Q5B4Z3.2 [Serine/threonine-protein kinase sepH] [3e-19]
OSYNE5|332R	140714	141094	127	a281R [hypothetical protein] [1e-38]	N/A
OSYNE5|333L	140725	142026	434	A278L [Protein kinase] [0.0]	N/A
OSYNE5|334L	141855	142043	63	N/A	N/A
OSYNE5|335L	142527	143525	333	A284L [Amindase] [8e-178]	P54966.1 [Uncharacterized protein A284L] [1e-174]
OSYNE5|336L	143411	143596	62	N/A	N/A
OSYNE5|337R	143433	144539	369	A286R [hypothetical protein] [0.0]	N/A
OSYNE5|338L	143777	143959	61	N/A	N/A
OSYNE5|339L	144084	144518	145	N/A	N/A
OSYNE5|340R	144517	144717	67	N/A	N/A
OSYNE5|341L	144593	145447	285	A289L [Protein kinase] [2e-160]	A8WYE4.1 [Serine/threonine-protein kinase par-1] [1e-27]
OSYNE5|342R	144747	145121	125	a290R [hypothetical protein] [2e-50]	N/A
OSYNE5|343L	144793	144984	64	N/A	N/A
OSYNE5|344R	145238	145480	81	a291R [hypothetical protein] [3e-30]	N/A
OSYNE5|345L	145533	146546	338	A292L [Chitosanase] [0.0]	O07921.1 [Chitosanase Flags: Precursor] [3e-14]
OSYNE5|346R	145615	145908	98	a293R [hypothetical protein] [2e-34]	N/A
OSYNE5|347R	145921	146121	67	a293R [hypothetical protein] [2e-30]	N/A
OSYNE5|348R	146289	146486	66	N/A	N/A
OSYNE5|349L	146558	147496	313	A295L [Fucose synthetase] [0.0]	Q9LMU0.1 [Putative GDP-L-fucose synthase 2] [1e-124]
OSYNE5|350R	146583	146930	116	N/A	N/A
OSYNE5|351L	147527	147772	82	N/A	N/A
OSYNE5|352R	147546	148025	160	A296R [hypothetical protein] [2e-46]	N/A
OSYNE5|353L	148060	149010	317	N/A	N/A
OSYNE5|354R	148110	148346	79	N/A	N/A
OSYNE5|355L	148866	149087	74	N/A	N/A
OSYNE5|356L	149271	149459	63	N/A	N/A
OSYNE5|357R	149856	150113	86	N/A	N/A
OSYNE5|358L	150151	150675	175	A297L [hypothetical protein] [6e-101]	N/A
OSYNE5|359R	150795	151256	154	N/A	N/A
OSYNE5|360L	151280	151954	225	A298L [hypothetical protein] [1e-143]	N/A
OSYNE5|361R	151752	151973	74	a299R [hypothetical protein] [3e-40]	N/A
OSYNE5|362L	151975	152709	245	A301L [hypothetical protein] [9e-105]	N/A
OSYNE5|363L	152088	152291	68	N/A	N/A
OSYNE5|364R	152764	153000	79	A304R [hypothetical protein] [1e-34]	N/A
OSYNE5|365L	153041	153655	205	A305L [Protein phosphatase] [6e-127]	Q9VVW5.2 [phosphatase] [8e-15]
OSYNE5|366R	153664	153945	94	a307R [hypothetical protein] [5e-27]	N/A
OSYNE5|367L	153680	153940	87	A306L [hypothetical protein] [1e-44]	N/A
OSYNE5|368L	153977	154330	118	A308L [hypothetical protein] [4e-58]	N/A
OSYNE5|369L	154459	154791	111	A180R [hypothetical protein] [1e-50]	O34693.1 [Uncharacterized protein YloA] [1e-14]
OSYNE5|370L	154854	155366	171	A310L [hypothetical protein] [2e-110]	N/A
OSYNE5|371L	155434	156150	239	A312L [hypothetical protein] [2e-167]	N/A
OSYNE5|372L	156358	156573	72	A313L [hypothetical protein] [3e-30]	N/A
OSYNE5|373L	156647	156973	109	N/A	N/A
OSYNE5|374R	156653	156895	81	A314R [hypothetical protein] [1e-44]	N/A

OSYNE5|375L	157432	157713	94	a317L [hypothetical protein] [9e-30]	N/A
OSYNE5|376R	158578	158943	122	A320R [hypothetical protein] [3e-59]	N/A
OSYNE5|377R	158980	159339	120	A321R [hypothetical protein] [2e-73]	N/A
OSYNE5|378L	159450	159980	177	A322L [hypothetical protein] [3e-97]	N/A
OSYNE5|379R	159760	159975	72	a323R [hypothetical protein] [1e-26]	N/A
OSYNE5|380L	160032	161366	445	A324L [hypothetical protein] [0.0]	Q5UQN9.1 [Uncharacterized protein R449] [8e-09]
OSYNE5|381R	160599	161282	228	N/A	N/A
OSYNE5|382L	161447	162085	213	A326L [hypothetical protein] [3e-144]	N/A
OSYNE5|383R	161525	161707	61	N/A	N/A
OSYNE5|384R	161715	161924	70	N/A	N/A
OSYNE5|385L	162116	163183	356	A328L [hypothetical protein] [0.0]	N/A
OSYNE5|386R	162183	162434	84	a327R [hypothetical protein] [2e-22]	N/A
OSYNE5|387L	163254	163577	108	N/A	N/A
OSYNE5|388R	163273	163563	97	A329R [hypothetical protein] [2e-36]	N/A
OSYNE5|389R	164860	165702	281	A267L [hypothetical protein] [9e-45]	Q5UQL9.1 [Uncharacterized protein R423] [9e-15]
OSYNE5|390L	165462	165680	73	N/A	N/A
OSYNE5|391L	165908	166210	101	N/A	N/A
OSYNE5|392L	165943	166131	63	N/A	N/A
OSYNE5|393R	165971	166195	75	N/A	N/A
OSYNE5|394R	166123	166359	79	a329cR [hypothetical protein] [8e-31]	N/A
OSYNE5|395L	166372	167553	394	A333L [hypothetical protein] [0.0]	C3PH19.1 [Translation initiation factor IF-2] [8e-07]
OSYNE5|396R	166432	166875	148	N/A	N/A
OSYNE5|397R	166695	167033	113	a335R [hypothetical protein] [4e-07]	N/A
OSYNE5|398R	167224	167481	86	a336R [hypothetical protein] [5e-29]	N/A
OSYNE5|399L	167594	168394	267	A337L [hypothetical protein] [3e-92]	N/A
OSYNE5|400R	167720	167944	75	N/A	N/A
OSYNE5|401L	168235	168471	79	N/A	N/A
OSYNE5|402L	168532	169083	184	A337L [hypothetical protein] [3e-54]	N/A
OSYNE5|403L	169145	169552	136	A341L [hypothetical protein] [4e-78]	N/A
OSYNE5|404L	169639	170493	285	A315L [hypothetical protein] [3e-79]	Q5UPT6.1 [Uncharacterized HNH endonuclease L245] [5e-08]
OSYNE5|405L	170390	170572	61	N/A	N/A
OSYNE5|406L	170573	172267	565	A342L [hypothetical protein] [0.0]	N/A
OSYNE5|407R	170579	170809	77	N/A	N/A
OSYNE5|408R	170959	171219	87	a343R [hypothetical protein] [1e-30]	N/A
OSYNE5|409L	171256	171762	169	a345L [hypothetical protein] [2e-29]	N/A
OSYNE5|410L	172365	172742	126	A349L [hypothetical protein] [1e-73]	N/A
OSYNE5|411L	172705	172911	69	A349L [hypothetical protein] [2e-17]	N/A
OSYNE5|412R	172787	173155	123	A350R [hypothetical protein] [1e-79]	N/A
OSYNE5|413L	172847	173104	86	N/A	N/A
OSYNE5|414R	173239	173448	70	N/A	N/A
OSYNE5|415L	173269	173892	208	A352L [hypothetical protein] [1e-133]	Q5UQF7.1 [Uncharacterized protein R489 Flags: Precursor] [5e-07]
OSYNE5|416L	173532	173735	68	N/A	N/A
OSYNE5|417R	173586	173792	69	a353R [hypothetical protein] [1e-05]	N/A
OSYNE5|418L	173956	174969	338	A357L [hypothetical protein] [4e-171]	N/A
OSYNE5|419R	174317	174937	207	a358R [hypothetical protein] [4e-08]	N/A
OSYNE5|420L	174612	174842	77	a359L [hypothetical protein] [2e-16]	N/A
OSYNE5|421R	175039	178653	1205	A363R [hypothetical protein] [0.0]	N/A
OSYNE5|422L	175996	176253	86	N/A	N/A
OSYNE5|423L	176778	176981	68	a364L [hypothetical protein] [3e-41]	N/A
OSYNE5|424R	177671	177859	63	N/A	N/A
OSYNE5|425L	178005	178277	91	N/A	N/A
OSYNE5|426L	178325	178576	84	N/A	N/A
OSYNE5|427R	178738	179202	155	A373R [hypothetical protein] [2e-55]	N/A
OSYNE5|428R	179322	180266	315	A007/008L [hypothetical protein] [4e-32]	Q8Q0U0.1 [Putative ankyrin repeat protein MM_0045] [4e-29]
OSYNE5|429L	179656	179943	96	N/A	N/A
OSYNE5|430L	179720	180004	95	N/A	N/A
OSYNE5|431L	180380	181105	242	A378L [hypothetical protein] [2e-112]	N/A
OSYNE5|432R	180751	180957	69	N/A	N/A
OSYNE5|433L	181129	181758	210	A379L [hypothetical protein] [3e-139]	N/A
OSYNE5|434R	181931	183391	487	A383R [Capsid protein] [0.0]	P30328.3 [Major capsid protein] [6e-39]
OSYNE5|435L	182383	182613	77	N/A	N/A
OSYNE5|436R	183111	183296	62	N/A	N/A
OSYNE5|437R	183415	183600	62	A384bL [hypothetical protein] [1e-27]	N/A
OSYNE5|438L	183683	185533	617	A384dL [Capsid protein] [0.0]	Q4JV51.1 [Translation initiation factor IF-2] [4e-06]
OSYNE5|439L	183706	183906	67	a385L [hypothetical protein] [4e-07]	N/A
OSYNE5|440R	184408	184740	111	N/A	N/A
OSYNE5|441R	184755	184991	79	N/A	N/A
OSYNE5|442R	184955	185428	158	N/A	N/A
OSYNE5|443R	184963	185208	82	N/A	N/A
OSYNE5|444R	185241	185540	100	a391R [hypothetical protein] [8e-32]	N/A
OSYNE5|445R	185624	186388	255	A392R [hypothetical protein] [2e-176]	Q196X2.1 [Uncharacterized protein 088R] [1e-42]
OSYNE5|446R	186424	186633	70	N/A	N/A
OSYNE5|447R	186645	187049	135	A394R [hypothetical protein] [3e-61]	N/A
OSYNE5|448L	187135	188619	495	N/A	N/A
OSYNE5|449R	188703	188954	84	A395R [hypothetical protein] [4e-47]	N/A
OSYNE5|450L	189107	189565	153	A396L [hypothetical protein] [1e-89]	N/A
OSYNE5|451R	189363	189551	63	N/A	N/A
OSYNE5|452L	189620	189976	119	A398L [hypothetical protein] [3e-76]	N/A
OSYNE5|453R	190049	190633	195	A399R [hypothetical protein] [7e-115]	N/A
OSYNE5|454L	190336	190722	129	N/A	N/A
OSYNE5|455L	190614	190805	64	N/A	N/A
OSYNE5|456R	190666	191019	118	A400R [hypothetical protein] [1e-79]	N/A
OSYNE5|457L	190996	191193	66	N/A	N/A
OSYNE5|458R	191057	191911	285	A401R [hypothetical protein] [0.0]	N/A
OSYNE5|459L	191576	192049	158	N/A	N/A
OSYNE5|460R	191790	192632	281	A402R [hypothetical protein] [1e-150]	N/A
OSYNE5|461R	192751	193032	94	A403R [hypothetical protein] [1e-64]	N/A
OSYNE5|462R	193063	193650	196	A404R [hypothetical protein] [1e-28]	N/A
OSYNE5|463R	193280	193627	116	N/A	N/A
OSYNE5|464L	193307	193555	83	N/A	N/A
OSYNE5|465R	193776	193967	64	N/A	N/A
OSYNE5|466L	194686	194886	67	N/A	N/A
OSYNE5|467L	194965	195153	63	N/A	N/A
OSYNE5|468L	195211	195411	67	a406L [hypothetical protein] [2e-29]	N/A

OSYNE5|469L	195448	196080	211	A407L [hypothetical protein] [5e-130]	N/A
OSYNE5|470L	196119	196886	256	A408L [hypothetical protein] [1e-145]	N/A
OSYNE5|471R	196423	196908	162	a409R [hypothetical protein] [5e-41]	N/A
OSYNE5|472L	196892	197221	110	A410L [hypothetical protein] [1e-62]	N/A
OSYNE5|473R	197309	197815	169	A411R [hypothetical protein] [3e-77]	N/A
OSYNE5|474R	197824	198390	189	A412R [hypothetical protein] [7e-125]	N/A
OSYNE5|475L	198198	198407	70	N/A	N/A
OSYNE5|476L	198391	199095	235	A413L [hypothetical protein] [9e-109]	N/A
OSYNE5|477R	198533	198925	131	N/A	N/A
OSYNE5|478R	199174	199392	73	A414R [hypothetical protein] [2e-39]	N/A
OSYNE5|479R	199468	200031	188	A416R [hypothetical protein] [5e-119]	Q197D1.1 [Putative kinase protein 029R] [6e-22]
OSYNE5|480L	200007	201296	430	A417L [hypothetical protein] [0.0]	A6UWR5.1 [Replication factor C large subunit large subunit] [6e-06]
OSYNE5|481R	200125	200370	82	N/A	N/A
OSYNE5|482R	201097	201333	79	a419R [hypothetical protein] [1e-32]	N/A
OSYNE5|483L	201328	201540	71	A420L [hypothetical protein] [3e-41]	N/A
OSYNE5|484R	201585	201884	100	A421R [hypothetical protein] [1e-46]	N/A
OSYNE5|485R	201865	202063	73	N/A	N/A
OSYNE5|486R	201908	202102	65	A422aR [hypothetical protein] [2e-35]	N/A
OSYNE5|487R	202113	202601	163	A423R [hypothetical protein] [5e-65]	N/A
OSYNE5|488R	202608	203012	135	N/A	N/A
OSYNE5|489R	202756	203028	91	N/A	N/A
OSYNE5|490L	202794	203036	81	N/A	N/A
OSYNE5|491R	203047	203361	115	A426R [hypothetical protein] [2e-58]	N/A
OSYNE5|492L	203388	203753	122	A427L [hypothetical protein] [2e-48]	P0A618.2 [Thioredoxin AltName: MPT46] [1e-06]
OSYNE5|493R	203397	203621	75	N/A	N/A
OSYNE5|494L	203803	204207	135	A428L [hypothetical protein] [3e-28]	P27951.1 [IgA FC receptor] [7e-07]
OSYNE5|495L	204233	205594	454	A429L [hypothetical protein] [0.0]	Q5ZUJ9.1 [E3 ubiquitin-protein ligase MIB2] [4e-06]
OSYNE5|496L	204793	204993	67	N/A	N/A
OSYNE5|497R	205745	206491	249	A315L [hypothetical protein] [4e-07]	N/A
OSYNE5|498L	206030	206293	88	N/A	N/A
OSYNE5|499L	206548	207861	438	A430L [Major capsid protein] [0.0]	P30328.3 [Major capsid protein AltName: VP54] [0.0]
OSYNE5|500R	207020	207241	74	N/A	N/A
OSYNE5|501L	207941	208759	273	A315L [hypothetical protein] [1e-22]	N/A
OSYNE5|502R	208202	208429	76	N/A	N/A
OSYNE5|503R	208877	209323	149	A432R [hypothetical protein] [6e-80]	N/A
OSYNE5|504R	209086	209304	73	a433R [hypothetical protein] [2e-13]	N/A
OSYNE5|505L	209327	209518	64	A436L [hypothetical protein] [3e-30]	N/A
OSYNE5|506L	209549	209875	109	A437L [hypothetical protein] [1e-63]	P15250.1 [Chromosomal protein MC1b] [5e-06]
OSYNE5|507L	209904	210140	79	A438L [Glutaredoxin] [8e-51]	Q1RHJ0.1 [Glutaredoxin-1] [6e-09]
OSYNE5|508R	210163	210501	113	A439R [hypothetical protein] [1e-69]	N/A
OSYNE5|509L	210658	211071	138	A441L [hypothetical protein] [1e-93]	N/A
OSYNE5|510R	210706	211056	117	a442R [hypothetical protein] [3e-66]	N/A
OSYNE5|511R	211212	212138	309	A443R [hypothetical protein] [0.0]	N/A
OSYNE5|512L	212153	212815	221	N/A	N/A
OSYNE5|513R	212394	212777	128	N/A	N/A
OSYNE5|514R	212600	212794	65	N/A	N/A
OSYNE5|515L	212870	213184	105	A444L [hypothetical protein] [1e-54]	N/A
OSYNE5|516R	212964	213203	80	N/A	N/A
OSYNE5|517L	213250	214638	463	A445L [hypothetical protein] [0.0]	Q98496.1 [Uncharacterized protein A445L] [0.0]
OSYNE5|518R	213284	213508	75	N/A	N/A
OSYNE5|519R	213554	214123	190	a446R [hypothetical protein] [6e-59]	N/A
OSYNE5|520R	214675	214950	92	N/A	N/A
OSYNE5|521L	214699	215019	107	A448L [Protein disulphide isomerase] [9e-70]	P52588.1 [Protein disulfide-isomerase Flags: Precursor] [3e-08]
OSYNE5|522R	214998	215753	252	A449R [hypothetical protein] [9e-126]	N/A
OSYNE5|523R	215302	215577	92	N/A	N/A
OSYNE5|524R	215987	216745	253	A450R [hypothetical protein] [1e-178]	N/A
OSYNE5|525L	216782	217606	275	A271L [hypothetical protein] [1e-149]	Q55EQ3.2 [Uncharacterized abhydrolase protein DDB_G0269066] [2e-08]
OSYNE5|526R	217244	217546	101	a272aR [hypothetical protein] [9e-14]	N/A
OSYNE5|527L	217722	218135	138	A452L [hypothetical protein] [1e-28]	N/A
OSYNE5|528L	218303	219169	289	A454L [hypothetical protein] [0.0]	N/A
OSYNE5|529L	219200	221161	654	A456L [hypothetical protein] [0.0]	N/A
OSYNE5|530R	219252	220103	284	a459R [hypothetical protein] [7e-57]	N/A
OSYNE5|531R	220313	220567	85	a460R [hypothetical protein] [4e-39]	N/A
OSYNE5|532R	220500	220688	63	N/A	N/A
OSYNE5|533R	220977	221168	64	N/A	N/A
OSYNE5|534L	221133	221912	260	a463L [hypothetical protein] [2e-70]	N/A
OSYNE5|535R	221250	221480	77	A461R [hypothetical protein] [6e-25]	N/A
OSYNE5|536R	221514	222326	271	A464R [Rnase III] [0.0]	Q98514.1 [Putative protein A464R] [2e-179]
OSYNE5|537R	222364	222720	119	A465R [hypothetical protein] [1e-78]	Q5UQV6.1 [Probable FAD-linked sulfhydryl oxidase R368] [9e-16]
OSYNE5|538Lc	222745	223683	313	A467L [hypothetical protein] [0.0]	N/A
OSYNE5|539L	223245	223451	69	N/A	N/A
OSYNE5|540R	223832	225157	442	A468R [hypothetical protein] [0.0]	N/A
OSYNE5|541R	225206	225796	197	A470R [hypothetical protein] [2e-127]	N/A
OSYNE5|542L	225283	225486	68	N/A	N/A
OSYNE5|543L	225523	225705	61	N/A	N/A
OSYNE5|544R	225848	226369	174	A471R [hypothetical protein] [3e-115]	Q5UQ75.1 [Uncharacterized protein L507] [7e-29]
OSYNE5|545R	226507	227481	325	A476R [hypothetical protein] [0.0]	P50650.1 [Ribonucleoside-diphosphate reductase small chain] [1e-135]
OSYNE5|546L	226639	226929	97	a477L [hypothetical protein] [5e-31]	N/A
OSYNE5|547R	226820	227155	112	N/A	N/A
OSYNE5|548L	227118	227372	85	N/A	N/A
OSYNE5|549L	227462	228346	295	A490L [hypothetical protein] [2e-149]	Q5UQL9.1 [Uncharacterized protein R423] [9e-33]
OSYNE5|550L	228382	228657	92	A480L [hypothetical protein] [5e-42]	N/A
OSYNE5|551R	228623	228895	91	N/A	N/A
OSYNE5|552L	228685	229368	228	A481L [hypothetical protein] [3e-126]	N/A
OSYNE5|553R	228883	229128	82	N/A	N/A
OSYNE5|554L	229326	229580	85	N/A	N/A
OSYNE5|555R	229447	230091	215	A482R [hypothetical protein] [3e-128]	N/A
OSYNE5|556L	229648	229971	108	N/A	N/A
OSYNE5|557L	230086	230553	156	A484L [hypothetical protein] [7e-98]	N/A
OSYNE5|558R	230189	230395	69	N/A	N/A
OSYNE5|559R	230636	231079	148	A485R [hypothetical protein] [7e-96]	N/A
OSYNE5|560R	231093	231359	89	N/A	N/A
OSYNE5|561R	231403	232362	320	A488R [hypothetical protein] [0.0]	Q5UQL4.1 [Uncharacterized protein L417] [9e-11]
OSYNE5|562R	232148	232378	77	a489R [hypothetical protein] [2e-12]	N/A

OSYNE5|563R	232412	232642	77	A491R [hypothetical protein] [1e-42]	N/A
OSYNE5|564L	232639	233184	182	A492L [hypothetical protein] [2e-93]	N/A
OSYNE5|565R	233226	234320	365	A494R [hypothetical protein] [0.0]	Q98544.1 [Putative transcription factor A494R] [0.0]
OSYNE5|566R	234374	234814	147	A497R [hypothetical protein] [1e-83]	Q9T1Q1.1 [Putative protein p47] [4e-08]
OSYNE5|567L	234429	234647	73	N/A	N/A
OSYNE5|568L	234864	235916	351	A500L [hypothetical protein] [7e-73]	Q8DQN5.1 [Zinc metalloprotease ZmpB Flags: Precursor] [6e-08]
OSYNE5|569R	234982	235203	74	N/A	N/A
OSYNE5|570L	235229	235573	115	N/A	N/A
OSYNE5|571R	235897	236427	177	N/A	N/A
OSYNE5|572L	235950	236237	96	A502L [hypothetical protein] [1e-59]	N/A
OSYNE5|573L	235982	236167	62	N/A	N/A
OSYNE5|574L	236283	237122	280	A503L [hypothetical protein] [0.0]	N/A
OSYNE5|575R	236536	236727	64	N/A	N/A
OSYNE5|576L	237059	237301	81	N/A	N/A
OSYNE5|577L	237201	238700	500	A505L [hypothetical protein] [0.0]	N/A
OSYNE5|578R	237585	237797	71	a507R [hypothetical protein] [6e-27]	N/A
OSYNE5|579R	237957	238145	63	a507R [hypothetical protein] [2e-18]	N/A
OSYNE5|580R	238231	238536	102	a508R [hypothetical protein] [1e-21]	N/A
OSYNE5|581R	238884	239447	188	N/A	N/A
OSYNE5|582L	238893	239093	67	N/A	N/A
OSYNE5|583R	238972	239160	63	N/A	N/A
OSYNE5|584L	239133	239363	77	N/A	N/A
OSYNE5|585L	239469	239717	83	A519L [hypothetical protein] [2e-49]	N/A
OSYNE5|586L	239722	240021	100	A520L [hypothetical protein] [2e-56]	N/A
OSYNE5|587L	240039	240587	183	A521L [hypothetical protein] [2e-89]	N/A
OSYNE5|588L	240616	241233	206	A521aL [hypothetical protein] [4e-133]	O55742.1 [Uncharacterized protein 136R] [9e-09]
OSYNE5|589R	240911	241123	71	a522R [hypothetical protein] [7e-35]	N/A
OSYNE5|590R	241288	241821	178	A523R [hypothetical protein] [1e-118]	N/A
OSYNE5|591L	241518	241718	67	a524L [hypothetical protein] [2e-37]	N/A
OSYNE5|592R	241864	242304	147	A526R [hypothetical protein] [2e-77]	N/A
OSYNE5|593L	242245	242439	65	N/A	N/A
OSYNE5|594R	242282	242596	105	A527R [hypothetical protein] [4e-56]	N/A
OSYNE5|595R	242711	242983	91	a528R [hypothetical protein] [1e-07]	N/A
OSYNE5|596L	242928	243146	73	a529L [hypothetical protein] [1e-39]	N/A
OSYNE5|597R	242947	243999	351	A530R [hypothetical protein] [0.0]	P36216.1 [Cytosine-specific methyltransferase CvtJI] [8e-98]
OSYNE5|598L	243595	243819	75	N/A	N/A
OSYNE5|599L	243996	244190	65	A531L [hypothetical protein] [2e-32]	N/A
OSYNE5|600L	244222	244461	80	A532L [hypothetical protein] [1e-50]	N/A
OSYNE5|601R	244740	246335	532	A533R [hypothetical protein] [0.0]	N/A
OSYNE5|602R	246193	246546	118	N/A	N/A
OSYNE5|603L	246337	246561	75	A535L [hypothetical protein] [3e-39]	N/A
OSYNE5|604L	246627	246881	85	A536L [hypothetical protein] [7e-22]	N/A
OSYNE5|605L	246886	247728	281	A537L [hypothetical protein] [4e-140]	N/A
OSYNE5|606R	247526	247711	62	N/A	N/A
OSYNE5|607R	247722	248330	203	A539R [hypothetical protein] [3e-108]	N/A
OSYNE5|608L	247725	248012	96	N/A	N/A
OSYNE5|609L	248346	251747	1134	A540L [hypothetical protein] [0.0]	N/A
OSYNE5|610R	249160	249354	65	N/A	N/A
OSYNE5|611R	250321	250527	69	N/A	N/A
OSYNE5|612L	250588	250770	61	N/A	N/A
OSYNE5|613R	250730	250921	64	N/A	N/A
OSYNE5|614R	250969	251217	83	N/A	N/A
OSYNE5|615R	251497	251688	64	N/A	N/A
OSYNE5|616R	251868	252764	299	A544R [ATP-dependent DNA ligase] [0.0]	P44121.2 [DNA ligase] [3e-11]
OSYNE5|617L	252239	252460	74	a545L [hypothetical protein] [4e-45]	N/A
OSYNE5|618L	252417	252767	117	N/A	N/A
OSYNE5|619L	252746	253975	410	A546L [hypothetical protein] [0.0]	N/A
OSYNE5|620R	253894	254148	85	N/A	N/A
OSYNE5|621L	253962	255332	457	A548L [hypothetical protein] [0.0]	Q91ZW3.1 [SWI/SNF actin-dependent regulator of chromatin] [4e-35]
OSYNE5|622L	254516	254731	72	N/A	N/A
OSYNE5|623R	254580	254751	64	N/A	N/A
OSYNE5|624R	254868	255125	86	a550R [hypothetical protein] [1e-40]	N/A
OSYNE5|625R	255019	255306	96	N/A	N/A
OSYNE5|626L	255424	255861	146	A551L [dUTP pyrophosphatase] [1e-77]	Q41033.1 [Deoxyuridine 5-triphosphate nucleotidohydrolase] [1e-74]
OSYNE5|627R	255988	256941	318	A552R [hypothetical protein] [0.0]	N/A
OSYNE5|628L	256826	257011	62	N/A	N/A
OSYNE5|629L	256956	258452	499	A554/555/557L [hypothetical protein] [0.0]	Q82VP4.1 [tRNA(Ile)-lysidine synthase] [2e-18]
OSYNE5|630R	256960	257361	134	N/A	N/A
OSYNE5|631R	257334	257816	161	a555R [hypothetical protein] [3e-41]	N/A
OSYNE5|632L	258552	259754	401	A558L [Capsid protein] [0.0]	P30328.3 [Major capsid protein AltName: VP54] [5e-77]
OSYNE5|633R	258568	258804	79	N/A	N/A
OSYNE5|634R	258904	259089	62	N/A	N/A
OSYNE5|635L	259867	260538	224	A559L [hypothetical protein] [5e-101]	N/A
OSYNE5|636R	259889	260365	159	a560R [hypothetical protein] [3e-32]	N/A
OSYNE5|637R	261824	262051	76	N/A	N/A
OSYNE5|638R	262271	262591	107	N/A	N/A
OSYNE5|639R	262485	263297	271	A287R [hypothetical protein] [3e-101]	O36580.1 [Probable intron-encoded endonuclease 1] [2e-08]
OSYNE5|640L	262505	262699	65	a288L [hypothetical protein] [3e-16]	N/A
OSYNE5|641L	262868	263158	97	N/A	N/A
OSYNE5|642R	263267	263539	91	N/A	N/A
OSYNE5|643R	264253	264528	92	N/A	N/A
OSYNE5|644R	264703	265260	186	A565R [hypothetical protein] [5e-112]	N/A
OSYNE5|645L	265273	265698	142	A567L [hypothetical protein] [1e-46]	N/A
OSYNE5|646R	265989	266243	85	a569R [hypothetical protein] [8e-25]	N/A
OSYNE5|647R	266262	266513	84	N/A	N/A
OSYNE5|648R	266377	266625	83	N/A	N/A
OSYNE5|649L	266389	266790	134	A570L [hypothetical protein] [4e-82]	N/A
OSYNE5|650L	266559	266783	75	N/A	N/A
OSYNE5|651R	266847	267197	117	A571R [hypothetical protein] [2e-73]	N/A
OSYNE5|652R	267211	267756	182	A572R [hypothetical protein] [1e-119]	N/A
OSYNE5|653L	267517	267765	83	N/A	N/A
OSYNE5|654L	267762	268508	249	A574L [hypothetical protein] [1e-164]	O41056.1 [Probable DNA polymerase sliding clamp 2] [2e-161]
OSYNE5|655R	267877	268095	73	N/A	N/A
OSYNE5|656L	268583	269089	169	A575L [hypothetical protein] [8e-115]	N/A

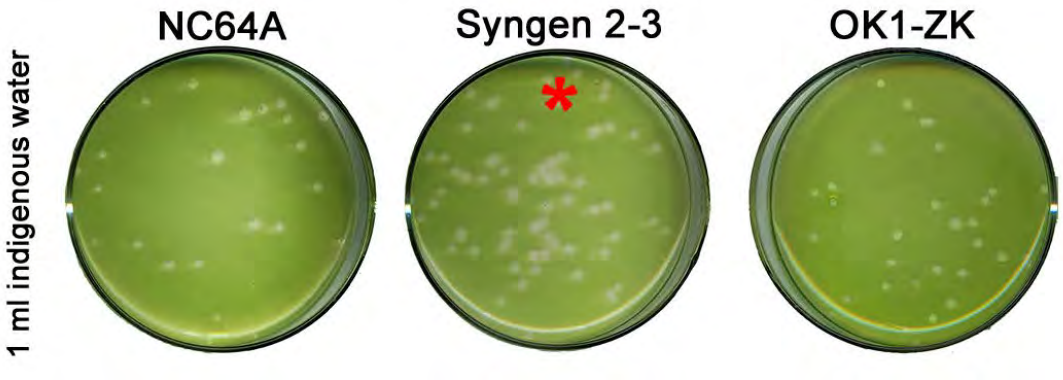
OSYNE5|657R	269070	269273	68	N/A	N/A
OSYNE5|658L	269196	269603	136	A577L [hypothetical protein] [3e-78]	N/A
OSYNE5|659L	269619	272804	1062	A583L [DNA topoisomerase II] [0.0]	P08096.2 [DNA topoisomerase 2] [0.0]
OSYNE5|660R	269980	270192	71	a584R [hypothetical protein] [5e-34]	N/A
OSYNE5|661R	270033	270641	203	a585R [hypothetical protein] [3e-19]	N/A
OSYNE5|662R	271074	271262	63	N/A	N/A
OSYNE5|663R	271350	271556	69	a587R [hypothetical protein] [1e-21]	N/A
OSYNE5|664R	271519	271719	67	N/A	N/A
OSYNE5|665R	272266	272499	78	N/A	N/A
OSYNE5|666L	272934	274043	370	A590L [hypothetical protein] [2e-66]	N/A
OSYNE5|667R	273788	274006	73	N/A	N/A
OSYNE5|668R	274259	274906	216	N/A	N/A
OSYNE5|669L	275353	275538	62	N/A	N/A
OSYNE5|670L	275459	275644	62	N/A	N/A
OSYNE5|671R	276121	276549	143	A598R [hypothetical protein] [1e-87]	Q9VWA2.1 [Probable deoxycytidylate deaminase] [6e-26]
OSYNE5|672R	276636	277202	189	N/A	N/A
OSYNE5|673L	277197	278351	385	A598L [hypothetical protein] [0.0]	P54772.1 [Histidine decarboxylase] [2e-61]
OSYNE5|674R	277503	277688	62	a599R [hypothetical protein] [6e-22]	N/A
OSYNE5|675L	277855	278052	66	N/A	N/A
OSYNE5|676R	277947	278273	109	a600aR [hypothetical protein] [1e-09]	N/A
OSYNE5|677R	278428	279345	306	N/A	Q3SYV9.1 [ADP-ribosylhydrolase 3] [6e-18]
OSYNE5|678R	279399	279698	100	A601R [hypothetical protein] [9e-47]	N/A
OSYNE5|679L	279688	279939	84	N/A	P51423.2 [Ubiquitin-60S ribosomal protein L40] [3e-46]
OSYNE5|680L	279954	280376	141	A602L [hypothetical protein] [5e-78]	N/A
OSYNE5|681R	280481	280795	105	a603R [hypothetical protein] [3e-60]	N/A
OSYNE5|682L	280983	281474	164	A604L [hypothetical protein] [2e-70]	N/A
OSYNE5|683L	281224	281481	86	N/A	N/A
OSYNE5|684L	281485	281961	159	A605L [hypothetical protein] [3e-80]	N/A
OSYNE5|685R	281990	282193	68	N/A	N/A
OSYNE5|686R	282021	283382	454	N/A	N/A
OSYNE5|687L	282682	282915	78	N/A	N/A
OSYNE5|688R	283412	284584	391	A607R [hypothetical protein] [0.0]	Q02357.2 [Ankyrin-1 AltName: Erythrocyte ankyrin] [9e-10]
OSYNE5|689L	284593	285762	390	A609L [UDP-glucose dehydrogenase] [0.0]	Q33952.1 [UDP-glucose 6-dehydrogenase dehydrogenase] [1e-144]
OSYNE5|690L	285844	286203	120	A612L [Histone H3K27 methylase] [1e-75]	Q9Y7Q6.1 [SET domain-containing protein 7] [5e-07]
OSYNE5|691L	286257	287969	571	A614L [Protein kinase] [0.0]	N/A
OSYNE5|692L	286579	286782	68	N/A	N/A
OSYNE5|693R	287568	287798	77	N/A	N/A
OSYNE5|694R	288040	289011	324	A617R [hypothetical protein] [0.0]	Q5UQJ6.1 [Putative serine/threonine-protein kinase R400] [1e-11]
OSYNE5|695L	289026	289436	137	A618L [hypothetical protein] [6e-65]	N/A
OSYNE5|696L	289453	290139	229	A619L [hypothetical protein] [3e-90]	N/A
OSYNE5|697R	289639	290061	141	N/A	N/A
OSYNE5|698R	289787	290182	132	N/A	N/A
OSYNE5|699L	290179	290430	84	A620L [hypothetical protein] [1e-52]	N/A
OSYNE5|700R	290222	290443	74	N/A	N/A
OSYNE5|701L	290450	290806	119	A621L [hypothetical protein] [3e-68]	N/A
OSYNE5|702L	290864	292441	526	A622L [Capsid protein] [0.0]	A7U6E9.1 [Major capsid protein] [9e-67]
OSYNE5|703L	292601	292792	64	A623aL [hypothetical protein] [1e-07]	N/A
OSYNE5|704R	292728	293096	123	A624R [hypothetical protein] [1e-65]	N/A
OSYNE5|705L	292997	293476	160	N/A	N/A
OSYNE5|706R	293115	294431	439	A627R [hypothetical protein] [0.0]	N/A
OSYNE5|707L	293325	293645	107	N/A	N/A
OSYNE5|708L	293765	293977	71	N/A	N/A
OSYNE5|709L	294450	294773	108	A628L [hypothetical protein] [9e-24]	N/A
OSYNE5|710L	294932	295195	88	N/A	N/A
OSYNE5|711R	294936	297239	768	A629R [hypothetical protein] [0.0]	Q03604.1 [Ribonucleoside-diphosphate reductase large subunit] [0.0]
OSYNE5|712L	296075	296302	76	a632L [hypothetical protein] [8e-37]	N/A
OSYNE5|713L	296238	296555	106	N/A	N/A
OSYNE5|714L	296609	296947	113	N/A	N/A
OSYNE5|715R	297277	297642	122	A633R [hypothetical protein] [3e-78]	N/A
OSYNE5|716L	297643	298062	140	A634L [hypothetical protein] [7e-84]	N/A
OSYNE5|717R	298128	299750	541	N/A	Q1DDB7.1 [CTP synthase] [0.0]
OSYNE5|718L	299771	300085	105	a634aL [hypothetical protein] [1e-08]	N/A
OSYNE5|719R	299913	300178	88	A635R [hypothetical protein] [3e-44]	N/A
OSYNE5|720R	300234	300671	146	A637R [hypothetical protein] [1e-63]	N/A
OSYNE5|721R	300765	302081	439	A643R [hypothetical protein] [0.0]	N/A
OSYNE5|722L	301256	301492	79	N/A	N/A
OSYNE5|723L	301709	301927	73	N/A	N/A
OSYNE5|724R	302120	302635	172	A644R [hypothetical protein] [3e-116]	Q5UQL1.1 [Uncharacterized protein R409] [3e-07]
OSYNE5|725L	302395	302646	84	N/A	N/A
OSYNE5|726R	302729	303106	126	A645R [hypothetical protein] [2e-71]	N/A
OSYNE5|727L	303041	303232	64	N/A	N/A
OSYNE5|728R	303394	304167	258	A649R [hypothetical protein] [2e-164]	N/A
OSYNE5|729L	303526	303849	108	a080L [hypothetical protein] [1e-19]	N/A
OSYNE5|730L	304174	304767	198	A654L [hypothetical protein] [2e-127]	N/A
OSYNE5|731L	304263	304577	105	A655L [hypothetical protein] [1e-32]	N/A
OSYNE5|732L	304835	305482	216	A656L [hypothetical protein] [6e-68]	N/A
OSYNE5|733L	305669	306241	191	A659L [hypothetical protein] [2e-96]	N/A
OSYNE5|734R	305854	306162	103	a660R [hypothetical protein] [1e-22]	N/A
OSYNE5|735L	306265	306780	172	A662L [hypothetical protein] [8e-96]	Q54FR4.1 [PXMP2/4 family protein 4] [2e-08]
OSYNE5|736L	306866	307327	154	A664L [hypothetical protein] [2e-70]	N/A
OSYNE5|737R	307363	307623	87	N/A	N/A
OSYNE5|738L	307468	308046	193	A348R [hypothetical protein] [2e-31]	N/A
OSYNE5|739L	308181	308648	156	A665L [hypothetical protein] [7e-88]	N/A
OSYNE5|740L	308683	311295	871	N/A	Q9M2L4.1 [Calcium-transporting ATPase plasma membrane] [2e-170]
OSYNE5|741R	309059	309508	150	N/A	N/A
OSYNE5|742R	309586	310048	161	N/A	N/A
OSYNE5|743R	310103	310603	167	N/A	N/A
OSYNE5|744R	310985	311173	63	N/A	N/A
OSYNE5|745R	311246	311548	101	N/A	N/A
OSYNE5|746R	311455	311898	148	N/A	N/A
OSYNE5|747L	311906	312189	88	N/A	N/A
OSYNE5|748L	312215	315058	948	A666L [hypothetical protein] [0.0]	O94489.1 [Elongation factor 3] [0.0]
OSYNE5|749R	312665	312994	110	a667R [hypothetical protein] [5e-44]	N/A
OSYNE5|750R	313164	313382	73	a669R [hypothetical protein] [3e-42]	N/A

OSYNE5|751L	313651	313836	62	N/A	N/A
OSYNE5|752R	313826	314170	115	N/A	N/A
OSYNE5|753R	314546	315073	176	a670R [hypothetical protein] [7e-67]	N/A
OSYNE5|754R	315104	315751	216	A672R [hypothetical protein] [2e-110]	Q8GXE6.2 [Potassium channel] [4e-18]
OSYNE5|755L	315954	316145	64	N/A	N/A
OSYNE5|756R	315987	316637	217	A674R [Thymidylate synthase X] [1e-153]	O41156.1 [Probable thymidylate synthase] [1e-150]
OSYNE5|757R	316258	316467	70	N/A	N/A
OSYNE5|758R	316676	317788	371	A676R [hypothetical protein] [0.0]	N/A
OSYNE5|759L	316884	317096	71	N/A	N/A
OSYNE5|760L	317337	317618	94	N/A	N/A
OSYNE5|761R	317885	320122	746	A330R [hypothetical protein] [1e-38]	P16157.3 [Ankyrin-1] [5e-48]
OSYNE5|762L	318046	318396	117	N/A	N/A
OSYNE5|763L	319685	319981	99	N/A	N/A
OSYNE5|764L	319711	319953	81	N/A	N/A
OSYNE5|765R	320171	321349	393	N/A	N/A

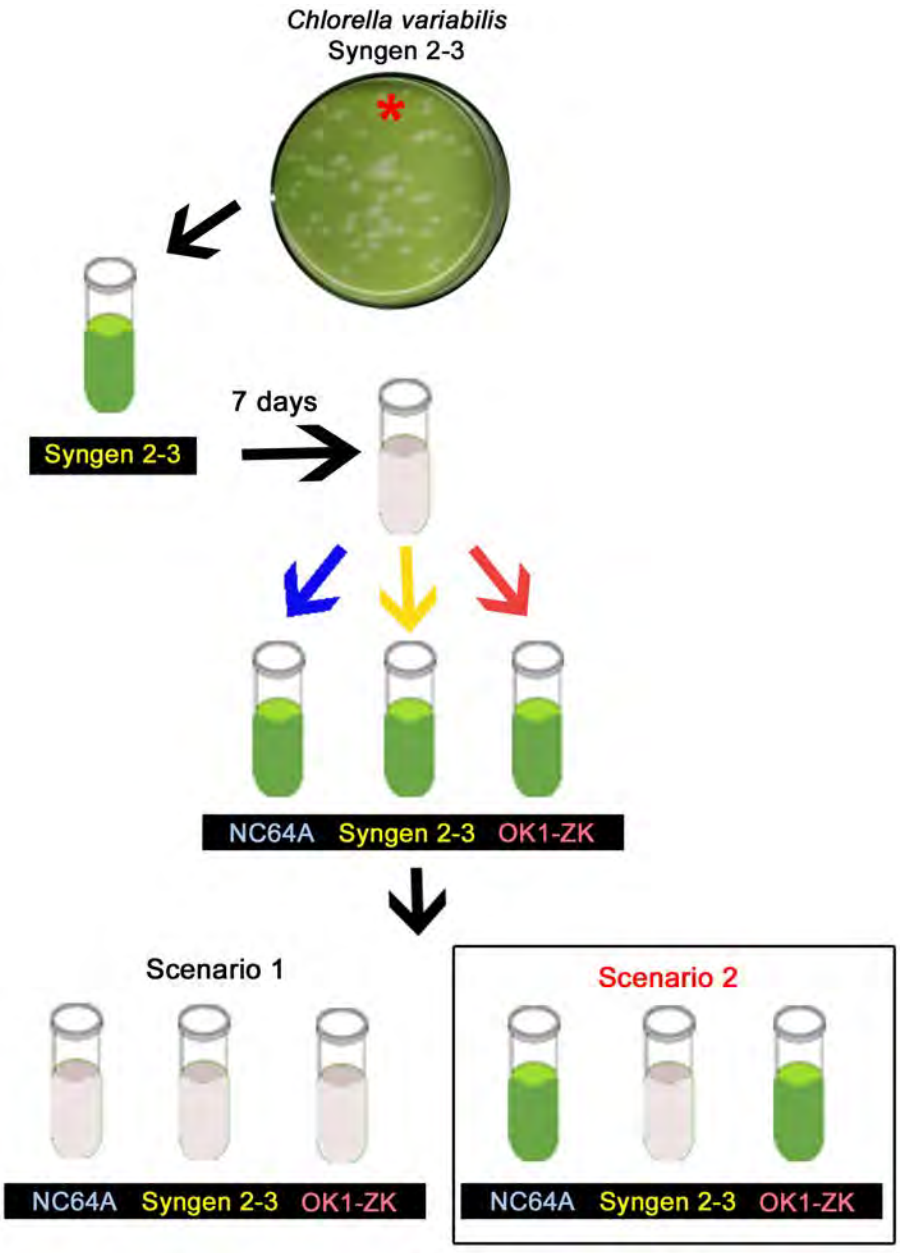
Table 6

Genomic region	CDCs	Region	Size	Query Cover	% Identity	Viral-hit	Non-viral	Most viruses are?	NC64A Virus Hit? If so, Best Hit?	PBCV-1?
A		1	142	57	50	Hypothetical	Nicotinate-nucleotide-dimethyl-benzimidazole phosphoribosyl transferase [Veillonella parvula]	Pbi-like	Yes, PBCV-1, Ny2A	Yes
A		2	1344	100	58	Chlorovirus glycoprotein repeat domain-containing protein	None	All, SAG-like	Yes, CviKI	No
A		3	245	100	83	Unknown function	None	All, Pbi-like	Yes, PBCV-1, AN69C	Yes
A		4	290	99	56	Hypothetical protein (mimivirus)	Emiliania huxleyi CCMP1516	Pbi-like	Yes, CviKI	No
A		5	952	99	65	Chlorovirus glycoprotein repeat domain-containing protein	None	All, SAG	Yes, AR158	No
A		6	446	98	71	Chlorovirus glycoprotein repeat domain-containing protein	None	All, Pbi-like	Yes, PBCV-1, CviKI	No
A		7	1483	98	67	Chlorovirus glycoprotein repeat domain-containing protein	None	All, Pbi-like	Yes, NY-2B	No
A		8	1506	100	65	Chlorovirus glycoprotein repeat domain-containing protein	None	All, Pbi-like	Yes, AN69C	No
A		9	411	99	92	Major capsid protein VP54	None	All, Pbi-like	Yes, CviKI	Yes
A		10	1033	99	58	Chlorovirus glycoprotein repeat domain-containing protein	None	All, Pbi-like	Yes, IL-3A	No
A		11	288	100	74	Serine/threonine-protein kinase	None	All, Pbi-like	Yes, AN69C	Yes
A		12	300	69	52	Aminoglycoside nucleotidyltransferase ANT9 [Ectothiorhodospira sp. PHS-2]	Ectotheiorhodospira sp. PHS-1	All, Pbi-like	Yes, NY2A	No
A		13	160	74	55	Hypothetical protein	None	Pbi-like	No	No
A		14	236	86	60	Phosphate ABC trasnporter substrate-binding protein [Isoptericola variabilis]	Isoptericola variabilis	All, Pbi-like	Yes, NY2A	No
A		15	190	36	53	Threonine aldolase [Desulfovibrio acrylicus]	Desulfovibrio acrylicus	All, Pbi-like	Yes, NY2A	No
B		1	377	100	71	Adenine-specific methyltransferase (PBCV)	Brachyspira pilosicoli	Pbi-like	Yes, NY-2B	No
B		2	338	99	40	D-alanyl-D-alanine dipeptidase [Mycobacterium sp. UNCCL9]	Mycobacterium sp. UNCCL9	All, Pbi-like	Yes, NY2A	Yes
B		3	192	100	62	Protein of unknown function	None	All, Pbi-like	Yes, AN69C	Yes
B		4	571	16	32	Endonuclease	Acanthamoeba castellanii str. Neff	Pbi-like	No	No
B		5	371	100	62	Adenine-specific methyltransferase [PBCV]	Thermogladius cellulolyticus	Pbi-like	Yes, NY2A	No
B		6	392	98	55	Adenine-specific methyltransferase (PBCV)	Brachyspira suanatina	Pbi-like	Yes, NY2A	No
B		7	247	98	86	Protein of unknown function (PBCV)	None	All, Pbi-like	Yes, CviKI	Yes
B		8	390	21	62	Large Ala/Glu-rich protein [Streptomyces sp. NRRL S-340]	Nocardia transvalensis	Only NC64A-like	Yes, IL-3A	Yes
B		9	191	43	43	Carbon catabolite-derepressing protein kinase [Rhizocotonia solani AG-1 IB]	Dictyostelium purpureum	Only NC64A-like	Yes, AR158	No
C		1	147	59	32	Pyrophosphorylase ModD [Sutterella sp. CAG:397]	Leisingera sp. ANG-Vp	Pbi-like	Yes, CviKI	No
C		2	166	43	43	Carbon catabolite-derepressing protein kinase [Rhizocotonia solani AG-1 IB]	Dictyostelium purpureum	Only NC64A-like	Yes, AR158	No
C		3	100	96	28	Phosphoribosylaminoimidazole-succinocarboxamide synthase [Bacillus sp. SB49]	Candidatus Entotheonella sp. TSY2]	N/A	No	No
C		4	109	88	65	beta-phosphoglucomutase [Butyrivibrio fibrisolvens]	Butyrivibrio fibrisolvens	Only NC64A-like	Yes, PBCV-1, AR158	Yes
C		5	114	65	34	inorganic polyphosphate kinase [Leisingera sp. ANG-Vp]	Ophiostoma picaea UAMH 11346	N/A	No	No
C		6	127	80	80	16S rRNA processing protein RimM [Candidatus Atelocyanobacterium thalassa]	Paenisporosarcina sp. TG20	Pbi-like	Yes, PBCV-1, AR158	Yes
C		7	881	100	98	elongation factor 3 [Paramecium bursaria Chlorella virus]	Rhizobium sp. AAP43	All, Pbi-like	Yes, KS1B	Yes
C		8	498	100	99	calcium-transporting ATPase, plasma membrane-type [Paramecium bursaria Chlorella virus]	Volvox carteri f. nagariensis	All, Pbi-like	Yes, KS1B	No

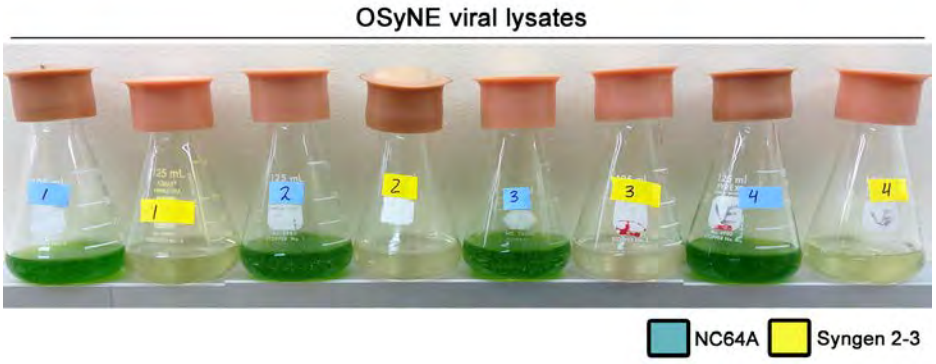
Supplementary Figure 1



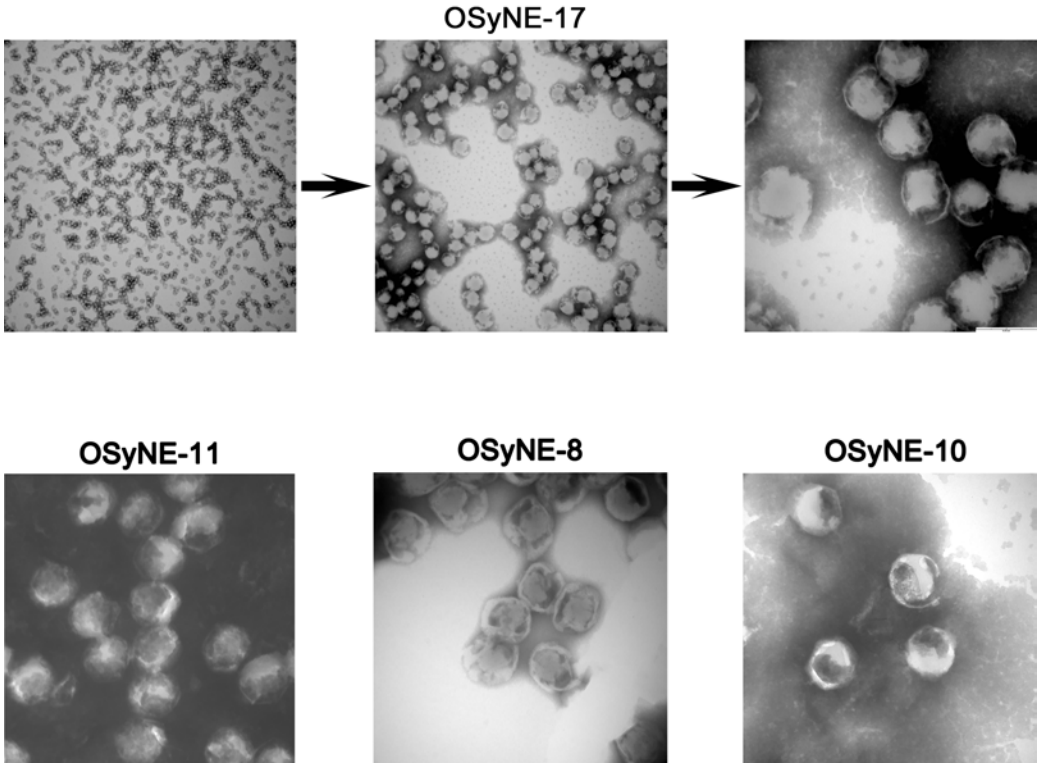
Supplementary Figure 2



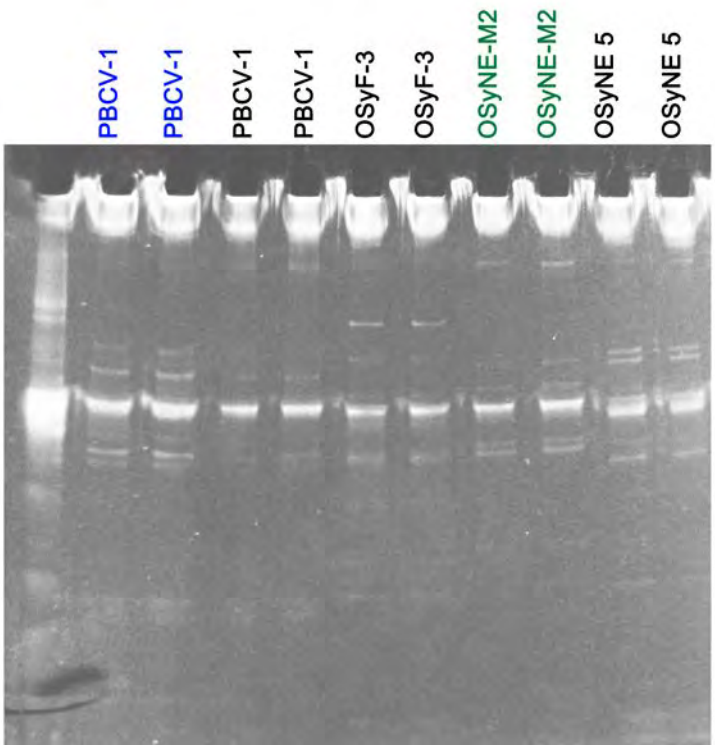
Supplementary Figure 3



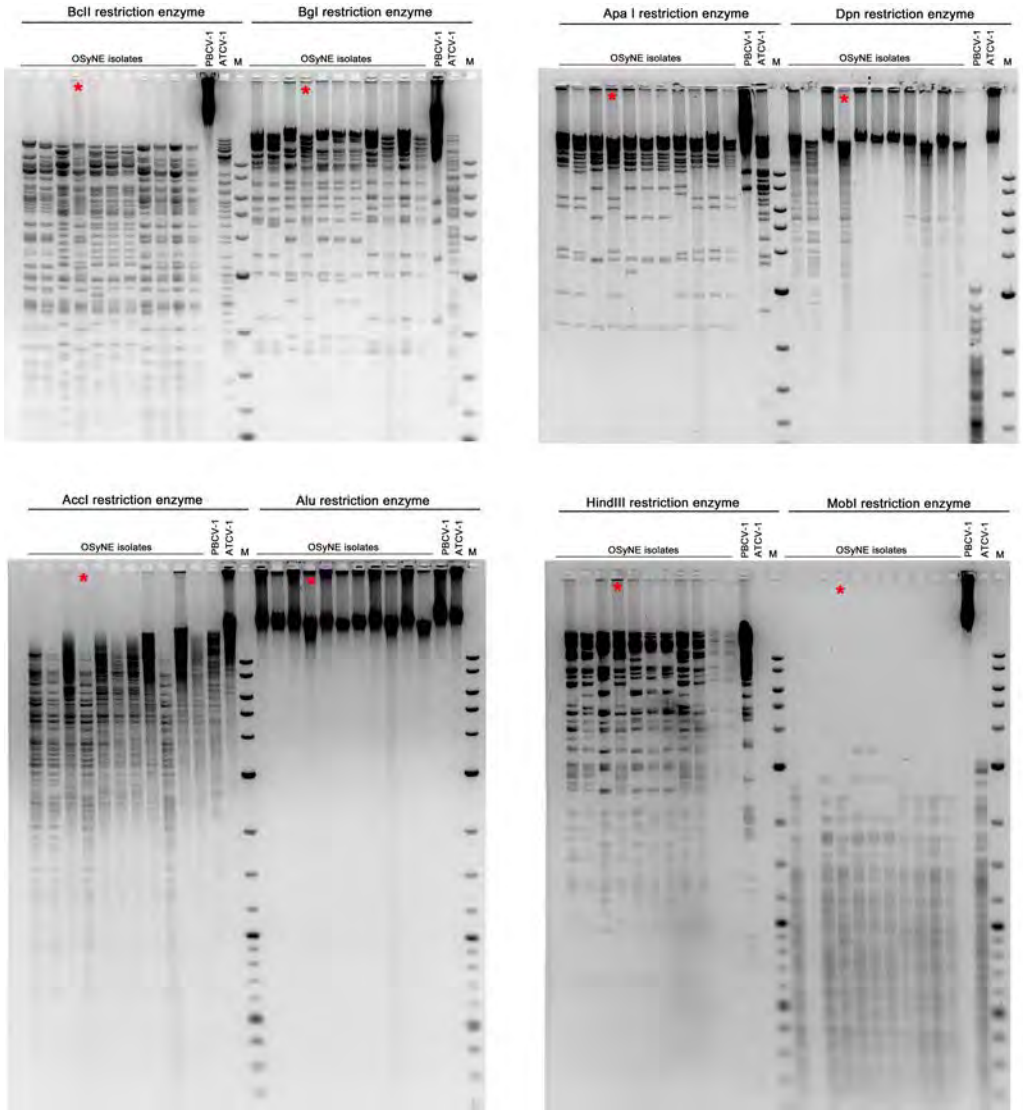
Supplementary Figure 4



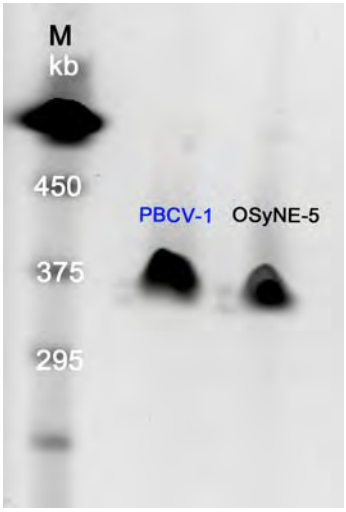
Supplementary Figure 5



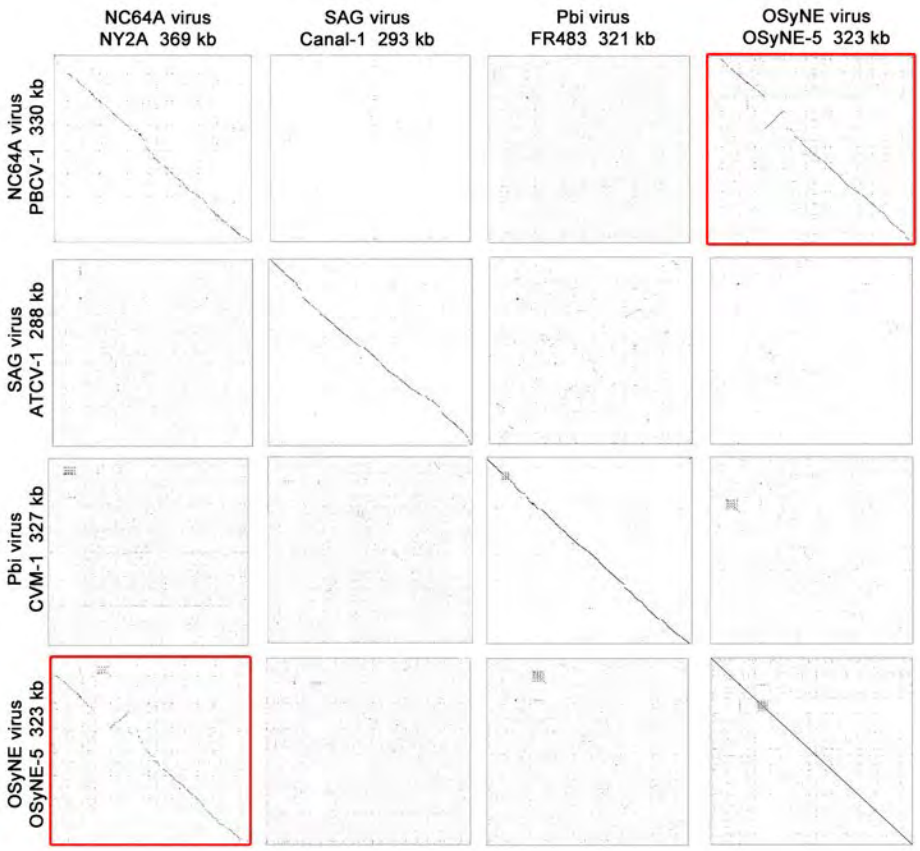
Supplementary Figure 6



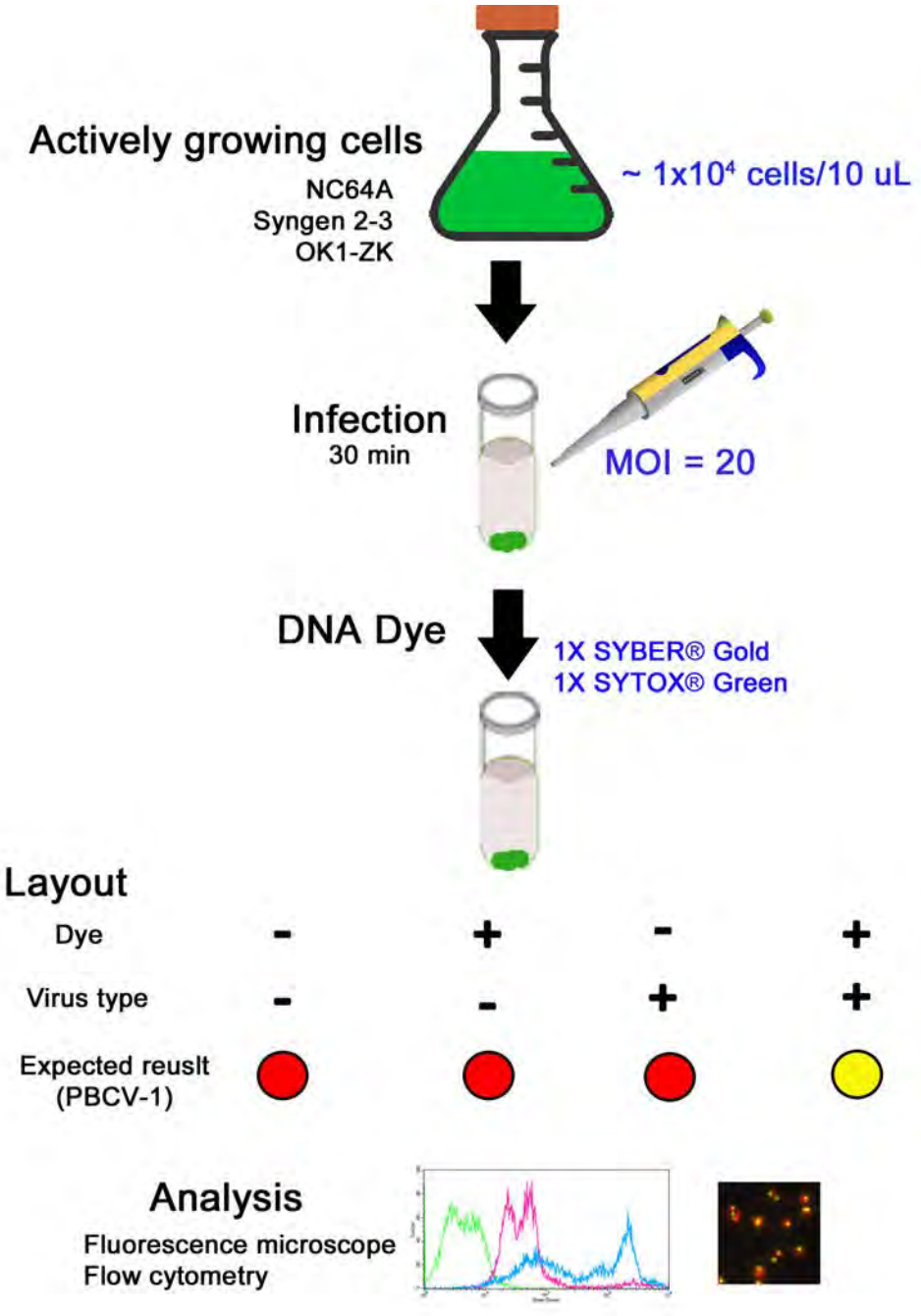
Supplementary Figure 7



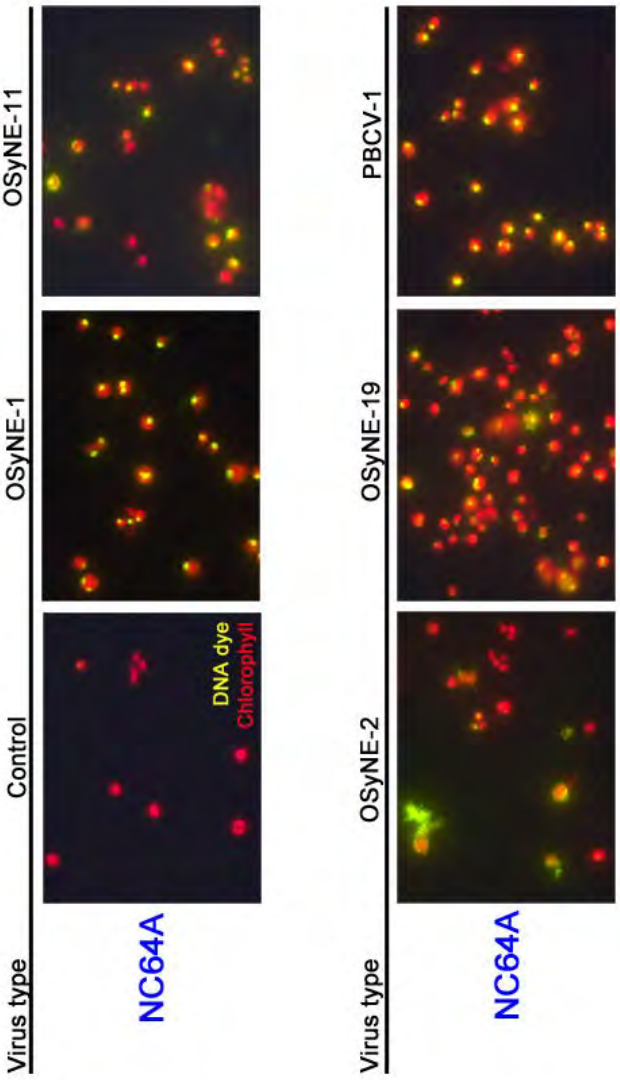
Supplementary Figure 8



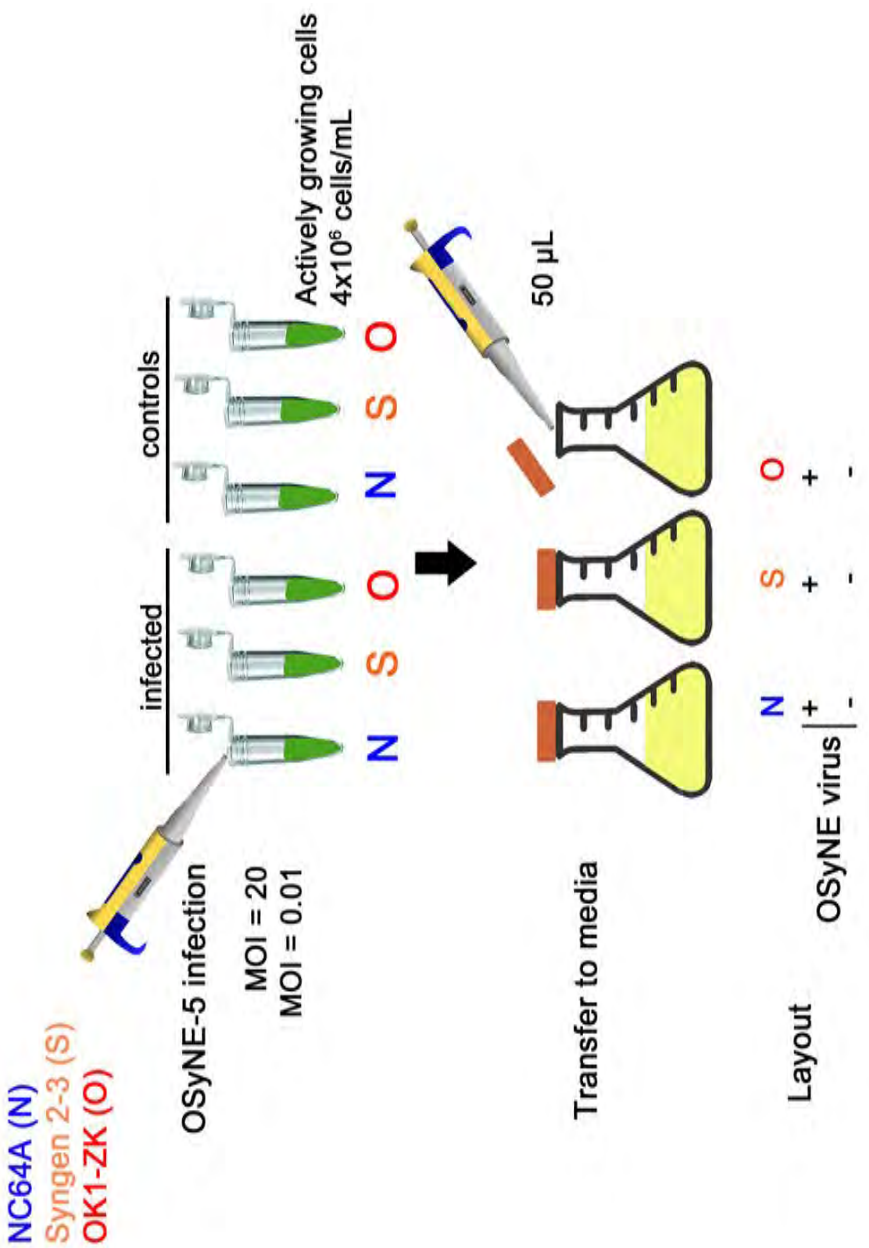
Supplementary Figure 9



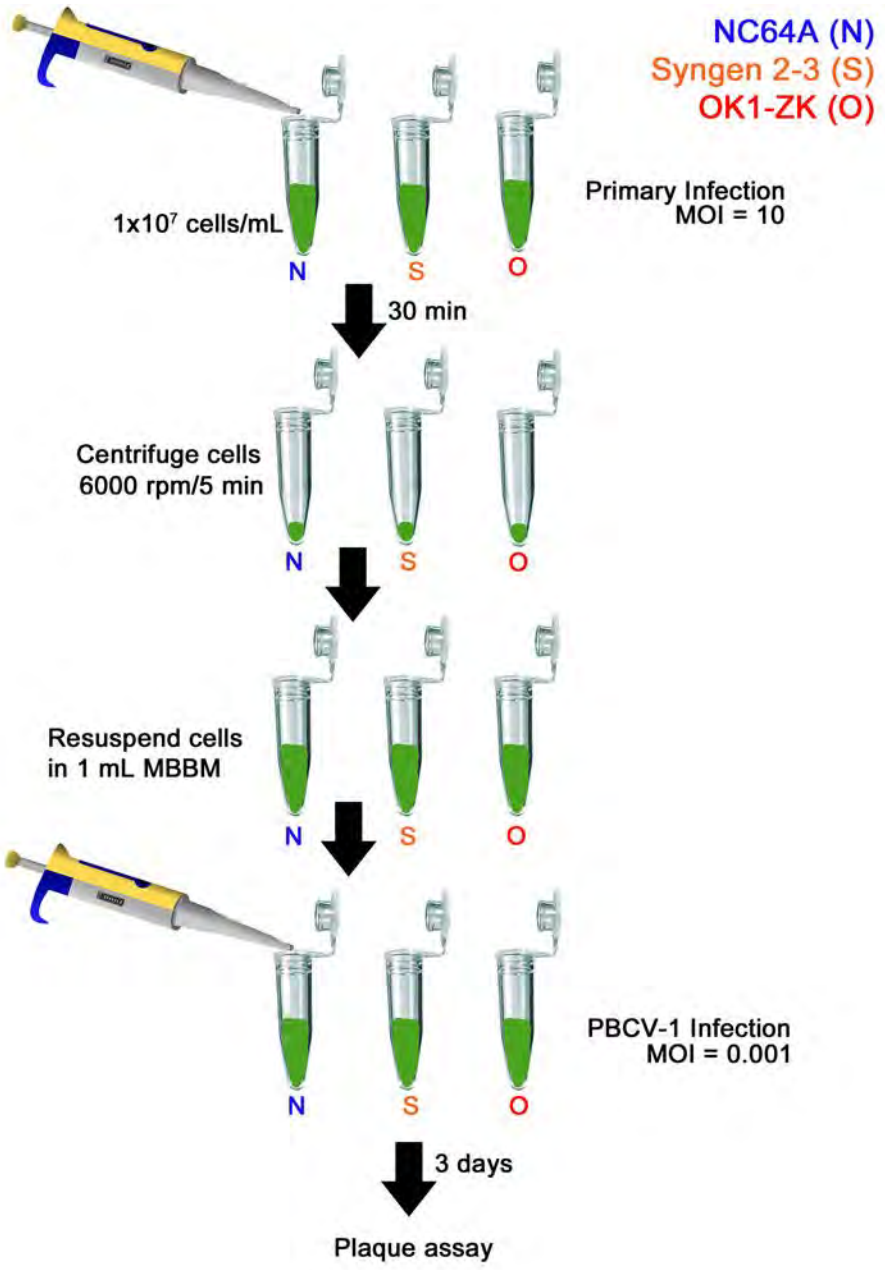
Supplementary Figure 10



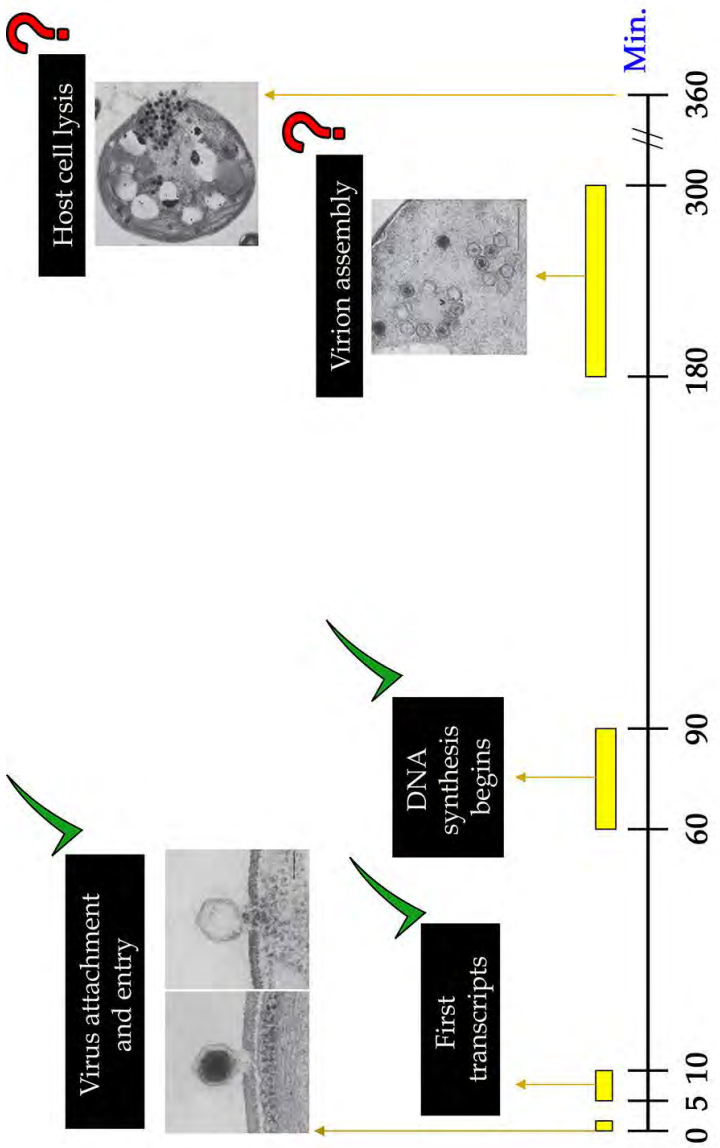
Supplementary Figure 11



Supplementary Figure 12



Supplementary Figure 13



Supplementary Table 1

Only Syngen virus isolates (OSy)

Virus	Number of Isolates	Collection Site
OSyNE	22	Lincoln, NE
OSyNE-M	9	Middle Loup River, NE
OSyNE-L	9	Loup River, NE
OSyF	3	Apalachicola River, FL

Supplementary Table 2

Virus	Host	Collection Site
OSYNE-1	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-2	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-3	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-4	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-5	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-6	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSYNE-7	Syngen 2-3	Lincoln, Nebraska, Feb 2013
OSY-F1	Syngen 2-3	Apalachicola River, Florida (USS-18 02359170)
OSY-F3	Syngen 2-3	Apalachicola River, Florida (USS-18 02359170)
OSYNE-M2	Syngen 2-3	North bend of middle Loup River, Nebraska, Summer 2013 (CC)
OSYNE-L2	Syngen 2-3	Marshy north bend of middle Loup River, Nebraska, Summer 2013 (CC#2)
KV1	F36-ZK	Lincoln, Nebraska, April 2014

CHAPTER IV

METABOLISM OF THE *CHLOROVIRUS* SYMBIOTIC CHLORELLA HOSTS

Comparative Genomics, Transcriptomics and Metabolism DistinguishSymbiotic from Free-living Chlorella

Cristian F. Quispe[1, 3], Olivia Sonderman[1, 2], Maya Khasin[3], Wayne Riekhof[3], Kenneth W. Nickerson[3] and James L. Van Etten[1, 2]

[1]Nebraska Center for Virology, [2]Department of Plant Pathology and [3]School of Biological Science, University of Nebraska-Lincoln, NE 68583-0900.

Short title: **Metabolic Reprograming as a Driver of Endosymbiosis in Chlorella Strains**

Correspondence: Cristian Quispe. School of Biological Sciences. University of Nebraska-Lincoln. 204 Morrison Center. Lincoln, NE 68583. Fax: 402.472.3323. Phone: (402) 472-5776. Email: quispecristian@gmail.com

Acknowledgements

Funding for this work was partially provided by the NSF-EPSCoR grant EPS-1004094 (JVE) and the COBRE program of the National Center for Research Resources Grant P20-RR15535 (JVE). OS was supported by UNL UCARE and ARD scholarships. We thank Steve Ladunga at UNL for statistical assistance and the UNL algal consortium.

Disclosure Statement: The authors declare no conflict of interest.

Author Contributions: C.F.Q., K.W.N and J.L.V.E. designed research; C.F.Q. and O.S., performed the experiments; C.F.Q., M.K, W.R., K.W.N and J.L.V.E. analyzed data; and C.F.Q., K.W.N and J.L.V.E. wrote the paper.

Comparative Genomics, Transcriptomics and Metabolism Distinguish Symbiotic from Free-living Chlorella

Abstract

Most animal-microbe symbiotic interactions must be advantageous to the host and provide nutritional benefits to the endosymbiont. When the host provides nutrients, it can gain the capacity to control the interaction, promote self-growth and increase its fitness. Chlorella-like green algae engage in symbiotic relationships with certain protozoans, a partnership which significantly impacts the physiology of both organisms. Consequently, it is challenging to grow axenic chlorella cultures after isolation from the host, as they are nutrient fastidious and susceptible to virus infection. We hypothesize that the establishment of a symbiotic relationship spurred natural selection on nutritional and metabolic traits that differentiate symbiotic algae from their free-living counterparts. Here, we compare metabolic capabilities of five symbiotic and four free-living *Chlorella* algae by determining growth levels on combinations of nitrogen and carbon sources. Data analysis by hierarchical clustering reveals clear separation of the symbiotic and free-living *Chlorella* into two distinct clades. Symbiotic algae cannot metabolize NO_3 but can utilize three symbiont-specific amino acids (Asn, Pro and Ser). These amino acids were exclusively affected by the presence/absence of Ca^{2+} in the medium, and differences were magnified if galactose but not sucrose or glucose was provided. Additionally, *Chlorella*

variabilis NC64A genomic and differential expression analysis confirm the presence of abundant amino acid transporter protein motifs, some of which the algae constitutively express axenically and within the host. Significantly, all five symbiotic strains exhibit similar metabolic phenotypes although they arise as protozoan symbionts from different origins. Such similarities indicate a parallel coevolution of shared metabolic pathways across multiple independent symbiotic events. Collectively, our results suggest that physiological changes drive the *Chlorella* symbiotic phenotype and contribute to their natural fitness.

Keywords: *Chlorella variabilis*, symbiosis, metabolism, amino acids, galactose

Introduction

Successful endosymbiosis provides advantages to both the host and the endosymbiont. Benefits may include better adaptation to nutrient limitation or reduction of mortality via protection against damage by UV light or pathogens (e.g. viruses). In such scenarios, symbionts increase their reproductive capacity and fitness within hosts relative to non-host environments (Bischoff and Bold, 1963; Johnson 2011; Karakashian 1975; Karakashian and Karakashian, 1965). For example, some protozoans harbor intracellular chlorella-like green algae in an inherited mutually beneficial symbiotic relationship, which serves as a well-recognized model for studying endosymbiotic relationships (Kovacevic, et al. 2007; Kovacevic 2012; Park et al. 1967; Siegel 1960).

Unicellular chlorella-like green algae inhabit the gastrodermal symbiosomes (perialgal vacuoles) of different protozoans and transfer a significant amount of their photosynthetically fixed carbon (e.g. maltose, fructose) to the non-photosynthetic partner (Cernichiari, et al. 1969; Karakashian 1975; Matzke et al. 1990). In this context, symbiotic chlorella require nutrients such as nitrogen from the host and their assimilation into the algal metabolome (McAuley 1987; Yellowlees, 2008). The mechanisms involved in this interaction have not yet been completely elucidated; however, the metabolic pathways involved in nitrogen and carbon utilization could be crucial physiological signatures of endosymbiosis (McAuley 1987). Endosymbiosis was essential during eukaryogenesis (Hom and Murray 2014; Lopez-Garcia and Moreira 2015);

therefore elucidating how such processes work would open new avenues of research in the understanding of the molecular, cellular, and organismal adaptations that allow successful mutualism.

Protozoan-chlorella intracellular interactions can be disrupted, and some attempts to isolate intact algae free of the host have been successful. These include algae that had been associated with several species of protozoans, including *Paramecium bursaria* (Kessler and Huss 1990; Siegel 1960), *Acanthocystis turfacea* (Kessler and Huss 1990), and *Hydra viridis* (Pardy and Muscatine 1973; Van Etten et al. 1981). Another approach to identify ex-symbiotic algal strains relied on their susceptibility to large DNA virus infections after the disruption of the host-chlorella interaction (Kvitko 1984; Kawakami and Kawakami 1978; Meints et al. 1981; Van Etten et al. 1983a). The only documented symbiotic-virus susceptible *Chlorella* species, which can be cultured axenically, include *Chlorella variabilis* NC64A (Van Etten et al. 1983b), *C. variabilis* Syngen 2-3 (Van Etten et al. 1983a), *C. variabilis* F36-ZK (Fujishima 2010; Kamako et al. 2005; Proschold et al. 2011), *C. variabilis* OK1-ZK (Fujishima 2010; Proschold et al. 2011), and *C. heliozoae* SAG 3.83 (Bubeck and Pfitzner 2005). Hereafter, for the purpose of this paper, symbiotic-virus susceptible algae strains will be referred to as symbiotic algae.

We are studying the chlorella-virus interaction for the past 35 years; however, we puzzled by the fastidious nutrient requirements that symbiotic algae strains possess (Albers et al. 1982; Kato et al. 2006; McAuley 1987). For

instance, unlike most *Chlorella* species, the symbiotic algal strains fail to grow on Bolds' Basal Medium (BBM), which has NO_3 as its sole N source. Thus, as a matter of convenience, 1% peptone was added when growing these symbiotic strains axenically (Jolley and Smith 1978; Kamako et al. 2005). We hypothesize that the establishment of a symbiotic relationship spurred intense natural selection on specific nutritional and metabolic features unique in symbiotic algae. In this study, we examine that idea by analyzing some physiological traits and growth requirements by comparing four free-living strains and five symbiotic *Chlorella* species.

Our physiological evaluation, which focused on alternative nitrogen (N) and carbon (C) sources, shows that symbiotic algae are better able to assimilate less preferred N and C sources. Significantly, they prefer organic N sources; inorganic N sources (e.g. NO_3 or NH_4), which are the primary sources of N in the environment, are poorly assimilated by symbiotic algae. Additionally, they also assimilate less preferred C sources (e.g. galactose over sucrose or glucose). Importantly, all symbiotic strains exhibit similar metabolic phenotypes although they are polyphyletic and arise as protozoan symbionts from different origins (Fujishima 2010). Namely, they derive from multiple independent symbiotic events. Such similarities denote a parallel coevolution of similar metabolic pathways across multiple independent symbiotic events. Taken together, in symbiotic *Chlorella* strains, ancient evolutionary genome plasticity and metabolic regulatory rewiring at the cellular level could come with a cost in nature (e.g.

metabolic adaptation for endosymbiosis, inability to survive as a free-living and virus susceptibility).

Materials and Methods

Algal strains

Symbiotic *C. variabilis* NC64A, *C. variabilis* Syngen 2-3 and *C. heliozoae* SAG 3.83 were maintained as slant stocks at 4°C. Symbiotic *C. variabilis* F36-ZK (NIES-2540) and *C. variabilis* OK1-ZK (NIES-2541) were obtained from the Japanese Culture Collection of the National Institute for Environmental Studies (http://www.nies.go.jp/index-e.html). Stock samples of free-living strains *C. sorokoniana* (UTEX-1230), *C. sorokoniana* (CS-01), *C. kessleri* (B228) and *C. protothecoides* (CP-29) were obtained from the Culture Collection of Algae at University of Texas at Austin (http://web.biosci.utexas.edu/utex/).

Cell cultures

Symbiotic and free-living strains were grown on BBM (Bischoff and Bold 1963) supplemented with 1% (w/v) peptone, 5% (w/v) sucrose and 0.001% (w/v) thiamine (complete MBBM) (Suppl. Fig. 8). Where indicated, 1 % peptone was replaced with 1% (w/v) casamino acids. The ability of algae to exploit different N and C sources was tested by adding to the N- C- deficient BBM (N-/C-BBM) the N and/or C source to be tested. Thus, 0.22 um filter-sterilized stock solutions of N and C sources were added to a final concentration of 10 mM. All flasks were

supplemented with 0.001% (w/v) thiamine. To test the effect of Ca^{2+} deprivation on algal growth, we followed a similar procedure but C-, N- and Ca^{2+} -deficient BBM (N-/C-/Ca^{2+} -BBM) were used.

125 mL narrow mouth pyrex Erlenmeyer flasks with 30 ml of supplemented BBM were prepared (Suppl. Fig. 9). For the inoculum, MBBM log-phase actively growing cells were pelleted and washed 3 times with either N-/C-BBM or N-/C-/Ca^{2+} -BBM medium. Flasks were inoculated to a final cell density of 1-5 x 10^5 cells/ml and shaken at 26°C/180 rpm in continuous light for variable time periods because symbiotic growth rates are slower compared to their free-living counterparts. Free-living strains were grown for 9 days on BBM with an added N source or for 7 days when both N and C sources were added. Similarly, symbiotic strains were grown for 12 days on BBM with added N source or 9 days when both N and C sources were included. MBBM and unsupplemented BBM were used as controls. Triplicate samples were run for the symbiotic algae, and duplicate samples were run for the free-living strains. Flask images were taken with a 12.1 mega pixel Sony Cyber-shot digital camera. Pictures were organized using Adobe Photoshop CS5.1.

Hierarchical Clustering analysis

We used Cluster 3.0 for Mac OS X (http://rana.lbl.gov/EisenSoftware.htm) and JavaTreeView Version 1.1.6r4 (http://jtreeview.sourceforge.net/) programs to analyze and visualize the growth experiments. The hierarchical clustering

algorithm was performed using the average-linkage method applied to the dataset. The objective of this algorithm is to compute a dendrogram that assembles all elements into a single tree, so the tree clusters the strains and treatments according to similarities in their growth patterns. The dataset consists of rows representing the nine algal strains and columns representing the numerical score for each media condition. The analysis was performed as bulk data and as a subset by treatment. The numerical score was assessed on individual flasks using a 0 to 5 scale, with 5 representing the best growth and 0 the absence of growth.

Display

The dataset is represented graphically in a hierarchical clustering by coloring each cell on the basis of the numerical flask score. Flasks with scores of 0 (no growth) are colored black and scores rise with reds of increasing intensity to denote growth. The dendrogram is attached on both axes to the colored graph to indicate the nature of the computed relationship among growth conditions and *Chlorella* species.

Comparative Genomics

Amino acid transporters from *A. thaliana* were identified using the expressed AA transporters in *C. variabilis* NC64A (Blanc et al. 2010) as queries. Sequences were examined by similarity search using BLASTP (Altschul et al.

1997) against the non-redundant (NR) database from the National Center for Biotechnology Information (NCBI: http://blast.ncbi.nlm.nih.gov/). Orthology of candidate sequences was verified using the *C. variabilis* NC64A KEGG database (Kanehisa and Goto 2000). Top hit sequences with more than 60% similarities and query coverage were considered as putative homologs. Additionally, members of a collection of characterized AA transporters from *A. thaliana* (Tegeder 2012) were used to perform a BLAST search against NC64A and UTEX-1230 (UNL algal consortium, *in preparation*) genomes, using an expected value of 1×10^{-10} as a cutoff. Each algal protein returned an *A. thaliana* AA transporter, and the gene designation and E-value for each gene is presented in Tables 1 and 2. Similarly, 35 putative AA transporters from NC64A (Blanc et al. 2010) were used to perform a BLAST search against UTEX-1230 proteome (Table 3).

RNAseq analysis

Datasets from RNAseq experiments were downloaded to the public Galaxy platform server (www.usegalaxy.org) and manipulated with data analysis tools as described below. For axenic *C. variabilis* NC64A, we used an uninfected control sample (NCBI SRA accession SRX316780) from a recently published viral infection experiment conducted in our lab (Rowe et al. 2013). For *C. variabilis* growing endosymbiotically within *P. bursaria*, we downloaded RNAseq datasets (NCBI SRA accessions DRX003053, DRX003054, and DRX003055)

(Kodama et al. 2014). These sequence files were reported to contain RNAseq reads mapping to the NC64A genome, thus providing potential information regarding genes that are differentially expressed when the alga is grown axenically on MBBM versus its natural endosymbiont stage with *P. bursaria*. The FASTQ files were converted to FASTQSANGER format with the FASTQ Groomer tool (Blankenberg et al. 2010) and Tophat (Kim et al. 2013) was used to align these datasets to the NC64A genome assembly (Blanc et al. 2010) with a minimum and maximum intron length of 50 and 5,000, respectively. Around 1% (~970,000) of the *P. bursaria* derived reads aligned to the *C. variabilis* genome, and these reads were taken to represent a snapshot of gene expression in endosymbiont cells. The same analysis pipeline was applied to the axenic *C. variabilis* NC64A data. Reads that mapped to the genomic intervals for each putative AA transporter denoted in Table 1 were counted using the Integrated Genome Browser software package (Nicol et al. 2009) and normalized as total mapped reads per gene in each condition per million mapped reads.

Results

High-throughput nutritional analysis identified distinct metabolic signatures for symbiotic and free-living Chlorella species

The 567 growth conditions depicted in Suppl. Figs. 1-6 were analyzed by a two-way heat map (Fig. 1). This average-linkage map displayed differences in metabolic capabilities of five symbiotic (green) and four free-living (blue)

Chlorella species (Suppl. Fig. 7). Combinations of organic and inorganic nitrogen (N) sources were tested with or without the addition of one of three sugars as a carbon (C) source. Both C and N sources were added at 10 mM concentration. The columns represent variations of 3 C and 12 N sources (2 complex mixtures, 7 organic, 3 inorganic) prepared on nitrogen-carbon-free (N-/C-) BBM. Purple labels identify inorganic N sources at only 1 mM concentration while orange labels represent media without Ca^{2+} both in N-/C- BBM. Rows represent the growth of nine *Chlorella* strains on the 64 media combinations. The five symbiotic strains include: *C. variabilis* NC64A, *C. variabilis* Syngen 2-3, *C. variabilis* F36-ZK, *C. variabilis* OK1-ZK, and *C. heliozoae* SAG 3.83. The four free-living strains are: *C. sorokoniana* (UTEX-1230), *C. sorokoniana* (CS-01), *C. kessleri* (B228) and *C. protothecoides* (CP-29). Over five-hundred (n=567) outcomes were plotted using cluster analysis. For each growth medium, triplicate tests were performed for the symbiotic *Chlorella* with duplicate analyses for the free-living *Chlorella*. Strains were compared with regard to their ability to assimilate the nutrient sources by assigning a score between 0 and 5 for algae growth based on color intensity in the flask after 9-12 days of growth for the symbiotic and 7-9 days of growth for the free-living strains. Gene Cluster 3.0 and Java TreeView clustered the data in a heat map layout (Fig. 1). Within the figure, red color represents robust growth and black color represents the absence of growth. A tree diagram is attached on both axes of the heat map to indicate the nature of the computed relationships among growth conditions and among the nine

Chlorella species. Importantly, two clusters clearly separate a symbiotic clade (top) from the free-living clade (bottom) based on their nutritional capabilities.

Casamino acids or peptone alone supply all the carbon and nitrogen molecules needed for symbiotic algal growth

Symbiotic algae possess fastidious nutrient requirements, and several reports confirm difficulties in the isolation and axenic growth of symbiotic algae (Albers et al. 1982). For example, *C. variabilis* NC64A did not grow on unsupplemented BBM but grew well on BBM with 1% peptone and 5% sucrose (MBBM) (Jolley and Smith 1978; Kamako et al. 2005; Van Etten et al. 1983b). Previous reports established that BBM with peptone alone was sufficient for growth of *C. variabilis* NC64A and F36-ZK (Kamako et al. 2005). Although complete, the heat map presented in Fig. 1 is cumbersome due to its large size. Therefore, we prepared 41 smaller analyses reflecting metabolic subsets of the data, and 6 of these subsets are presented. In each subdivision, five symbiotic and four free-living *Chlorella* were compared based on their abilities to grow on modifications of BBM and MBBM.

In the first case, we analyzed the sub-group based on complex polypeptide mixtures (peptone and casamino acids) with or without sucrose. The symbiotic and free-living strains formed two distinct clades (Fig. 2) while two media clusters also appeared, one for casamino acids alone and a second which included

peptone + sucrose (MBBM), peptone alone, and casamino acids + sucrose. The free-living strains did not grow well on casamino acids alone.

This analysis confirms that BBM with peptone is sufficient for growth of *C. variabilis* NC64A and F36-ZK and extended the analysis to include three additional symbiotic strains (*C. variabilis* OK1-ZK and Syngen 2-3 as well as *C. heliozoae* SAG 3.83). These five *Chlorella* strains formed the symbiotic algal clade. All the symbiotic strains grew slightly better on 1% casamino acids than on 1% peptone and, for both peptone and casamino acids, removal of sucrose and NO_3 from the control MBBM had no effect on their growth (Fig. 2). Although the differences in growth were slight, the data suggest that organic N sources rich in amino acids (AA), such as casamino acids, might be better assimilated by the symbiotic over the free-living group. In contrast, stronger growth was observed among most free-living species upon the addition of sucrose to either casamino acids or peptone, suggesting that they are better adapted to the presence of a sugar source in the media.

Asparagine, serine, and proline are better assimilated by symbiotic strains

Fig. 2 suggests that simpler organic N sources such as free amino acids (AA) might be better assimilated by symbiotic strains. The juxtaposition of most symbiotic stains growing comparatively better on casamino acids along with prior data that NC64A could not utilize NO_3 drew our attention to the N metabolism of these organisms (Suppl. Fig. 11). Thus, we tested the metabolic capabilities of the nine *Chlorella* species on ten organic and inorganic N sources.

The dendogram (Fig. 3) shows a clear separation of symbiotic and free-living strains based on their N assimilation patterns. All of the *Chlorella* species grew robustly with Arg, urea, Gln and Gly as the sole N source in the media with the proviso that the symbiotic strains grew slightly better. A separate cluster was formed with Asn, Ser, and Pro with the symbiotic algae growing consistently better on these 3 AA as compared to the free-living set. Within the symbiotic-specific group, Asn prompted better growth than Ser, which in turn exceeded growth on Pro. Thus, we confirm that organic N sources in general are better assimilated by symbiotic than free-living algae. Following this trend, Asn, Ser, and Pro appear to be symbiont-specific in that they were used poorly, if at all by free-living Chlorella. Hence, Asn, Ser and Pro metabolism might be important during axenic and symbiotic growth. The only inorganic N source that clustered within the organic group was NH_4 acetate.

The other two inorganic N sources, NH_4 tartrate and sodium NO_3, clustered in a different clade, which exhibited poor or no growth for all strains. This finding was surprising because NH_4 and NO_3 are the primary sources of N in most environments. In plants, NO_3 reductase is one of the few substrate-inducible enzymes known and regulates external and internal signals that include light, phytohormones, circadian rhythms, plastidic factors, N and C metabolites and CO_2 (Vidal et al. 2015); hence the importance of N metabolism in cell homeostasis.

Thus, in symbiotic algae, we encountered a remarkable duality in the form of extracellular repression for NH_4 and NO_3 coupled with an intracellular ability to use NH_4 after uptake of Arg, Gln, Asn, and urea. These observations also confirm previous reports on the Japanese *Chlorella variabilis* symbiont F36-ZK regarding loss of NO_3 assimilation coupled with an enhanced ability to uptake certain AA (Kato et al. 2006). A general AA transport system is responsible for the ability of F36-ZK to rapidly up take most AA, and the system has an acid-proton symport mechanism that is pH-dependent (Kamako et al. 2005).

In plants, fungi and bacteria, most systems for AA transport are constitutively expressed or derepressed only during N starvation (Sauer, 1984). Thus, N metabolism in symbiotic algae is acting in starvation mode, hence the rapid assimilation of some AA by the constitutively active general AA transport system.

Low ammonium concentrations are better assimilated by symbiotic strains

We examined the inorganic N sources in greater detail (Fig. 4). Fungal growth media commonly include 10 mM concentrations of inorganic N salts; however, these levels could be inhibitory or toxic to most *Chlorella* species. Thus, the NH_4 and NO_3 sources were compared at two concentrations (1 mM and 10 mM). NH_4 tartrate was chosen because it is well known that it does not acidify medium for fungal growth as much as NH_4 sulfate or chloride (Hornby et al, 2001).

The 1 mM data are shown in purple in Fig. 4. Four of the five symbiotic strains clustered tightly except for Syngen 2-3, which clustered with one of the free-living

groups. Two clusters were formed when analyzing the media treatments, one included both NO_3 concentrations and another all the NH_4 salts.

The lower NO_3 level (1 mM) did not support any algal growth while at the higher level (10 mM) three of the free-living and Syngen 2-3 strain (symbiotic group) had minimal growth. In contrast, most symbiotic strains did not grow on NO_3. On NO_3 at 10 mM, only the Syngen 2-3 strain (symbiotic group) showed minimal growth (Fig. 4). This is the first cluster analysis where at least one symbiotic strain did not group within the symbiotic cluster.

Lower levels of NH_4 acetate and NH_4 tartrate (1 mM) supported growth of all symbiotic strains while higher concentrations (10 mM) were inhibitory. Thus, we conclude that most symbiotic strains grow better on low concentrations of NH_4 compared to free-living strains and that NH_4 acetate (1 mM) is better than NH_4 tartrate (1 mM) for both symbiotic and free-living strains. The acetate helps stimulate algal growth.

In summary, symbiotic algae possess an efficient system to import and metabolize many AA and small oligopeptides. Intriguingly, they cannot efficiently utilize NO_3 or NH_4 as sole N sources.

Galactose is a better carbon source than glucose or sucrose for symbiotic growth

All of the algal growth rates observed in Figs. 3-4 were slower than those observed on MBBM, which contains added sucrose and peptone. Thus, to further

clarify growth rates and patterns, we supplemented all N sources with 3 C sources: sucrose (present in MBBM), glucose, and galactose. Galactose was chosen because in other microbes it does not exert catabolite repression (Gancedo 1998). Results for the inorganic and organic N sources are shown in Figs. 5A and B, respectively.

All symbiotic strains did not grow with any combination of sugar and NO_3 (Fig. 5A) except for Syngen 2-3, which grew minimally in all NO_3/sugar treatments but the 10 mM NO_3-galactose. Thus, even after the addition of sugars, the symbiotic strains were unable to utilize NO_3 as a sole N source (Fig. 5A).

Similar to the phenotypes observed in Fig. 4, 1 mM NH_4 salts were better than 10 mM regardless of the sugar used, and consistently better growth was achieved with 1 mM NH_4 acetate as opposed to 1 mM NH_4 tartrate. Additionally, better growth was observed when galactose or sucrose was supplemented rather than glucose.

Although sugar supplementation increased the growth rates of all symbiotic strains, those rates never reached the growth levels observed on MBBM with any inorganic N source tested. Therefore, the remaining difference is undoubtedly due to the AA derived from the peptone on MBBM.

In contrast, the free-living strains had similar or better growth rates than on MBBM after the addition of sugars, and they preferred sucrose and glucose. Intriguingly, growth rates were significantly compromised when galactose was used, except for the strain B228, which grew on galactose. Similar to the results

in Fig. 4, most free-living strains consistently grew better on 10 mM salts than on 1 mM levels. Importantly, all free-living *Chlorella* grew on NO_3, NH_4 tartrate (10 mM), and NH_4 acetate.

Galactose and organic nitrogen sources improve growth for symbiotic strains but galactose is inhibitory for free-living strains

Addition of galactose and organic N sources to symbiotic strains improved their growth rates to levels similar to MBBM (Fig. 5B). For the symbiont-specific organic N sources (Asn, Ser, and Pro), better growth rates were observed for Asn than for Ser and Pro.

In contrast, galactose had inhibitory effects on the growth of free-living strains with most organic N sources, including the symbiont-specific N sources as well as urea, Gln, and Arg. Minimal effects were observed for Gly. Strain B228 was the only free-living strain that was able to utilize multiple organic N sources in the presence of galactose. NC64A appears to have more enzymes involved in carbohydrate metabolism than other sequenced chlorophytes (Blanc et al. 2010), including some related to galactose metabolism (Suppl. Table 3). In addition, in free-living strains, sucrose and glucose increased the growth rate of some organic N sources; however, the phenotypes might be strain-specific.

Addition of sucrose also improved the assimilation of organic N sources on symbiotic strains but not at levels similar to those observed for galactose.

Interestingly, addition of glucose inhibited symbiotic growth on most organic N sources, with the most glucose-sensitive strain being NC64A (Suppl. Fig. 10).

Differential calcium regulation in the assimilation of organic N sources by symbiotic and free-living species

Amino acid transport is coupled to movement of ions, including Na^+, H^+, K^+, Ca^{2+} and/or Cl^- as well as movement of sugars. Thus, we investigated the possible role of Ca^{2+} regulation in transport that might differ between symbiotic and free-living *Chlorella* species.

We compared the effectiveness of organic N sources with (Fig. 6, black) or without Ca^{2+} (Fig. 6, orange). We found that the assimilation of organic N sources was either enhanced or inhibited by the absence of Ca^{2+} ions in the medium. For instance, NC64A had decreased assimilation of Pro in the absence of Ca^{2+} while Syngen 2-3 had decreased assimilation of Asn and Pro. In contrast, OK1-ZK, F36-ZK and SAG 3.83 grew better when Asn, Pro, and Ser were present in the absence of Ca^{2+} ions; these three strains showed the strongest response to the manipulation of Ca^{2+} in the medium. No appreciable differences were observed for urea, Gly, Arg or Gln for the five symbiotic algae. Thus, although Ca^{2+} regulation plays a role in the assimilation of the symbiont-specific organic N sources in symbiotic strains, the phenotypes were strain-specific (Fig. 6). Addition of a C source masked the effect of the absence/presence of Ca^{2+} in the medium (data not shown).

Intriguingly, Ca^{2+} ions in the medium also affected organic N assimilation in most free-living Chlorella. The absence of Ca^{2+} had no appreciable effects on Asn, Ser, and Pro (symbiont-specific organic N sources) in most free-living algae, but strain-specific significant differences were observed for those N sources for which symbiotic strains did not appear to be affected (e.g. urea, Arg, Gly, and Gln).

Together, Ca^{2+} ions interfere with the assimilation of Asn, Ser or Pro exclusively in symbiotic algae. In contrast, differential Arg, Gly and Gln assimilation was observed only in free-living algae species. Although these phenotypes are relatively strain-specific, the dichotomy in which N sources were affected between symbiotic and free-living strains differentiates both groups.

Overrepresented amino acid transporter protein domains in the NC64A genome are constitutively expressed in axenic NC64A cells

In 2010, we sequenced the NC64A genome to 9X coverage with 89% of the genome on 413 scaffolds (46 Mb) (Blanc et al. 2010). While the overall GC content is very high (67.2%), genomic islands with significantly lower GC content that have greater ETS coverage existed throughout the genome. Additionally, a significant (X^2 test, $\alpha = 0.05$ after Bonferroni correction) expansion of some protein families (PFAM) was identified in NC64A that could have participated in adaptation to symbiosis. A similar subset of PFAM domains was found overrepresented in organisms that have intracellular or symbiotic life styles. We

hypothesize that the corresponding proteins in NC64A play a role in the chlorella intracellular interaction with the protozoan *Paramecium bursaria.* These PFAM domains include proteins containing protein–protein interaction motifs (F-box and MYND), adhesion domains (fasciclin), Cys-rich GCC2_GCC3 signatures, trypsin-like protease domains, class 3 lipase motifs and amino acid (AA) transporters domains (Blanc et al. 2010).

In all domains of life, AA transporters act as extracellular or intracellular nutrient sensors and as carriers of cellular nutrient supplies. Amino acids do not readily diffuse across lipid membranes; rather, membrane-spanning transporter proteins are required to move AAs in and out of a cell and between intracellular compartments (Fischer et al. 1995). The significant increase in the number of AA transporters in NC64A (35 proteins, Suppl. Table 1) caught our attention since fifteen of them have ESTs, suggesting that they are constitutively expressed not only in axenically growing cells (Fig. 7) (Blanc et al. 2010) but also within *P. bursaria* cells (Fig. 8 and Suppl. Table 2) (Kodama et al. 2014). Therefore, *Chlorella* symbionts (including NC64A) might have an efficient system for importing AA from the *P. bursaria* host and could use some AA as a source of N instead of other inorganic N sources (Albers et al. 1982; Blanc et al. 2010; Karakashian 1975; Karakashian and Karakashian 1965; Kato et al. 2006). Taken together, our genomic and transcriptomic analysis suggests that NC64A is able to better assimilate unusual organic N and C sources likely by the overrepresentation and expression of some trypsin-like proteases (which

degrade peptides into AA) and AA transporter domains in its genome. This evidence suggests a strong relationship between lifestyle and genomic expansion of functional domains in the NC64A strain.

Bioinformatic and transcriptome analysis of amino acid transporter orthologs in Chlorella species

Physiological observations led us to hypothesize that in nature the protozoan host regulates the population of symbiotic *Chlorella* by restricting their N supply with AA. Keeping the N supply low and the chlorophyll content 5-10 fold higher is consistent with the symbiont functioning to provide excess photosynthate to the host as secreted maltose (Rees, 1991). To test this hypothesis we compared genes for AA transporter proteins between the symbiont *C. variabilis* NC64A and the free-living algae *C. sorokoniana* UTEX-1230 as well as analyzed gene expression of symbiont Chlorella growing within *P. bursaria* and NC64A growing in axenic culture.

Given the demonstrated variation in capacity for utilization of AA as sole N source between the free-living and symbiotic *Chlorella* strains, we sought to assess whether the presence or absence of AA transporter encoding genes could explain variation in this trait. Members of a collection of characterized AA transporters from *A. thaliana* (Table 1; reviewed in Tegeder 2012) were used to perform BLAST searches against both *C.* v*ariabilis* NC64A (Blanc et al. 2010) and *C. sorokiniana* UTEX-1230 (UNL algal consortium, *in preparation*) genomes,

using an expected value of 1×10^{-10} as a cutoff (Suppl. Fig. 12). The major predicted isoform for each identified locus was selected and used to perform a BLAST search against the *Arabidopsis thaliana* genome using an expected value of 1×10^{-10} as a cutoff. Each algal protein returned an *A. thaliana* AA transporter, and the gene designation and E-value is presented in Tables 1 and 2.

Results of the initial and reciprocal BLAST searches demonstrated that the UTEX 1230 genome encodes 25 putative orthologs to some *A. thaliana* AA transporter genes (Table 1), and NC64A encodes 16 putative AA transporter orthologs (Table 2). UTEX 1230 contains two GABA transporter-like proteins and two lysine-histidine-like transporters, whereas NC64A contained two of the former and one of the latter. Multiple isoforms of broad-specificity AA permeases and AA transporter genes were present in both genomes. Taken together, reciprocal BLAST using *A. thaliana* proteins identified more AA transporter orthologs in the UTEX 1230 genome than in the NC64A genome.

We compared AA transporter orthologs in NC64A and UTEX 1230 in more detail. Thus, we used the set of 35 proteins present in NC64A (Suppl. Fig. 1) to do reciprocal BLAST searches to the UTEX 1230 genome. We identified 27 AA transporter orthologs shared between both genomes and 8 genes that are only present in NC64A. From these, 3 are expressed in axenic cultures: EFN53131, EFN50622 and EFN54340 (Table 3). Comparative genomics established that the 27 AA transport orthologs in UTEX 1230 are generally longer proteins that their NC64A counterparts. They have protein identities that are lower that 65% (Table

3). Even in outputs with the highest e-value (e-value=0), the expected bit score is substantially higher that the calculated bit score between the orthologs proteins. These results suggest that NC64A and UTEX 1230 orthologs had substantially different sequence composition (Suppl. Table 4). We conclude that an increase in gene duplications and diversification of AA transport genes could have been a major enhancer for successful mutualistic in the NC64A strain.

In line with previous studies using genomic and transcriptomic approaches, we estimated the expression levels of the *C. variabilis* AA transporter in the endosymbiont state and found that 14 AA transporters orthologs (Fig. 8, Suppl. Table 2, Suppl. Fig. 14) were expressed at detectable levels in the *P. bursaria* assemblies. Of the expressed isoforms, several were differentially expressed between *C. variabilis* growing axenically and as endosymbiont in *P. bursaria*, but the sample size and differences in sequencing platform precluded a formal statistical analysis of the significance of the differential expression. Of the 14 genes, two (EFN51990 and EFN50622) accounted for the vast majority (>80%) of mapped reads in both conditions, indicating that these transporters likely provide the majority of AA uptake capabilities in these symbiotic organisms. Thus, mining genomes and transcriptomes we identified some AA transporter genes that might play a role in symbiotic metabolism, however future studies need to address the specific function(s) of the AA transporter proteins.

Analysis of clustering by inferred models of evolution (CLIME) indicates that amino acid transporters are evolutionarily absent in some protist species

In order to map any evolutionary and functional relationships between the overrepresentation of AA transporter genes in *C. variabilis*, we carried out a statistical comparison of genomic content across species. We hypothesized that the input set could contain modules with highly informative patterns of evolutionary gains and losses that could shed light on the mechanisms underlying the molecular evolution of genes and genomes.

Firstly, we applied CLIME using eight AA transporter gene homologues from *A. thaliana*. CLIME partitions the input set into evolutionarily conserved modules (ECMs) and an expansion set (ECM+) that includes other genes that likely have arisen under similar inferred models of evolution. Our results included seven singletons ECMs, indicating that most input genes do not share similar (common) history of gains and losses across eukaryotic evolution (unrelated evolutionary histories). The ECM with the highest strength ($\phi = 4.3$) contained 2 genes (AT1G80510 and AT3G30390) from the input set. AA transporters genes included in ECMs 1, 3, 4, 5 and 7 showed an ancestral node gene gain coming early from the last common ancestor (LCA).

Surprisingly the ECM gene queries were broadly conserved across eukaryotes but lost in most protist species (Suppl. Fig. 16). ECM 7, which contained the amino acid permease 2 gene (APP2), was absent in *Paramecium tetraurelia* in

particular. Importantly, seven out of 15 ESTs are AAP2-like AA transporters in *C. variabilis*. The analysis suggests a strong degree of gene coevolution that included a coordinated gene loss of the AA transporters in protists. Although, AA transporter gene losses were sporadic across other species, AAP2 (ECM 7) showed the most dynamic gene losses across lineages, where multiple events were observed not only in protists but also in plants, fungi and metazoan genomes. As a result, a lineage-specific gene family expansion of AA transporters by gene duplication occurred in symbiotic *Chlorella*, probably to complement the endosymbiotic interaction with their protist host(s).

Secondly, we sought to identify other cellular processes that might be under common selective constraints to the AA transporters. They could provide an evolutionary indication that the collective activity of the gene group might be relevant for their overall function. We analyzed all expansion ECMs+ that contained non-paralogs genes with significant log-likelihood ratio (LLR > 10). We found that most genes with common selective constraints were involved in carbon, nitrogen and nucleotide metabolism (46%), DNA and RNA processes (19%), cell cycle regulation (10%), defense response (8%), cell death (6%), lipid metabolism (6%) and ion transport (5%) (Suppl. Fig. 15, Suppl. Table 6). ECM7 showed the least amount of genes within the ECM+, with only 3 additional gene paralogs in *A. thaliana* (APP3, APP4 and lysine histidine transporter LHT7), suggesting a dynamic and unique evolutionary history across species.

Specific genes under common selective constraints to the AA trasnporters were involved in gluconeogenesis (succinate-semialdehyde dehydrogenase, 6-phosphofructo-2-kinase/fructose-2,6-biphosphate, uridine kinase-like 5), galactose metabolism (UDP-arabinose 4-epimerase, bifunctional UDP-glucoes 4-epimerase and UDP-xylose 4-epimerase 1), AA metabolism (serine racemase, pyridoxal phosphate-dependent transferase, pyridoxine/pyridoxamine 5'-phosphate oxidase 1), sucrose metabolism (UDP-glucose pyrophosphorylase), TOR signaling (Raptor2), calcium and potassium ion transport (calcineurin B-like protein, quiescin-sulfphydryl oxidase) and those involved in defense response to virus, bacteria and fungus.

Taken together, the observed differences in overall levels of gene gains and losses between protist lineages and their potential symbionts implicate specific gene inventory flux as an important symbiotic-associated process in nature. Additionally, metabolic gene fluxes might have implications in metabolic rewiring at the cellular level.

In silico network reconstruction analysis of Arabidopsis amino acid transporters suggest an overall cell metabolic reprogramming

To identify cellular interactions and functions affected by the constitutive expression of AA transporters, we analyzed the *A. thaliana* microarray database with the ATTED-II tool to search for genes that are co-expressed with the counterpart *A. thaliana* homologues. Using six *A. thaliana* AA transporter

genes we identified that they co-regulated important subnetworks related with central metabolism. Astoundingly, the gene network agreed with the functional relationships identified by CLIME, suggesting that some cellular functions might be influenced by common selective constraints and the collective activity of the gene group might be relevant for their function.

We grouped them in four main subnetworks which contained significantly enriched gene ontology terms for biological processes involved in nucleotide, N and C metabolism (56%), defense regulation (20%), DNA and RNA processes (15%) and cell death (9%) (Suppl. Fig. 13, Suppl. Table 5). Specifically evident genes were those related to glucose and galactose metabolism (reversibly-glycosylated protein 5, PMT5 polyol transporter, STP4 glucose transporter protein, UDP-galactose transporter), nitrate metabolism (glutamine synthetase, beta-fructofuranosidase, high-affinity nitrate transporter), autophagy and defense regulation (G18D autophagy-related protein, NiaP nicotinate transporter, major facilitator protein).

Together by reconstructing and analyzing a gene regulatory network of AA transporters in *A. thaliana*, we identified coexpressed gene pairs mainly involved in central metabolism and defense response that could potentially be drivers orchestrating unique symbiotic growth advantage. This genome-wide gene-expression data approach has potentially important implications for signatures for endosymbitoic life styles. Our analysis suggests that the overrepresentation and active gene expression of some AA transporters in *C. variabilis* could potentially

trigger ancient evolutionary genome plasticity and metabolic reprogramming at the cellular level where distinct metabolic networks might become increasingly interwoven and interdependent throughout evolution.

Discussion

Chlorellla symbionts are polyphyletic (i.e. they have arisen as protozoan symbionts from different origins) (Fujishima 2010). These strains have sustained long-standing associations with their protozoan hosts and have evolved unique interactions that are not random but, rather, the result of natural selection operating to ensure the survival of both the symbiont and the host. Frequently, these constraints include blocks, such as the inability to use NO_3, designed to force the two partners to stay together. Nutrient utilization must be highly regulated during the symbiotic growth phase in order for the algae to respond appropriately to available nutrient conditions and to limit nutrient acquisition from the host cell. For instance, studies on *Hydra viridissima* suggest that algae cells are in N-limiting conditions while growing in their host (Pardy, 1974, McAuley 1987b). The host can restrict the supply of N to the algae and control the symbiont proliferation by manipulating the N supply, either by regulating N uptake or by controlling the supply of AA derived from the host metabolism.

In this study, we conclude that symbiotic algae have physiological signatures that are conserved. Fig. 1 depicts the resulting heat map and relatedness trees. This analysis confirms the robust nutritional/metabolic differences between the

symbiotic and free-living *Chlorella* as well as the generalization that the free-living Chlorella are better adapted for inorganic N sources while the symbiotic Chlorella are adapted for organic N sources. The five symbiotic *Chlorella* are polyphyletic; namely, they derive from multiple independent symbiotic events. Yet, they are all similar in their inability to use NO_3 but they exhibit rapid uptake and utilization of certain AA as sole N sources. Additionally their growth rates were slower compared to their free-living counterparts and they all are susceptible to double stranded DNA virus infections (Suppl. Table. 7) (Van Etten and Dunigan 2012; Quispe et al. unpublished). Thus, our data suggest that these phenotypic differences reflect a major cellular and metabolic reprogramming at the structural and molecular level. These evolutionary changes could include previous observations related to cell wall structure, energy balance, cell cycle regulation, and decreased cellular defenses in NC64A. The results also provide possible connection between the endosymbiotic life style and virus susceptibility, illustrating the trade-offs endosymbiotic *Chlorella* must make in nature.

References

Albers D, Reisser W, Wiessner W. 1982. Studies on the nitrogen supply of endosymbiotic chlorellae in greem *Paramecium bursaria*. Plant Science Letters 25(1):85-90.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25(17):3389-402.

Bischoff, H. W. & Bold, H. C. 1963. Phycological studies IV. some soil algae from enchanted rock and related algal species. No. 6318 ed. Austin, Texas University of Texas Publication.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, et al. 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. The Plant Cell 22(9):2943-55.

Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy Team. 2010. Manipulation of FASTQ data with galaxy. Bioinformatics 26(14):1783-5.

Bubeck JA and Pfitzner AJ. 2005. Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. The Journal of General Virology 86(Pt 10):2871-7.

Cernichiari E, Muscatine L, Smith D. 1969. Maltose excretion by the symbiotic algae of *Hydra viridis*. Proceedings of the Royal Society of London.Series B.Biological Sciences 173(1033):557-76.

Fischer WN, Kwart M, Hummel S, Frommer WB. 1995. Substrate specificity and expression profile of amino acid transporters (AAPs) in arabidopsis. J Biol Chem 270(27):16315-20.

Fujishima M. 2010. Endosymbionts in paramecium. .

Gancedo JM. 1998. Yeast carbon catabolite repression. Microbiology and Molecular Biology Reviews 62(2):334-61.

Hom EFY and Murray AW. 2014. Niche engineering demonstrates a latent capacity for fungal-algal mutualism. Science 345(6192):94-8.

Johnson MD. 2011. The acquisition of phototrophy: Adaptive strategies of hosting endosymbionts and organelles. Photosynth Res 107(1):117-32.

Jolley E and Smith DC. 1978. The green hydra symbiosis. I. isolation, culture and characteristics of the chlorella symbiont of 'european' hydra viridis. New Phytol 81(3):637-45.

Kvitko BVG. 1984. New finding of a titratable infectious zoochlorella virus. Dokl. Akad. Nauk. SSSR (279):998-9.

Kamako S, Hoshina R, Ueno S, Imamura N. 2005. Establishment of axenic endosymbiotic strains of japanese *Paramecium bursaria* and the utilization of carbohydrate and nitrogen compounds by the isolated algae. Eur J Protistol 41(3):193-202.

Kanehisa M and Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28(1):27-30.

Karakashian MW. 1975. Symbiosis in *Paramecium bursaria*. Symp Soc Exp Biol (29)(29):145-73.

Karakashian SJ and Karakashian MW. 1965. Evolution and symbiosis in the genus chlorella and related algae. Evolution 19(3):368-77.

Kato Y, Ueno S, Imamura N. 2006. Studies on the nitrogen utilization of endosymbiotic algae isolated from japanese *Paramecium bursaria*. Plant Science 170(3):481-6.

Kawakami H and Kawakami N. 1978. Behavior of a virus in a symbiotic system, *Paramecium bursaria zoochlorella*. J Protozool 25(2):217-25.

Kessler E and Huss V. 1990. Biochemical taxonomy of symbiotic chlorella strains from paramecium and acanthocystis*. Botanica Acta 103(2):140-2.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4):R36,2013-14-4-r36.

Kodama Y, Suzuki H, Dohra H, Sugii M, Kitazume T, Yamaguchi K, Shigenobu S, Fujishima M. 2014. Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts. BMC Genomics 15:183,2164-15-183.

Kovacevic G. 2012. Value of the hydra model system for studying symbiosis. Int J Dev Biol 56(6-8):627-35.

Kovacevic G, Kalafatic M, Ljubesic N. 2007. New observations on green hydra symbiosis. Folia Biol (Krakow) 55(1-2):77-9.

Lopez-Garcia P and Moreira D. 2015. Open questions on the origin of eukaryotes. Trends Ecol Evol 30(11):697-708.

Matzke B, Schwarzmeier E, Loos E. 1990. Erratum : Maltose excretion by the symbiotic chlorella of the heliozoan *Acanthocystis turfacea.* Planta 182(2):312.

McAuley P. 1987. Nitrogen limitation and amino-acid metabolism of chlorella symbiotic with green hydra. Planta 171(4):532-8.

Meints RH, Van Etten JL, Kuczmarski D, Lee K, Ang B. 1981. Viral infection of the symbiotic chlorella-like alga present in *Hydra viridis*. Virology 113(2):698-703.

Nicol JW, Helt GA, Blanchard SG,Jr, Raja A, Loraine AE. 2009. The integrated genome browser: Free software for distribution and exploration of genome-scale datasets. Bioinformatics 25(20):2730-1.

Pardy RL and Muscatine L. 1973. Recognition of symbiotic algae by *Hydra viridis*. A quantitative study of the uptake of living algae by aposymbiotic *H. viridis*. Biol Bull 145(3):565-79.

Park HD, Greenblatt CL, Mattern CFT, Merril CR. 1967. Some relationships between chlorohydra, its symbionts and some other chlorophyllous forms. J Exp Zool 164(2):141-61.

Proschold T, Darienko T, Silva PC, Reisser W, Krienitz L. 2011. The systematics of zoochlorella revisited employing an integrative approach. Environ Microbiol 13(2):350-64.

Rees 1991. Are symbiotic algae nutrient deficient? Proceedings of the Royal Society B Biological Sciences 243(1308):227.

Rowe JM, Dunigan DD, Blanc G, Gurnon JR, Xia Y, Van Etten JL. 2013. Evaluation of higher plant virus resistance genes in the green alga, *Chlorella variabilis* NC64A, during the early phase of infection with *Paramecium bursaria* chlorella virus-1. Virology 442(2):101-13.

Siegel RW. 1960. Hereditary endosymbiosis in *Paramecium bursaria*. Exp Cell Res 19(2):239-52.

Tegeder M. 2012. Transporters for amino acids in plant cells: Some functions and many unknowns. Curr Opin Plant Biol 15(3):315-21.

Van Etten JL, Burbank DE, Kuczmarski D, Meints RH. 1983a. Virus infection of culturable chlorella-like algae and dlevelopment of a plaque assay. Science 219(4587):994-6.

Van Etten JL and Dunigan DD. 2012. Chloroviruses: Not your everyday plant virus. Trends in Plant Science 17(1):1-8.

Van Etten JL, Burbank DE, Xia Y, Meints RH. 1983b. Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. Virology 126(1):117-25.

Van Etten JL, Meints RH, Burbank DE, Kuczmarski D, Cuppels DA, Lane LC. 1981. Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis.* Virology 113(2):704-11.

Vidal EA, Alvarez JM, Moyano TC, Gutierrez RA. 2015. Transcriptional networks in the nitrate response of arabidopsis thaliana. Curr Opin Plant Biol 27:125-32.

Yellowlees. 2008. Metabolic interactions between algal symbionts and invertebrate hosts. Plant Cell and Environment 31(5):679.

Figure Legends

Figure 1. Hierarchical heat map (average-linkage) clusters symbiotic and free-living strains based on their metabolic capabilites. Columns represent combinations of organic and inorganic nitrogen (N) sources with or without the addition of a carbon (C) source. All were added at 10 mM concentrations. MBBM is the control. Purple labels identify inorganic N sources at 1 mM concentration, and orange labels represent media without Ca^{2+}. Rows represent the five symbiotic strains (green) and four free-living strains (blue). Tree diagrams indicate the nature of the computed relationship among growth conditions and among *Chlorella* species. For each growth medium, triplicates were performed for the symbiotic *Chlorella* strains and duplicates for the free-living *Chlorella* species. A color scale indicates relative growth: red represents robust growth and black represents absence of growth. Flask tests were performed for 9-12 days for the symbiotic and 7-9 days for the free-living strains. Subsequent heat map figures follow a similar layout.

Figure 2. Heat map subgroup from Fig. 1 displays variations of MBBM (sucrose + peptone). Peptone is replaced by 1% casamino acids. A color scale indicates relative growth. Flask tests were performed for 9 days for the symbiotic and 7 days for the free-living strains.

Figure 3. Heat map subgroup from Fig. 1 compares inorganic and organic N sources at 10 mM concentrations as the sole N source. Inorganic sources include NH_4 tartrate, sodium NO_3, and NH_4 acetate. Organic sources include arginine (Arg), urea, glutamine (Gln), glycine (Gly), asparagine (Asn), proline (Pro), and serine (Ser). A color scale indicates relative growth. Flask tests were performed for 12 days for the symbiotic and 9 days for the free-living strains.

Figure 4. Heat map subgroup from Fig. 1 displays growth on NO_3 at 1 mM (purple) and 10 mM concentrations. Inorganic N sources are listed in the columns. A color scale indicates relative growth. Flask tests were performed for 12 days for the symbiotic and 9 days for the free-living strains.

Figure 5. Heat map subgroups from Fig. 1. Inorganic (A) and organic (B) N sources supplemented with glucose, sucrose, and galactose. Purple labels identify N sources at 1 mM concentration. A color scale indicates relative growth. Flask tests were performed for 9 days for the symbiotic and 7 days for the free-living strains.

Figure 6. Heat map subgroup from Fig. 1 displays removal of Ca^{2+} (orange) from media with organic N sources. Flask tests were performed for 12 days for the symbiotic and 9 days for the free-living strains.

Figure 7. *C. variabilis* NC64A mRNA of AA transporter genes during axenic growth**.** Normalized mRNA abundance of 15 AA transporter genes.

Figure 8. Comparison of relative expression of AA transporter genes as Log_2 fold changes between axenic *C. variabilis* NC64A and *P. bursaria* harboring symbiotic *C. variabilis*. Expression data for each gene represents manual counts of reads aligning to the corresponding genomic interval normalized per million mapped reads. The relative expression fold change for genes identified as being differentially expressed are represented, so genes that are upregulated in the *P. bursaria* harboring symbiotic alga have a negative Log_2 fold change value (white bars).

Figure. 9. Maximum-likelihood phylogenetic tree of expressed AA transporters in *C. variabilis* NC64A (blue circles) extracted from transcriptomic analysis from axenic cultures. Green circles indicate *A. thaliana* orthologs.

Table Legends

Table 1. Accession numbers of putative *C. variabilis* NC64A orthologs to *A. thaliana* proteins involved in AA transport. AAP=amino acid permeases, AAT= amino acid transporter, LHT= lysine histidine transporter.

Table 2. Scaffold numbers of putative *C. sorokiniana* UTEX-1230 orthologs to *A. thaliana* proteins involved in AA transport. AAP=amino acid permeases, AAT= amino acid transporter, LHT= lysine histidine transporter.

Table 3. Accession and scaffold numbers of putative *C. variabilis* NC64A orthologs to *C. sorokiniana* proteins involved in AA transport.

Supplementary Figure Legends

Supplementary Figure 1. In-vitro flask test identifies metabolic differences between symbiotic and free-living *Chlorella* strains grown on variations of MBBM (sucrose + peptone). Columns represent combinations of complex N (1% peptone or 1% casamino acids) with or without the addition of sucrose (10 mM). MBBM is the control. Rows represent the nine strains. The five symbiotic strains include *Chlorella variabilis* NC64A, *Chlorella heliozoae* SAG 3.83, *Chlorella variabilis* Syngen 2-3, *Chlorella variabilis* F36-ZK, and *Chlorella variabilis* OK1-ZK. The four free-living strains are *Chlorella sorokiniana* UTEX 1230, *Chlorella sorokiniana* CS-01, *Chlorella kessleri* B228, and *Chlorella protothecoides* 29. Flasks were shaken at 200 rpm and 26^{o}C in constant light. Symbiotic strains were incubated for 9 days and free-living strains were incubated for 7 days. For each growth medium, triplicates were performed for the symbiotic *Chlorella* strains and duplicates for the free-living *Chlorella* species. Flasks were evaluated based on the color scale included. Subsequent supplementary figures have similar layouts.

Supplementary Figure 2. In-vitro flask test of organic N (10 mM). Nitrogen sources include arginine (Arg), urea, glutamine (Gln), glycine (Gly), asparagine (Asn), proline (Pro), and serine (Ser). Photos were taken after 12 days for the symbiotic and after 9 days for the free-living strains.

Supplementary Figure 3. In-vitro flask test of inorganic N (1 or 10 mM). Nitrogen sources include NH_4 tartrate, sodium NO_3, and NH_4 acetate. Photos were taken after 12 days for the symbiotic and after 9 days for the free-living strains.

Supplementary Figure 4. In-vitro flask test of inorganic N (1 or 10 mM) supplemented with C source (10 mM). Carbon sources include (A) glucose, (B) sucrose, and (C) galactose. Nitrogen sources include NH_4 tartrate, sodium NO_3, and NH_4 acetate. Photos were taken after 9 days for the symbiotic and after 7 days for the free-living strains.

Supplementary Figure 5. In-vitro flask test of organic N (1 or 10 mM) supplemented with C source (10 mM). Carbon sources include (A) glucose, (B) sucrose, and (C)

galactose. Nitrogen sources include Arg, Urea, Gln, Gly, Asn, Pro, and Ser. Photos were taken after 9 days for the symbiotic and after 7 days for the free-living strains.

Supplementary Figure 6. In-vitro flask test displays removal of Ca^{2+} from media with organic N sources (10 mM). Nitrogen sources include Arg, Urea, Gln, Gly, Asn, Pro, and Ser. Photos were taken after 12 days for the symbiotic and after 9 days for the free-living strains.

Supplementary Figure 7. Phylogenetic tree displays relationship between *Chlorella* strains.

Supplementary Figure 8. MBBM and FES media components.

Supplementary Figure 9. Experimental flow chart of liquid nutritional test.

Supplementary Figure 10. Growth of symbiotic NC64A, SAG 3.83, and Syngen 2-3 cells on minimal defined media. MBBM is the control.

Supplementary Figure 11. Sequence comparative analysis of nitrate metabolic genes in green algae.

Supplementary Figure 12. Genetic regulatory network reconstruction using AA transporters genes from *Arabidopsis thaliana*.

Supplementary Figure 13. Genetic regulatory network reconstruction chart of characterized genes from *Arabidopsis thaliana*.

Supplementary Figure 14. *C. variabilis* NC64A mRNA levels of 15 AA transporter genes after PBCV-1 infection**.** Raw read counts shown in black, and DESeq normalized read counts shown in green.

Supplementary Figure 15. Functional distribution of expansion ECMs+ that contained non-paralogs genes with significant log-likelihood ratio (LLR > 10).

Supplementary Figure 16. Clustering by inferred models of evolution (CLIME) indicates that AA transporters are evolutionarily absent in some protist species.

Supplementary Table Legends

Supplementary Table 1. Thirty-five putative AA transporter proteins (Pfam PF01490.9) in *C. variabilis* NC64A.

Supplementary Table 2. Expression summary of predicted AA transporter genes in *C. variabilis* NC64A. Expression data for each gene represents manual counts of reads aligning to the corresponding genomic interval normalized per million mapped reads.

Supplementary Table 3. Carbohydrate-active (CAZy) enzymes with galactose activity that are overrepresented in the *C. variabilis* NC64A genome when compared to other chlorophyte green algae.

Supplementary Table 4. Putative *C. variabilis* (cvr) NC64A orthologs to *C. sorokiniana* (UTEX 1230) proteins involved in AA transport. Table includes the number of KEGG orthologs present in other organisms: *A. thaliana* (ath), *S. cerevisiae* (cse), *Chlamydomonas reinhardtii* (cre) *Ostreococcus lucimarinus* (olu), *Oryzias latipes* (ola), *Ostreococcus tauri* (ota), *Micromonas* sp. RCC299 (mis), *Coccomyxa subellipsoidea* (csl), *Micromonas pusilla* (mpp), and *Volvox carteri* (vcn).

Supplementary Table 5. Genetic regulatory network reconstruction gene list after using 6 AA transporter genes from *Arabidopsis thaliana* to predict the output.

Supplementary Table 6. Clustering by inferred models of evolution (CLIME) analyzes using eight AA transporter genes from *A. thaliana*. CLIME partitions the input set into evolutionarily conserved modules (ECMs) and an expansion set (ECM+) that includes other genes that likely have arisen under similar inferred models of evolution.

Supplementary Table 7. Growth rates of symbiotic NC64A, SAG 3.83, and Syngen 2-3 cells on various nitrogen and carbon sources with MBBM as control.

Figure 1

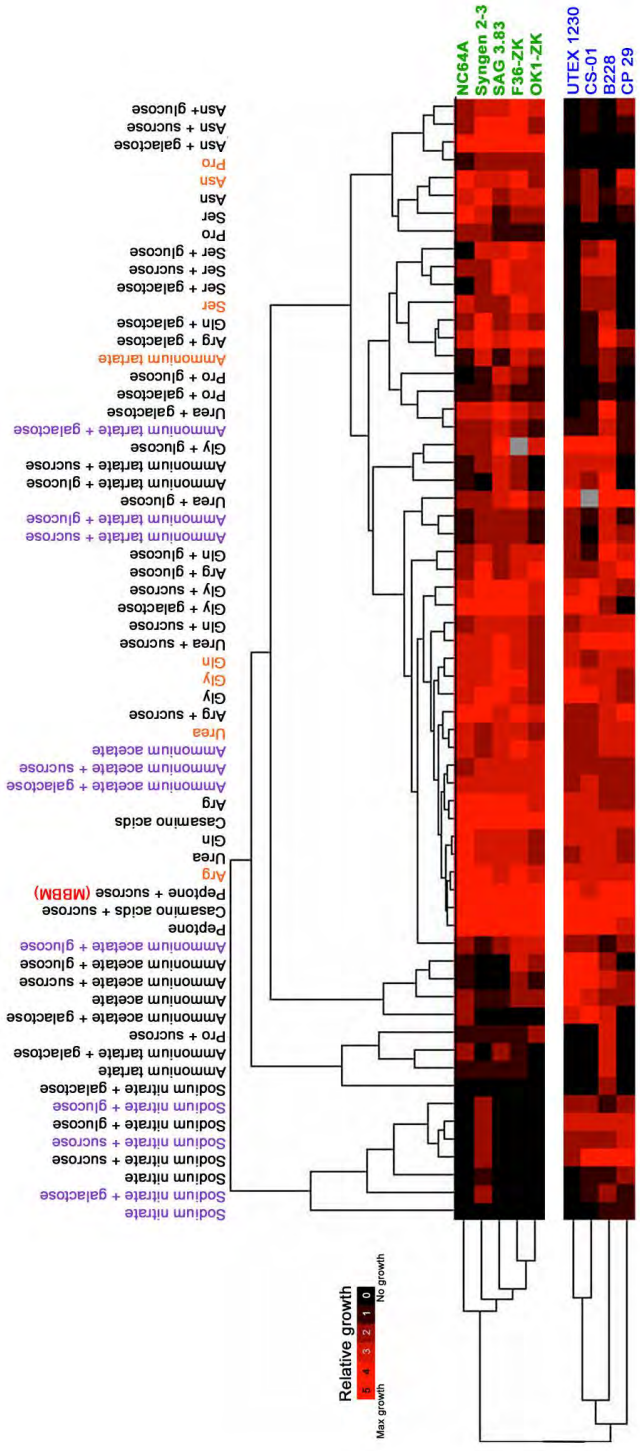

NC64A
Syngen 2-3
SAG 3.83
F36-ZK
OK1-ZK
UTEX 1230
CS-01
B228
CP 29
Asn+ glucose
Asn + sucrose
Asn + galactose
Pro
Asn
Asn
Ser
Pro
Ser + glucose
Ser + sucrose
Ser + galactose
Ser
Gln + galactose
Arg + galactose
Ammonium tartate
Pro + glucose
Pro + galactose
Urea + galactose
Ammonium tartate + galactose
Gly + glucose
Ammonium tartate + sucrose
Ammonium tartate + glucose
Urea + glucose
Ammonium tartate + glucose
Ammonium tartate + sucrose
Gln + glucose
Arg + glucose
Gly + sucrose
Gly + galactose
Gln + sucrose
Urea + sucrose
Gln
Gly
Gly
Arg + sucrose
Urea
Ammonium acetate
Ammonium acetate + sucrose
Ammonium acetate + galactose
Arg
Casamino acids
Gln
Urea
Arg
Peptone + sucrose (MBBM)
Casamino acids + sucrose
Peptone
Ammonium acetate + glucose
Ammonium acetate + glucose
Ammonium acetate + sucrose
Ammonium acetate
Ammonium acetate + galactose
Pro + sucrose
Ammonium tartate + galactose
Ammonium tartate
Sodium nitrate + galactose
Sodium nitrate + glucose
Sodium nitrate + glucose
Sodium nitrate + sucrose
Sodium nitrate
Sodium nitrate + sucrose
Sodium nitrate
Relative growth
5
4
3
2
1
0
Max growth
No growth

Figure 2

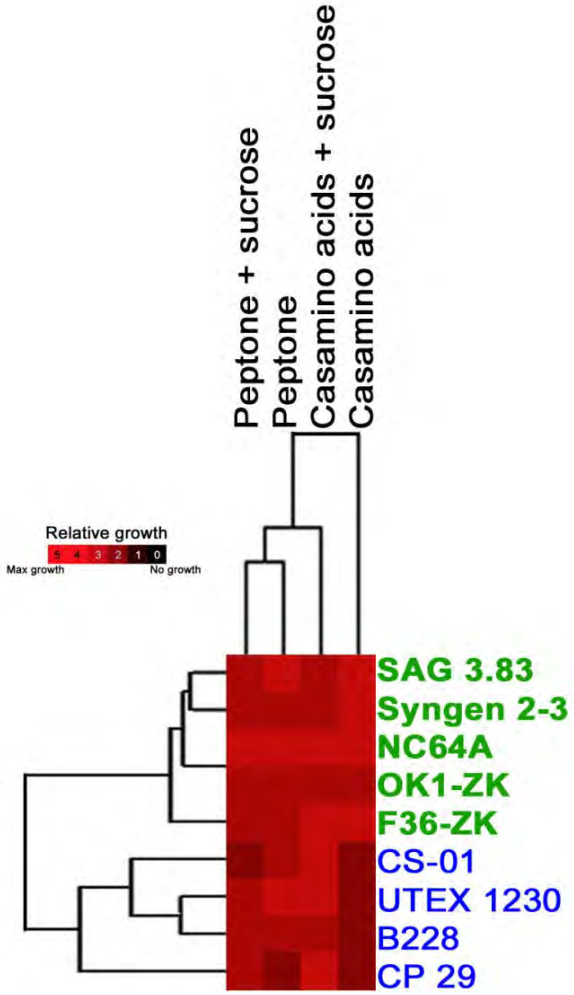

Peptone + sucrose
Peptone
Casamino acids + sucrose
Casamino acids
Relative growth
5 4 3 2 1 0
Max growth
No growth
SAG 3.83
Syngen 2-3
NC64A
OK1-ZK
F36-ZK
CS-01
UTEX 1230
B228
CP 29

Figure 3

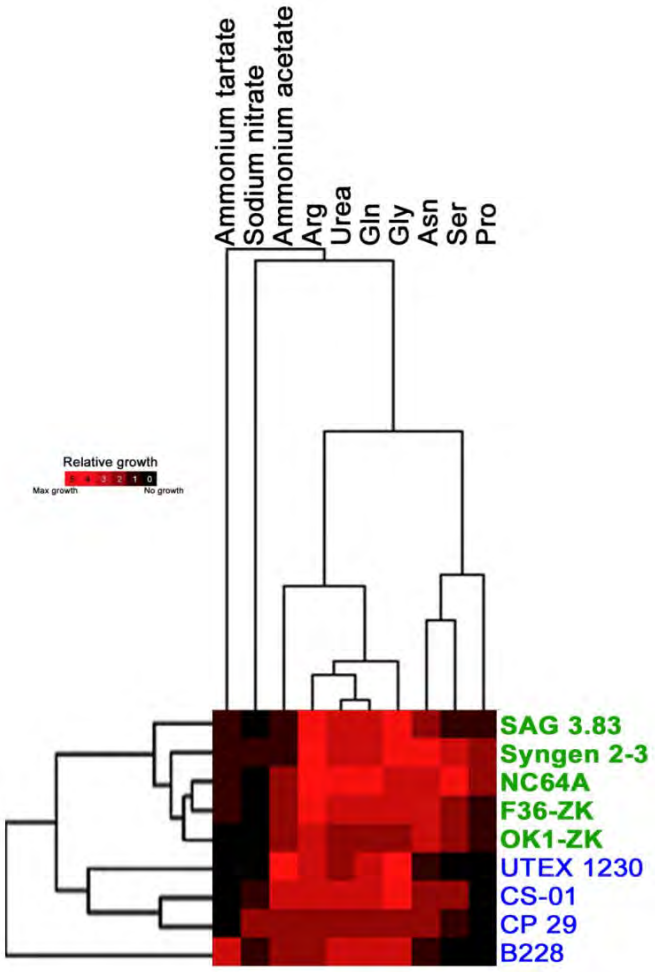

Ammonium tartate
Sodium nitrate
Ammonium acetate
Arg
Urea
Gln
Gly
Asn
Ser
Pro
Relative growth
4
3
2
1
0
Max growth
No growth
SAG 3.83
Syngen 2-3
NC64A
F36-ZK
OK1-ZK
UTEX 1230
CS-01
CP 29
B228

Figure 4

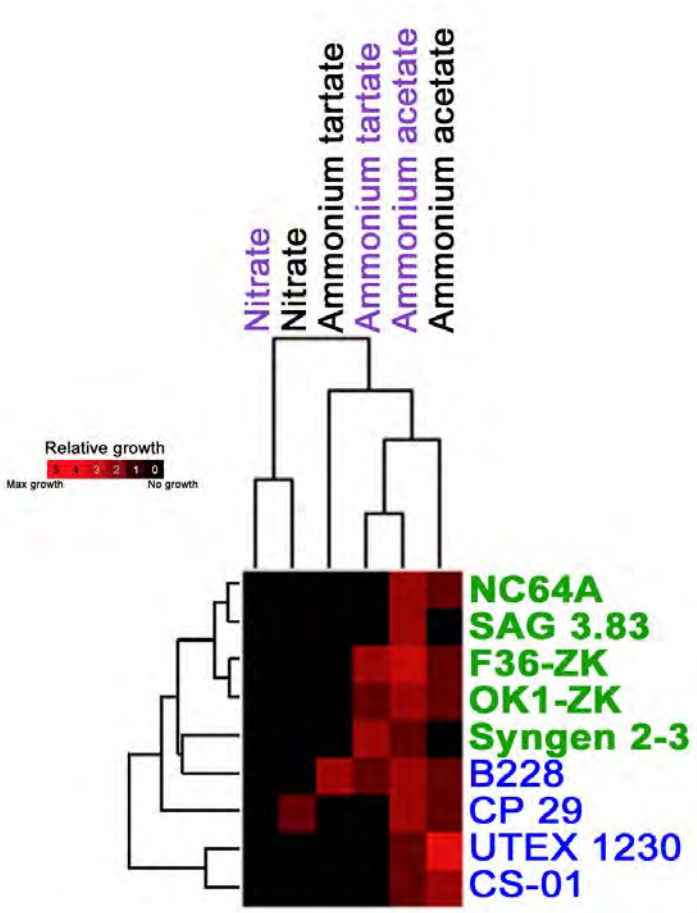

Nitrate
Nitrate
Ammonium tartate
Ammonium tartate
Ammonium acetate
Ammonium acetate
Relative growth
5 4 3 2 1 0
Max growth
No growth
NC64A
SAG 3.83
F36-ZK
OK1-ZK
Syngen 2-3
B228
CP 29
UTEX 1230
CS-01

Figure 5A

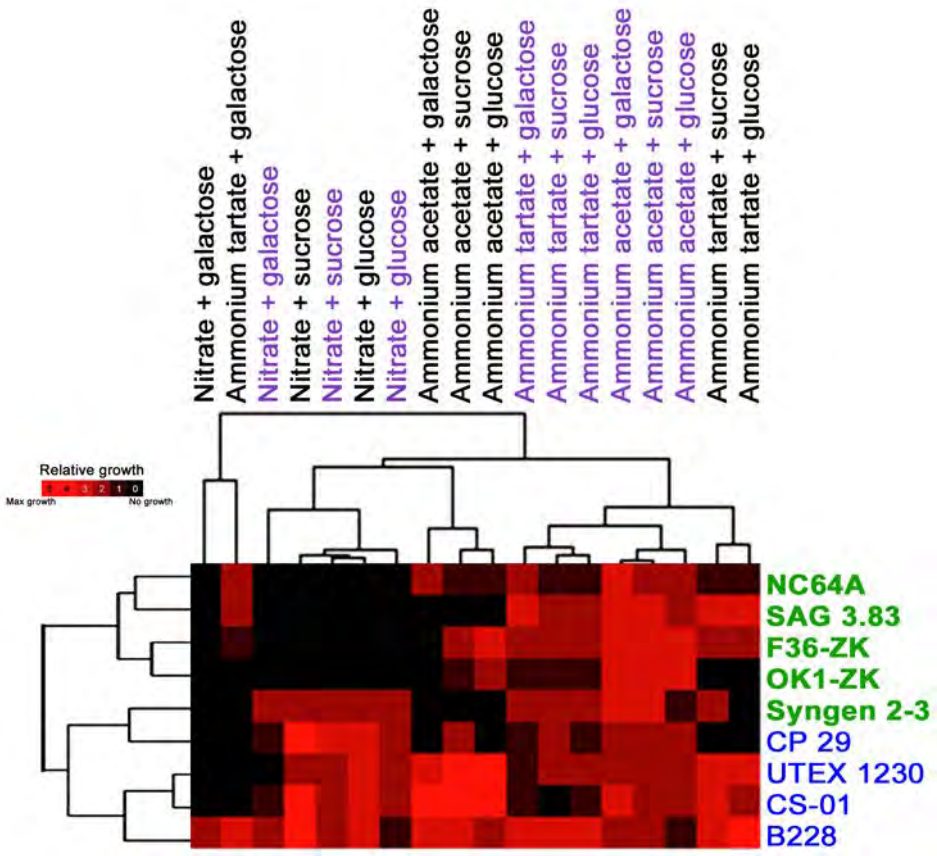

Nitrate + galactose
Ammonium tartate + galactose
Nitrate + galactose
Nitrate + sucrose
Nitrate + sucrose
Nitrate + glucose
Nitrate + glucose
Ammonium acetate + galactose
Ammonium acetate + sucrose
Ammonium acetate + glucose
Ammonium tartate + galactose
Ammonium tartate + sucrose
Ammonium tartate + glucose
Ammonium acetate + galactose
Ammonium acetate + sucrose
Ammonium acetate + glucose
Ammonium tartate + sucrose
Ammonium tartate + glucose
Relative growth
Max growth
No growth
NC64A
SAG 3.83
F36-ZK
OK1-ZK
Syngen 2-3
CP 29
UTEX 1230
CS-01
B228

Figure 5B

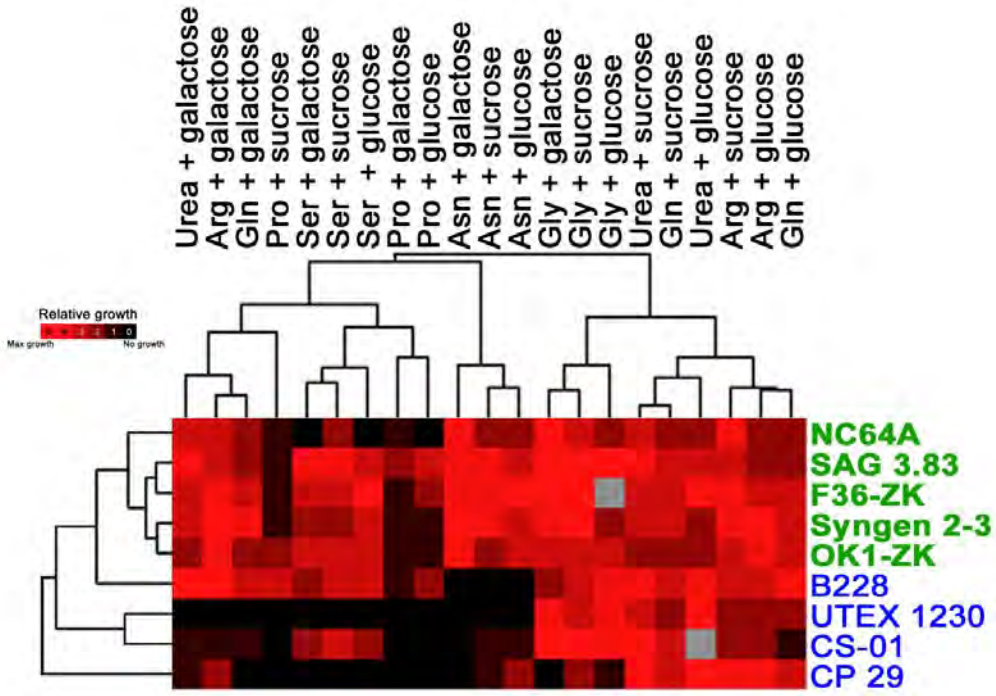

Urea + galactose
Arg + galactose
Gln + galactose
Pro + sucrose
Ser + galactose
Ser + sucrose
Ser + glucose
Pro + galactose
Pro + glucose
Asn + galactose
Asn + sucrose
Asn + glucose
Gly + galactose
Gly + sucrose
Gly + glucose
Urea + sucrose
Gln + sucrose
Urea + glucose
Arg + sucrose
Arg + glucose
Gln + glucose
Relative growth
Max growth
No growth
NC64A
SAG 3.83
F36-ZK
Syngen 2-3
OK1-ZK
B228
UTEX 1230
CS-01
CP 29

Figure 6

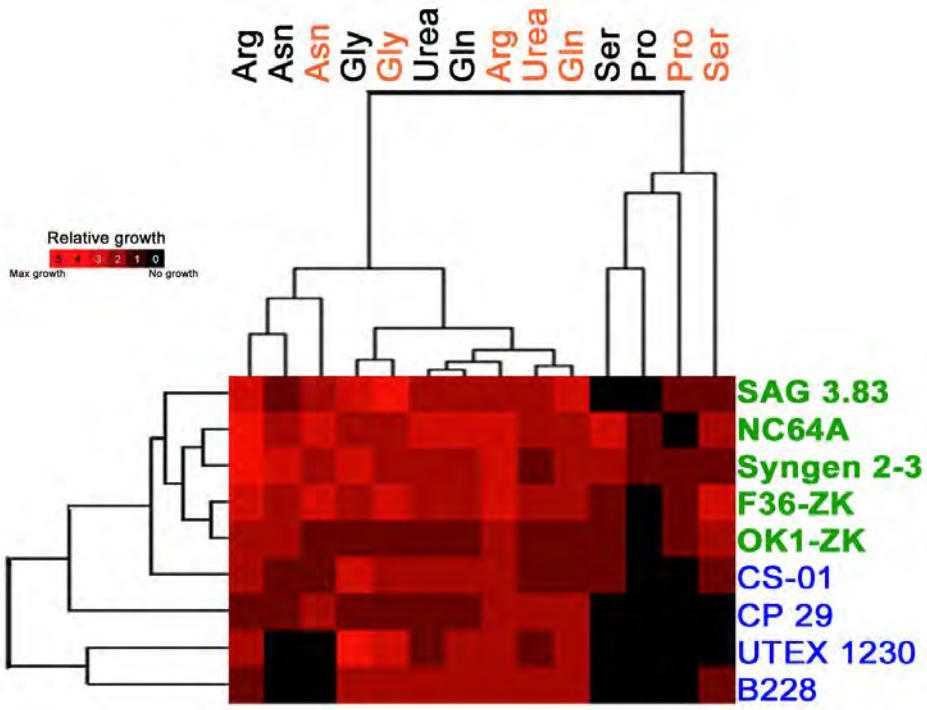

Arg
Asn
Asn
Gly
Gly
Urea
Gln
Arg
Urea
Gln
Ser
Pro
Pro
Ser
Relative growth
5 4 3 2 1 0
Max growth
No growth
SAG 3.83
NC64A
Syngen 2-3
F36-ZK
OK1-ZK
CS-01
CP 29
UTEX 1230
B228

Figure 7

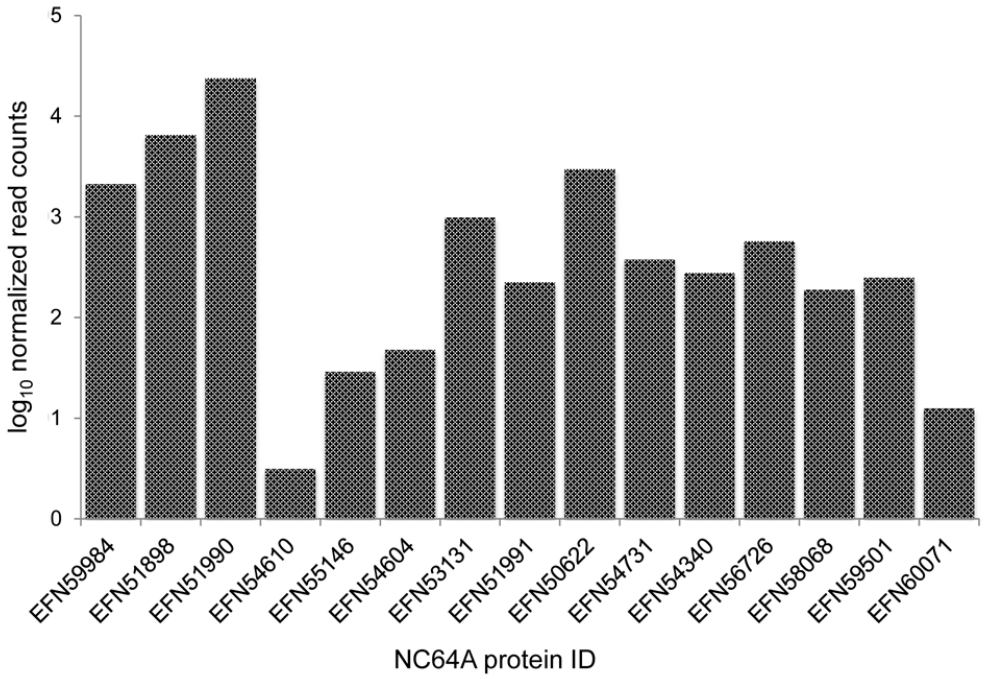

log10 normalized read counts
0
1
2
3
4
5
EFN59984
EFN51898
EFN51990
EFN54610
EFN55146
EFN54604
EFN53131
EFN51991
EFN50622
EFN54731
EFN54340
EFN56726
EFN58068
EFN59501
EFN60071
NC64A protein ID

Figure 8

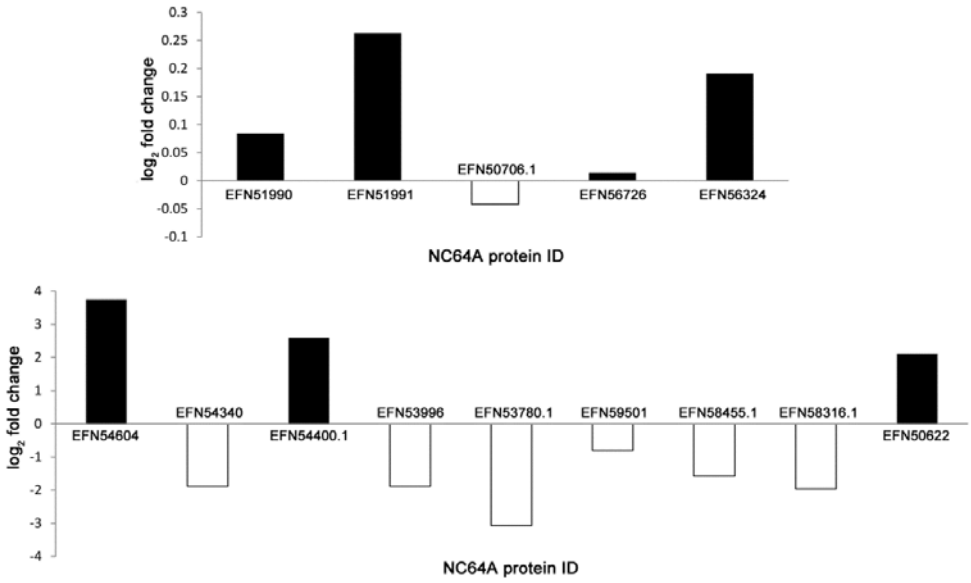

0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
log2 fold change
EFN51990
EFN51991
EFN50706.1
EFN56726
EFN56324
NC64A protein ID
4
3
2
1
0
-1
-2
-3
-4
log2 fold change
EFN54604
EFN54340
EFN54400.1
EFN53996
EFN53780.1
EFN59501
EFN58455.1
EFN58316.1
EFN50622
NC64A protein ID

Figure 9

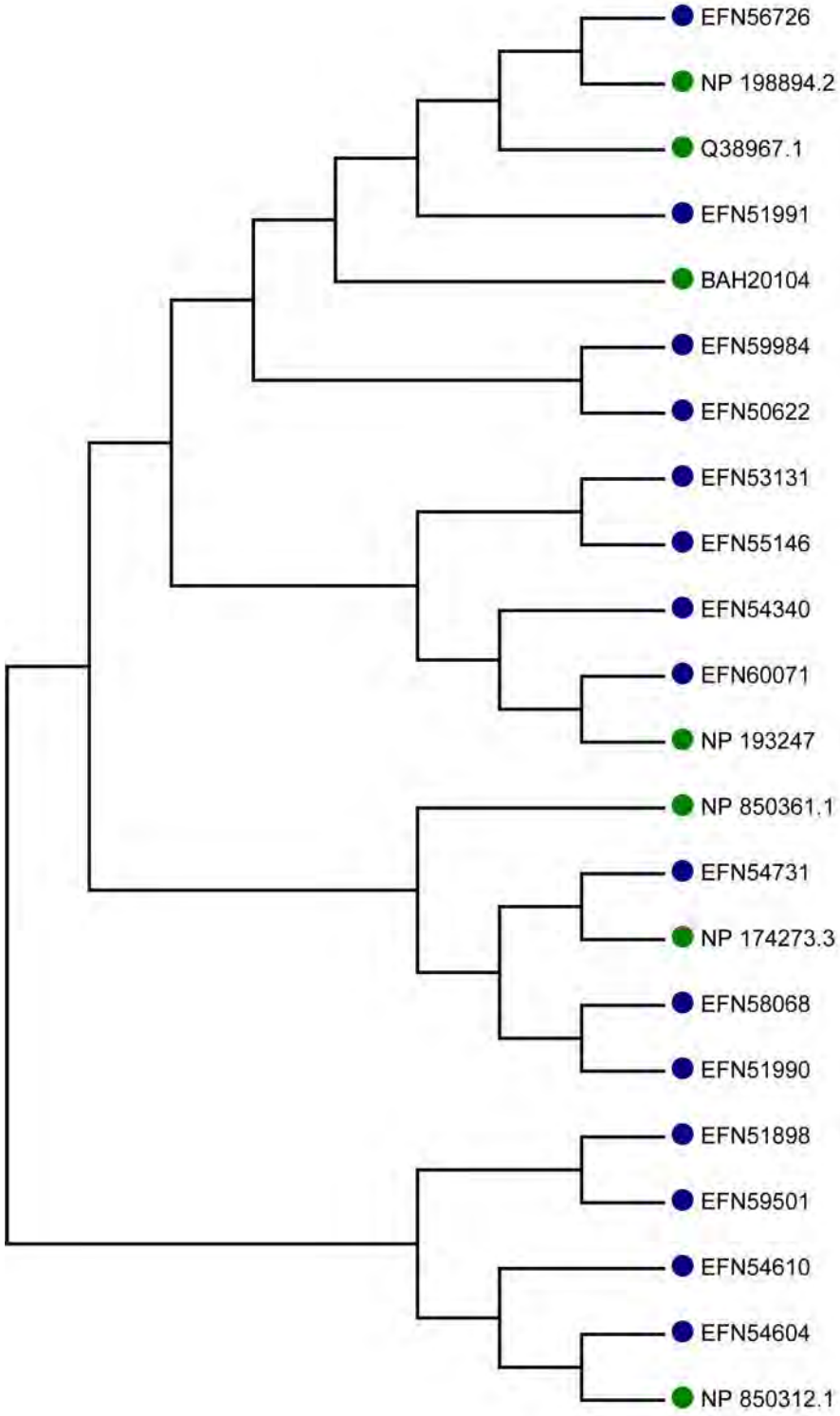

EFN56726
NP 198894.2
Q38967.1
EFN51991
BAH20104
EFN59984
EFN50622
EFN53131
EFN55146
EFN54340
EFN60071
NP 193247
NP 850361.1
EFN54731
NP 174273.3
EFN58068
EFN51990
EFN51898
EFN59501
EFN54610
EFN54604
NP 850312.1

Table 1

AA transporter ortholog in NC64A	*A. thaliana* best hit	e-value
37093	AAP2	2E-76
58128	AAP2	9E-72
32765	AAP2	1E-52
57473	LHT 1	2E-83
138133	AAP2	2E-49
59057	AAP2	2E-45
53357	AAP	3E-45
59479	AAP2	3E-29
24724	AAP2	1E-21
144770	GABA transporter 1	7E-31
17797	AAP8	1E-19
142334	AAP or GABA permease	4E-158
140447	AAT1	1E-34
7483	AAT1	5E-112
58448	AAT1	2E-61
36103	AAT1	2E-54

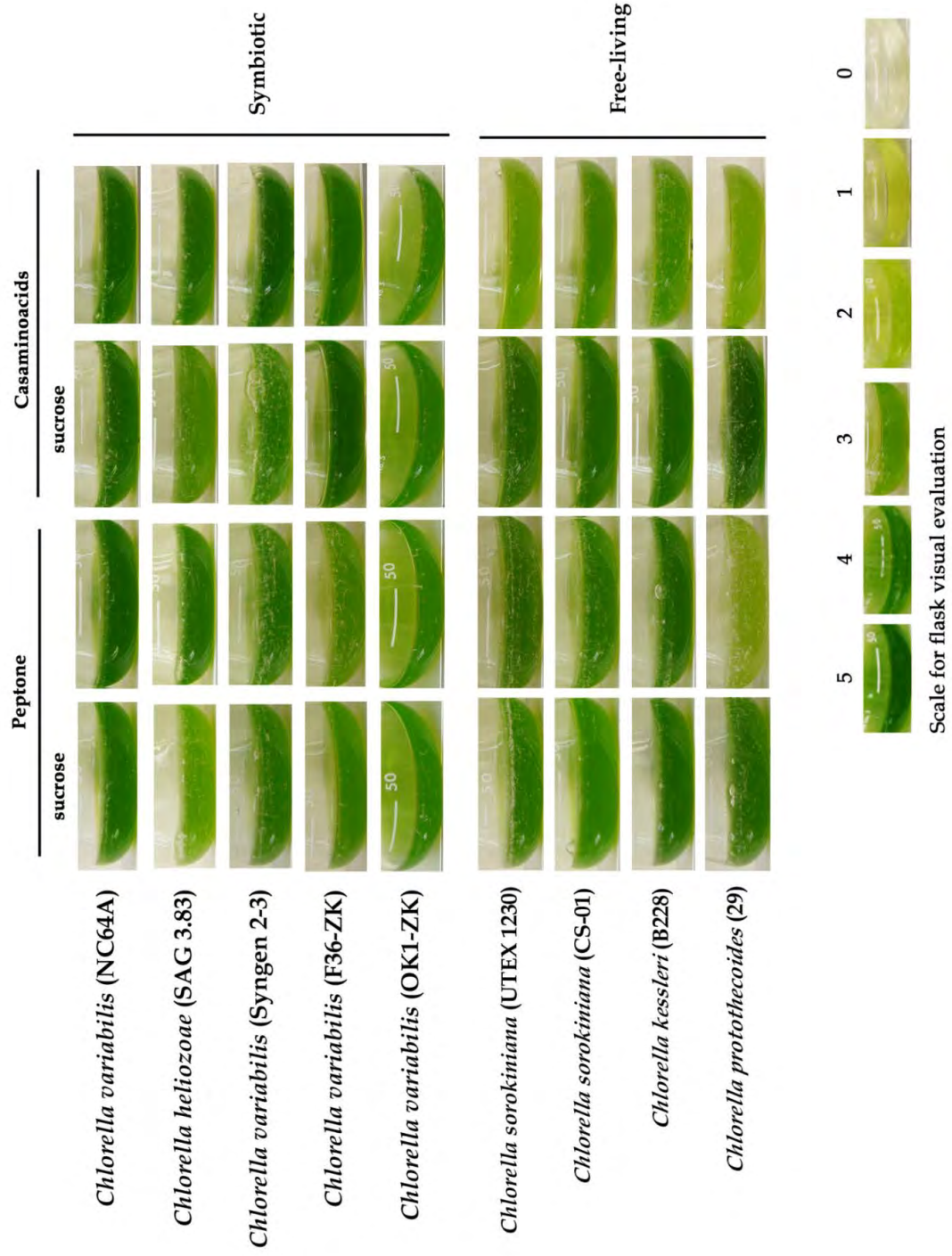
Table 2

AA transporter ortholog in UTEX-1230	*A. thaliana* best hit	e-value
scaffold82.g49.iso1	AAP2	4E-81
scaffold181.g27.iso1	AAP5	1E-73
scaffold172.g106.iso1	AAP3	1E-76
scaffold15.g150.iso1	AAP2	4E-69
scaffold99.g53.iso4	AAP2	1E-67
scaffold106.g243.iso1	AAP2	4E-63
scaffold34.g191.iso1	AAP CAA54632.1	1E-63
scaffold91.g67.iso3	LHT1	6E-83
scaffold35.g114.iso1	AAP2	2E-51
scaffold13.g237.iso1	AAP2	2E-58
scaffold35.g117.iso2	AAP2	2E-43
scaffold6.g13.iso1	AAP2	5E-54
scaffold270.g17.iso1	LHT1	6.5E-60
scaffold56.g4.iso2	GABA transporter 1	3E-34
scaffold124.g21.iso1	AAP8	1E-34
scaffold57.g99.iso1	AAT	9E-64
scaffold76.g4.iso1	AAP2	2E-45
scaffold132.g58.iso1	AAP AAB71468.1	9E-29
scaffold1.g418.iso1	GABA transporter 1	1E-87
scaffold110.g43.iso1	AAP2	4E-36
scaffold6.g14.iso2	AAP3	2E-11
scaffold13.g234.iso1	AAP4	9E-13
scaffold92.g116.iso1	AAP1	3E-117
scaffold98.g9.iso1	AAP1	2E-60
scaffold13.g249.iso1	AAT1	8E-125

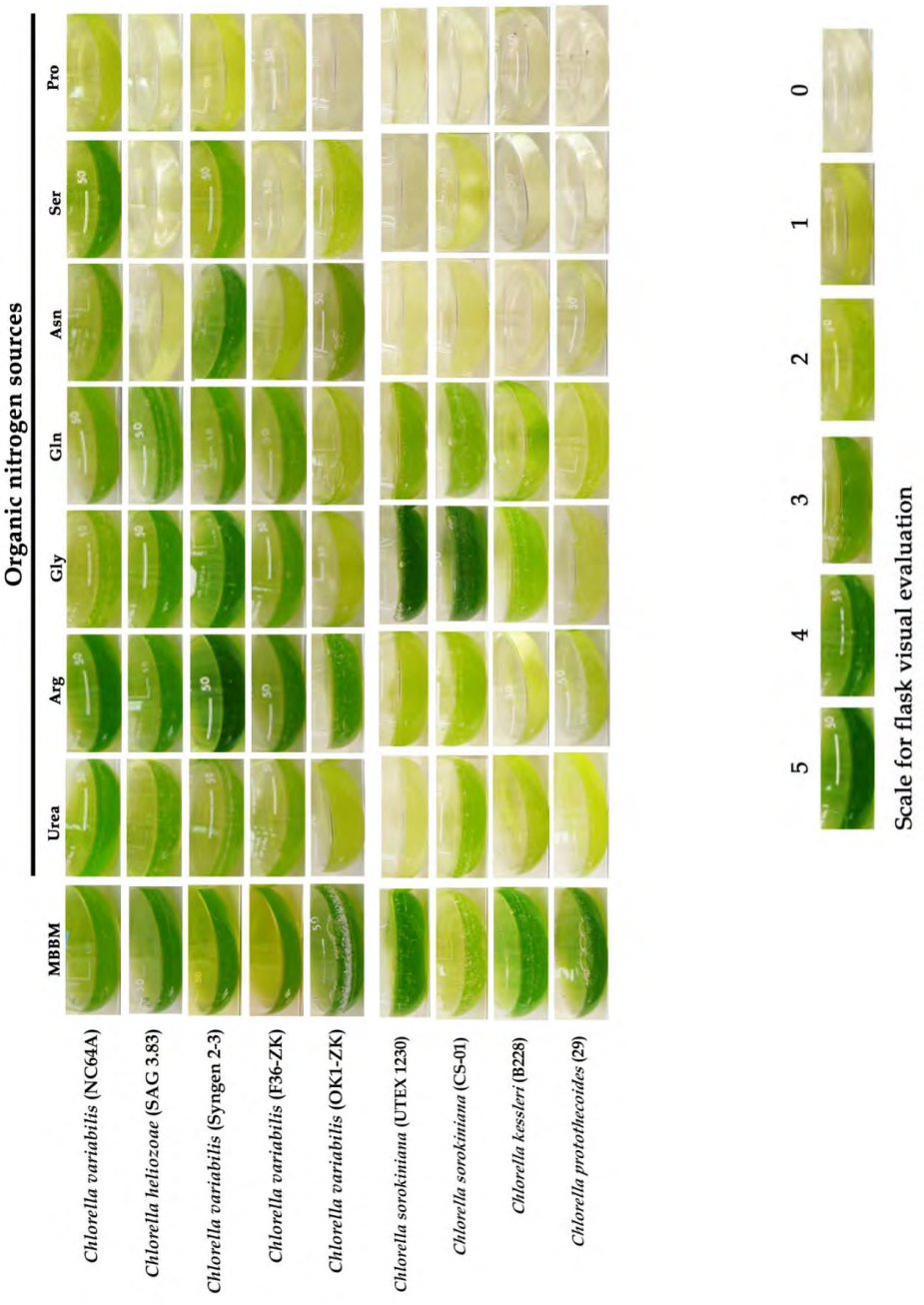
Table 3

	NC64A gene ID	Protein ID	Scaffold in *C. Sorokoniana*	e-value	Bit score	Identity	Protein length *C. variabilis*	Protein length *C. Sorokoniana*
1	138810	EFN52376	18.g93.iso1	3E-73	231 bits (590)	126/181 (70%)	183	504
2	50436	EFN58616	4.g163.iso1	5E-88	284 bits (726)	186/451 (41%)	410	690
3	138809	EFN52375	18.g93.iso1	2E-113	340 bits (871)	190/306 (62%)	287	504
4	144770	EFN56324	56.g4.iso2	7E-128	380 bits (977)	193/361 (53%)	471	405
5	144819	EFN56345	4.g37.iso3	2E-120	364 bits (935)	213/389 (55%)	489	469
6	142340	EFN58068	34.g94.iso1	1E-134	426 bits (1096)	219/305 (72%)	695	1070
7	37093	EFN51991	172.g106.iso1	2E-141	431 bits (1108)	244/446 (55%)	519	815
8	135113	EFN54610	170.g37.iso1	3E-134	414 bits (1063)	246/309 (80%)	932	461
9	145403	EFN55845	4.g163.iso1	2E-134	410 bits (1053)	255/547 (47%)	535	690
10	133029	EFN59609	28.g135.iso1	5E-139	409 bits (1052)	257/443 (58%)	431	449
11	134730	EFN54961	57.g52.iso1	2E-156	455 bits (1171)	261/424 (62%)	453	469
12	134234	EFN55146	33.g185.iso1	4E-142	420 bits (1079)	263/484 (54%)	473	480
13	142091	EFN57962	8.g116.iso2	2E-158	461 bits (1186)	266/417 (64%)	471	466
14	135437	EFN54731	3.g27.iso1	4E-161	475 bits (1222)	272/399 (68%)	518	606
15	51413	EFN57306	84.g138.iso1	1E-174	503 bits (1294)	282/452 (62%)	452	490
16	49669	EFN60144	4.g158.iso1	8E-179	533 bits (1372)	294/470 (63%)	726	742
17	133360	EFN60071	174.g65.iso1	0	552 bits (1422)	296/405 (73%)	692	685
18	56488	EFN59984	53.g43.iso1	0	585 bits (1509)	300/495 (61%)	973	515
19	138717	EFN52627	21.g78.iso1	0	547 bits (1410)	317/540 (59%)	498	547
20	133351	EFN60067	27.g54.iso1	0	600 bits (1548)	331/494 (67%)	489	473
21	138133	EFN59501	110.g43.iso1	0	594 bits (1532)	336/587 (57%)	576	512
22	58128	EFN54604	15.g150.iso1	0	637 bits (1644)	337/516 (65%)	522	484
23	57473	EFN56726	91.g67.iso3	0	681 bits (1758)	346/471 (73%)	476	468
24	138505	EFN52920	43.g21.iso1	0	659 bits (1701)	353/546 (65%	544	560
25	32765	EFN51990	35.g117.iso2	0	659 bits (1701)	366/650 (56%)	605	635
26	59057	EFN51898	35.g117.iso2	0	645 bits (1664)	373/698 (53%)	742	635
27	133392	EFN60084	39.g82.iso1	0	688 bits (1775)	378/516 (73%)	516	496
28	59479	EFN50622	No hits found					
29	133952	EFN55688	No hits found					
30	133953	EFN55689	No hits found					
31	24724	EFN54340	No hits found					
32	53357	EFN53996	No hits found					
33	137496	EFN53131	No hits found					
34	138482	EFN52813	No hits found					
35	17797	EFN50713	No hits found					

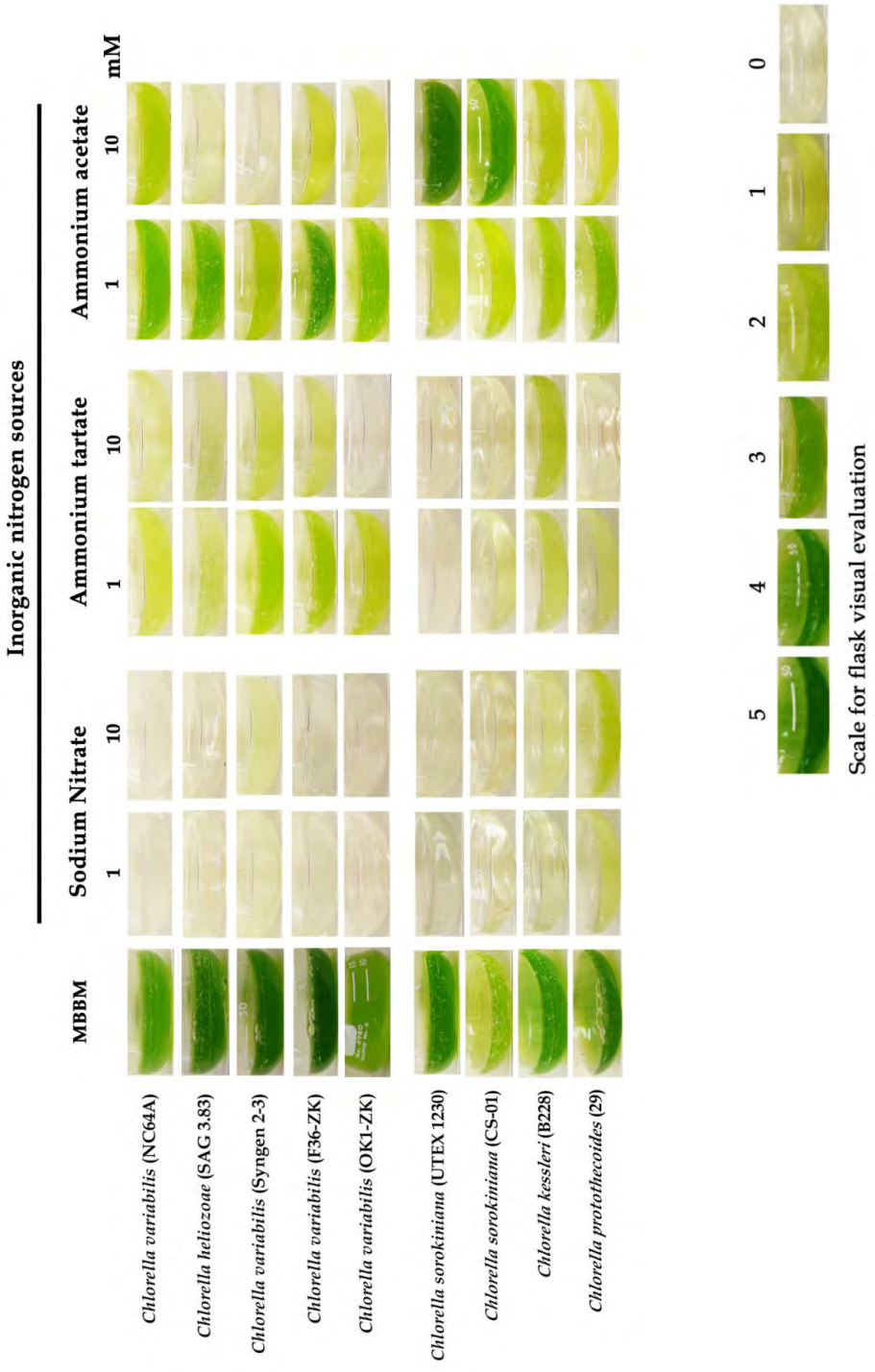
Supplementary Figure S1



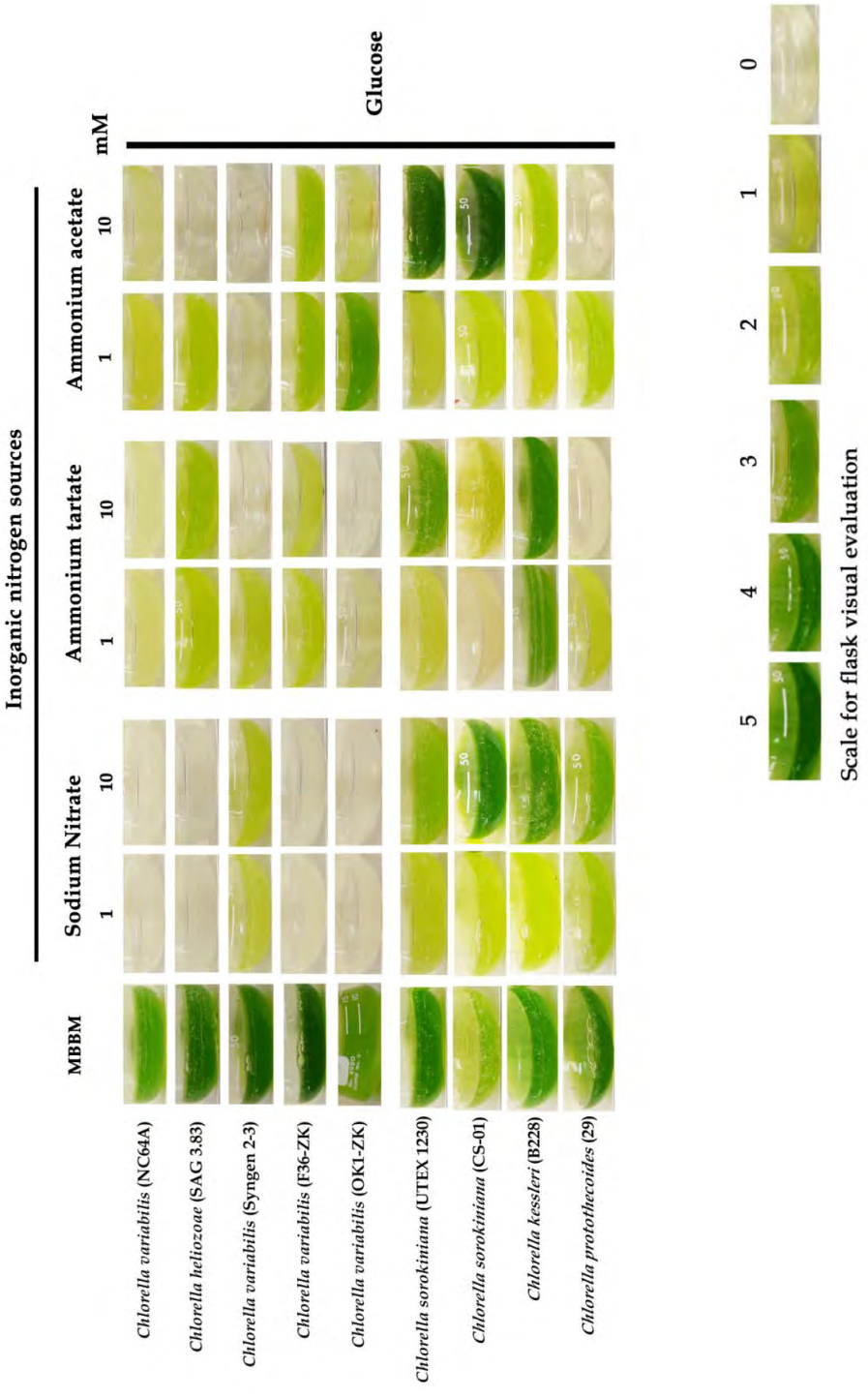
Supplementary Figure S2



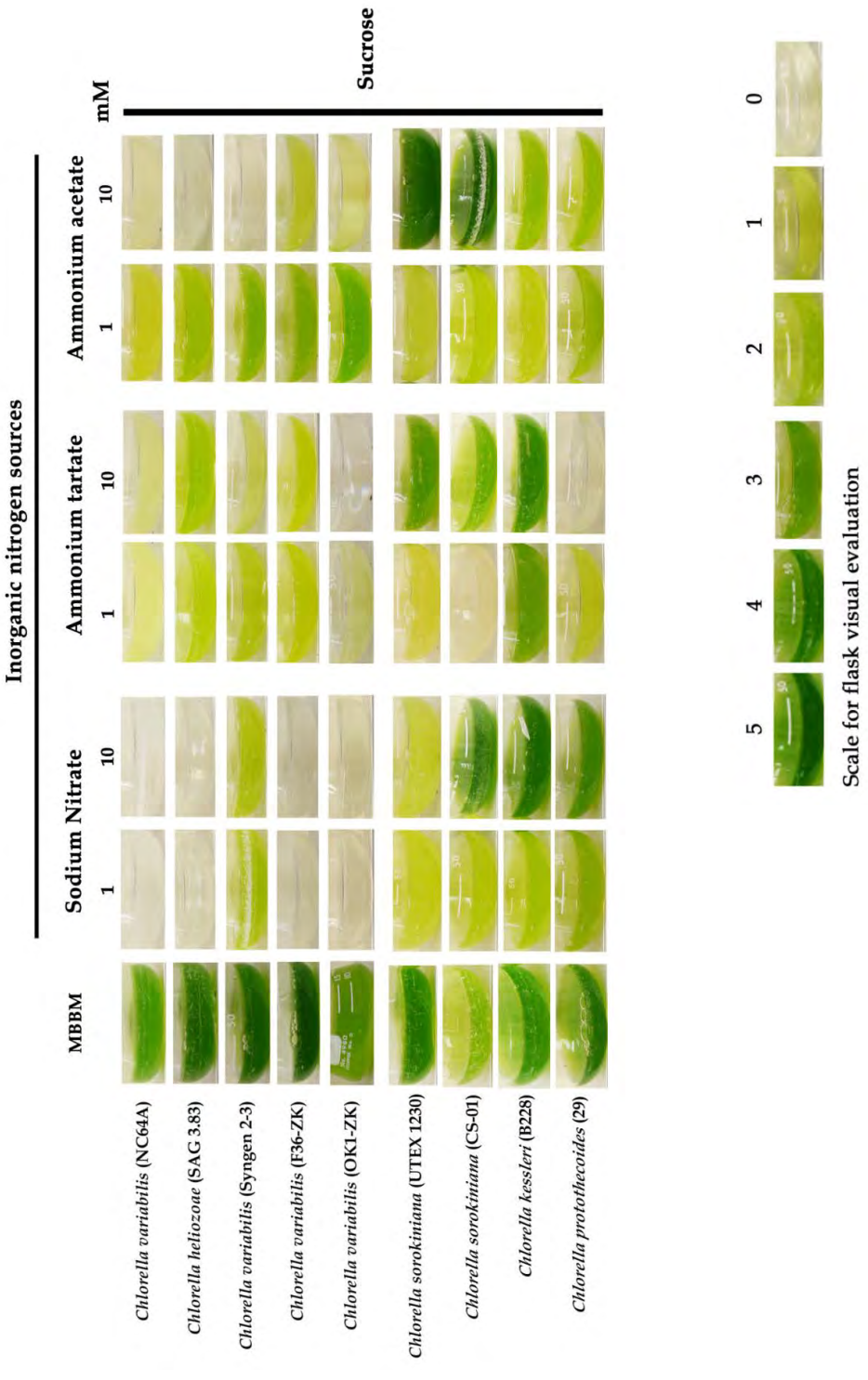
Supplementary Figure S3



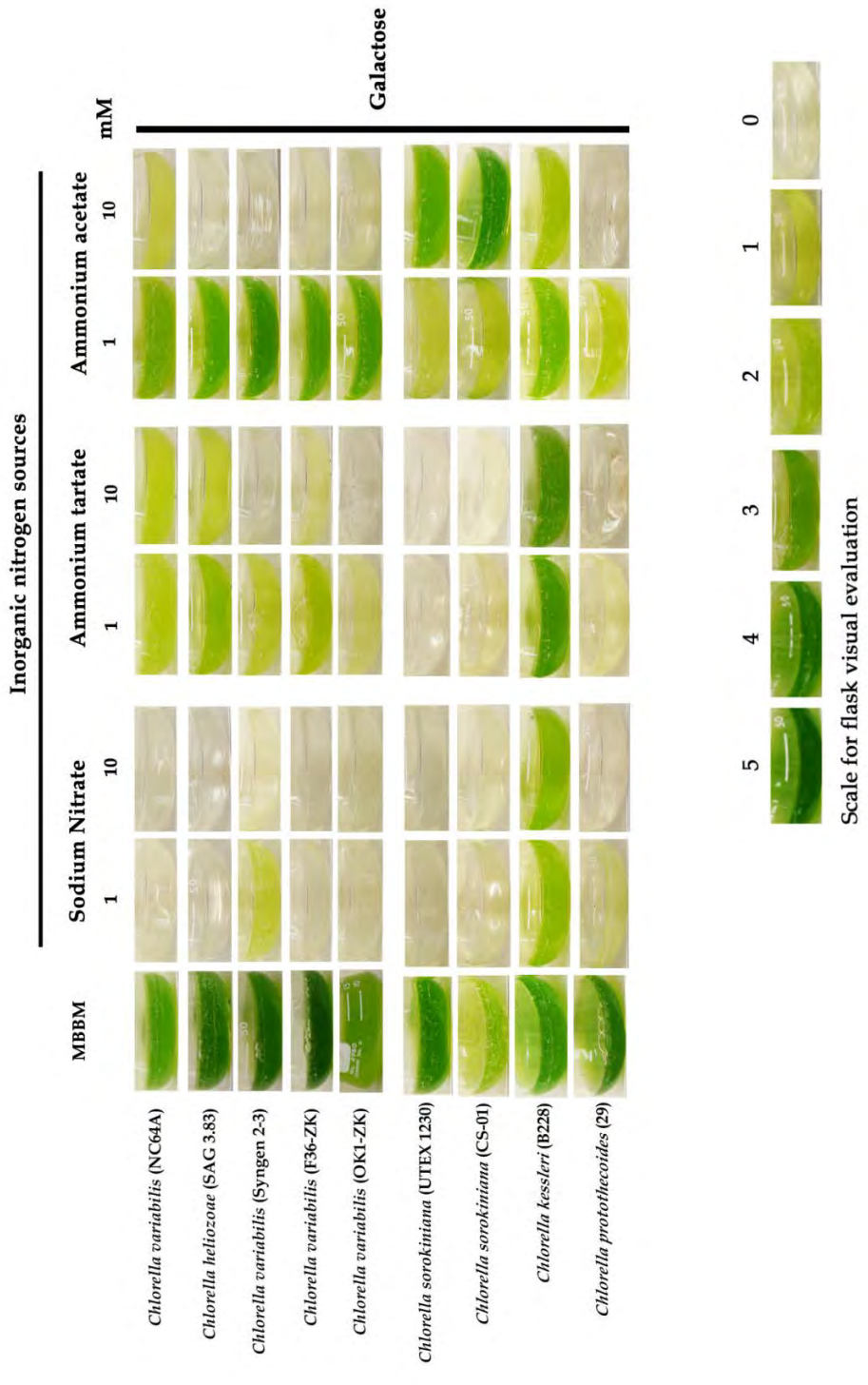
Supplementary Figure S4A



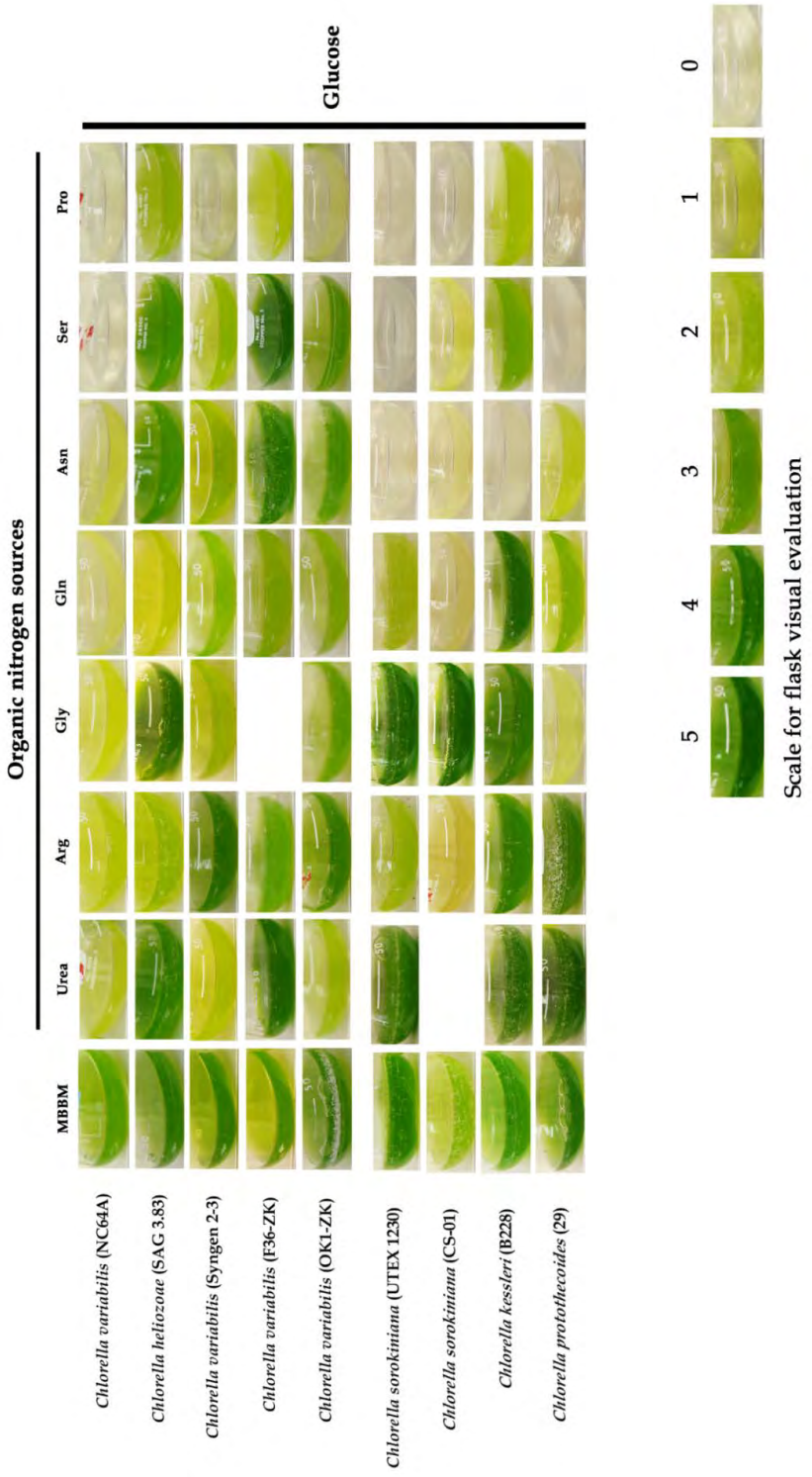
Supplementary Figure S4B



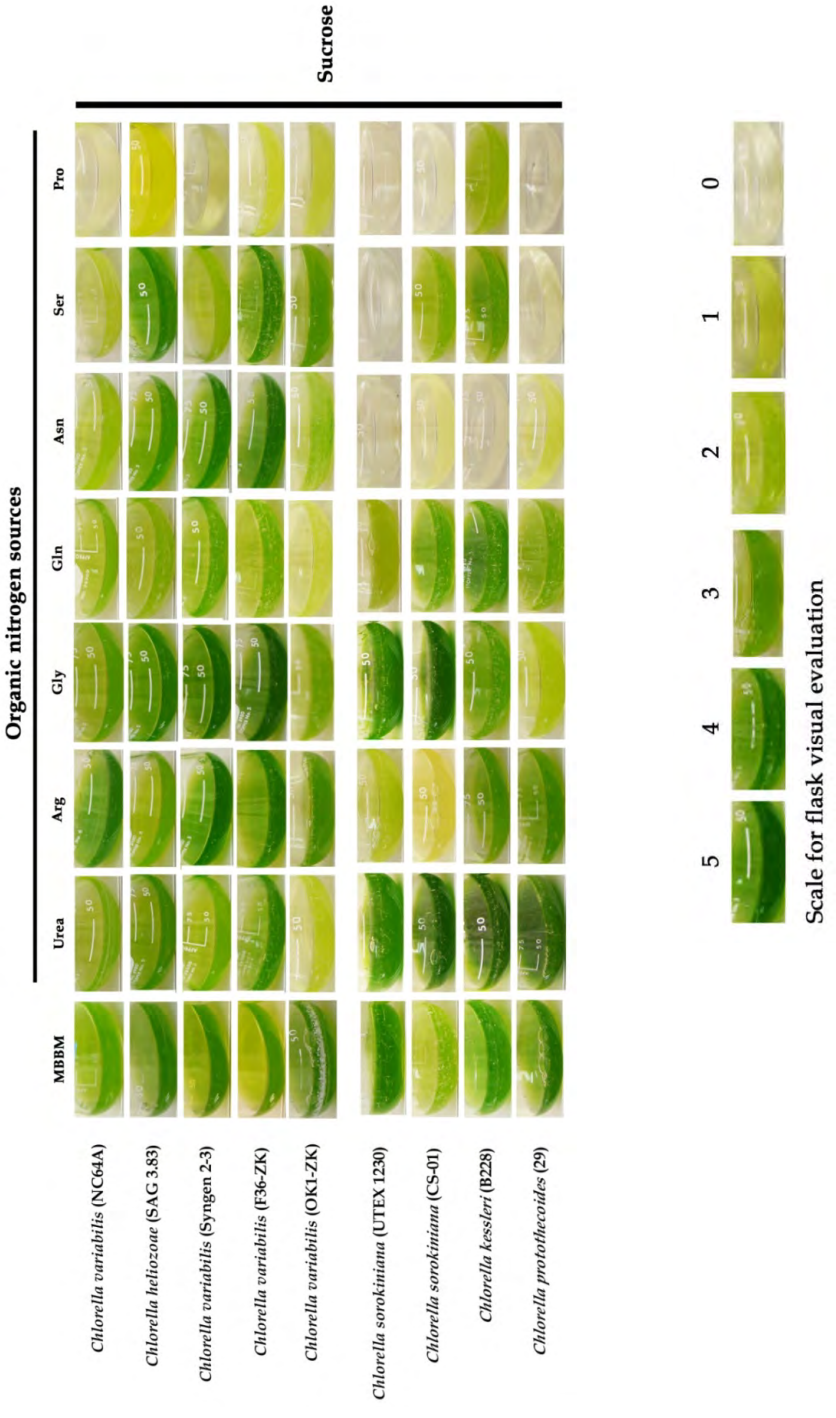
Supplementary Figure S4C



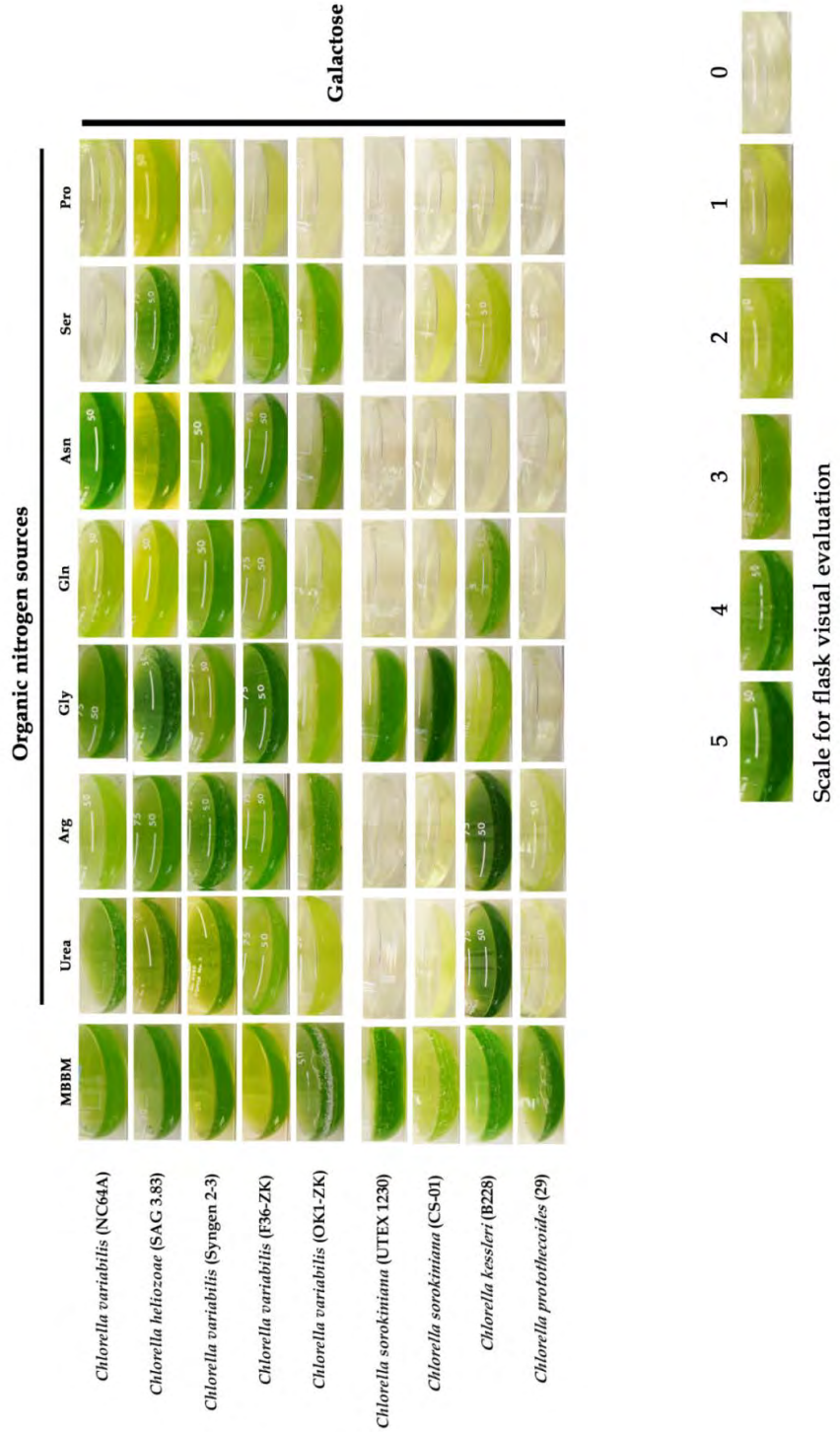
Supplementary Figure S5A



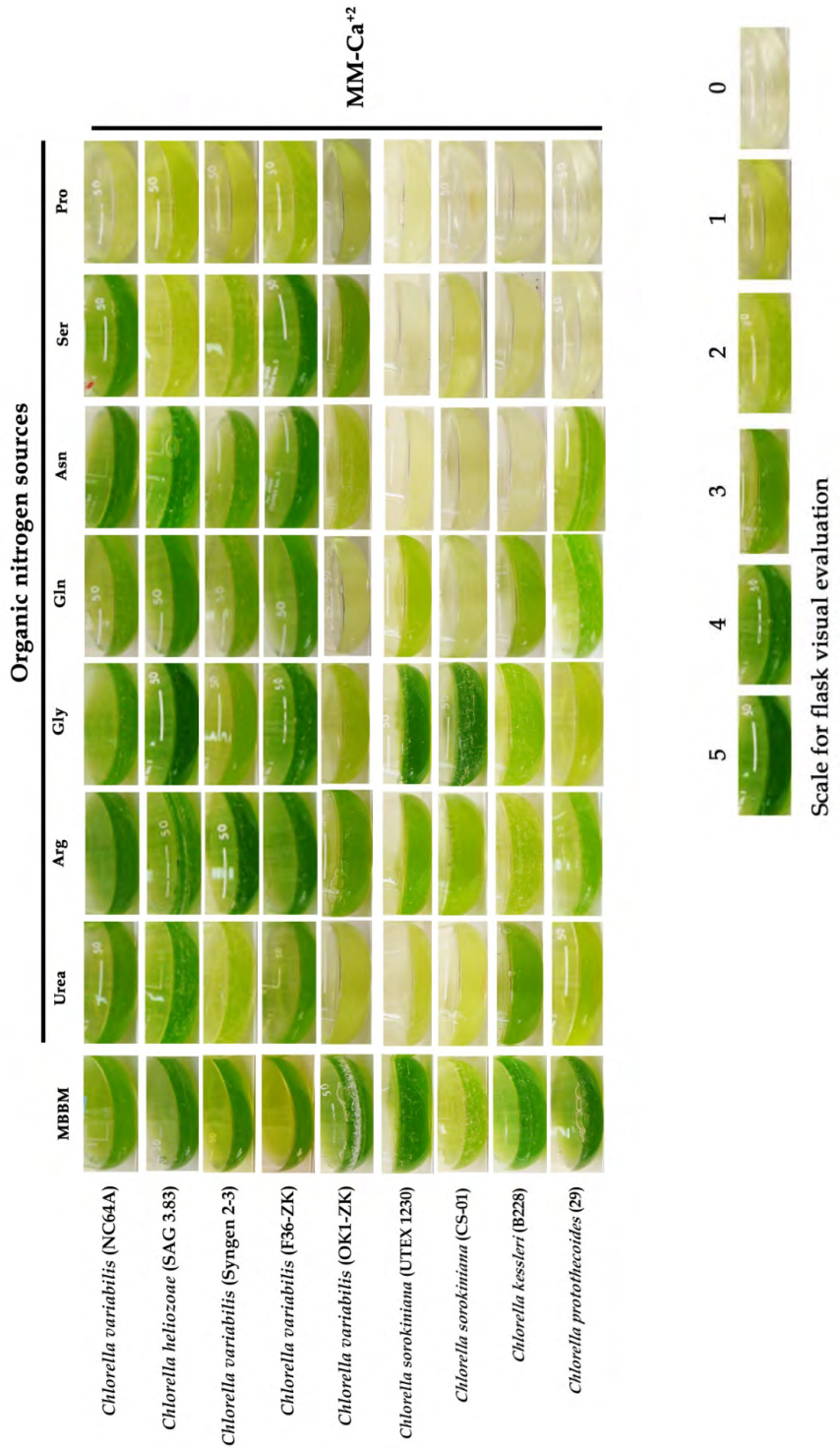
Supplementary Figure S5B



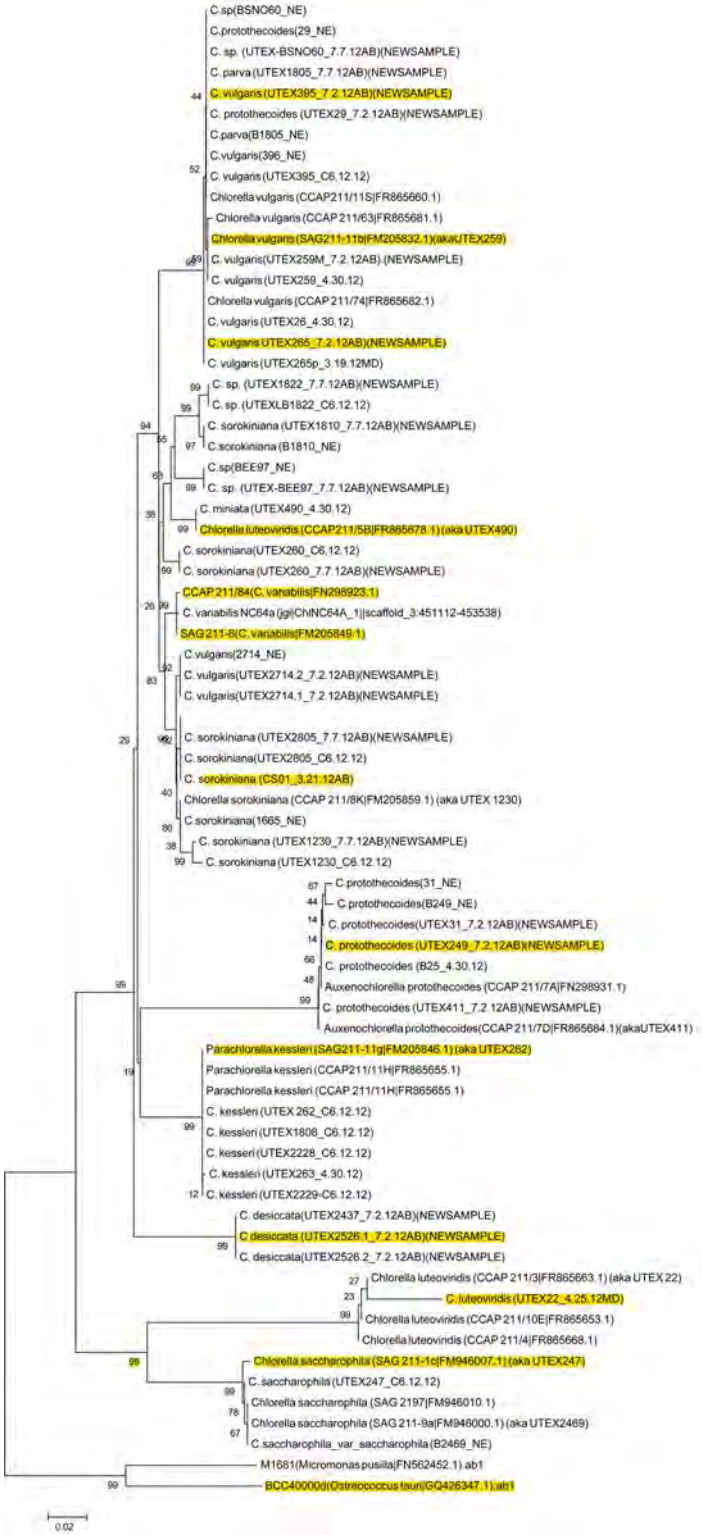
Supplementary Figure S5C



Supplementary Figure S6



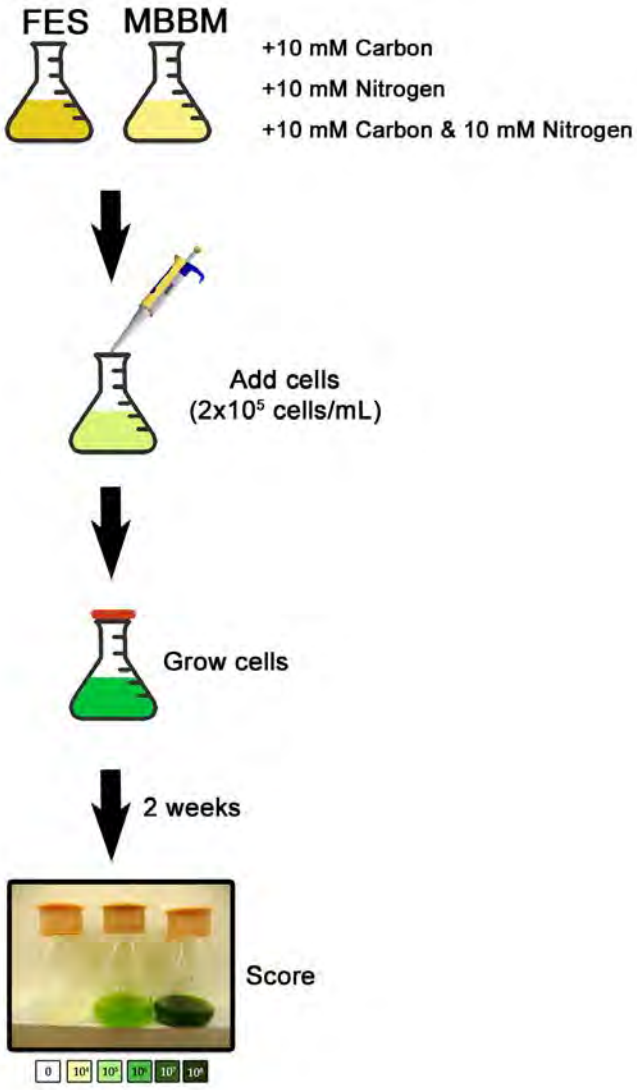
Supplementary Figure S7



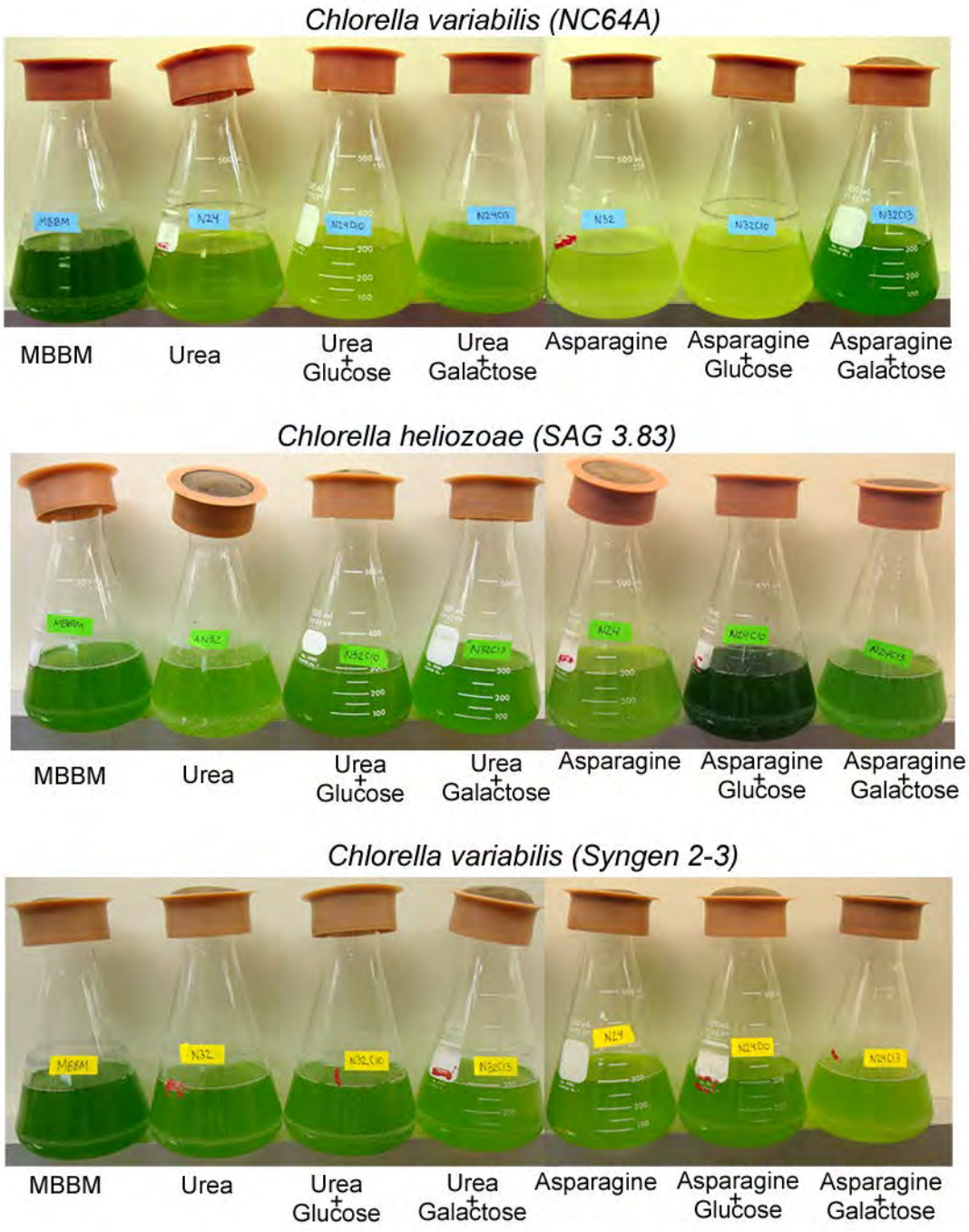
Supplementary Figure S8

		MBBM	FES
Salts	Common	Disodium EDTA (50 g) KOH (31 g) H_3BO_3 (11.42 g) $FESO_4 7H_2O$ (4.98 g) $ZnSO_4 7H_2O$ (8.82 g) $MnCl_2 4H_2O$ (1.44 g) MoO_3 (0.71 g) $CuSO_4 5H_2O$ (1.57 g) $CoNO_3 6H_2O$ (0.49 g) /L	
		$MgSO_4 7H_2O$ (7.5g) K_2HPO_4 (7.5 g) /L	$MgSO_4 7H_2O$ (10 g) K_2HPO_4 (1 g) /L
	Exclusive	$NaNO_3$ (25 g) $CaCl_2$ 2 H_2O (2.5 g) KH_2PO_4 (17.5 g) NaCl (2 g) /L	KNO_3 (1 g) /L
Nitrogen Source		Peptone (1 g/L)	Peptone (1 g/L) Lemco Powder (2 g/L)
Carbon Source		Sucrose (5 g/L)	Sucrose (5 g/L)
pH		6.8	9.4

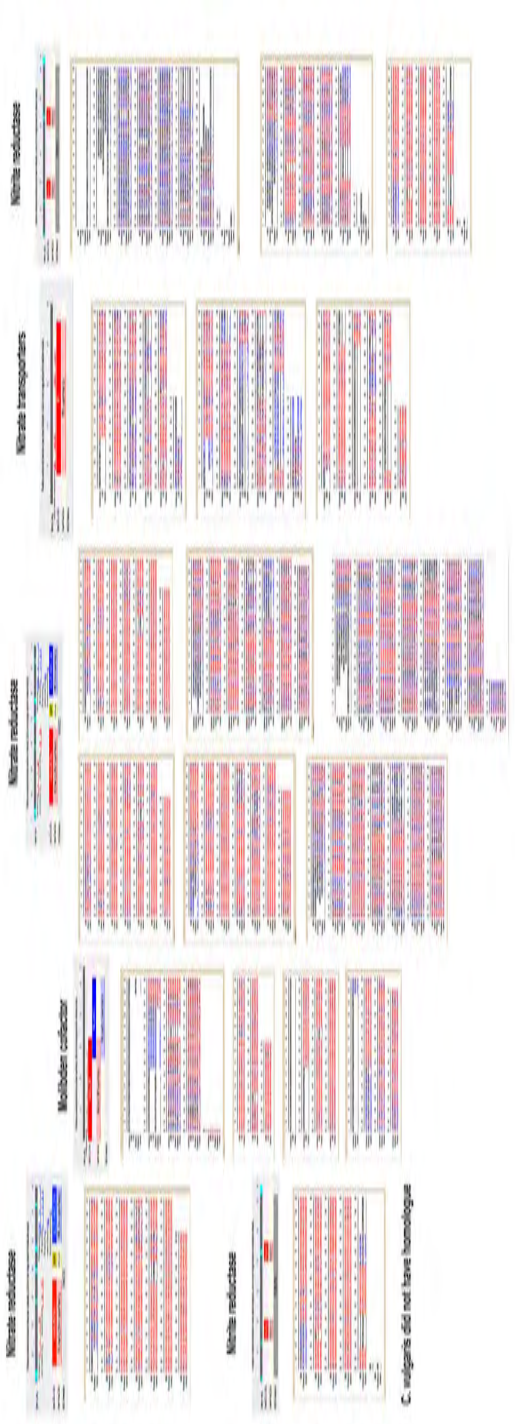
Supplementary Figure S9



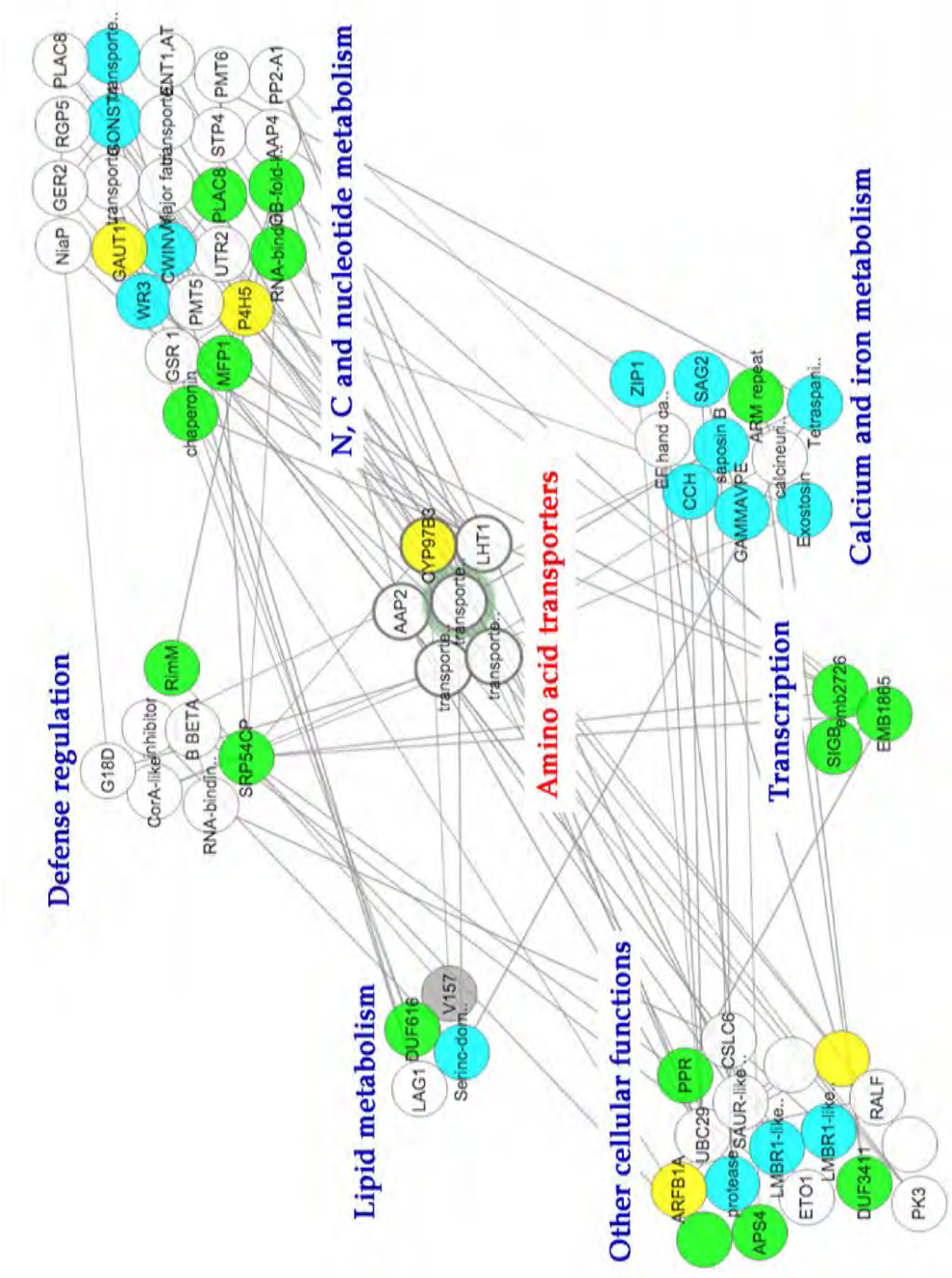
Supplementary Figure S10



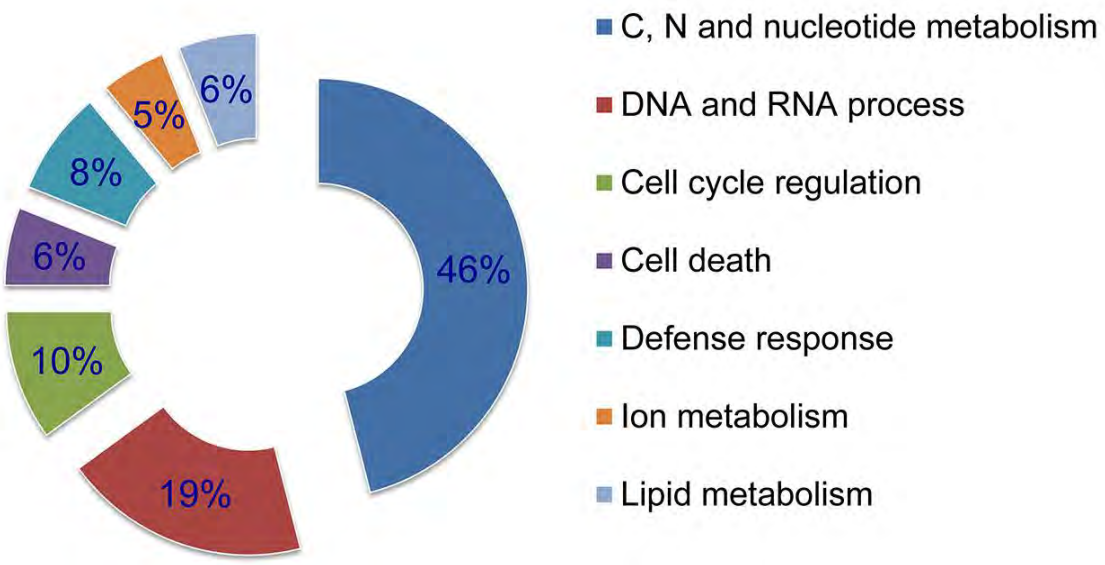
Supplementary Figure S11



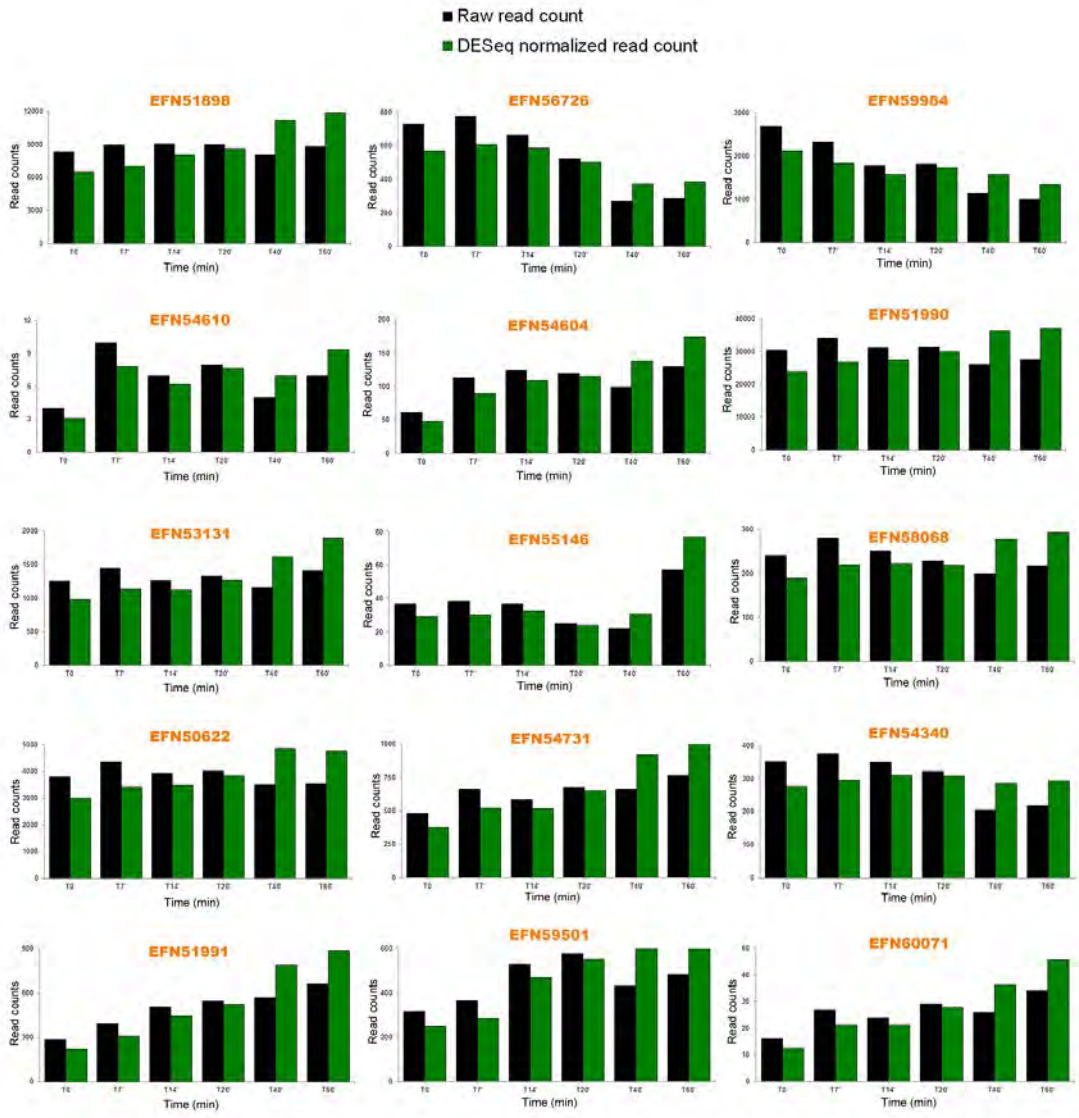
Supplementary Figure S12



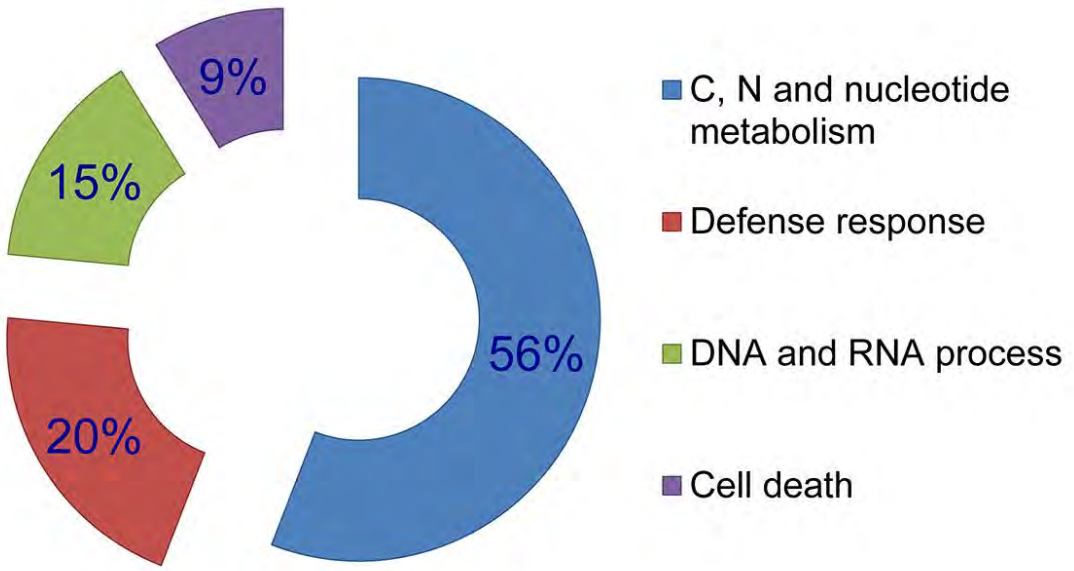
Supplementary Figure S13



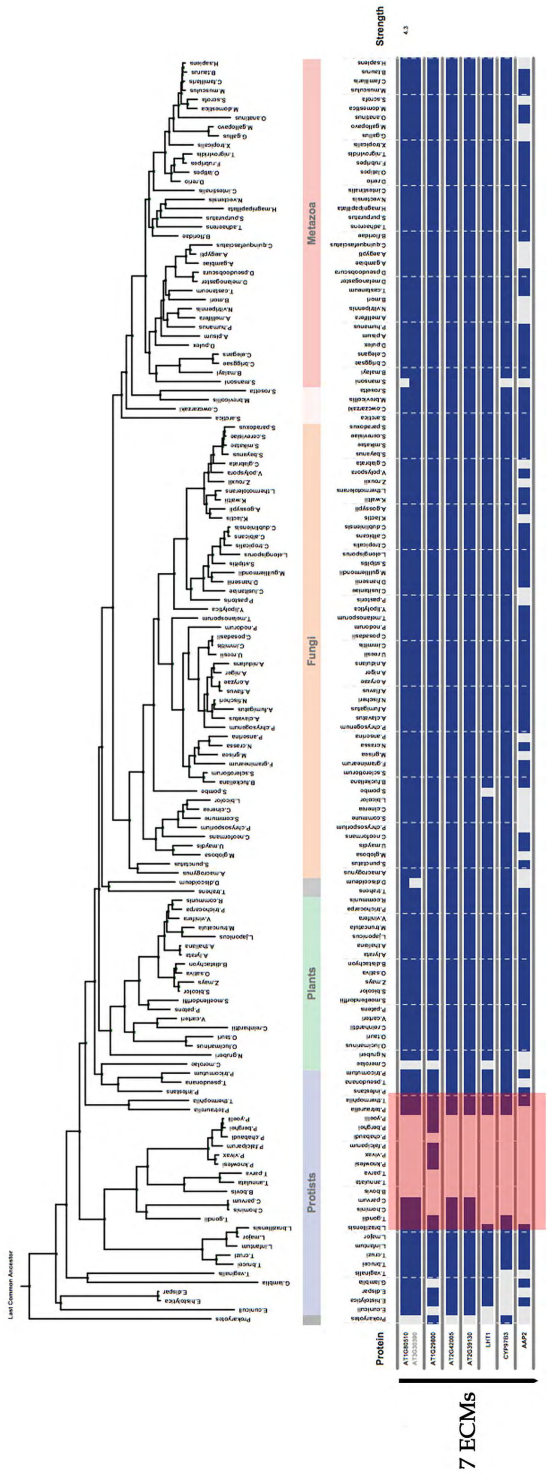
Supplementary Figure S14



Supplementary Figure S15



Supplementary Figure S16



Supplementary Table 1

NC64A gene ID	Protein ID	e-value for AA domain
133029	EFN59609	6.7E-09
56488	EFN59984	0.000000089
133351	EFN60067	0.00057
133360	EFN60071	0.00007
133392	EFN60084	4.6E-18
49669	EFN60144	0.00000078
138133	EFN59501	1.7E-51
50436	EFN58616	1.1E-09
142091	EFN57962	2.7E-11
142340	EFN58068	8.1E-09
51413	EFN57306	1.3E-13
57473	EFN56726	2.6E-78
144770	EFN56324	1.6E-31
144819	EFN56345	5.7E-19
145403	EFN55845	6.2E-28
133952	EFN55688	1E-14
133953	EFN55689	0.0002
134234	EFN55146	6.4E-14
134730	EFN54961	9.9E-11
58128	EFN54604	8.1E-63
135113	EFN54610	0.00000039
135437	EFN54731	0.000000077
24724	EFN54340	4.3E-28
53357	EFN53996	8.4E-11
137496	EFN53131	0.00000053
138482	EFN52813	0.00000011
138505	EFN52920	0.0035
138717	EFN52627	1.2E-21
138809	EFN52375	2.3E-14
138810	EFN52376	0.000027
32765	EFN51990	4.3E-77
37093	EFN51991	1E-58
59057	EFN51898	2.2E-61
17797	EFN50713	5.4E-12
59479	EFN50622	4.1E-20

Supplementary Table 2

NC64A Gene ID	Scaffold	Axenic	Symbiont	Fold change	Log_2 FC
58128	13:78798-84676 (-)	7	99	13.46	3.7
24724	14:418634-420275 (-)	30	8	0.27	-1.8
36103	14:833000-835270 (-)	3	17	6.01	2.5
53357	15:290406-294216 (-)	30	8	0.27	-1.8
58448	16:626917-630844 (-)	52	6	0.11	-3
138133	2:2150538-2155669 (-)	16	9	0.57	-0.8
32765	25:222544-226977 (-)	2635	2794	1.06	0.08
37093	25:227152-230758 (-)	35	42	1.19	0.26
59057	25:231079-235117 (+)			No Exp	
140447	3:553777-556173 (+)	95	32	0.33	-1.57
142334	4:1840877-1844482 (-)	48	12	0.25	-1.95
7483	43:40210-42680 (+)	8	8	0.97	-0.04
17797	43:87189-87664 (+)			No Exp	
57473	7:287118-291257 (-)	83	84	1	0
59479	79:1541-9952 (+)	1124	4846	4.31	2.1
144770	8:392270-395765 (-)	25	29	1.14	0.19
AAT-All	Sum of all reads	4193	7996	1.9	0.93

Supplementary Table 3

Carbohydrate active (CAZy) enzymes involved in galactose metabolism in green chlorophyte algae									
CAZy family	NC64A	C. reinhardtii	Micromonas sp. RCC299	Micromonas sp. CCMP	O. lucimarinus	O. tauri	PFAM	DESC	FUNCTION
Carbohydrate Esterases									
CE04	25						PF01522	Polysacc_deac_1	acetyl xylan esterase (EC 3.1.1.72); chitin deacetylase (EC 3.5.1.41); chitooligosaccharide deacetylase (EC 3.5.1.-); peptidoglycan GlcNAc deacetylase (EC 3.5.1.-); peptidoglycan N-acetylmuramic acid deacetylase (EC 3.5.1.-).
CE14	2	1	1	1	1	1	PF02585	PIG-L	N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (EC 3.5.1.89); diacetylchitobiose deacetylase (EC 3.5.1.-); mycothiol S-conjugate amidase (EC 3.5.1.-)
Glycoside Hydrolases									
GH01	2	2		1	1	1	PF00232	Glyco_hydro_1	β-glucosidase (EC 3.2.1.21); β-galactosidase (EC 3.2.1.23); β-mannosidase (EC 3.2.1.25); β-glucuronidase (EC 3.2.1.31); β-D-fucosidase (EC 3.2.1.38); phlorizin hydrolase (EC 3.2.1.62); exo-β-1,4-glucanase (EC 3.2.1.74); 6-phospho-β-galactosidase (EC 3.2.1.85); 6-phospho-β-glucosidase (EC 3.2.1.86); strictosidine β-glucosidase (EC 3.2.1.105); lactase (EC 3.2.1.108); amygdalin β-glucosidase (EC 3.2.1.117); prunasin β-glucosidase (EC 3.2.1.118).
GH02	2	1	1	1	1	1	PF00703	Glyco_hydro_2	β-galactosidase (EC 3.2.1.23); β-mannosidase (EC 3.2.1.25); β-glucuronidase (EC 3.2.1.31); mannosylglycoprotein endo-β-mannosidase (EC 3.2.1.152); exo-β-glucosaminidase (EC 3.2.1.165)
GH03	4		1	1			PF00933	Glyco_hydro_3	β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37); β-N-acetylhexosaminidase (EC 3.2.1.52); glucan 1,3-β-glucosidase (EC 3.2.1.58); glucan 1,4-β-glucosidase (EC 3.2.1.74); exo-1,3-1,4-glucanase (EC 3.2.1.-); alphα-L-arabinofuranosidase (EC 3.2.1.55).
GH05	7	6	3	4	5	5	PF00150	Cellulase	chitosanase (EC 3.2.1.132); β-mannosidase (EC 3.2.1.25); Cellulase (EC 3.2.1.4); glucan 1,3-β-glucosidase (EC 3.2.1.58); licheninase (EC 3.2.1.73); glucan endo-1,6-β-glucosidase (EC 3.2.1.75); mannan endo-β-1,4-mannosidase (EC 3.2.1.78); endo-β-1,4-xylanase (EC 3.2.1.8); cellulose β-1,4-cellobiosidase (EC 3.2.1.91); endo-β-1,6-galactanase (EC 3.2.1.-); β-1,3-mannanase (EC 3.2.1.-); xyloglucan-specific endo-β-1,4-glucanase (EC 3.2.1.151); mannan transglycosylase (EC 2.4.1.-)
GH09	6	3					PF00759	Glyco_hydro_9	endoglucanase (EC 3.2.1.4); cellobiohydrolase (EC 3.2.1.91); β-glucosidase (EC 3.2.1.21)
GH16	6	2	1		1	1	PF00722	Glyco_hydro_16	xyloglucan:xyloglucosyltransferase (EC 2.4.1.207); keratan-sulfate endo-1,4-β-galactosidase (EC 3.2.1.103); endo-1,3-β-glucanase (EC 3.2.1.39); endo-1,3(4)-β-glucanase (EC 3.2.1.6); licheninase (EC 3.2.1.73); β-agarase (EC 3.2.1.81); κ-carrageenase (EC 3.2.1.83); xyloglucanase (EC 3.2.1.151)
GH27	4	1					PF02065	Melibiase	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); isomalto-dextranase (EC 3.2.1.94); β-L-arabinopyranosidase (EC 3.2.1.88)
GH28	1	1					PF00295	Glyco_hydro_28	polygalacturonase (EC 3.2.1.15); exo-polygalacturonase (EC 3.2.1.67); exo-polygalacturonosidase (EC 3.2.1.82); rhamnogalacturonase (EC 3.2.1.-); endo-xylogalacturonan hydrolase (EC 3.2.1.-); rhamnogalacturonan α-L-rhamnopyranohydrolase (EC 3.2.1.40)
GH31	6	2	4	4	4	4	PF01055	Glyco_hydro_31	α-glucosidase (EC 3.2.1.20); α-1,3-glucosidase (EC 3.2.1.84); sucrase-isomaltase (EC 3.2.1.48) (EC 3.2.1.10); α-xylosidase (EC 3.2.1.-); α-glucan lyase (EC 4.2.2.13); isomaltosyltransferase (EC 2.4.1.-).

GH32	3	4					PF00251	Glyco_hydro_32N	invertase (EC 3.2.1.26); endo-inulinase (EC 3.2.1.7); β-2,6-fructan 6-levanbiohydrolase (EC 3.2.1.64); endo-levanase (EC 3.2.1.65); exo-inulinase (EC 3.2.1.80); fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153); fructan β-(2,6)-fructosidase/6-exohydrolase (EC 3.2.1.154); sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99); fructan:fructan 1-fructosyltransferase (EC 2.4.1.100); sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10); fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243); levan fructosyltransferase (EC 2.4.1.-)
GH35	1						PF01301	Glyco_hydro_35	β-galactosidase (EC 3.2.1.23); exo-β-glucosaminidase (EC 3.2.1.165)
GH36	5	1	2	2	1	1	PF05691	Raffinose_syn	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); stachyose synthase (EC 2.4.1.67); raffinose synthase (EC 2.4.1.82)
GH37	2	1	1	1	1	1	PF01204	Trehalase	α,α-trehalase (EC 3.2.1.28).
GH42	3	2					PF02449	Glyco_hydro_42	β-galactosidase (EC 3.2.1.23) : sucre sur protein ?
GH46	4						PF01374	Glyco_hydro_46	chitosanase (EC 3.2.1.132)
GH47	4	1	3	4	2	2	PF01532	Glyco_hydro_47	α-mannosidase (EC 3.2.1.113)
GH51	2						PF06964	Alpha-L-AF_C	α-L-arabinofuranosidase (EC 3.2.1.55); endoglucanase (EC 3.2.1.4)
GH77	2	2	2	2	2	2	PF02446	Glyco_hydro_77	amylomaltase or 4-α-glucanotransferase (EC 2.4.1.25)
GlycosylTransferases									
GT01	4	1					PF00201	UDPGT	UDP-glucuronosyltransferase (EC 2.4.1.17); 2-hydroxyacylsphingosine 1-β-galactosyltransferase (EC 2.4.1.45); N-acylsphingosine galactosyltransferase (EC 2.4.1.47); flavonol 3-O-glucosyltransferase (EC 2.4.1.91); indole-3-acetate β-glucosyltransferase (EC 2.4.1.121); sterol glucosyltransferase (EC 2.4.1.173); ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-); zeaxanthin glucosyltransferase (EC 2.4.1.-); zeatin O-β-glucosyltransferase (EC 2.4.1.203); zeatin O-β-xylosyltransferase (EC 2.4.2.40); limonoid glucosyltransferase (EC 2.4.1.210); sinapate 1-glucosyltransferase (EC 2.4.1.120); anthocyanin 3-O-galactosyltransferase (EC 2.4.1.-); anthocyanin 5-O-glucosyltransferase (EC 2.4.1.-); anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115); dTDP-β-2-deoxy-L-fucose: α-L-2-deoxyfucosyltransferase (EC 2.4.1.-); UDP-β-L-rhamnose: α-L-rhamnosyltransferase (EC 2.4.1.-); UDP-glucose: 4-hydroxybenzoate 4-O-β-glucosyltransferase (EC 2.4.1.194); flavonol L-rhamnosyltransferase (EC 2.4.1.159); salicylic acid β-glucosyltransferase (EC 2.4.1.-)
GT08	5		1		1		PF01501	Glyco_transf_8	lipopolysaccharide α-1,3-galactosyltransferase (EC 2.4.1.44); UDP-Glc:(glucosyl)lipopolysaccharide α-1,2-glucosyltransferase (EC 2.4.1.-); lipopolysaccharide glucosyltransferase 1 (EC 2.4.1.58); glycogenin glucosyltransferase (EC 2.4.1.186); inositol 1-α-galactosyltransferase (galactinol synthase) (EC 2.4.1.123); homogalacturonan α-1,4-galacturonosyltransferase (EC 2.4.1.43)
GT31	5	1	1	1			PF01762	Galactosyl_T	N-acetyllactosaminide β-1,3-N-acetylglucosaminyltransferase (EC 2.4.1.149); Glycoprotein-N-acetylgalactosamine 3-β-galactosyltransferase (EC 2.4.1.122); fucose-specific β-1,3-N-acetylglucosaminyltransferase (EC 2.4.1.-); globotriosylceramide β-1,3-GalNAc transferase (EC 2.4.1.79); chondroitin synthase (β-1,3-GlcUA and β-1,4-GalNAc transferase (EC 2.4.1.175); chondroitin β-1,3-glucuronyltransferase (EC 2.4.1.226); chondroitin β-1,4-N-acetylgalactosaminyltransferase (EC 2.4.1.-).
GT41	3	1	1	1	2	1			UDP-GlcNAc: peptide β-N-acetylglucosaminyltransferase (EC 2.4.1.94)
GT47	16	29	3	1	3	2	PF03016	Exostosin	heparan β-glucuronyltransferase (EC 2.4.1.225); xyloglucan β-galactosyltransferase (EC 2.4.1.-); heparan synthase (EC 2.4.1.-); arabinan alphα-L-arabinosyltransferase (EC 2.4.2.-).
GT64	4	1	2	2	2		PF09258	EXTL2	UDP-GlcNAc: heparan α-N-acetylhexosaminyltransferase (EC 2.4.1.224)
GT77	14	11	9	6	9	7			α-xylosyltransferase (EC 2.4.2.39); α-1,3-galactosyltransferase (EC 2.4.1.37); arabinosyltransferase (EC 2.4.2.-)
GT90	7	10							UDP-Xyl: (mannosyl) glucuronoxylomannan/galactoxylomannan β-1,2-xylosyltransferase (EC 2.4.2.-)
TOTAL	**149**	**84**	**36**	**30**	**36**	**29**			

GH32	3	4					PF00251	Glyco_hydro_32N	invertase (EC 3.2.1.26); endo-inulinase (EC 3.2.1.7); β-2,6-fructan 6-levanbiohydrolase (EC 3.2.1.64); endo-levanase (EC 3.2.1.65); exo-inulinase (EC 3.2.1.80); fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153); fructan β-(2,6)-fructosidase/6-exohydrolase (EC 3.2.1.154); sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99); fructan:fructan 1-fructosyltransferase (EC 2.4.1.100); sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10); fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243); levan fructosyltransferase (EC 2.4.1.-)
GH35	1						PF01301	Glyco_hydro_35	β-galactosidase (EC 3.2.1.23); exo-β-glucosaminidase (EC 3.2.1.165)
GH36	5	1	2	2	1	1	PF05691	Raffinose_syn	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); stachyose synthase (EC 2.4.1.67); raffinose synthase (EC 2.4.1.82)
GH37	2	1	1	1	1	1	PF01204	Trehalase	α,α-trehalase (EC 3.2.1.28).
GH42	3	2					PF02449	Glyco_hydro_42	β-galactosidase (EC 3.2.1.23) : sucre sur protein ?
GH46	4						PF01374	Glyco_hydro_46	chitosanase (EC 3.2.1.132)
GH47	4	1	3	4	2	2	PF01532	Glyco_hydro_47	α-mannosidase (EC 3.2.1.113)
GH51	2						PF06964	Alpha-L-AF_C	α-L-arabinofuranosidase (EC 3.2.1.55); endoglucanase (EC 3.2.1.4)
GH77	2	2	2	2	2	2	PF02446	Glyco_hydro_77	amylomaltase or 4-α-glucanotransferase (EC 2.4.1.25)
GlycosylTransferases									
GT01	4	1					PF00201	UDPGT	UDP-glucuronosyltransferase (EC 2.4.1.17); 2-hydroxyacylsphingosine 1-β-galactosyltransferase (EC 2.4.1.45); N-acylsphingosine galactosyltransferase (EC 2.4.1.47); flavonol 3-O-glucosyltransferase (EC 2.4.1.91); indole-3-acetate β-glucosyltransferase (EC 2.4.1.121); sterol glucosyltransferase (EC 2.4.1.173); ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-); zeaxanthin glucosyltransferase (EC 2.4.1.-); zeatin O-β-glucosyltransferase (EC 2.4.1.203); zeatin O-β-xylosyltransferase (EC 2.4.2.40); limonoid glucosyltransferase (EC 2.4.1.210); sinapate 1-glucosyltransferase (EC 2.4.1.120); anthocyanin 3-O-galactosyltransferase (EC 2.4.1.-); anthocyanin 5-O-glucosyltransferase (EC 2.4.1.-); anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115); dTDP-β-2-deoxy-L-fucose: α-L-2-deoxyfucosyltransferase (EC 2.4.1.-); UDP-β-L-rhamnose: α-L-rhamnosyltransferase (EC 2.4.1.-); UDP-glucose: 4-hydroxybenzoate 4-O-β-glucosyltransferase (EC 2.4.1.194); flavonol L-rhamnosyltransferase (EC 2.4.1.159); salicylic acid β-glucosyltransferase (EC 2.4.1.-)
GT08	5		1		1		PF01501	Glyco_transf_8	lipopolysaccharide α-1,3-galactosyltransferase (EC 2.4.1.44); UDP-Glc:(glucosyl)lipopolysaccharide α-1,2-glucosyltransferase (EC 2.4.1.-); lipopolysaccharide glucosyltransferase 1 (EC 2.4.1.58); glycogenin glucosyltransferase (EC 2.4.1.186); inositol 1-α-galactosyltransferase (galactinol synthase) (EC 2.4.1.123); homogalacturonan α-1,4-galacturonosyltransferase (EC 2.4.1.43)
GT31	5	1	1	1			PF01762	Galactosyl_T	N-acetyllactosaminide β-1,3-N-acetylglucosaminyltransferase (EC 2.4.1.149); Glycoprotein-N-acetylgalactosamine 3-β-galactosyltransferase (EC 2.4.1.122); fucose-specific β-1,3-N-acetylglucosaminyltransferase (EC 2.4.1.-); globotriosylceramide β-1,3-GalNAc transferase (EC 2.4.1.79); chondroitin synthase (β-1,3-GlcUA and β-1,4-GalNAc transferase (EC 2.4.1.175); chondroitin β-1,3-glucuronyltransferase (EC 2.4.1.226); chondroitin β-1,4-N-acetylgalactosaminyltransferase (EC 2.4.1.-).
GT41	3	1	1	1	2	1			UDP-GlcNAc: peptide β-N-acetylglucosaminyltransferase (EC 2.4.1.94)
GT47	16	29	3	1	3	2	PF03016	Exostosin	heparan β-glucuronyltransferase (EC 2.4.1.225); xyloglucan β-galactosyltransferase (EC 2.4.1.-); heparan synthase (EC 2.4.1.-); arabinan alphα-L-arabinosyltransferase (EC 2.4.2.-).
GT64	4	1	2	2	2		PF09258	EXTL2	UDP-GlcNAc: heparan α-N-acetylhexosaminyltransferase (EC 2.4.1.224)
GT77	14	11	9	6	9	7			α-xylosyltransferase (EC 2.4.2.39); α-1,3-galactosyltransferase (EC 2.4.1.37); arabinosyltransferase (EC 2.4.2.-)
GT90	7	10							UDP-Xyl: (mannosyl) glucuronoxylomannan/galactoxylomannan β-1,2-xylosyltransferase (EC 2.4.2.-)
TOTAL	**149**	**84**	**36**	**30**	**36**	**29**			

Supplementary Table 4

NC64A gene ID	Protein ID	AA size	KEGG orthologs in other oganisms (number)	*A. thaliana* orthologs	*S. cerevisiae* orthologs	*C. sorokoniana* transcripts	*C. sorokoniana* scaffold
56488	EFN59984	973	cre(1), olu(2), ola(4), ota (1), mis (1), csl (2), cvr(35)	at3g30390		Yes, low	53:146447..157039
59057	EFN51898	742	ath(23), mpp (5), cvr (12), cse (1), csl (5)	AT5G09220	YJR001W	Yes, very low	35:498378..505259
32765	EFN51990	605	ath(23), mpp (5), cvr (12), cse (1), csl (5)	AT5G09220	YJR001W	Yes, very low	35:498544..505357
135113	EFN54610	932	ath(220), cse(5),cre (28), vcn(17), olu (17), ola (50), mis(14), mpp(9), csl(33), cvr (30)	AT4G15110	YMR015C	Yes, very low present	170:163406..171046
134234	EFN55146	473	ath (14), sce (1), cre(1), vcn (1), olu (2), ota (1), mis (1), mpp(2), csl (7), cvr (2)	AT2G39130	YJR001W	No present	
58128	EFN54604	522	ath((35), sce (3), cre (1), vcn (2), olu (1), miss (2), mpp (1), csl (7), cvr (13)	AT5G09220	YEL064C	Yes low	15:688606..674984
137496	EFN53131	1010	ath((5), sce (3), cre (3), vcn (2), olu (2), ota(1), miss (2),, csl (1), cvr (8)	AT1G80510	YBL089W	No present	
37093	EFN51991	519	ath((28), sce (2), vcn (1), csl (6), cvr (14)	AT5G09220	YER119C	No present	
59479	EFN50622	210	ath((20),csl (5), cvr (10)	AT5G09220	YIL088C	No present	
135437	EFN54731	518	ath((6), sce (3), cre (3), vcn (2), olu (1),ola (1), miss (2), csl (1), cvr (12)	AT1G80510	YER119C	Yes, very low	87934..93569
24724	EFN54340	227	ath((17), csl (4), cvr (9)	AT5G09220		No present	
57473	EFN56726	476	ath((41), sce (2), cre (2), vcn (2), olu (2), ota (2), miss (1), mpp (2), csl (10), cvr (14)	AT5G40780	YJR001W	Yes, very low	91:290538..297435
142340	EFN58068	695	ath((18), sce (5), cre (5), vcn (5), olu (3),ota (2), miss (5), mpp (1), csl (4), cvr (19)	AT1G29800	YIL088C	Yes very low	34:367174..380401
138133	EFN59501	576	ath((24), miss (1), csl (5), cvr (11)	AT5G09220		No express	76:9231..19276
133360	EFN60071	692	ath((3), sce (1), cre (2), vcn (a), olu (2), ota(1), csl (3), cvr (4)	AT2G42005	YJR001W	Yes, very low	174:286706..292612

Supplementary Table 5

	Gene name	ID	Function	Process
C, N and nucleotide metabolism	CSLC6 putative xyloglucan glycosyltransferase 6	AT3G07330	cellulose synthase activity	carbohydrate biosynthetic process
	GAUT11 putative galactouronosyltransferase	AT1G18580	polygalacturonate 4-alpha-galacturonosyltransferase activity	cell wall organization
	PK3 serine/threonine protein kinase 3	AT5G08160	ATP binding	positive regulation of translation
	RGP5 reversably-glycosylated protein 5	AT5G16510	NOT UDP-arabinopyranose mutase activity	glucose catabolic process
	GONST4 golgi nucleotide sugar transporter 4	AT5G19960	None listed	carbohydrate transport, nucleotide-sugar transport
	GER2 putative GDP-L-fucose synthase 2	AT1G17890	GDP-L-fucose synthase activity, catalytic activity	GDP-L-fucose biosynthetic process
	transporter nucleotide-sugar transporter family protein	AT4G32390	organic anion transmembrane transporter activity	carbohydrate transport
	ENT1 equilibrative nucleotide transporter 1	AT1G70330	nucleoside transmembrane transporter activity	nucleoside transmembrane transport
	PMT5 polyol transporter 5	AT3G18830	D-ribose transmembrane transporter activity	glucose import
	GSR 1 glutamine synthetase 1.1	AT5G37600	ATP binding, copper ion binding	nitrate assimilation
	CCH copper chaperone	AT3G56240	copper chaperone activity	cellular modified amino acid biosynthetic process
	UTR2 UDP-galactose transporter 2	AT4G23010	UDP-galactose transmembrane transporter activity	UDP-galactose transmembrane transport
	STP4 sugar transport protein 4	AT3G19930	glucose transmembrane transporter activity	glucose import
	PMT6 putative polyol transporter 6	AT4G36670	glucose transmembrane transporter activity	oligopeptide transport
	AAP4 amino acid permease 4	AT5G63850	amino acid transmembrane transporter activity	amino acid transport
	PP2-A1 protein PHLOEM protein 2-LIKE A1	AT4G19840	carbohydrate binding	regulation of plant-type hypersensitive response
	RimM-like 16S rRNA processing protein	AT5G4620	ribosome binding	maltose metabolic process
	MFP1 MAR-binding filament-like protein 1	AT3G16000	DNA binding	starch metabolic process
	RNA recognition motif-containing protein	AT3G20930	RNA binding	starch biosynthetic process
Defense response	ETO1 Ethylene-overproduction protein 1	AT3G51770	protein binding, bridging	regulation of ethylene biosynthetic process
	NiaP nicotinate transporter	AT3G13050	ATP binding, N-methylnicotinate transporter activity	salicylic acid mediated signaling pathway
	P4H5 prolyl 4-hydroxylase 5	AT2G17720	L-ascorbic acid binding, iron ion binding	response to endoplasmic reticulum stress
	major facilitator protein	AT2G39210	molecular function	systemic acquired resistance
	CWINV1 beta-fructofuranosidase	AT3G13790	beta-fructofuranosidase activity	nitrate transport, respiratory burst involved in defense response
	ALB1 magnesium-chelatase subunit chlD	AT1G08520	ATP binding	cytokinin metabolic process
	hypothetical protein	AT5G04080	molecular function	regulation of plant-type hypersensitive response
DNA and RNA process	SIG2 RNA polymerase sigma subunit 2	AT1G08540	DNA binding, DNA-directed RNA polymerase activity	rRNA processing
	chaperonin trigger factor type chaperone family protein	AT5G55220	peptidyl-prolyl cis-trans isomerase activity	mRNA modification
	UBC29 ubiquitin-conjugating enzyme E2 29	AT2G16740	ATP binding, acid-amino acid ligase activity	DNA endoreduplication
	Nucleic acid-binding, OB-fold-like protein	AT1G12800	RNA binding	ncRNA metabolic process
	CPSRP54 chloroplast signal recognition particle 54	AT5G03940	7S RNA binding, GTP binding	production of miRNAs involved in gene silencing by miRNA
Cell death	G18D autophagy-related protein 18D	AT3G56440	phosphatidylinositol-3,5-bisphosphate binding	autophagy
	CorA like Magnesium transporter CorA-like family protein	AT2G04305	metal ion transmembrane transporter activity	negative regulation of defense response
	WR3 high-affinity nitrate transporter 3.1	AT5G50200	nitrate transmembrane transporter activity	negative regulation of programmed cell death

Supplementary Table 6

	ECM	LLR	ECM+	gene ID	Function	Process
Carbon, nitrogen and nucleotide metabolism	1	18.41	HMG (high mobility group) box protein	AT4G23800	DNA binding	acetyl-CoA metabolic process
	1	10.41	uridine kinase-like protein 3	UKL3	ATP binding, GTP binding	gluconeogenesis
	1	10.41	succinate-semialdehyde dehydrogenase	ALDH5F1	3-chloroallyl aldehyde dehydrogenase activity, NAD binding	glutamate decarboxylation to succinate
	2	19.26	UDP-arabinose 4-epimerase	MUR4	UDP-arabinose 4-epimerase activity	galactose metabolic transduction
	2	19.26	UDP-arabinose 4-epimerase	MUR4	UDP-arabinose 4-epimerase activity	galactose metabolic transduction
	2	16.62	methylthioribose-1-phosphate isomerase	AT2G05830	GTP binding, S-methyl-5-thioribose-1-phosphate	L-methionine biosynthetic process
	2	16.21	organic cation/carnitine transporter 5	4Z280	ATP binding, carbohydrate transmembrane transporter activity	nitrate transport
	2	15.9	phosphoglycolate phosphatase 3	AT1G23190	magnesium ion binding	galactose catabolic process, gluconeogenesis
	2	15.85	Nitrilase/cyanide hydratase	AT5G12040	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds)	nitrogen compound metabolic process
	2	15.85	Pyridoxal phosphate (PLP)-dependent transferases protein	AT1G33320	cystathionine gamma-synthase activity	methionine biosynthetic process
	2	15.85	pyridoxal-5'-phosphate-dependent enzyme family protein	AT1G55880	lyase activity	cellular amino acid metabolic process
	2	14.56	pyridoxine/pyridoxamine 5'-phosphate oxidase 1	PPOX	FMN binding, NADHX epimerase activity	pyridoxal metabolic process
	2	12.96	ATP citrate lase subunit B 2	ACLB-2	ATP binding, ATP citrate synthase activity	acetyl-CoA biosynthetic process
	2	12.88	hydroxypyruvate reductase	HPR	NAD binding	cysteine biosynthetic process
	2	11.6	high mobility group protein B1	HMGB1	DNA binding, chromatin binding	glucose catabolic process
	2	11.48	transketolase	AT3G60750	metal ion binding	glucose catabolic process
	2	10.9	prenyltransferase-like protein	AT3G04300	transferase activity	glucose catabolic process
	2	10.32	serine racemase	SR	ATP binding, D-serine ammonia-lyase activity	serine family amino acid metabolic process
	2	10.32	glutamate--glyoxylate aminotransferase 2	AGT2	L-alanine:2-oxoglutarate aminotransferase activity	L-alanine catabolic process
	3	17.31	succinate-semialdehyde dehydrogenase	ALDH5F1	3-chloroallyl aldehyde dehydrogenase activity, NAD binding	glutamate decarboxylation to succinate
	3	14.44	threonine dehydratase	OMR1	L-threonine ammonia-lyase activity	gluconeogenesis
	3	14.21	protein ILITYHIA	ILA	None listed	gluconeogenesis
	3	13.54	transducin/WD40 domain-containing protein	AT5G11240	nucleotide binding	pyrimidine ribonucleoside biosynthetic process
	3	13.11	UDP-GLUCOSE PYROPHOSPHORYLASE 1	UGP1	UTP:glucose-1-phosphate uridylyltransferase activity	sucrose metabolic process
	3	10.81	UXS6 UDP-XYL synthase 6	AT2G28760	UDP-glucuronate decarboxylase activity	nucleotide sugar metabolic process
	3	10.75	VTC4 L-galactose-1-phosphate phosohatase	AT3G02870	3'(2'),5'-bisphosphate nucleotidase activity, L-galactose-1-phosphate phosphatase activ	methylglyoxal catabolic process to D-lactate
	4	17.32	UKL3 uridine kinase-like protein 3	AT1G55810	ATP binding, GTP binding	gluconeogenesis
	4	17.32	ALDH5F1 succinate-semialdehyde dehydrogenase	AT1G79440	3-chloroallyl aldehyde dehydrogenase activity, NAD binding	glutamate decarboxylation to succinate
	4	14.46	OMR1 threonine dehydratase	AT3G10050	L-threonine ammonia-lyase activity	gluconeogenesis
	4	14.23	ILA protein ILITYHIA	AT1G64790	None listed	gluconeogenesis
	4	13.59	transducin/WD40 domain-containing protein	AT5G11240	nucleotide binding	pyrimidine ribonucleoside biosynthetic process
	4	13.14	UGP1 UDP-GLUCOSE PYROPHOSPHORYLASE 1	AT3G03250	UTP:glucose-1-phosphate uridylyltransferase activity	UDP-glucose metabolic process
	5	13.12	GAD2 glutamate decarboxylase 2	AT1G65960	calmodulin binding	glutamate metabolic process
	5	12.66	NHO1 glycerol kinase	AT1G80460	ATP binding	carbohydrate metabolic process
	5	12.19	PFK7 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase	AT1G07110	6-phosphofructo-2-kinase activity, ATP binding	fructose 2,6-bisphosphate metabolic process
	5	11.2	UGE1 bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1	AT1G12780	UDP-arabinose 4-epimerase activity, UDP-glucose 4-epimerase activity	galactose metabolic transduction
	5	10.96	tyrosine aminotransferase	AT5G36160	L-phenylalanine:2-oxoglutarate aminotransferase activity	L-phenylalanine and tyrosine catabolic process
	5	10.68	GGT1 glutamate glyoxylate aminotransferase	AT1G23310	L-alanine:2-oxoglutarate aminotransferase activity	L-alanine and cysteine catabolic process
	5	10.58	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	AT2G24270	3-chloroallyl aldehyde dehydrogenase activity	cysteine biosynthetic process
	5	10.08	AOAT2 glutamate--glyoxylate aminotransferase 2	AT1G70580	L-alanine:2-oxoglutarate aminotransferase activity	L-alanine catabolic process
	5	10.08	SIP serine racemase	AT4G11540	ATP binding, D-serine ammonia-lyase activity	L-serine metabolic process
	5	10.37	RAPTOR2 regulatory-associated protein of TOR2	AT5G01770	nucleotide binding	TOR signaling
	5	10.05	UKL5 uridine kinase-like 5	AT3G27440	ATP binding, GTP binding	gluconeogenesis
	6	23.03	ETFQO electron-transfer flavoprotein--ubiquinone oxidoreductase	AT2G43400	4 iron 4 sulfur cluster binding, catalytic activity	leucine catabolic process
	8	23.82	BCAT-2 branched-chain-amino-acid aminotransferase 2	AT1G10070	L-isoleucine transaminase activity, L-leucine transaminase activity	branched-chain amino acid metabolic process
	8	23.28	probable 3-hydroxyisobutyryl-CoA hydrolase 2	AT2G30550	3-hydroxyisobutyryl-CoA hydrolase activity	valine catabolic process
DNA and RNA process	1	14.53	U2 small nuclear ribonucleoprotein A'	U2A'	RNA binding	mRNA splicing
	2	14.84	probable splicing factor 3A subunit 1	AT1G14650	RNA binding	RNA processing
	2	14.24	U2 small nuclear ribonucleoprotein B'' 2	AT1G09960	RNA binding	mRNA splicing
	2	11.34	ADP-ribosylation factor GTPase-activating protein AGD2	AGD2	GTPase activator activity	production of miRNAs involved in gene silencing by miRNA
	2	11.28	translation initiation factor eIF-4E	eIF4E3	RNA binding	translational initiation
	3	13.77	E3 ubiquitin-protein ligase ORTHRUS 1	VIM3	ligase activity, methyl-CpG binding	chromatin modification
	3	12.95	BRCA1/BRCA2-containing complex subunit 36B	AT3G06820	molecular function	DNA repair
	3	12.4	Pre-mRNA-processing-splicing factor	AT4G38780	None listed	meiotic DNA double-strand break formation
	3	12.4	putative splicing factor Prp8	SUS2	U5 snRNA binding	production of ta-siRNAs involved in RNA interference
	3	12.39	Y-family DNA polymerase H	POLH	DNA-directed DNA polymerase activity	DNA-dependent DNA replication
	4	11.14	nucleolar complex-associated protein domain-containing protein	AT1G79150	None listed	mRNA export from nucleus
	4	13.81	VIM3 E3 ubiquitin-protein ligase ORTHRUS 1	AT5G39550	ligase activity, methyl-CpG binding	chromatin modification
	4	12.97	BRCA1/BRCA2-containing complex subunit 36B	AT3G06820	molecular function	DNA repair
	4	12.42	Pre-mRNA-processing-splicing factor	AT4G38780	None listed	meiotic DNA double-strand break formation
	4	12.42	SUS2 putative splicing factor Prp8	AT1G80070	U5 snRNA binding	production of miRNAs involved in gene silencing by miRNA
	4	12.4	POLH Y-family DNA polymerase H	AT5G40740	DNA-directed DNA polymerase activity	DNA-dependent DNA replication
	4	11.48	EDA7 embryo sac development arrest 7 protein	AT3G58990	molecular function	mRNA export from nucleus
	4	11.01	high mobility group B protein 15	AT1G04880	sequence-specific DNA binding transcription factor activity	chromatin assembly or disassembly
	5	10.33	probable splicing factor 3A subunit 1	AT1G14650	RNA binding	RNA processing
Cell cycle regulation	1	14.93	splicing factor U2af small subunit A	ATU2AF35A	DNA binding, RNA binding	mitotic nuclear division
	1	14.08	cyclin-A1;1	CYCA1;1	cyclin-dependent protein serine/threonine kinase regulator activity	regulation of cell cycle
	2	10.17	anaphase-promoting complex subunit 2	APC2	protein binding	gene silencing by miRNA
	3	20.99	cyclin a2;1	CYCA2;1	cyclin-dependent protein serine/threonine kinase regulator activity	regulation of cell cycle
	3	13.46	double-strand break repair protein MRE11	MRE11	protein binding	meiotic cell cycle
	3	12.05	cell division cycle protein 27-B	HBT	protein binding	regulation of cell division
	3	11.08	SGA2 putative histone chaperone ASF1A	AT1G66740	protein binding	mitotic cell cycle control checkpoint
	4	.21	CYCB3;1 cyclin B3-1	AT1G16330	cyclin-dependent protein serine/threonine kinase activity	regulation of cell cycle
	4	12.18	HBT cell division cycle protein 27-B	AT2G32000	protein binding	regulation of cell division
	5	10.75	CYCB3;1 cyclin B3-1	AT1G34460	cyclin-dependent protein serine/threonine kinase regulator	regulation of cell cycle
Cell death	1	14.08	receptor like protein 30	RLP30	kinase activity	negative regulation of programmed cell death
	2	24.4	autophagy-related protein 18D	ATG18D	phosphatidylinositol 3,5-bisphosphate	autophagy
	2	16	cysteine protease ATCys	AT2G44140	peptidase activity	autophagy
	4	.21	RLP30 receptor like protein 30	AT3G05360	kinase activity	negative regulation of programmed cell death
	5	11.45	AtMYB103 myb domain protein 103	AT3G62770	phosphatidylinositol-3,5-bisphosphate	autophagic cell death
Defense response	1	11.97	syntaxin-121	SYP121	SNAP receptor activity	negative regulation of cellular defense response
	1	10.4	epithiospecifier protein	ESP	enzyme regulator activity	defense response to bacterium
	2	18.94	putative flavin monooxygenase	AT1G12200	N,N-dimethylaniline monooxygenase	defense response to fungus
	3	20.99	receptor like protein 9	RLP9	None listed	response to molecule of bacterial origin
	3	18.18	syntaxin-121	SYP121	SNAP receptor activity	negative regulation of cellular defense response
	3	14.73	translation initiation factor 4E-2	LSP1	RNA 7-methylguanosine cap binding, RNA binding	negative regulation of defense response to virus
	4	18.18	SYP121 syntaxin-121	AT3G11820	SNAP receptor activity	negative regulation of cellular defense response
	4	13.05	SFTASE11 Type I inositol-1,4,5-trisphosphate 5-phosphatase 11	AT1G47510	hydrolase activity, inositol-1,3,4,5-tetrakisphosphate 5-phosphatase activity	response to jasmonic acid
	5	17.24	putative TIR-NBS-LRR class disease resistance protein	AT5G17680	ADP binding	defense response, signal transduction
Ion metabolism	2	13.06	calcineurin B-like protein 7	CBL7	calcium ion binding	calcium-mediated signaling
	3	11.99	cation efflux family protein	AT2G04620	cation transmembrane transporter	cation transport
	4	12.03	cation efflux family protein	AT2G04620	cation transmembrane transporter	cation transmembrane transport
	4	10.78	QSOX1 quiescin-sulfhydryl oxidase 1	AT1B15020	thiol oxidase activity	positive regulation of potassium ion transport
	5	11.55	QSOX1 quiescin-sulfhydryl oxidase 1	AT1G15020	thiol oxidase activity	positive regulation of potassium ion transport
Lipid metabolism	2	10.32	glutamate glyoxylate aminotransferase	GGT1	L-alanine:2-oxoglutarate aminotransferase activity	cellular lipid catabolic process
	3	11.62	RNA polymerase Rpb7 N-terminal domain-containg protein	AT1G06790	DNA-directed RNA polymerase activity	fatty acid catabolic process
	4	13.69	KCR2 beta-ketoacyl reductase 2	AT1G24470	acetoacetyl-CoA reductase activity	sphingolid biosynthetic process
	4	11.04	RNA polymerase Rpb7 N-terminal domain-containg protein	AT1G06790	DNA-directed RNA polymerase activity	fatty acid catabolic process
	5	10.76	SR-beta domain-containing protein	AT2G18770	signal recognition particle binding	sterol biosynthetic process
	6	25.29	ACX3 acyl-coenzyme A oxidase 3	AT1G06290	acyl-CoA dehydrogenase activity	fatty acid beta-oxidation

Supplementary Table 7

Source		*C. variabilis* NC64A		*C. variabilis* Syngen 2-3		*C. heliozoae* SAG 3.83	
Nitrogen	Carbon	Doubling time (days)	Growth rate	Doubling time (days)	Growth rate	Doubling time (days)	Growth rate
Peptone	Sucrose	1.57	0.44	1.39	0.50	1.22	0.57
	-	1.78	0.39	1.20	0.58	1.51	0.46
Urea	Sucrose	0.92	0.76	1.27	0.54	0.99	0.70
	Glucose	0.90	0.77	1.25	0.55	0.54	1.28
	Galactose	1.00	0.69	1.29	0.54	0.98	0.70
Asparagine	-	1.43	0.49	1.49	0.46	1.64	0.42
	Sucrose	1.01	0.69	1.31	0.53	0.72	0.96
	Glucose	0.93	0.74	1.33	0.52	0.68	1.01
	Galactose	0.80	0.87	1.54	0.45	0.77	0.90
-N		5.22	0.13	9.43	0.07	8.39	0.08

APPENDIX

Appendix Figure Legends

Appendix 1. Systematic salt removal using BBM and FES media on symbiotic NC64A and SAG 3.83 cells. Media is supplemented with galactose (10 mM) and urea (10 mM). Growth was evaluated after 9 days using the visual scale shown in left.

Appendix 2. pH effect in NC64A growth on minimal defined FES and BMM medium. Growth was evaluated after 15 days using the visual scale shown in left.

Appendix 3. Light microscopy observations of SAG 3.83 cells growing on various nitrogen and carbon sources under white and UV light at 3 and 7 days.

Appendix 4. Light microscopy observations of Syngen 2-3 cells growing on various nitrogen and carbon sources under white and UV light at 3, 7, 9, and 11 days.

Appendix 5. Light microscopy observations of NC64A cells growing on various nitrogen and carbon sources under white and UV light at 4 and 7 days.

Appendix 6. Experimental flow chart to test colony induction on symbiotic *Chlorella* species.

Appendix 7. Seventy days old SAG 3.83 colonies formed on various carbon and nitrogen sources.

Appendix 8. SAG 3.83 colonies formed on various carbon and nitrogen sources with a range of starting cell concentrations.

Appendix 9. Experimental flow for high-throughput solid media nutritional screening of *Chlorella* species.

Appendix 10. Growth of 9 *Chlorella* species growing on solid asparagine, urea, and glucosamine minimal media.

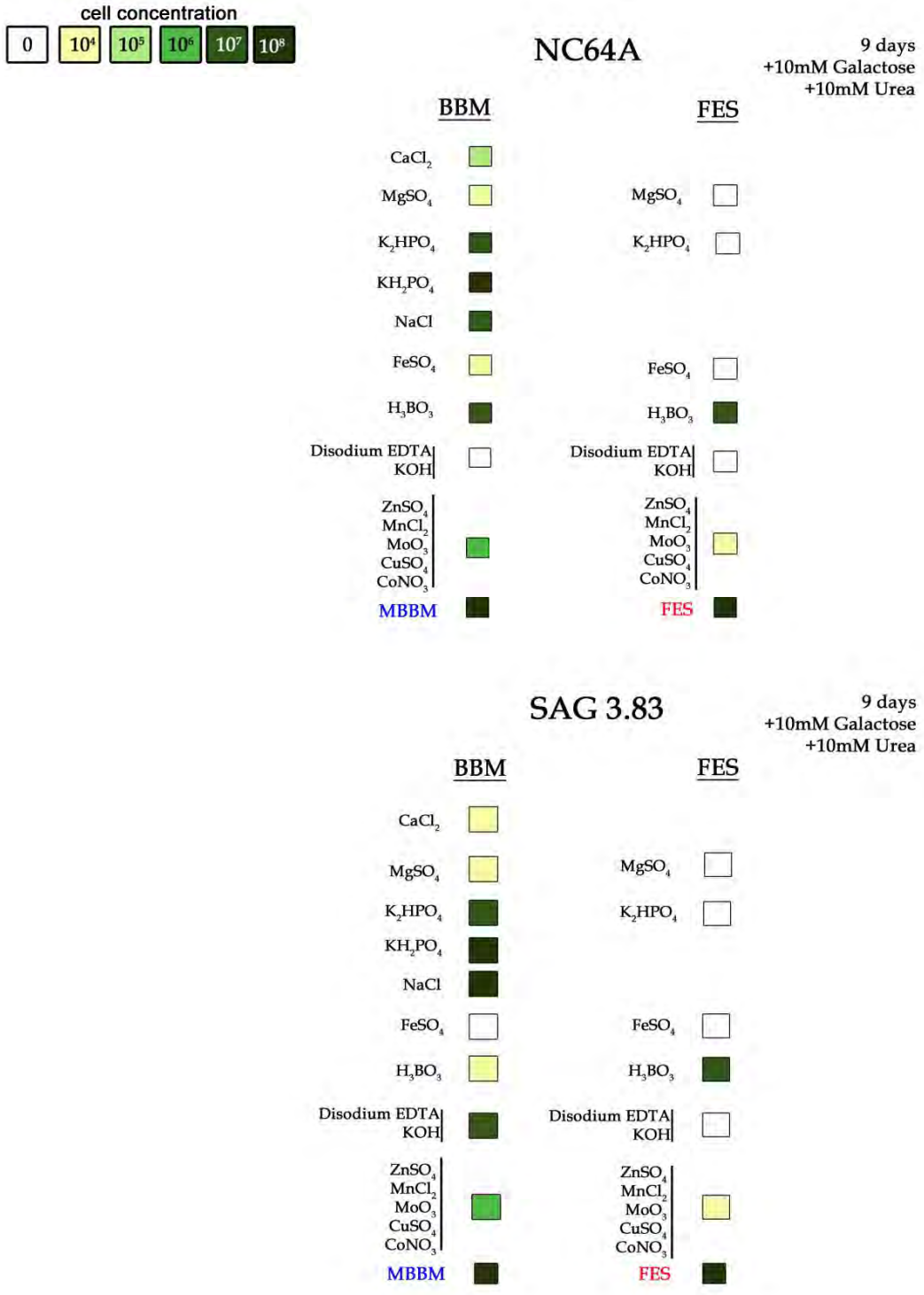
Appendix 11. Pictures representing the range of nitrogen and carbon assimilation in the high-throughput solid media nutritional screening with nine *Chlorella* species.

Appendix 12. Pulsed field gel electrophoresis of genomic DNAs from SAG 3.83, F36-ZK, Syngen 2-3, NC64A, OK1-ZK and *C. sorokoniana* (UTEX 1230). Markers and running conditions labeled.

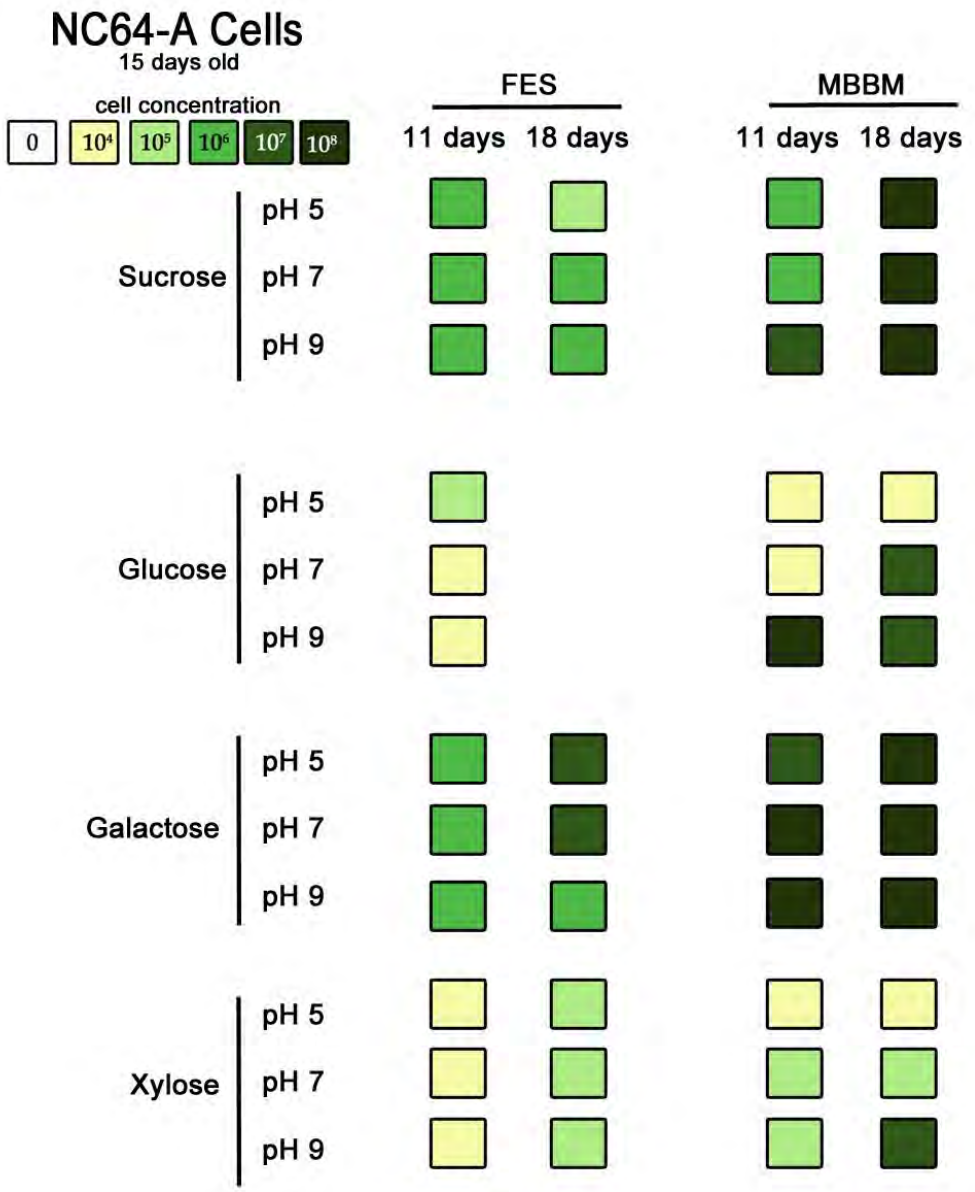
Appendix 13. Full genome alignment of symbiotic *C. variabilis* NC64A and free-living *C. sorokoniana* 1230.

Appendix 14. Comparison of chloroplast and mitochondrial genomes of *C. variabilis* NC64A and *C. sorokiniana* (UTEX 1230).

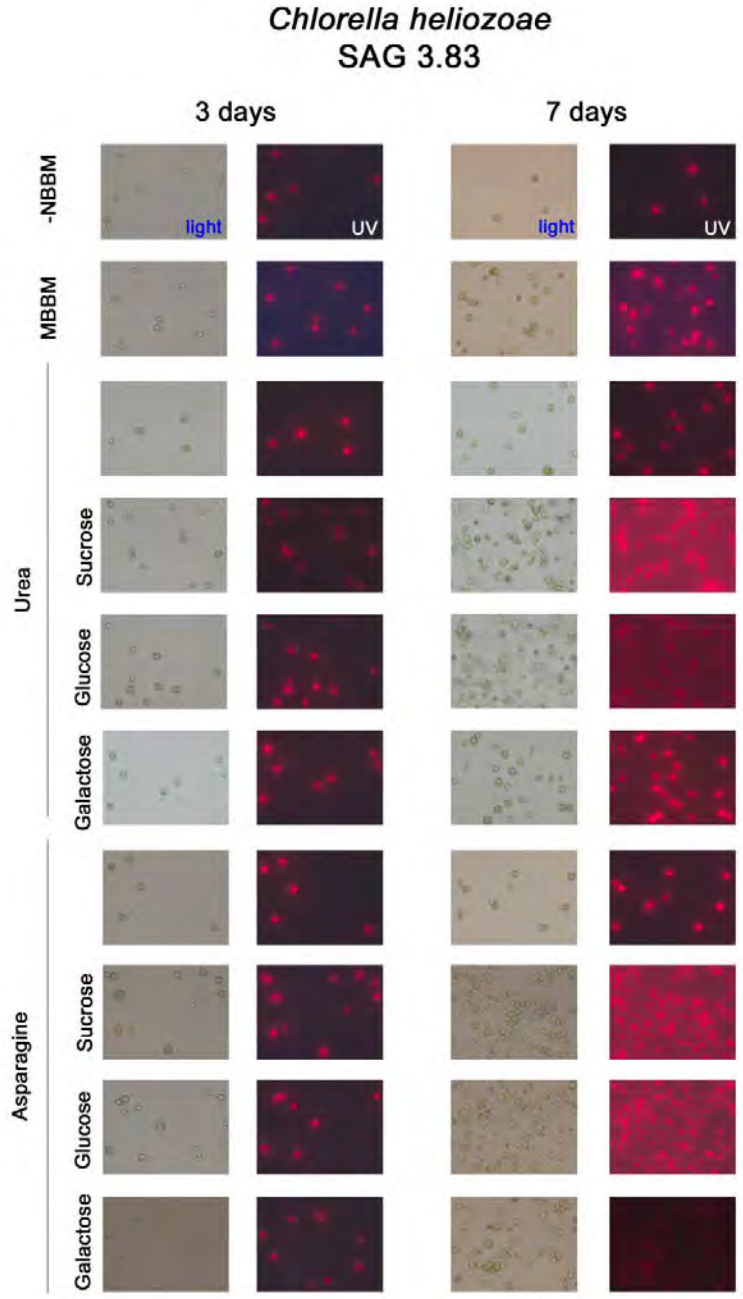
Appendix 1


cell concentration
0
10⁴
10⁵
10⁶
10⁷
10⁸
NC64A
9 days
+10mM Galactose
+10mM Urea
BBM
FES
CaCl2
MgSO4
K2HPO4
KH2PO4
NaCl
FeSO4
H3BO3
Disodium EDTA
KOH
ZnSO4
MnCl2
MoO3
CuSO4
CoNO3
MBBM
FES
SAG 3.83
9 days
+10mM Galactose
+10mM Urea
BBM
FES

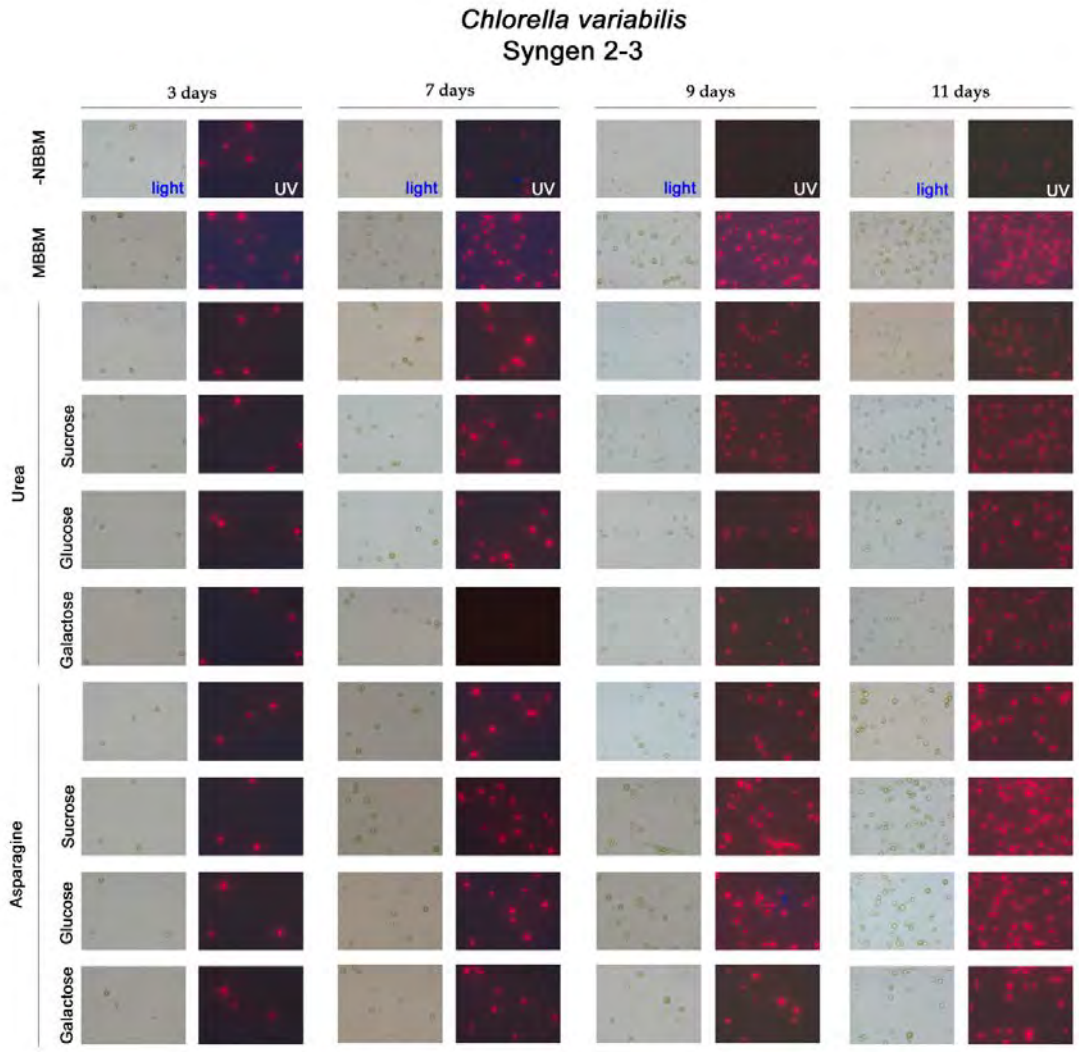
Appendix 2


NC64-A Cells
15 days old
cell concentration
0
10⁴
10⁵
10⁶
10⁷
10⁸
FES
MBBM
11 days
18 days
11 days
18 days
Sucrose
pH 5
pH 7
pH 9
Glucose
pH 5
pH 7
pH 9
Galactose
pH 5
pH 7
pH 9
Xylose
pH 5
pH 7
pH 9

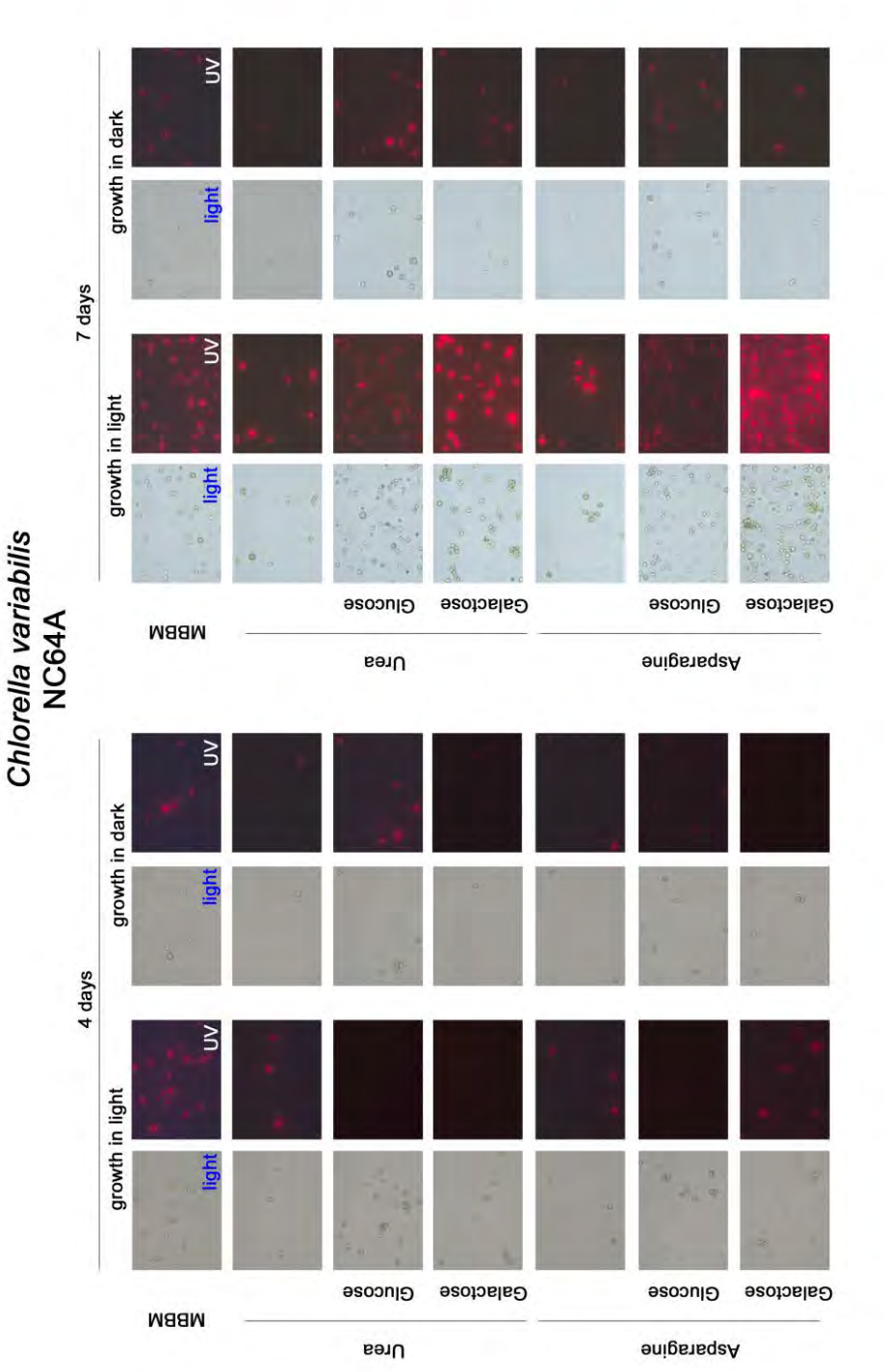
Appendix 3


Chlorella heliozoae
SAG 3.83
3 days
7 days
light
UV
light
UV
-NBBM
MBBM
Urea
Sucrose
Glucose
Galactose
Asparagine
Sucrose
Glucose
Galactose

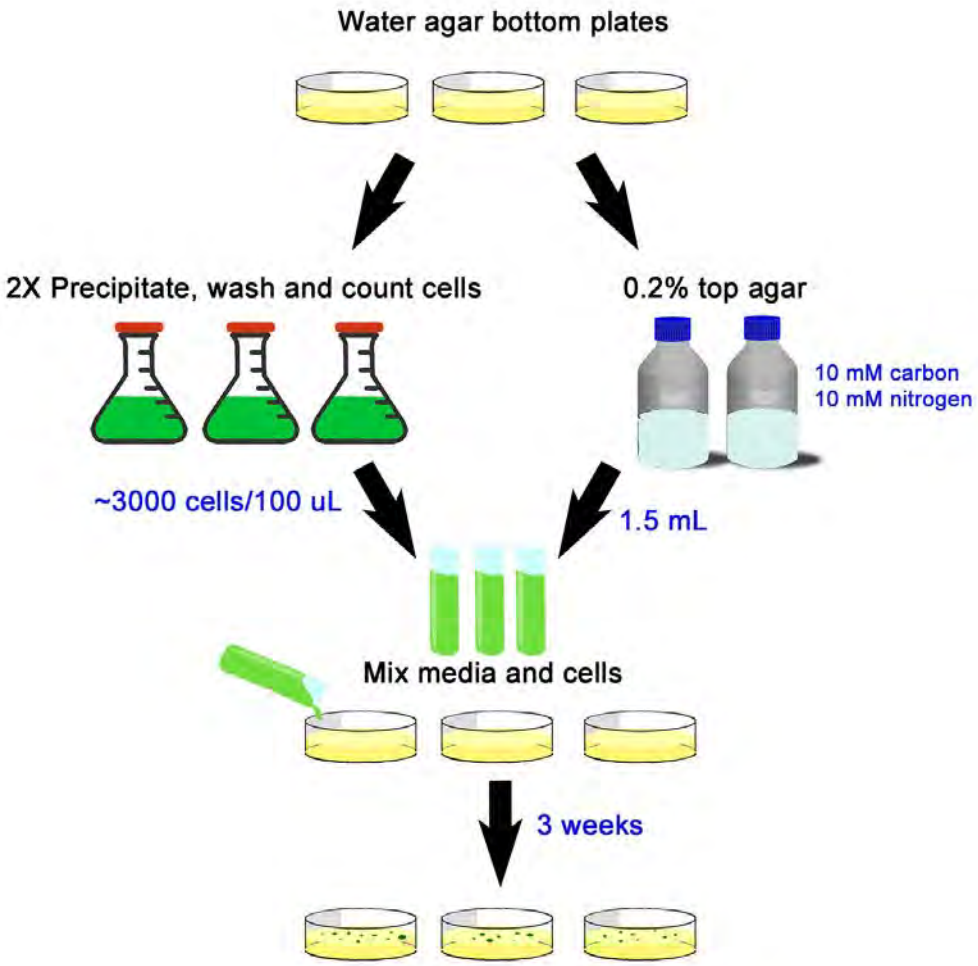
Appendix 4


Chlorella variabilis
Syngen 2-3
3 days
7 days
9 days
11 days
light
UV
-NBBM
MBBM
Urea
Sucrose
Glucose
Galactose
Asparagine
Sucrose
Glucose
Galactose

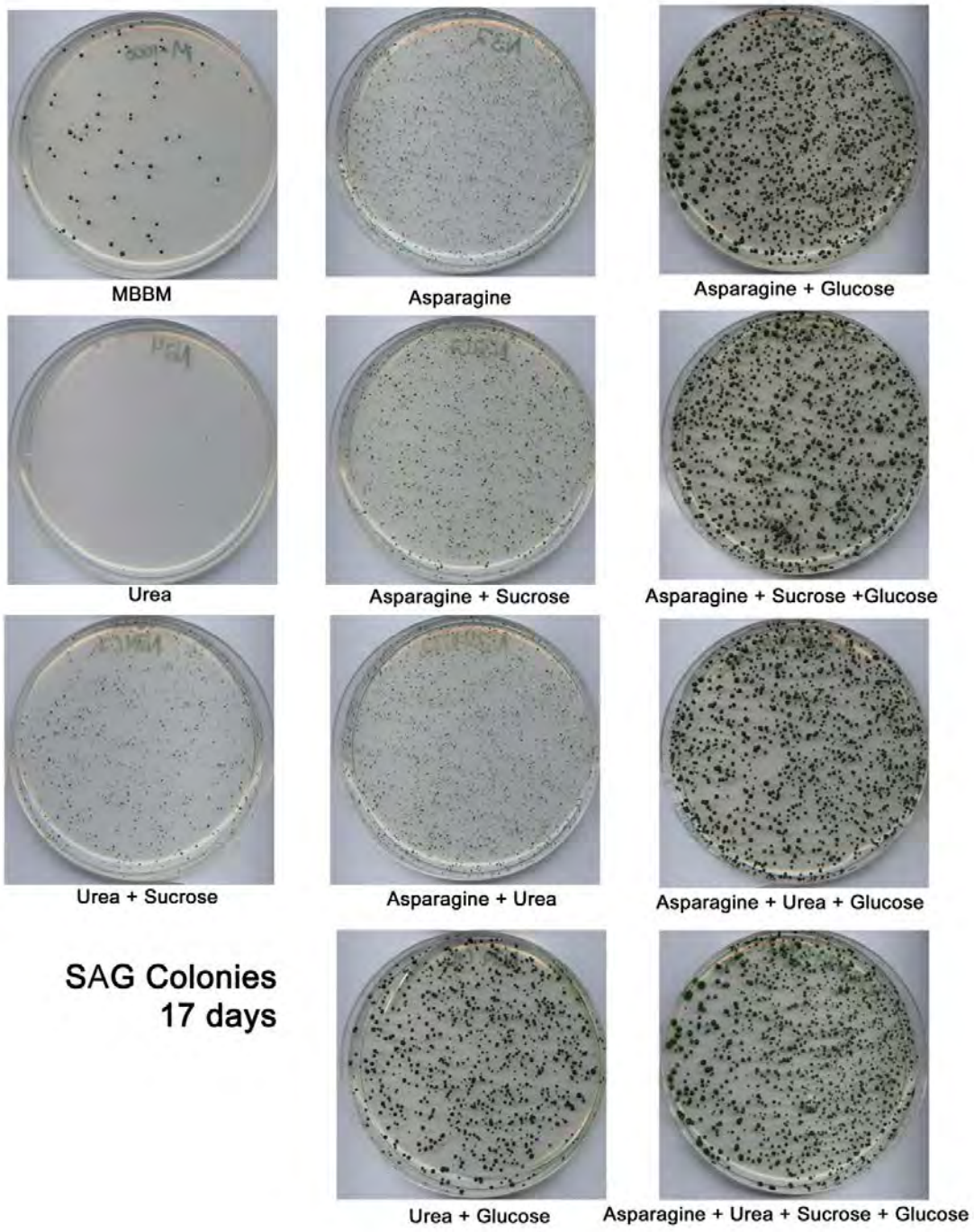
Appendix 5


Chlorella variabilis
NC64A
4 days
7 days
growth in light
growth in dark
light
UV
MBBM
Urea
Asparagine
Galactose
Glucose

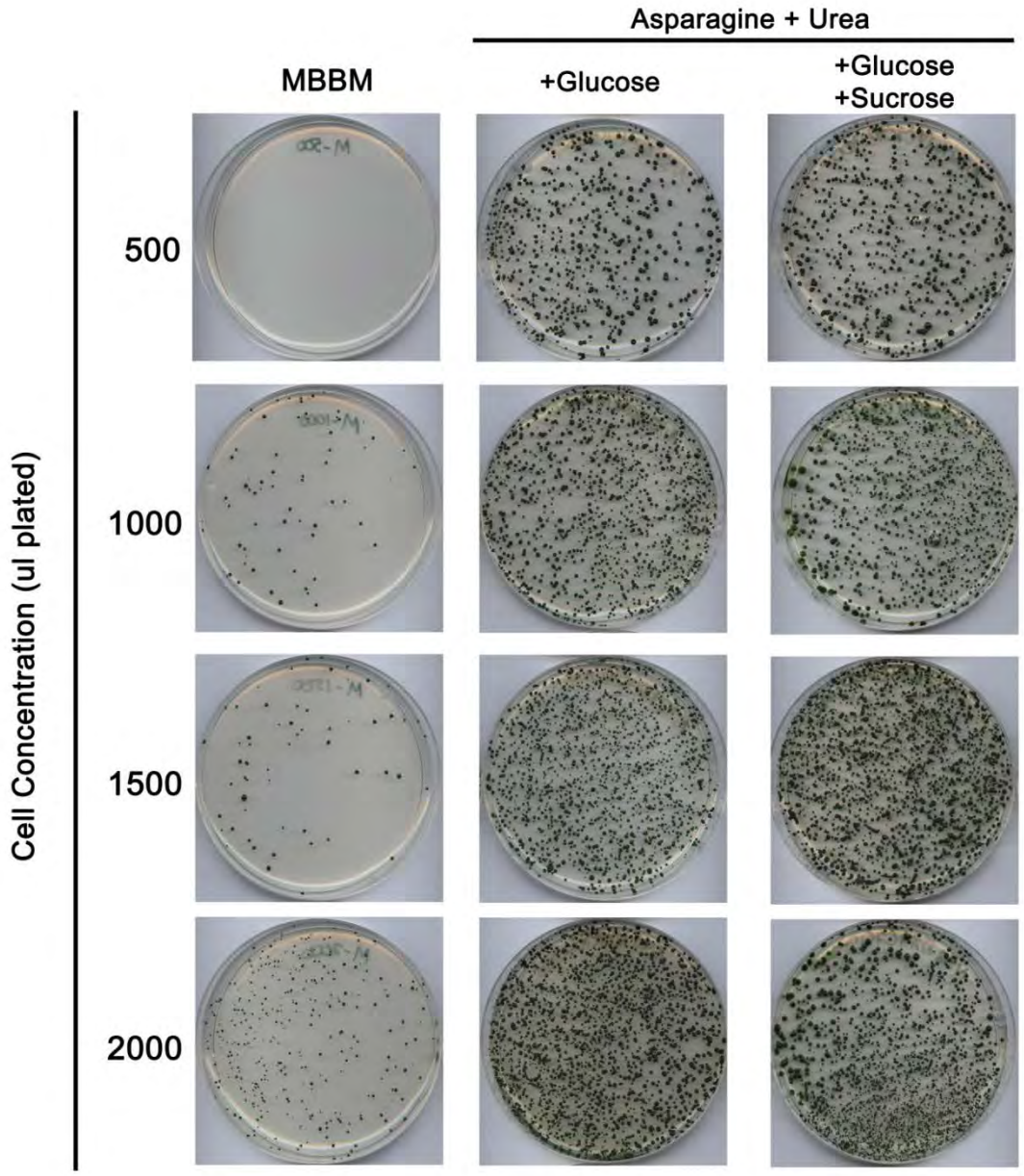
Appendix 6


Water agar bottom plates
2X Precipitate, wash and count cells
0.2% top agar
10 mM carbon
10 mM nitrogen
~3000 cells/100 uL
1.5 mL
Mix media and cells
3 weeks

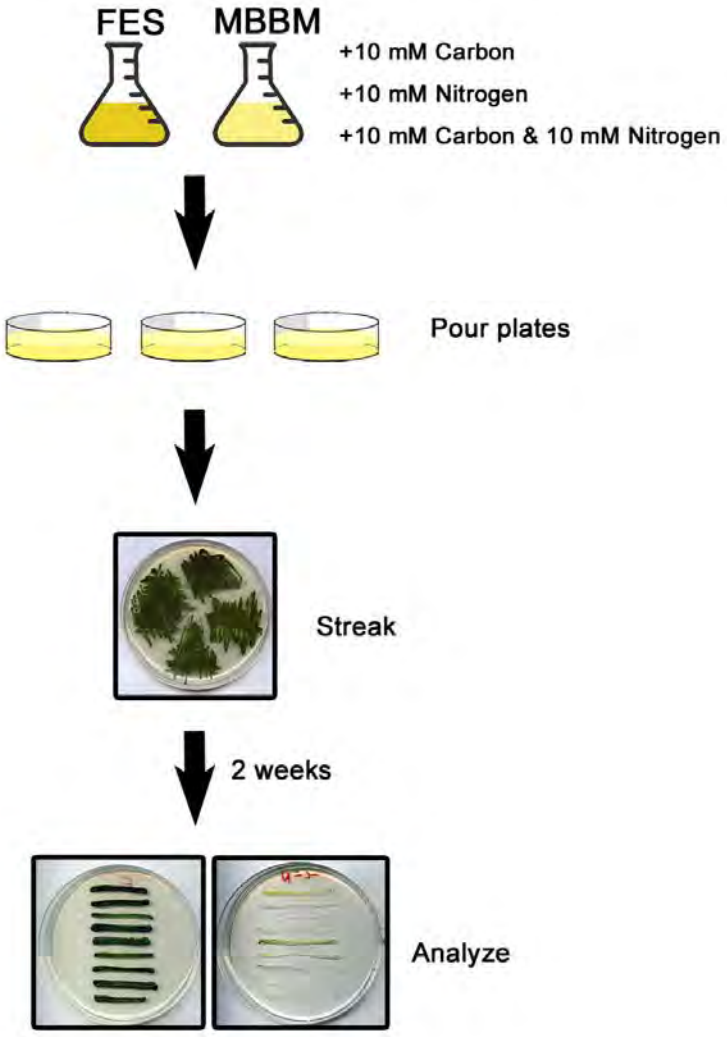
Appendix 7


MBBM
Asparagine
Asparagine + Glucose
Urea
Asparagine + Sucrose
Asparagine + Sucrose +Glucose
Urea + Sucrose
Asparagine + Urea
Asparagine + Urea + Glucose
SAG Colonies
17 days
Urea + Glucose
Asparagine + Urea + Sucrose + Glucose

Appendix 8

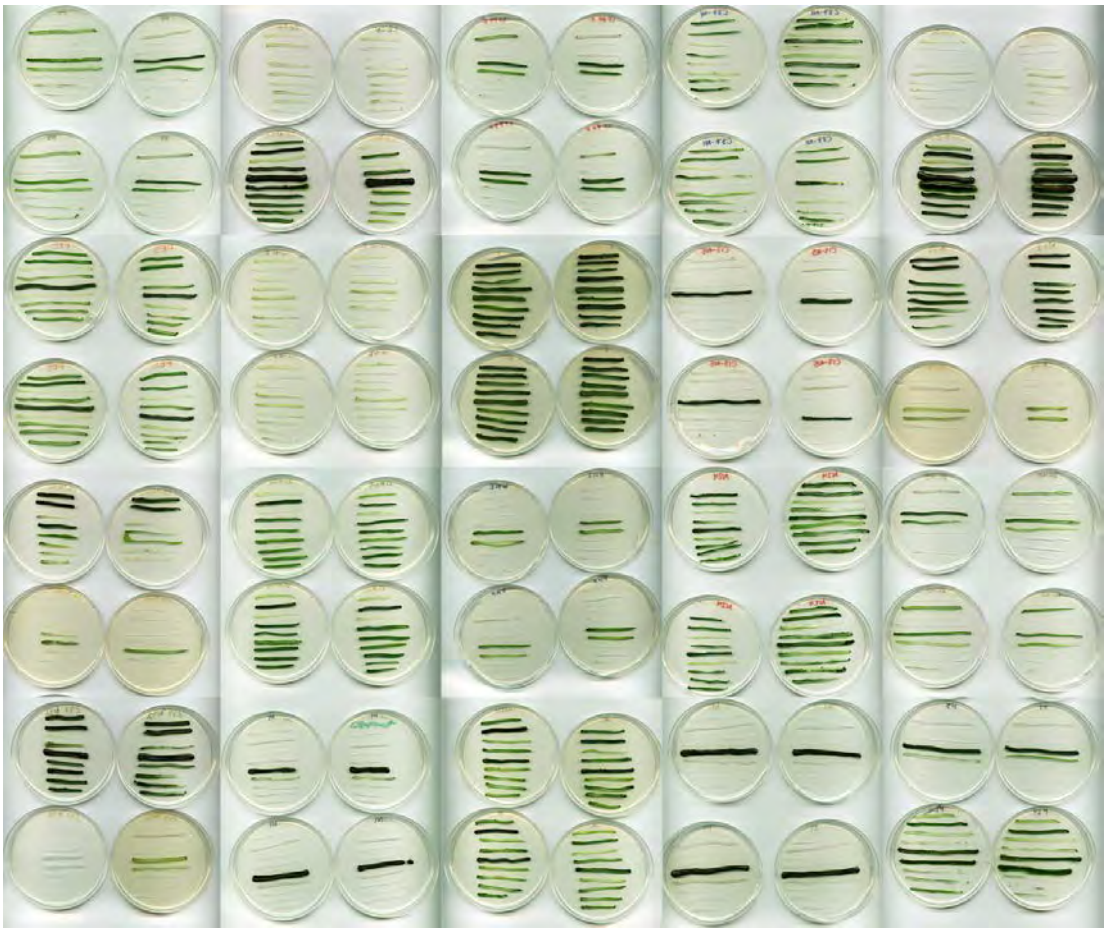

Asparagine + Urea
MBBM
+Glucose
+Glucose
+Sucrose
Cell Concentration (ul plated)
500
1000
1500
2000

Appendix 9

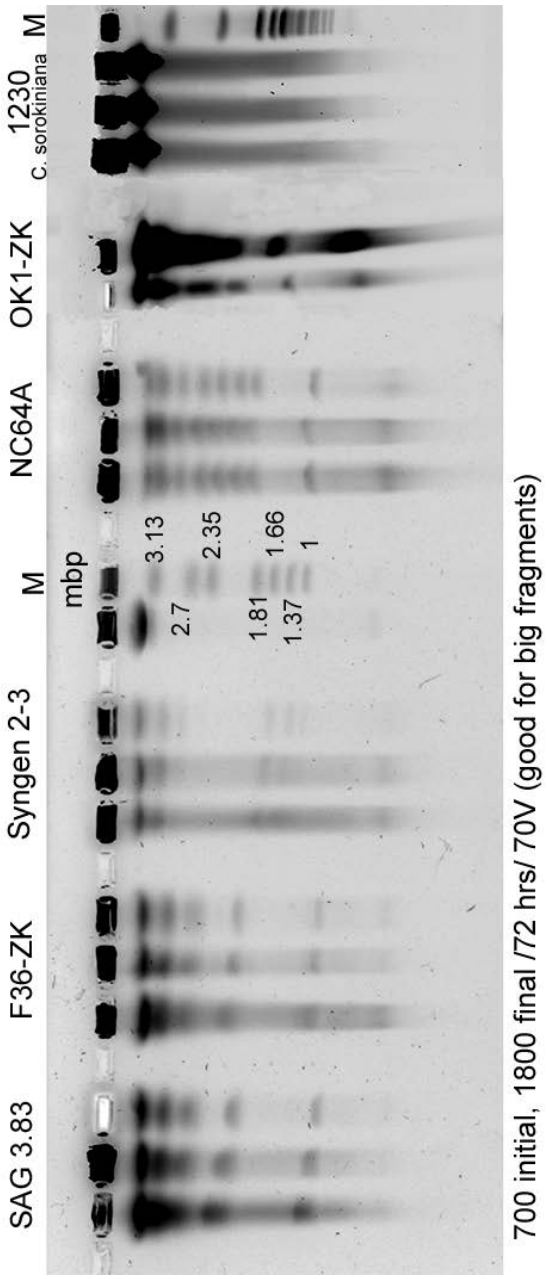

FES
MBBM
+10 mM Carbon
+10 mM Nitrogen
+10 mM Carbon & 10 mM Nitrogen
Pour plates
Streak
2 weeks
Analyze

Appendix 10

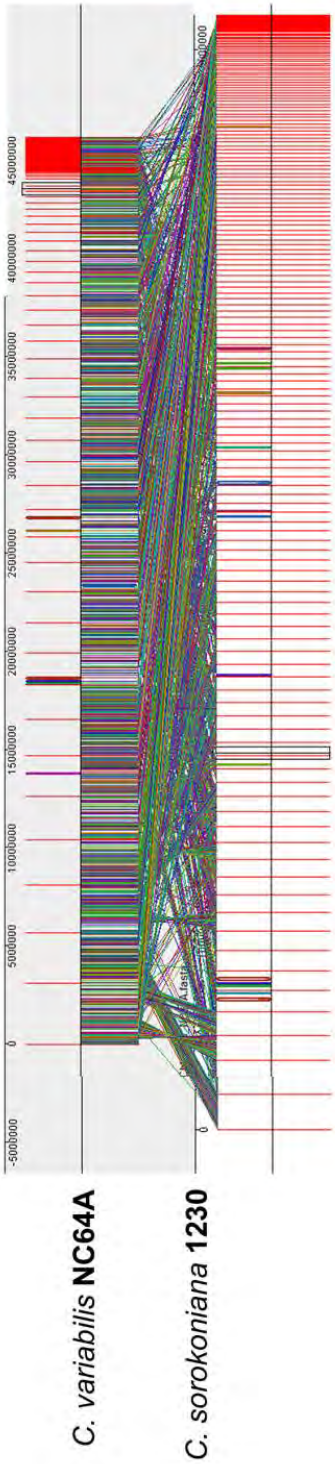
Appendix 11



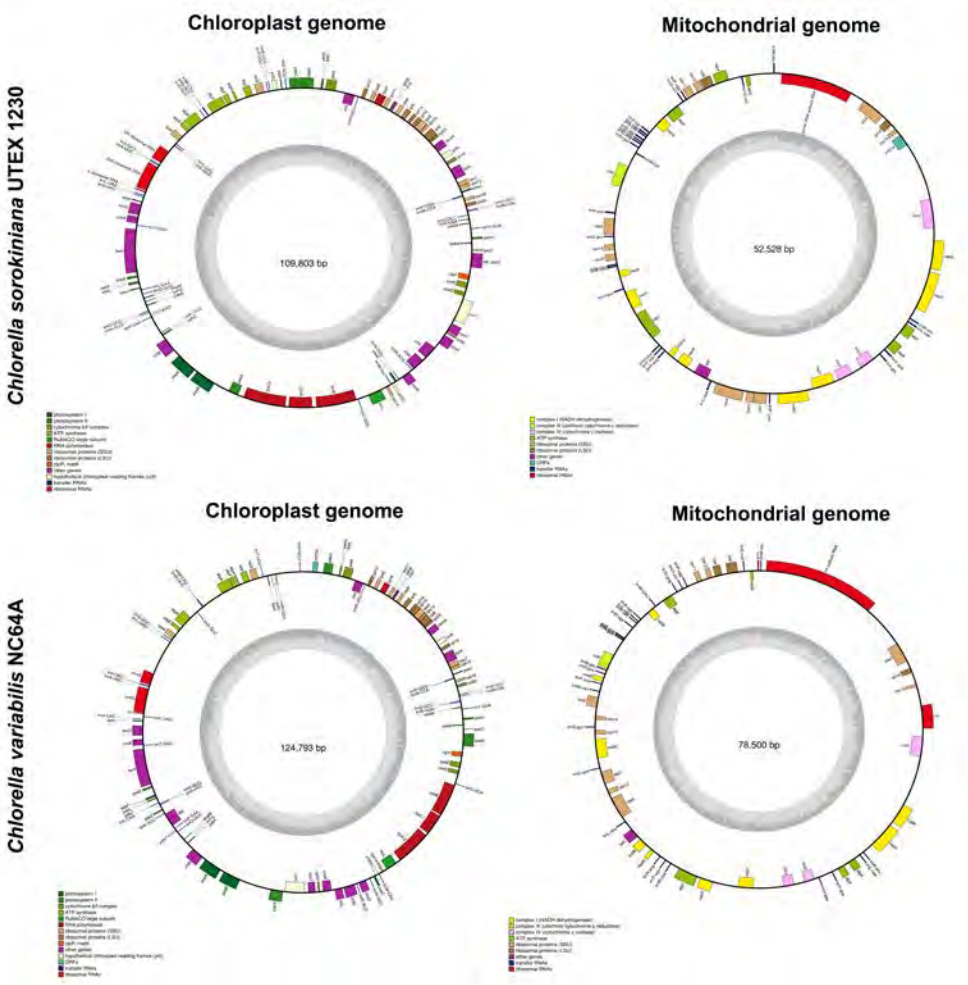
Appendix 12


SAG 3.83
F36-ZK
Syngen 2-3
M
NC64A
OK1-ZK
1230
C. sorokiniana
M
mbp
3.13
2.7
2.35
1.81
1.66
1.37
1
700 initial, 1800 final /72 hrs/ 70V (good for big fragments)
1% 0.5 TBE gel

Appendix 13


45000000
40000000
35000000
30000000
25000000
20000000
15000000
10000000
5000000
0
-500000
C. variabilis NC64A
C. sorokoniana 1230

Appendix 14


Chloroplast genome
Mitochondrial genome
Chlorella sorokiniana UTEX 1230
109,803 bp
52,528 bp
Chloroplast genome
Mitochondrial genome
Chlorella variabilis NC64A
124,793 bp
78,500 bp