AN EARLY INSIGHT INTO THE BIOLOGY AND VIRULENCE TRAITS OF THE ALGAL GENUS *PROTOTHECA*

by

CHRISTOPHER DAVID SHAVE

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection
School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
September 2023

Abstract

The aim of this thesis is to provide a deeper understanding of the basis of pathology in the genus *Prototheca*, an enigmatic group of pathogenic algae known to infect humans, cattle, and other mammalian hosts. These infections represent burdens on public health, as a result of chronic infection and inefficient treatment, and economic output, as a result of reduced milk production. The mechanisms that allow algae to infect mammalian hosts, and their interactions with mammalian immunity, are currently unknown.

Prototheca is also taxonomically complex, and pathogenic species are distantly related within the genus. Additionally, *Prototheca* is paraphyletic. The monophyletic clade containing all *Prototheca* species must also include at least the genera *Auxenochlorella* and *Helicosporidium*, forming the AHP lineage.

Much of the existing work on *Prototheca* is taxonomically restricted, covering a handful of species and largely ignoring the larger AHP lineage. This makes it difficult to meaningfully identify features important for the pathology of *Prototheca* species, or whether pathogenic AHP species share a common basis for pathology. Thus, this thesis takes a broader view, looking at the wider AHP lineage rather than one or two species. This thesis also considers multiple aspects of pathology, considering that pathology in *Prototheca* species may be accidental rather than as a result of specific adaptation to infection.

The methods and results of this thesis are varied. A method based on phenol DNA extraction was developed to detect *Prototheca* in milk, and was used to detect the first reported cases of bovine protothecosis from the UK. A method based on growth curves identified that ability to grow at 37°C separates known pathogenic AHP species from non-pathogenic AHP species. Exposure of AHP species to macrophages revealed that capacity for phagocytosis varies between the AHP species, without an obvious link to pathology. Genome sequencing produced complete organelle genomes for members of the AHP lineage and revealed that *Prototheca* have lost genes core to the Chlorophyta lineage. Finally, phylogenetics have clarified several relationships within the AHP lineage, revealing two major sub-lineages of the AHP lineage and up to seven boundaries for new genera.

This thesis provides a theoretical and experimental basis for future investigation, allowing for specific questions to be asked of the AHP lineage and more valid comparisons to be made within more closely related members of the lineage.

Acknowledgements

There are too many individuals to thank by name and acknowledge their contribution to this thesis without turning this section into a chapter of its own. Thus, I will have to express my gratitude to general groups with broad statements.

Thanks to my supervisors, mentors, and teachers during my PhD across the University of Birmingham, University of Warwick, and internships at the Association of the British Pharmaceutical Industry and MicrobesNG for their guidance, advice, and perspective over the last four years without which I could not have finished my PhD. Additional thanks to the supervisors, mentors, and teachers from before my PhD, without whom I could not have started my PhD.

Thanks to my students for helping me understand and internalise my own personal development over the last four years. Thanks to my lab mates, friends, and family for their mental and emotional support, words of wisdom, and (seemingly) interested ears as I spoke for hours at a time about algae and/or infection. Special thanks to Kai Briscoe for their constant support and encouragement.

Finally, thanks to all my collaborators for providing the many strains required for the broad scope of this thesis. Thanks to all the support and administrative staff at the University of Birmingham and the University of Warwick, who kept everything running. And thanks to the Midlands Integrative Biosciences Training Programme for providing funding (Grant number

BB/M01116X/1), adding extra universities to my CV, and for providing the opportunities to become more well-rounded as a scientist.

Table of Contents

Figures and Tables

Figures

Tables

Appendices

Abbreviations

3N-BBM+V	Bold Basal Medium with 3-fold Nitrogen and Vitamins
5FC	5-fluorocytosine
95% CI	95% confidence interval
AHP (lineage)	*Auxenochlorella*, *Helicosporidium*, *Prototheca*
AIDS	acquired immunodeficiency syndrome
ANOVA	analysis of variance
AP (lineage)	*Auxenochlorella*, *Prototheca*
ATCC	American Type Culture Collection
BMDM	bone marrow derived macrophage
CCAP	Culture Collection of Algae and Protozoa
CFU	colony forming unit
CHAMP (lineage)	*Chlorella*, *Helicosporidium*, *Auxenochlorella*, *Micractinium*, *Prototheca*
CHAP (lineage)	*Chlorella*, *Helicosporidium*, *Auxenochlorella*, *Prototheca*
CTAB	cetyltrimethylammonium bromide
EB	elution buffer
EDTA	ethylenediaminetetraacetic acid
GO	gene ontology
GWAS	genome-wide association studies
HAPI (lab)	Host and Pathogen Interaction lab
HIV	human immunodeficiency virus
HMW	high molecular weight
HP (lineage)	*Helicosporidium*, *Prototheca*
HPC	high performance computing

IUPAC	International Union of Pure and Applied Chemistry
LTR	lysotracker red
MAMP	microbe-associated molecular pattern
ML	maximum likelihood
MOI	multiplicity of infection
Na-PP	modified Protease Peptone Medium
NCBI	National Center for Biotechnology Information
NIB	nuclear isolation buffer
NML	National Milk Laboratories
OD	optical density
PAMP	pathogen-associated molecular pattern
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCR-RFLP	polymerase chain reaction-restriction fragment length polymorphism
PFGE	pulsed-field gel electrophoresis
PIM	*Prototheca* isolation medium
PMA	phorbol 12-myristate 13-acetate
PMN	polymorphonuclear
qPCR	quantitative polymerise chain reaction
rDNA	ribosomal DNA
RFLP	restriction fragment length polymorphism
RNS	reactive nitrogen species
ROS	reactive oxygen species
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase

SAB	Sabouraud dextrose
SAR (lineage)	Stramenopiles, Alveolates, Rhizaria
SDS	sodium dodecyl sulphate
UK	United Kingdom
YPD	yeast, peptone, dextrose

1) Overall Introduction

This thesis focuses on the pathology of the members of the *Prototheca* genus, and their relatives. Unusually for a thesis focused on mammalian pathogens, *Prototheca* is a genus of green algae. The overlap between phycology and medical microbiology is minimal. Thus, some theory regarding the evolution and mechanics of pathology in general as well as an introduction to algae is necessary before the genus in question can be addressed.

Theory of Pathology

Defining a Pathogen

Simplistically, a pathogen is an organism that can infect a host at the expense of the host. That expense, which is usually manifest as damage, is recognised by humans as disease (Pirofski and Casadevall, 2012; Balloux and van Dorp, 2017). The extent of damage associated with pathogen is considered its virulence (Balloux and van Dorp, 2017; Brown et al., 2006). For simplicity, this thesis will assume that pathogens are microorganisms, as the organisms investigated are all microscopic. This is not necessarily true; helminths are parasitic worms, some of which are non-microscopic, that invade a host and cause damage – meaning they are acting like pathogens (Hotez et al., 2008). However, some other parasitic relationships, including lice and the larvae of parasitoid wasps with their respective hosts, that also colonise a host and cause damage would not generally be considered pathogenic. The boundary at which a host-pathogen relationship transitions into other exploitative relationships is not crucial to the investigation of the *Prototheca* genus, and will therefore not be considered further.

Pathogens can be classified in several ways. Often, those that are able to cause pathology in healthy hosts are considered primary pathogens (Méthot and Alizon, 2014) and those that must do so to replicate are called obligate pathogens (Balloux and van Dorp, 2017). These terms are often used interchangeably, and in contrast to an opportunistic or facultative pathogen (also sometimes used interchangeably), which does not require a host to complete its life cycle and is unable to do so unless the host has a weakness that can be exploited (Méthot and Alizon, 2014; Balloux and van Dorp, 2017). Of course, requiring infection in a host is not the same as the ability to infect a host – so we might consider that the aforementioned terms describe two axes that pathogens may be found upon. A pathogen that requires infection might be considered an obligate pathogen, in contrast to a facultative pathogen; a pathogen that can infect a healthy host might be considered a primary pathogen, in contrast to a secondary pathogen (Figure 1-1).

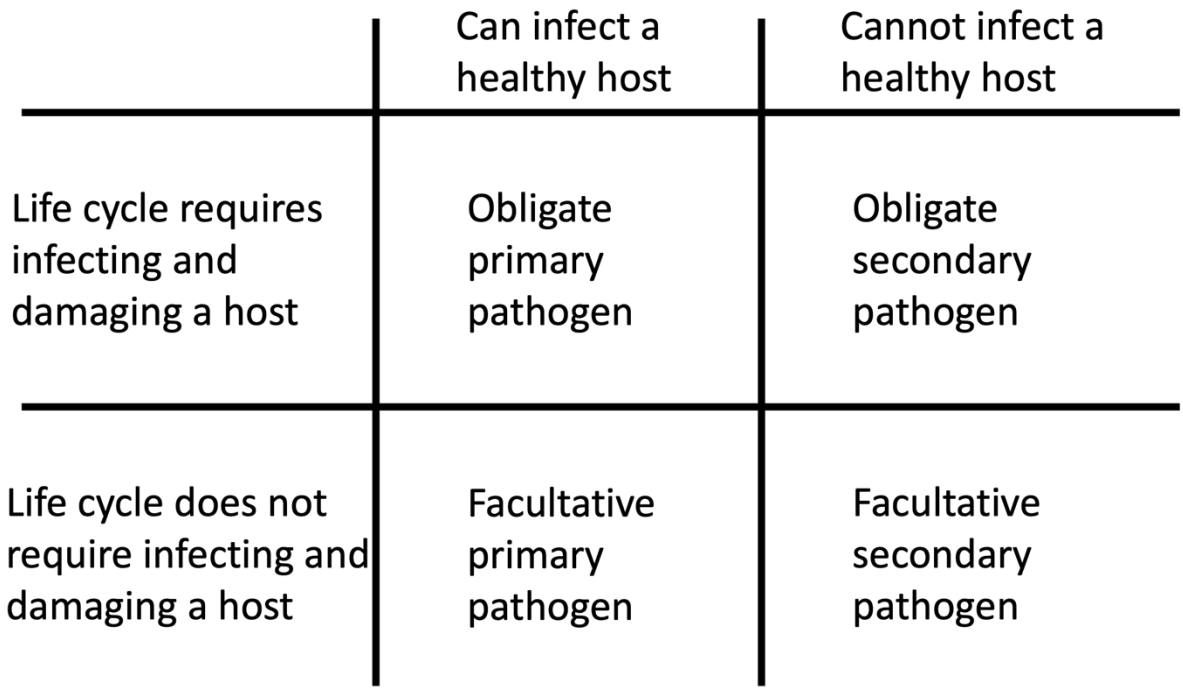


Figure 1-1. Illustration of the two axes of pathology. The terms "obligate" or "primary" are often used interchangeably and in contrast to "facultative" or "opportunistic" pathogens, despite these terms not being equivalent. The former usually refers to what would be considered an obligate primary pathogen in this diagram, but may occasionally include obligate secondary pathogens or facultative primary pathogens. Likewise, the latter usually refers to facultative secondary pathogens, but may occasionally include obligate secondary pathogens or facultative primary pathogens.

There are controversies regarding the definition of a pathogen. For example, the ability to infect a susceptible host is not equivalent to the necessity of doing so. Additionally, it is not clear what exactly should count as a susceptible host (Pirofski and Casadevall, 2012). Some have suggested that we abandon the term "pathogen", as the ability of some microorganisms to cause disease depends on the host status rather than characteristics of the microorganism (Casadevall and Pirofski, 2014). Instead, we are moving towards a paradigm in which pathology is a trait that is emergent from the interaction between the microbe and host, rather than as a result of trait acquisition in the microbe (Pirofski and Casadevall, 2012; Casadevall and Pirofski, 2001).

However, even if "pathogen" is not a clearly defined, essential characteristic of a microorganism, it is still a role that can be played by one.

Challenges Faced by Pathogens

If a microorganism is to include the role of pathogen as a regular part of its life history, no matter how small, we might expect four key steps in its infection process. These four steps are: entry into the host, surviving the host's immune response, growth within the host, and escape either into a new host or the environment (Köhler et al., 2015; Brown et al., 2012). These challenges are faced again on the level of an individual host cell, if a pathogen's life cycle includes an intracellular phase (Casadevall, 2008).

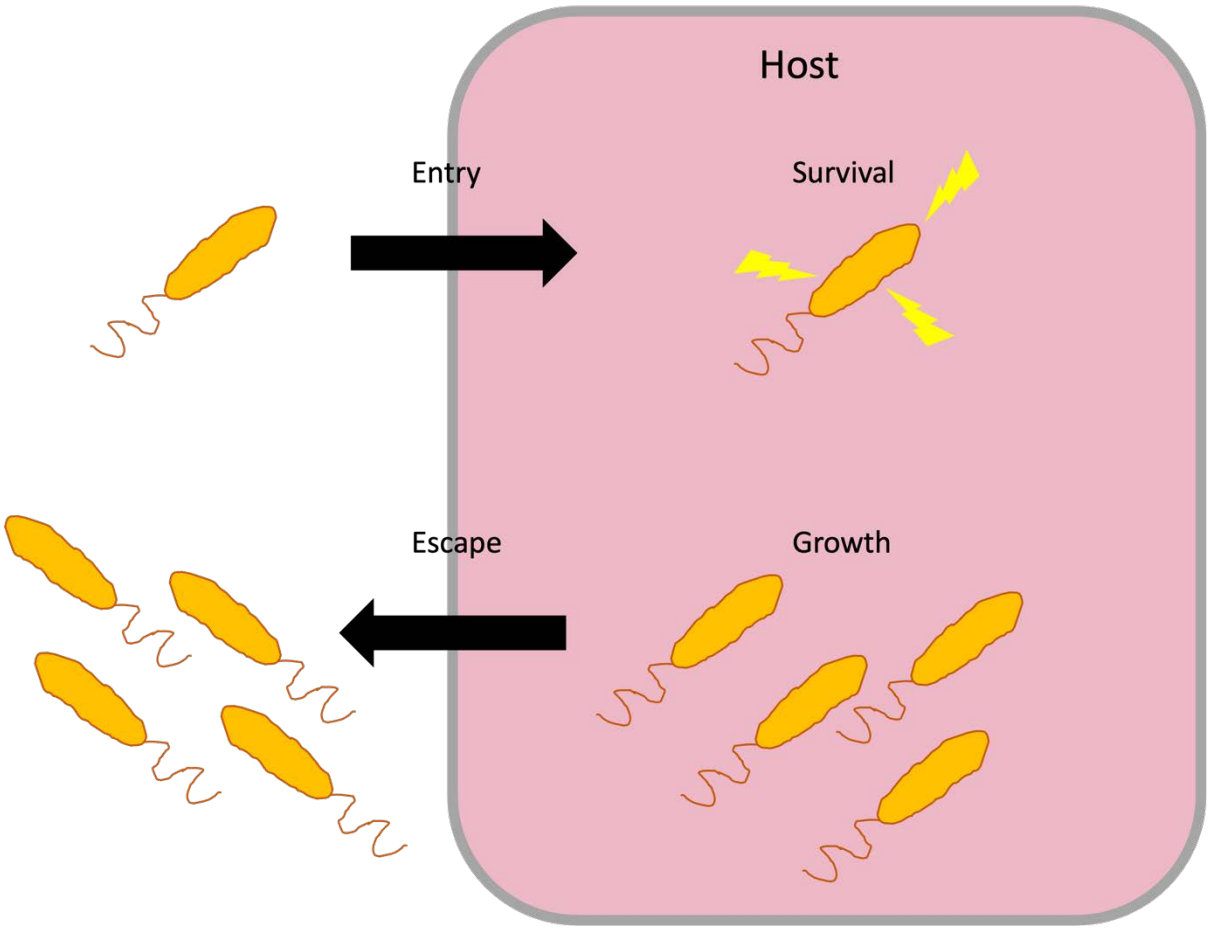


Figure 1-2. Illustration of the four essential processes for a pathogen during infection. The space from which pathogens enter and to which they escape is not defined; this space may be a non-host environment or (particularly in the case of obligate pathogens) another host. The processes of growth and survival are presented as two separate steps, but are in reality happening concurrently, likely for the entire duration the pathogen exists within the host. Pathogens are represented here by orange bacteria, but these steps apply to all pathogens. Host immunity is represented by yellow lightning, which includes all components of innate and adaptive immunity.

To resolve these challenges, pathogens use virulence factors. A virulence factor could be considered to be any aspect of a pathogen that enables it to cause disease (Cramer and Perfect, 2009; Casadevall and Pirofski, 2009). The classical view of virulence factors (referred to hereafter as "classical virulence factors") is that the role of a virulence factor is predominantly or exclusively virulence. An extrapolation of this view is that they can be lost, preventing virulence, without significantly impacting the viability of the microorganism in a suitable culture medium (Brown et al., 2012; Cramer and Perfect, 2009).

However, it has been noted that sometimes there will be some aspects of virulence that are simply components of normal microbial growth (Cramer and Perfect, 2009; Casadevall and Pirofski, 2001). These cannot be removed, as classical virulence factors can, and yet play an important role in pathology. These aspects arise as a result of environmental selection pressures, are often present in opportunistically pathogenic fungi (sometimes alongside a total absence of classical virulence factors), and are often insufficient to facilitate invasion of a healthy host. It has been suggested that these be referred to as "virulence attributes" (Cramer and Perfect, 2009).

The Origin of Pathology

The presence of classical virulence factors presents something of a paradox – where did they come from?

Virulence factors are necessary to invade a host. Without suitable virulence factors, a pathogen cannot survive in the host (Brown et al., 2012; Cramer and Perfect, 2009). At the same time, the host responses that virulence factors counter are very complex, suggesting that virulence factors are the product of evolution in response to immune stresses. However, without exposure to, and survival of, the stressful environment within the host, a pathogen should not be able to evolve adapted virulence factors.

Features cannot develop based on the possibility that they will improve fitness at some point in the future (Gregory, 2009). Additionally, infections are often evolutionary dead ends (Adiba et al., 2010; Boamah et al., 2017). Fitness in this instance is not the same as a straightforward "reproductive success" – there is no difference, in terms of evolution of a

lineage, between a microorganism that dies immediately on invading a host and a pathogen that replicates very successfully within a host for many generations but cannot escape into a new host or an environmental niche. Thus, the paradox.

One such opportunity for emergence is an initially non-antagonistic, long-standing association between microorganism and host, i.e. mutualism or commensalism. This non-antagonistic relationship provides an interface for microbes to safely become adapted to a host, while allowing multiple opportunities for pathogenic invasion. Thus, mutualism, commensalism, and parasitism exist along a continuum (Méthot and Alizon, 2014; Drew et al., 2021). A transition to parasitism might involve translocation to a new, less benign site (Sitkiewicz, 2018) or a change in conditions in an established site (Baker et al., 2018; Lesser et al., 2013).

A second solution, already alluded to while defining virulence traits, is that virulence factors or virulence attributes can arise in the environment, as a result of environmental rather than host-derived selection pressures. We might consider these virulence traits to have emerged “accidentally”.

Accidental Pathology

The basic premise of accidental pathology is simple: that a pathogenic phenotype emerges by chance accumulation of adaptations to the environment, and not as a result of specific evolution to infect a host (Sitkiewicz, 2018). So called “virulence” factors or traits are therefore developed and maintained due to environmental selection pressures, not selection pressures derived from a host. It is an argument of exaptation, considering the

host's body as just another environmental niche to be colonised (Brosius, 2019), and exaptation is known to allow colonising new niches (Azua-Bustos et al., 2012). The core idea of accidental pathology has also been referred to as "pre-adaptation", "coincidental selection", or "coincidental evolution" (Brown et al., 2012; Adiba et al., 2010), and could equally be described as "incidental" instead of "accidental".

A commonly speculated origin for these "virulence" factors or traits is predation by amoebae, which bear similarities to phagocytic immune cells. Amoebal-grazing resistance has been associated with resistance to phagocytes in fungi and bacteria (Steenbergen et al., 2001; Adiba et al., 2010; Chrisman et al., 2010). There may also be environmental origins for behaviour that seems highly specific to infection and virulence. A possible example of this may be the parasitic protist *Naegleria fowleri*, colloquially known as the "brain-eating amoeba" has a receptor for acetylcholine, which it uses for chemotaxis to neurons (Jahangeer et al., 2020). This would appear to be a highly specialised virulence factor. However, *N. fowleri* predates on bacteria (Miller et al., 2018) and bacteria, along with many other organisms, produce acetylcholine (Horiuchi et al., 2003; Kawashima et al., 2007), providing a plausible alternative origin for this chemotaxis as a mechanism to identify prey outside of a mammalian host.

Additionally, some forms of host protection are passive or non-specific, which might make it easier for non-specific resistances to develop. These include the higher body temperatures of mammals and birds, which has been suggested to protect against fungal infection (Cramer and Perfect, 2009; Casadevall, 2005), and the sequestration of essential nutrients within the body, known as nutritional immunity (Hennigar and McClung, 2016). A

polysaccharide capsule and production of melanin are likely defences against desiccation and ultraviolet light, but are both virulence factors for the opportunistic fungal pathogens in the *Cryptococcus* genus (May et al., 2016).

It is worth noting that trait loss can be an important aspect of evolution (Johnson et al., 2012). There are examples of genes that, when present in a microorganism, prevent it from acting as a pathogen, or at least reduce virulence or fitness when acting as a pathogen. In plant pathogens these are called avirulence or Avr genes (Laugé and De Wit, 1998), while for mammalian pathogens they are referred to as antivirulence genes (Bliven and Maurelli, 2012; Siscar-Lewin et al., 2019). Accidental loss of these factors may, in some cases, be as important as acquisition of relevant virulence factors.

The term "accidental pathogen" exists within the literature, but the distinction between accidental and opportunistic or facultative pathogens is not always clear. Some consider a facultative pathogen to be one that can infect a host, but does not need to (which would be described as a primary, facultative pathogen using the axes described earlier), compared to an accidental pathogen which can only infect weakened hosts, and only does so occasionally (which would be described as a secondary, facultative pathogen using the axes described earlier) (Balloux and van Dorp, 2017). Others consider an opportunistic pathogen to be one that is adapted to a given host, but doesn't always cause disease, whereas an accidental pathogen is not specifically adapted to the host while also not always causing disease, as might be the case in a zoonotic infection (Brown et al., 2006). Here, an accidental pathogen is considered to be a pathogen capable of infecting a host (i.e. a primary or secondary facultative pathogen), but for which none of the traits facilitating pathology evolved in

response to selection pressures imposed by a host. It is therefore possible that an accidental pathogen derives no or negligible fitness benefits from infection.

Identifying an accidental pathogen

The four general features of a pathogen's life provided above are unlikely to restrict an accidental pathogen in the same way as a non-accidental pathogen. Chiefly, it is not necessary for an accidental pathogen to be able to escape their host. The factors that allow them to infect a host are formed in response to environmental selection pressures, so there is no need for individuals who have survived the selection pressures present within a host to return to the environment. Indeed, a lack of an exit mechanism may provide support for the accidental origin of pathology.

Likewise, an accidental pathogen may not require specialised invasion mechanisms. Instead, an accidental pathogen may "rely" on happenstance to infect hosts. Thus, we might expect infections by accidental pathogens may be more likely in the gastrointestinal tract, due to regular introduction of material which might be contaminated, or to simply exploit pre-existing wounds and might be expected to be secondary infections, following a primary infection that disrupts normal defences. A lack of invasion mechanisms will make infections less common, but this may also provide support for an accidental origin of pathology as extremely rare infections would provide less of an advantage to adapting to pathology. Necessarily, accidental pathogens will not be obligate pathogens.

The most important challenge for an accidental pathogen to overcome is survival within the host. A microorganism that is eliminated promptly after finding itself accidentally within a

wound is unlikely to impact a host to any meaningful extent. Additionally, without causing harm to the host, an accidental pathogen could not be considered a pathogen. In the absence of active mechanisms to cause harm, the accidental pathogen should facilitate damage through some other mechanisms, for example through immunopathology.

On that note, it is necessary to explicitly make a distinction between pathogenicity and virulence (Cramer and Perfect, 2009; Pirofski and Casadevall, 2012). Pathogenicity is a qualitative trait – whether an organism is capable of being pathogenic (i.e. to cause disease). Virulence, on the other hand, is a quantitative trait of a pathogen – the extent of the damage an organism causes its host during an infection. An accidental pathogen must possess the quality of pathogenicity, but the extent of its virulence is irrelevant.

Finally, growth is not technically an essential characteristic of an accidental pathogen, as an accidental pathogen does not need to derive a fitness benefit from infection. However, it would likely be challenging for a sufficient quantity of material to enter a host to cause damage without growing within a host, particularly in the absence of specialised invasion mechanisms. Thus, the ability to grow within a host can be considered practically essential. However, as with the extent of its virulence, the amount of growth may be minimal. Alternatively, if an organism can produce immunogenic material inside a host, it may not matter if the population of the pathogen increases. Thus, productivity (i.e. production of biofilm or other secreted products) may substitute growth for an accidental pathogen.

It is beneficial to investigate the relatives of a potentially accidental pathogen to truly identify accidental pathology. Ideally, traits that would, when present in pathogenic species,

be seen as evidence of specific adaptations for virulence (e.g. evasion of phagocytosis, surviving the terminal complement complex (membrane attack complex), removal of antibodies, interference with signalling mechanisms, etc) will be present in organisms that are related to the pathogen, but are incapable of causing disease themselves. Additionally, the traits that allow pathogens to infect that are not present in non-pathogenic relatives should have plausible environmental origins.

Merely finding a pathogen in the environment is insufficient to distinguish accidental from opportunistic pathogens. Additionally, demonstrating that a factor (or factors) relevant for virulence confers a benefit outside of a host is necessary for an argument of accidental virulence, but such evidence is not sufficient for a compelling argument. Finally, accidental pathogens need not be evolutionarily stable, as pathology can be lost following adaptation to a different niche (Welter et al., 2021), but instability is not sufficient to demonstrate accidental pathology. The presence of some mechanisms to interfere with vertebrate adaptive immunity, e.g. antigen presentation or cytokine signalling, might be considered evidence for specialised adaptation, as there are no clear non-immune analogues (May et al., 2016).

Introduction to *Prototheca*

A Brief Introduction to Algae

"Algae" are a hugely diverse, polyphyletic group (Heimann and Huerlimann, 2015; Lewis and McCourt, 2004). Some phycologists consider any organism with chlorophyll a to be an alga (Hachicha et al., 2022; Baurain et al., 2010), while others restrict algae to photosynthetic eukaryotes excluding land plants (Cock and Coelho, 2011) or have additional requirements

like possession of chlorophyll b and other pigments (Lewis and McCourt, 2004). As such, algae exist within many of the major eukaryotic lineages (Figure 1-3): Alveolates, which include dinoflagellates (Moore et al., 2008; Stoecker, 1999); Stramenopiles (Heterokonts), which include diatoms (Simpson and Eglit, 2016; Guéguen et al., 2021); Rhizaria (Simpson and Eglit, 2016; Rogers et al., 2007); Excavates (Simpson and Eglit, 2016; Triemer and Zakryś, 2015); as well as taxa that are not clearly members of one of these larger lineages (Burki et al., 2012). If the definition is not restricted to eukaryotes, then cyanobacteria, which possess chlorophyll a, could also be considered algae (Myers and Kratz, 1955; Cláudia Barros et al., 2020).

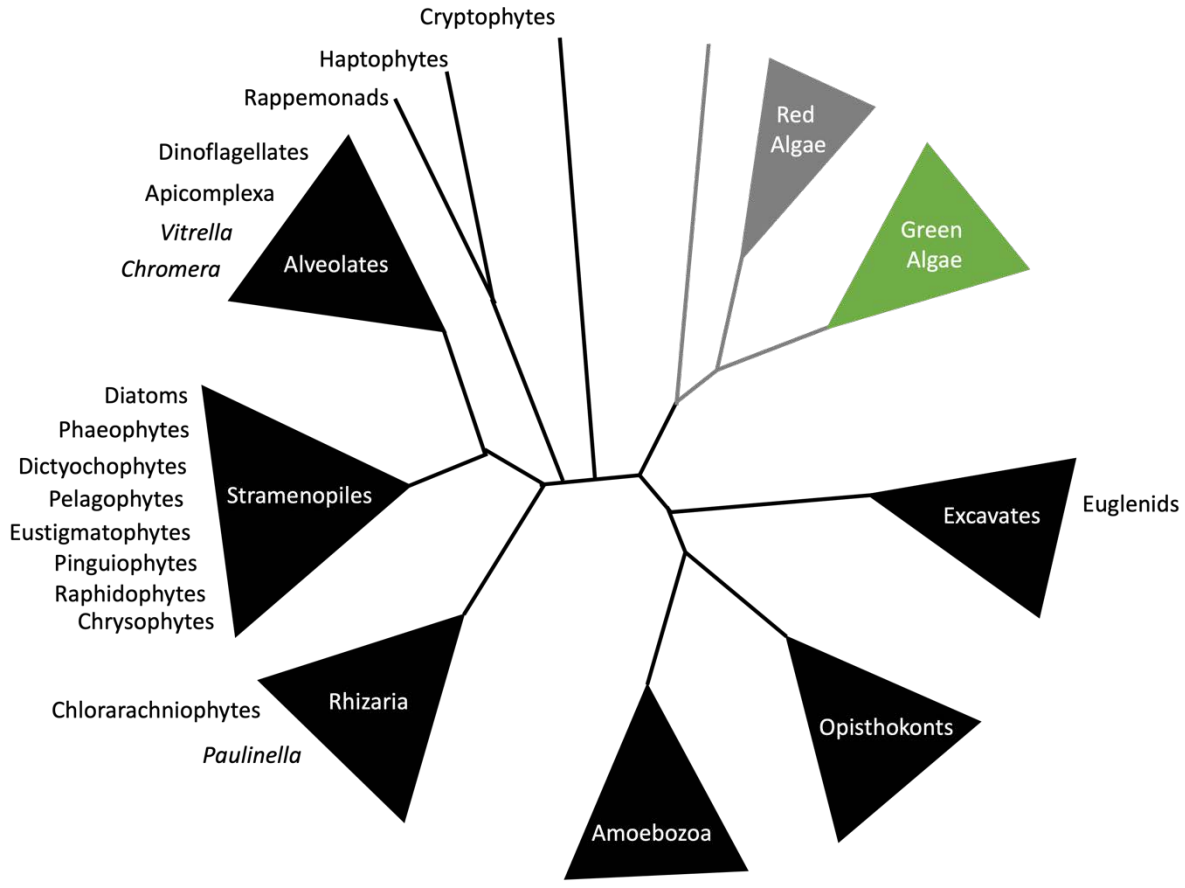


Figure 1-3. Schematic of relationships between the major lineages of Eukaryotic life, based on Figure 1 of Burki (2014). Radiations of major lineages are represented as black triangles, with the exception of the Archaeplastida lineage which is instead indicated as grey lines. Within the Archaeplastids, the radiation of green algae, which contains Prototheca*, is represented as a green triangle. Taxa outside the Archaeplastids that contain photosynthetic algae, or are descended from photosynthetic algae in the case of the Apicomplexa, are labeled approximately according to their relationships in their respective eukaryotic lineages.*

In addition, non-photosynthetic lineages descended from photosynthetic ancestors may have lost the ability to photosynthesise, but are still considered algae; many still have organelles derived from their ancestral chloroplasts that often retain important biochemical functions (Gawryluk et al., 2019; Kayama et al., 2020; Figueroa-Martinez et al., 2017; Hadariová et al., 2018; Nair and Striepen, 2011).

In *Plasmodium*, the causative apicomplexan parasite of malaria, the most significant biochemical role may be the production of isopentenyl pyrophosphate, an essential precursor for isoprenoid synthesis (Nair and Striepen, 2011). Apicomplexans also retain heme, iron sulfur cluster assembly, and fatty acid synthesis pathways in their apicoplasts (Nair and Striepen, 2011). A non-photosynthetic chlamydomonad strain was identified to produce carotenoid pigments and encodes plastoquinol antioxidants and other components of an electron sink system, which may provide protection from redox stress (Kayama et al., 2020). Multiple functions exist in the plastid of *Prototheca wickerhamii*, including carbohydrate, amino acid, lipid, tetrapyrrole, and isoprenoid metabolism (Borza et al., 2005). Of course, it is also necessary to maintain the machinery for replication, transcription and translation, and membrane translocation in plastids with other essential biochemical functions.

Photosynthetic genes are often lost in their entirety in non-photosynthetic algae, but sometimes components remain. Subunits of Rubisco are present in multiple lineages of algae that have lost the ability to photosynthesise, though some of these subunits are highly diverged (Hadariová et al., 2018). Genes for the Calvin cycle are present in the plastid genome of *Helicosporidium*, but chlorophyll synthesis and light harvesting complexes have

been lost (Pombert et al., 2014). The plastid genomes of *P. wickerhamii* and *P. cutis* contain several genes for ATP synthase which have been lost in *P. stagnora* (Knauf and Hachtel, 2002; Suzuki et al., 2018).

"Green algae" are a monophyletic group, within the Archaeplastida lineage of eukaryotes (Burki, 2014). Green algae possess chloroplasts, derived from a single primary endosymbiotic event which they may share with the red algae and glaucophyte lineages (Burki, 2014). They can be divided into three lineages: the Embryophytes, which includes all land plants, the Charophytes, and the Chlorophytes, which includes the well-studied genera *Chlamydomonas* and *Chlorella* (Lewis and McCourt, 2004). Algae that are microscopic are frequently referred to as microalgae.

It is a common assumption that algae do not consume carbon sources from their environments, i.e. they are obligate autotrophs, but some microalgae are able to use both photosynthesis and environmental carbon sources for energy, an ability called mixotrophy (Azaman et al., 2020; Martínez et al., 1991). In some organisms (e.g. *Chlorella*), mixotrophy can completely substitute photosynthesis, providing all the nutrients required for growth (Killam and Myers, 1956); in others (e.g. *Nephroselmis* and *Isochrysis*), mixotrophy is used as a survival mechanism, enabling survival in the face of nutrient limitation but without facilitating growth (Anderson et al., 2018). Mixotrophy may facilitate the loss of photosynthesis, and a transition into obligate heterotrophy, as non-photosynthetic dinoflagellate, stramenopile, cryptophyte, and chlorophyte (i.e. *Prototheca*) algae have mixotrophic relatives (Figueroa-Martinez et al., 2015).

It is another common assumption that algae exclusively live in water, which may be fresh or salt water, but many algae are also members of soil microbial communities (Gärtner, 1996; Metting, 1981) (even in deserts in the case of *Chlorella ohadii* (Treves et al., 2013)). *Apatococcus*, *Chlorella*, *Klebsormidium* and *Trentepohlia* are able to form crusts or biofilms on or in building materials like brick, limestone and concrete (Hofbauer, 2021; Nowicka-Krawczyk et al., 2022), and can be dispersed through the air (Tesson et al., 2016). Green microalgae exist on and in animals: *Trichophilus*, *Acrochaetium*, and *Basicladia* colonise external surfaces of sloths, otters and turtles respectively to form part of their microbiota (Yang et al., 2022) and *Chlorella* and *Auxenochlorella* can be found intracellularly as invertebrate (and protist) endosymbionts (Kodama and Fujishima, 2005; Esteban et al., 2010; Huss et al., 1994).

Finally, green algae can also be parasitic. *Cephaleuros*, *Chlorochytrium* and *Phyllosiphon* are green algae genera that can parasitise plants (Joubert and Rijkenberg, 1971; Aboal and Werner, 2011). *Gastrodia confusa* and *Monotropa uniflora* are plants that parasitise fungi (Ogura-Tsujita et al., 2009). Animals are not safe from algal parasitism: *Helicosporidium* is a genus of green algae that can infect insects (Tartar, 2013; Keilin, 1921) and *Prototheca*, a closely related genus, is capable of infecting mammals, including humans.

Although *Prototheca* are occasionally described as the only pathogenic algae, *Plasmodium*, along with all other Apicomplexa, are pathogenic algae; the lineage possesses an organelle derived from a chloroplast (the apicoplast, though some taxa (e.g. *Cryptosporidium* and possibly *Gregarina*) have subsequently lost it (Toso and Omoto, 2007)) and have photosynthetic relatives (Moore et al., 2008; Keeling et al., 2021). Similarly, some

dinoflagellate endosymbionts appear to reduce the fitness of their hosts, making them parasitic (Lesser et al., 2013). However, the Apicomplexa and dinoflagellates are not green algae, instead being part of the Alveolate lineage (Figure 1-3).

Prototheca (and *Helicosporidium*) are also not the only green algae for which there are cases of parasitic infection. There are isolated cases of animal infection by other green algae (*Coccomyxa* in mussels (Rodríguez et al., 2008; Syasina et al., 2012; Zuykov et al., 2018); *Scenedesmus* in mammals and cichlid fish (Hafner et al., 2013; Yanong et al., 2002; Oliveira et al., 2020)), with some (*Desmodesmus* (Westblade et al., 2015; Fujimoto et al., 2018); "*Chlorella*" (Yu et al., 2009), which may now be classified as *Chloroidium* (Darienko et al., 2016); and *Chlorella* (Hart et al., 2014; Jones et al., 1983)) cases of infections in humans. But it does appear true that *Prototheca* and its relatives are the most significant pathogenic green algae, and it may be true that they are the only green algae to reliably cause infections (though "reliably" may not necessarily mean "frequently").

Phylogenetics and Taxonomy of *Prototheca* and Related Organisms

Historically, algal taxonomy was dependent on morphology, biochemistry and "growth habit" (Lewis and McCourt, 2004). This was unfortunate, as algal morphology can be both variable within closely related taxa, even within the same species (El-Ani, 1967), and not well distinguishable between distantly related taxa (El-Ani, 1967; Lewis and McCourt, 2004). This has led to many historical algal taxa being para- or polyphyletic (Lewis and McCourt, 2004; Pröschold et al., 2011). This may have been particularly unfortunate for *Prototheca*, as an important trait of the genus is its lack of photosynthesis – a rare feature within green algae but relatively common for other eukaryotic lineages. Its inclusion within green algae

was itself somewhat controversial, with requiring identification as algae on at least two different occasions (Kruger, 1894; Chodat, 1913; Nadakavukaren and McCracken, 1977). Even so, some historical species identified as *Prototheca* turned out to be fungi (Nadakavukaren and McCracken, 1975) or Opisthokont protists (Baker et al., 1999).

As a result, there is an extended history of species level taxonomic revisions for *Prototheca*. These revisions include: the classification of species that were subsequently discovered to be synonymous with other species, and thus discarded (Sudman and Kaplan, 1973; Arnold and Ahearn, 1972; Ueno et al., 2002; Jagielski et al., 2019a); the separation of existing species into new species on the basis of new data (Roesler et al., 2006; Masuda et al., 2016; Satoh et al., 2010; Jagielski et al., 2019a); the revival of species classifications that were proposed, thought to be synonymous and discarded, and subsequently determined to be valid (Jagielski et al., 2019a); and the reclassification of individual strains from one valid species to another (Hirose et al., 2018).

Additionally, two incompatible and non-comprehensive systems of classification below species level were suggested for the historical species *P. zopfii* (genotypes (Roesler et al., 2006) and variants (Pore, 1985)), both of which were in use until 2019 (genotypes: (Milanov et al., 2016; Suvajdžić et al., 2017); variants: (Nagatsuka et al., 2017; Fernández et al., 2019)). A similar genotyping system was proposed for *P. wickerhamii*, with genotypes 1 and 2 being represented by strains IFM 56379 and JCM 9643 respectively, but this system was not widely adopted (Hirose et al., 2013).

In recent years, the world of algal taxonomy has been in a state of upheaval, with many genera being revisited with molecular data resulting in reclassification of individual strains and the creation, recreation, or redefining of genus or species level classifications (Bock et al., 2011; Darienko and Pröschold, 2015; Škaloud et al., 2016; Darienko et al., 2016; Pröschold et al., 2001; Caisová and Lenka, 2011; Pröschold et al., 2011; Darienko and Pröschold, 2019; Ueno et al., 2005). *Prototheca* has not been excluded from this upheaval (Figure 1-4), with a revision in 2019 classifying 14 species from the previous 8 (Jagielski et al., 2019a). Since this revision, several new species have been discovered (Kunthiphun et al., 2019; Iskra et al., 2020; Jagielski et al., 2022). Currently, 18 species of *Prototheca* are widely accepted (Kunthiphun et al., 2019; Jagielski et al., 2019a, 2022). Of these, 7 were originally classified as *P. zopfii* and 4 were originally classified as *P. wickerhamii* (including the current *P. zopfii* and *P. wickerhamii*) (Jagielski et al., 2019a; Masuda et al., 2016; Satoh et al., 2010).

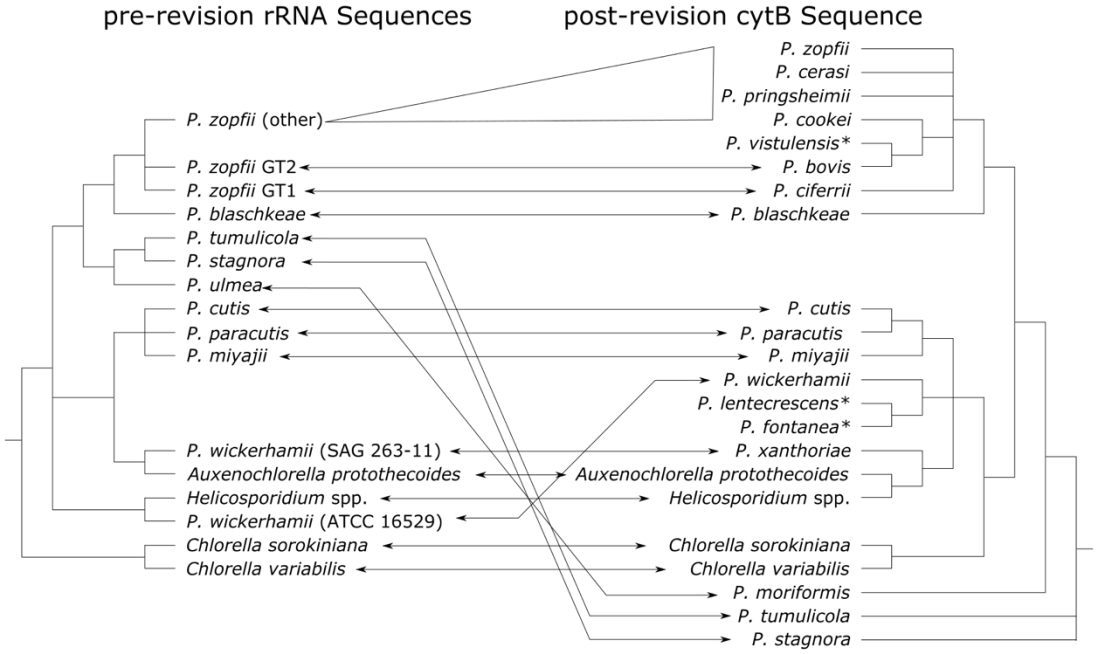


Figure 1-4. Cladograms illustrating relationships between AHP species, before (left) and after (right) the revision of the genus. The cladogram before the revision is based on analyses using predominantly ribosomal RNA (small subunit, internal transcribed spacer region, and D1/D2 region of the large subunit) sequence data. The cladogram after the revision is based on the most recent phylogenies based on the partial sequence of mitochondrial cytb. Arrows between the trees are used to indicate equivalent species. All nodes shown are supported by bootstrap values greater than 70 in at least one analysis. Species discovered after the revision but not as part of the revision are indicated with asterisks and are not included in the pre-revision cladogram. Updated from Figure 1 of Shave et al. (2021).

This history of frequent taxonomic revision impacts the contemporary literature, often making it unclear what an author is referring to when with some species names, especially "*P. zopfii*". In addition, the revision of *Prototheca* is incomplete; it is well known that the current "genus" *Prototheca* is not monophyletic, with at least two other genera, *Auxenochlorella* and *Helicosporidium*, included in the lineage, though the extent of the paraphyly remains unresolved (Conte and Pore, 1973; Ueno et al., 2005). The lineage containing *Auxenochlorella*, *Helicosporidium*, and *Prototheca* is often referred to as the AHP lineage (Ueno et al., 2005). The lineage may also contain the genus *Chlorella* (Mancera et al., 2012; Jagielski et al., 2019a), which might accordingly be referred to as the CHAP lineage (Shave et al., 2021). The genus *Chlorella* may itself be paraphyletic, including the genus *Micractinium* (Wu et al., 2019), which could lead to the lineage being referred to as the CHAMP lineage. It has been proposed that *Prototheca* species should be reclassified into different genera, which may include the establishment of new genera, to resolve the paraphyly. However, it is not currently clear where new genera boundaries should fall, thus the uncomfortable paraphyly of *Prototheca* is temporarily being intentionally maintained to avoid the need for additional revisions in the future (Todd et al., 2018).

All *Prototheca* species are currently defined and identified according to the sequence of the mitochondrial cytochrome b (*cytb* or *cob*) gene (Jagielski et al., 2019a; Dziurzyński et al., 2021). Species can also be identified without sequencing through polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), where digestion with the enzymes RsaI and TaiI is sufficient to distinguish between most species (Jagielski et al., 2019a, 2022). There are currently no classifications below species level.

Life history of *Prototheca*

As already noted, *Prototheca* cannot photosynthesise and have lost the ability to produce chlorophyll. All *Prototheca* species are therefore obligate heterotrophs. Additionally, *Prototheca* is not green (Figure 1-5), which likely contributed to the historical uncertainty regarding whether *Prototheca* species are fungi or algae. However, *Auxenochlorella* and *Chlorella* species retain the ability to photosynthesise, despite being clades within the paraphyletic "genus" *Prototheca* (according to all phylogenies and phylogenies based on *cytb* respectively). This indicates that photosynthesis has likely been lost multiple times within the AHP lineage. The extent of this loss is variable within the lineage: ATP synthase genes are present in plastid genomes of *P. wickerhamii* and *P. cutis*, but not *P. stagnora* or *Helicosporidium* (Suzuki et al., 2018). *Prototheca* species are assumed to be saprophytic, probably existing in rotting plant matter or human waste (Iskra et al., 2020; Schöniger et al., 2016; Buendía et al., 1998).

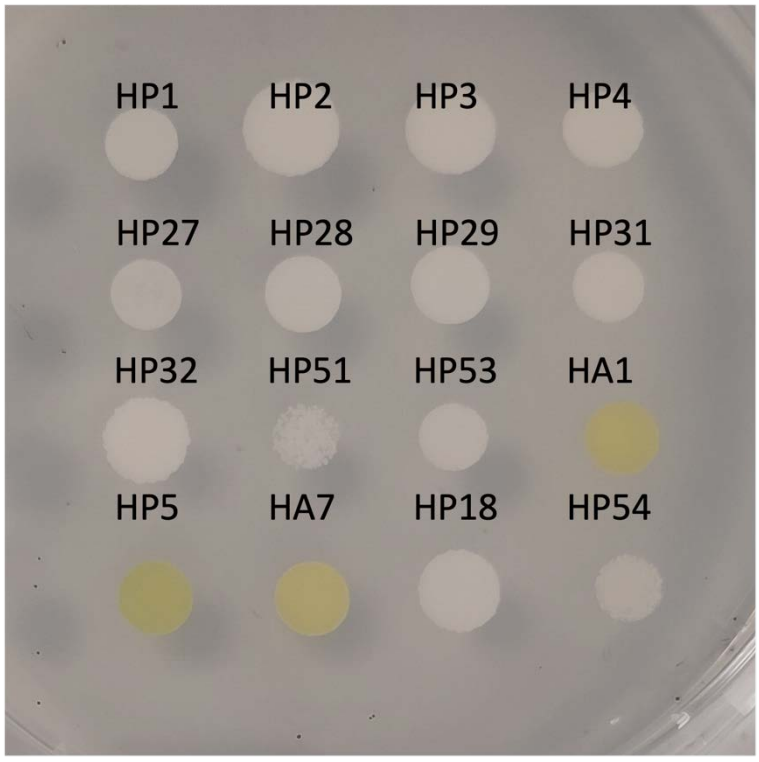


Figure 1-5. Photograph of Prototheca and Auxenochlorella species on Prototheca isolation medium (PIM) agar. 14 species are represented by these 16 strains: HP1 – P. wickerhamii*; HP2* – P. ciferrii*; HP3 and HP18* – P. bovis*; HP4* – P. blaschkeae*; HP27* – P. miyajii*; HP28* – P. cutis*; HP29* – P. tumulicola*; HP31* – P. paracutis*; HP32* – P. cookei*; HP51* – P. moriformis*; HP53 and HP54* – P. xanthoriae*; HA1* – A. protothecoides*; HA5* – A. symbiontica*; HA7* – Auxenochlorella *sp. 2.*

Prototheca does not replicate by budding, as yeast do. Instead, a growing algal cell, also called a dauer cell, divides internally to produce daughter cells, becoming a sporangium or mother cell (Figure 1-6). These daughter cells may be referred to as aplanospores, autospores, endospores, or sporangiospores (Pore, 1985; Jánosi et al., 2001). Up to 50 spores can be produced per sporangium (Leimann et al., 2004). The endospores of *P. wickerhamii* tend to be arranged symmetrically within the sporangium, while the endospores of species formerly identified as *P. zopfii* are arranged more randomly (Lass-Flörl and Mayr, 2007). At all life stages, cells of the species formerly identified as *P. zopfii* (at 7-30 µm in diameter) are larger than cells of *P. wickerhamii* (at 3-10 µm in diameter) (Leimann et al., 2004; Lass-Flörl and Mayr, 2007). Endospores are released passively by the sporangia bursting. A small proportion of endospores, referred to as hypnospores, may be produced with a thicker cell wall and are considered a resting cell stage (Jánosi et al., 2001).

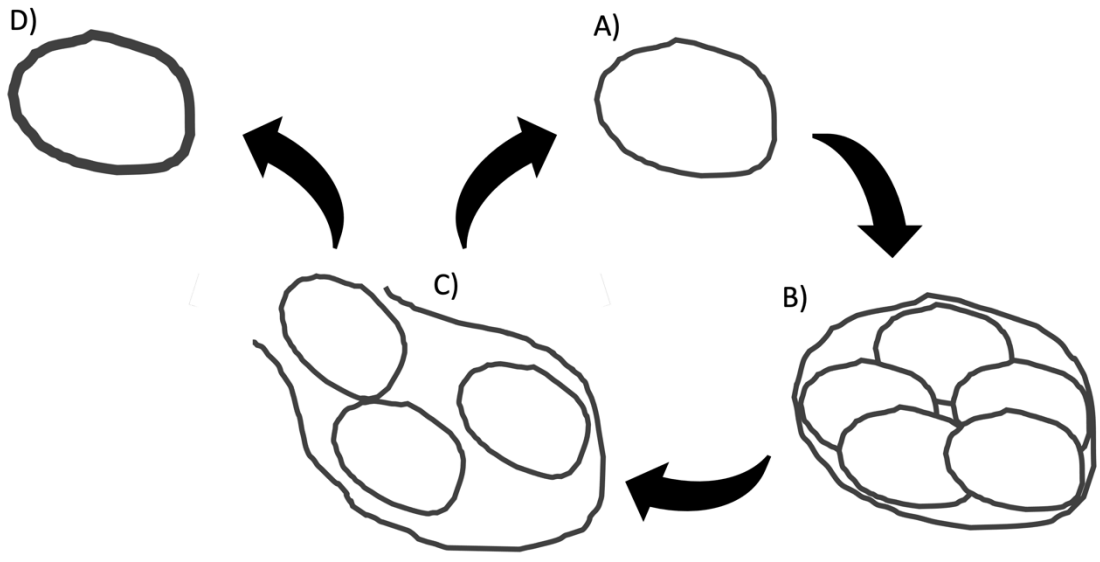


Figure 1-6. Illustration of the life cycle of Prototheca. A) Growing algal cell or dauer cell. B) Dauer cells grow into sporangia. Divisions within the sporangia produce endospores. C) Lytic release of endospores from the mother cell. D) Arrested growth as a hypnospore.

Prototheca shares this method of replication with *Auxenochlorella* and *Chlorella* (Pröschold et al., 2011). *Helicosporidium* can replicate this way, but can also transition into a filamentous cell that invades host tissue to form a cyst (Tartar, 2013). This cyst can also divide internally and burst to release daughter cells. Reproduction in *Prototheca* is assumed to be asexual (Jánosi et al., 2001; Pore, 1985).

Pathology of *Prototheca* (and Related Organisms) – Pathogenic Species

Some *Prototheca* species are pathogenic, and some are almost certainly not. However, identifying pathogenic strains or species of *Prototheca* can prove problematic, due to an absence of an animal infection model. At time of writing, of the 18 accepted *Prototheca* species, six have been documented as agents of disease in clinical case reports in which the isolate was identified to species level using the current definition: *P. blaschkeae*, *P. bovis*, *P. ciferrii*, *P. cutis*, *P. miyajii*, and *P. wickerhamii*. Some *Prototheca* species have confirmed cases of infection in multiple host species. The host ranges of each known pathogenic species of *Prototheca* are provided in **Error! Reference source not found.**.

Within the literature there are recent (i.e. occurring since the genus was revised) references to *P. zopfii* (Abu Ali et al., 2021; Palaniappan et al., 2021; Joerger et al., 2020; Fernández et al., 2019; Sim et al., 2019; Ely et al., 2019; Sari et al., 2018; Geisen et al., 2020; Wesselowski et al., 2022; Herold et al., 2021; Hsieh et al., 2020; Whipple et al., 2020; Teh et al., 2021; Katwilat et al., 2019; Krukowski et al., 2020; Falcaro et al., 2021; Riet-Correa et al., 2021; Miura et al., 2020; Shahid et al., 2020b; Kim et al., 2022; Asiag et al., 2022; Chen et al., 2023; Di Prinzio et al., 2023; Góes et al., 2021) and *P. paracutis* (Masuda et al., 2021) as being pathogens. However, as a result of the taxonomic changes within *Prototheca*, it is not clear

that these species are pathogenic, as discussed below. They are therefore not included in **Error! Reference source not found.**.

With regards to *P. zopfii*, confusion is arising from the fact that the species currently identified as *P. bovis* was, until recently, previously identified as *P. zopfii* genotype 2. Many of the cited references to pathogenic *P. zopfii* are themselves referencing earlier work that identified a pathogenic strain as *P. zopfii* before the revision that reclassified most (if not all) pathogenic *P. zopfii* strains to either *P. bovis* or *P. ciferrii*. Some describe isolates taken prior to the species revision, which may have been identified before classifications like *P. bovis* existed.

However, some are cases where species have been identified after the revision of the genus (Abu Ali et al., 2021; Palaniappan et al., 2021; Joerger et al., 2020; Fernández et al., 2019; Sim et al., 2019; Ely et al., 2019; Sari et al., 2018; Geisen et al., 2020; Wesselowski et al., 2022; Herold et al., 2021; Hsieh et al., 2020; Krukowski et al., 2020; Kim et al., 2022; Asiag et al., 2022; Chen et al., 2023). The misidentification of an isolate as *P. zopfii* instead of *P. bovis* may occur for many reasons: using methods that no longer provide species level resolution; using a reference which has not been updated to reflect the current nomenclature; or the continued use of "*P. zopfii*" as an (incorrect) species name in order to connect to existing literature. Where *P. zopfii* is used to connect to existing literature, the work may explicitly link their *P. zopfii* isolate(s) to the current *P. bovis* species in the body of the work, but choosing to initially present a strain as "*P. zopfii*". *P. zopfii*, as defined by the sequence of *cytb*, will not be considered to be pathogenic in this work.

Table 1-1. Host species that have been infected by Prototheca, and which Prototheca species were responsible for the infection. Host species are arranged in descending order of approximate number of infections; Prototheca species are arranged in alphabetical order. References are used to indicate presence or absence of infection, not frequency, with the oldest reference to a given species causing infection being cited. Infections diagnosed as P. zopfii genotype 1 are assumed to be equivalent to the current P. ciferrii, and infections diagnosed as P. zopfii genotype 2 are assumed to be equivalent to the current P. bovis. Entries followed by one asterisk were diagnosed as P. zopfii prior to the establishment of P. ciferrii and P. bovis but not investigated to genotype level. Entries followed by two asterisks were identified to species level, but not to a currently valid species. Entries followed by three asterisks were the result of experimental inoculation, and thus do not represent a case of natural infection. An entry is only provided in the unspecified column where there are no cases where species level identification was done – there are unspecified cases of Prototheca infection in some hosts that are not presented here because there are specified cases for that host.

Host	*P. blaschkeae*	*P. bovis*	*P. ciferrii*	*P. cutis*	*P. miyajii*	*P. wickerhamii*	Un-specified
Human	(Roesler et al., 2006)	(Ahrholdt et al., 2012)	(Hirose et al., 2018)	(Satoh et al., 2010)	(Masuda et al., 2016)	(Klintworth et al., 1968)	
Cattle	(Thompson et al., 2009)	(Jagielski et al., 2011)	(Ahrholdt et al., 2012)			(Marques et al., 2006)	
Dog		(Ahrholdt et al., 2012)	(Silveira et al., 2018)			(Ginel et al., 1997)	
Cat		(Huth et al., 2015)		(Maboni et al., 2021)		(Kessell et al., 2017)	
Goat		(Ely et al., 2023)***				(Macedo et al., 2008)	
Horse		(Schöniger et al., 2016)					
Buffalo		(Capra et al., 2014)				(Capra et al., 2014)	
Deer		(Frese and Gedek, 1968)*					
Bat		(Stockinger and Doster, 2017)*				(Mettler, 1975)	
Beaver							(Sileo and Palmer, 1973)
Cape Hyrax							(Noda et al., 2003)
Corn Snake							(Crispens and Marion, 1975)
Carp						(Jagielski et al., 2017a)	
Platypus							(Munday and Peel, 1983)

Salmon							(Gentles and Bond, 1977)**

With regards to *P. paracutis*, no case reports are available describing cases of infection (Masuda et al., 2021). Instead, *P. paracutis* has been deemed unlikely to be pathogenic due to an inability to grow at mammalian physiological temperatures (Kunthiphun et al., 2019). *P. paracutis* is closely related to *P. cutis* (as indicated by the naming of *P. paracutis* after *P. cutis*), which is a pathogenic species (Satoh et al., 2010). This close relationship may account for the belief that *P. paracutis* is probably also pathogenic, but in the absence of any case reports it will also not be regarded as pathogenic in this work.

It is interesting that there is not a single pathogenic clade within the overall AHP lineage, at least for the pathogenic species currently known. The non-pathogenic *P. paracutis* is closely related to the pathogenic *P. cutis* and *P. miyajii* (Jagielski et al., 2019a). Similarly, *P. cerasi*, *P. cookei*, *P. pringshemii*, and *P. zopfii* are all non-pathogenic and are closely related to the pathogenic *P. blaschkeae*, *P. bovis*, and *P. ciferrii* (Jagielski et al., 2019a).

Pathology of *Prototheca* (and Related Organisms) – Host Range

Pathogenic *Prototheca* species vary in their apparent host range, and hosts vary in their apparent susceptibility to *Prototheca*, presented in **Error! Reference source not found.**. Of the pathogenic *Prototheca* species, *P. bovis* and *P. wickerhamii* appear to have similarly wide host ranges.

However, **Error! Reference source not found.** contains no information regarding the number of infections in each host. The host range of a species of *Prototheca* will be dependent on its presence in the environment, as this will determine the number of interactions between a pathogen and a potential host. *P. miyajii* and *P. cutis* were not identified as environmental isolates in aquatic environments in Poland (Jagielski et al., 2022), and the only isolates that exist of these species come from infections. This may suggest that these species are rare in the environment, and that their restricted host range reflects their abundance rather than their capacity to infect.

Additionally, some *Prototheca* species are considered pathogenic with very few cases of infection (e.g. *P. cutis* (Maboni et al., 2021; Satoh et al., 2010), *P. miyajii* (Masuda et al., 2016; Ramanan et al., 2018)), and most hosts have only a handful of cases. Given that *Prototheca* infections appear to be rare, and typically progress slowly (as described below), it is likely that *Prototheca* infections may not be detected, limiting the known host ranges of pathogenic *Prototheca* species. It is telling that pets appear to be infected by a wider range of *Prototheca* species than wild or domesticated farm animals, and that humans are apparently the host species infected by the widest range of *Prototheca* species. Thus, host ranges provided in **Error! Reference source not found.** are only "apparent" for some species of *Prototheca*.

Within the literature, there is reference to sheep being a potential host (Jagielski and Lagneau, 2007). However, no case reports of *Prototheca* infection in sheep exist, unlike case reports for other economically important animals like goats, horses, and buffalo. However, reports of chlorellosis in sheep (Cordy, 1973; Riet-Correa et al., 2021), an algal infection

caused by an alga that has not lost chlorophyll (usually designated as either *Chlorella* or *Zoochlorella* (the latter now an invalid classification (Pröschold et al., 2011))) do exist. Additionally, *Prototheca* infections are occasionally referred to as chlorellosis (Krcméry, 2000), presumably for historical reasons – either because most algae found in associations with animals were, for a time, considered *Zoochlorella* or because *Chlorella* does genuinely appear to be closely related to *Prototheca*. Thus, infections with other algae may have been misattributed to *Prototheca*.

It has been suggested that *Prototheca* species may be commensals as well as pathogens. *Prototheca* have been isolated from healthy humans, cattle and one pigeon (Jinatham et al., 2021; Rosario Medina et al., 2017; Jagielski et al., 2019b), demonstrating that transient colonisation with *Prototheca* can occur without health consequences. Additionally, strains have been isolated from a rat, three dogs and eleven pigs of unknown health status (Sobukawa et al., 2011; Pore et al., 1983). However, when the possibility of persistent commensal colonisation was explored, it was found that *Prototheca* were only found in pig and rat faeces when *Prototheca* were also present in the diet (Pore and Shahan, 1988). When the dietary source of *Prototheca* was removed, *Prototheca* was no longer excreted. *Prototheca* have been found to contaminate the food of humans and cattle, providing a possible source for *Prototheca* isolated from faeces (Pore et al., 1983; Jagielski et al., 2019b). While this does not rule out the possibility of commensal colonisation, evidence suggesting such a relationship is limited.

Of the host species, two are of predominant interest, as they account for the majority of infections and are affected by the widest range of species: humans and cattle. Infections of

both of these hosts occur globally, on every continent except Antarctica (Asfour and El-Metwally, 2010; Shave et al., 2021).

Human Protothecosis

The first clear case of human protothecosis was reported in 1964 (Davies et al., 1964), though *Prototheca* species were associated with humans before then (Ashford et al., 1930). Cases are infrequent. Estimates for the total number of cases (from 2020 and 2021) range from 219 to 269 (Godofredo et al., 2020; Minato et al., 2020; Lu et al., 2021). However, the incidence of infection is increasing over time, particularly in immunocompromised individuals (Todd et al., 2018).

Human protothecosis tends to be caused by *P. wickerhamii* (Cullen et al., 2023; Lass-Flörl and Mayr, 2007). Infections can occur in both immunocompromised and seemingly immunocompetent individuals, but are more frequent in the former (Yamashita et al., 2022; Rajan et al., 2020; Melo et al., 2022; Trespalacios Sierra et al., 2022; Ahn et al., 2017; Lu et al., 2021; Todd et al., 2018). Not all forms of immunodeficiency are equally susceptible. In particular, patients with organ transplants, diabetes, and cancer appear particularly susceptible (Chen et al., 2023; Khan et al., 2018; Torres et al., 2003; Góes et al., 2021; Godofredo et al., 2020; Di Prinzio et al., 2023; Trespalacios Sierra et al., 2022) as do patients that use topical steroids (Perrotti et al., 2023). However, patients infected by the human immunodeficiency virus (HIV) seem to be less affected than expected (Carey et al., 1997). Interestingly, in at least one case a *Prototheca* infection led to the discovery of a previously unknown immunodeficiency (Sari et al., 2018). *Prototheca* infections are often associated

with other infections, but it is not clear whether this a product of *Prototheca* tending to infect immunocompromised individuals (Narita et al., 2008).

Prototheca infections do occur in children (Kuo et al., 1987; Sari et al., 2018), but typically tend to affect older individuals (Todd et al., 2018).

The route of entry for *Prototheca* species is unclear. It is currently believed that *Prototheca* enters a host through direct inoculation into a wound, either at the time of trauma or into a pre-existing wound (Lass-Flörl and Mayr, 2007). Diagnosis is often made on the basis of culture or histology (Todd et al., 2018), but infections can easily be misdiagnosed as fungal infections (Yamashita et al., 2022; Melo et al., 2022) or cancer (Yamashita et al., 2022).

It is generally accepted that there are three main presentations for *Prototheca* infections in humans: cutaneous, articular (often specifically referred to as olecranon bursitis), and systemic or disseminated disease (Lass-Flörl and Mayr, 2007; Jagielski and Lagneau, 2007). A majority of infections (54%) are cutaneous, and are restricted to a single site (Todd et al., 2018). Most cases of disseminated infection occur in immunocompromised individuals, but being immunocompromised does not guarantee dissemination of infection (Ramírez et al., 2016; Carey et al., 1997; Piyophirapong et al., 2002; Yagnik et al., 2019). Additionally, immunocompetency is no guarantee of infection remaining local, as there are known cases of systemic protothecosis in immunocompetent individuals (Cox et al., 1974; Chan et al., 1990).

Infections are considered systemic if *Prototheca* is present in at least two non-contiguous locations (Lass-Flörl and Mayr, 2007). *Prototheca* cells are found in the blood in approximately half of systemic cases (Lass-Flörl and Mayr, 2007). However, the systemic categorisation may occlude a number of different presentations, which may become recognised in time. Sites of infection that may be included in "systemic" presentations include: central nervous system (Herold et al., 2021; Takaki et al., 1996; Zhang et al., 2007; Joerger et al., 2020; Zak et al., 2012; Ahn et al., 2017; Hariprasad et al., 2005; Rao et al., 2018), eye (Hariprasad et al., 2005; Minato et al., 2020), gallbladder (Chan et al., 1990), gastrointestinal system (Sari et al., 2018; Raz et al., 1998; Meinke et al., 2017; Telkes et al., 2018; Chan et al., 1990; Matsuda and Matsumoto, 1992; Melville et al., 1999), heart (Buendía et al., 1998; Lass-Flörl et al., 2004), kidney (Lass-Flörl et al., 2004; Herold et al., 2021), liver (Cox et al., 1974; Chan et al., 1990; Matsuda and Matsumoto, 1992; Lass-Flörl et al., 2004; Herold et al., 2021), lungs (Lass-Flörl et al., 2004; Torres et al., 2003; Herold et al., 2021), lymph nodes (Zhang et al., 2010; Matsuda and Matsumoto, 1992), nails (Zaitz et al., 2006; Gandham et al., 2015; Galán et al., 1997), peritoneum (O'Connor et al., 1986; Gibb et al., 1991; Sands et al., 1991; Cox et al., 1974), spleen (Torres et al., 2003), throat (nasopharynx, oropharynx, oesophagus) (Iacoviello et al., 1992; Yamashita et al., 2022), urinary tract (van Bezooijen and Newling, 2002), and vagina (Józsa and Méhes, 1990).

The most reliable treatment for *Prototheca* infection is excising the affected tissue to physically remove the infection (Todd et al., 2018). Antifungal drugs are inconsistently effective, with amphotericin B treatment being the most reliable (Todd et al., 2012, 2018; Leimann et al., 2004). *In vitro* susceptibility testing has also not been a reliable indicator of drug efficacy *in vivo* (Zhang et al., 2007; Lass-Flörl et al., 2004; Zhao et al., 2020; Boyd et al.,

1995). Unreliable treatment is likely a result of there being no drugs designed to kill or inhibit the growth of algal pathogens, unlike fungal, viral and bacterial pathogens. Other potential treatments, including silver nanoparticles (Jagielski et al., 2018a; Ely et al., 2022), polyhexamethylene biguanide (Fidelis et al., 2022), blue light (dos Anjos et al., 2019), 3-bromopyruvate (Jagielski et al., 2018c), essential oils (Nardoni et al., 2018; Grzesiak et al., 2016, 2018), and herbicides (Morello et al., 2020) are being explored, but *in vivo* efficacy remains to be demonstrated.

Typically, *Prototheca* infections are only diagnosed once infection has been present for a while, often on the scale of a year or more, but has not cleared (Yamashita et al., 2022; Melo et al., 2022). Early presentations are therefore unknown. It is also unknown if acute but self-resolving infections exist.

There is no established exit route for *Prototheca* cells from a human infection, though some case reports of cutaneous protothecosis mention exudate or discharge from the site of infection (Yun et al., 2016; Leimann et al., 2004). It is possible that this exudate may contain *Prototheca* cells. There is only one report of potential person-to-person transmission (Khan et al., 2018), but it is not clear that there was not some environmental source responsible for the outbreak.

Prototheca can also be identified in association with healthy humans. *Prototheca* have been detected in the stool of seemingly healthy individuals (Jinatham et al., 2021), is frequently present in human sewage (Pore et al., 1983), and is more prevalent in urban waterways than rural waterways (Jagielski et al., 2022).

Cattle Protothecosis

There appear to be more cases of *Prototheca* infection in cattle than for any other host: with reports regularly identifying tens of cattle or farms being infected (Huilca-Ibarra et al., 2022; Ricchi et al., 2013; Osumi et al., 2008), though it can be difficult to get counts of infected individuals due to differences in methodology and use of bulk milk. The first case of cattle protothecosis was reported in 1952 (Lerche, 1952). This first case was an instance of mastitis, inflammation of the mammary tissue, as are the overwhelming majority of cattle protothecosis cases. *Prototheca* cells are present in milk from infected udder quarters, from cattle (Jagielski et al., 2019b; Osumi et al., 2008; Bozzo et al., 2014) and other species which suffer from *Prototheca* mastitis (currently, only buffalo) (Abdelhameed, 2016). Systemic infection does occur, but is exceedingly rare (Thompson et al., 2009; Taniyama et al., 1994; Jánosi et al., 2001). It has been suggested that persistent intestinal infections exist, and contribute to mastitis via introduction of pathogenic *Prototheca* into the environment (Kurumisawa et al., 2018). However, evidence for intestinal infection, persistent or otherwise, in cattle is limited.

The vast majority of cattle protothecosis is caused by *P. bovis*, with some cases caused by *P. blaschkeae* (Toyotome and Matsui, 2022; Tashakkori et al., 2022; Jagielski et al., 2019b). This is despite *P. ciferrii*, a species that is closely related to *P. bovis* (and, to a lesser extent, *P. blaschkeae*) often being similarly abundant in the environment of cattle (Jagielski et al., 2019b; Osumi et al., 2008) and also being pathogenic (Hirose et al., 2018; Ahrholdt et al., 2012; Silveira et al., 2018). Given that more cases are reported in cattle than any other host

and is *P. bovis* responsible for most cattle infections, it would be reasonable to identify *P. bovis* as the most pathogenic species of *Prototheca*.

The prevalence of mastitis caused by *Prototheca* is unclear. Some estimates suggest protothecosis is rare, accounting for 0.5-1% of mastitis cases (Toyotome and Matsui, 2022; Krukowski et al., 2020). Others report a much higher prevalence, with the highest estimating that *Prototheca* accounts for 38.1% of clinical mastitis cases (Jagielski et al., 2019b). This variability is discussed further in Chapter 3) Identification of Algal Strains, but may be due to poorly standardised detection methods and inconsistent reporting statistics.

There are no known cases of spontaneous clearance in cattle (Milanov et al., 2016), though antibodies specific to *Prototheca* have been found in adult cows (Thiele and Bergmann, 2002). There is no economically viable cure, and infected cattle are culled (Jánosi et al., 2001; Milanov et al., 2016). The challenges of treating cattle protothecosis, as with human protothecosis, is likely due to a lack of antialgal therapies, but is probably exacerbated by mastitis treatment often being unreliable (Melchior et al., 2006).

It is somewhat unclear how cattle become infected. Risk factors appear to include poor farm hygiene practices and excessive antibiotic use, as well as the cow being in the first few weeks of lactation (Milanov et al., 2016; Jánosi et al., 2001). Cattle udders are split into quarters, which have no direct connections to each other making them relatively independent (Nickerson and Akers, 2011). *Prototheca* infections tend to be restricted to one or two quarters (Milanov et al., 2006, 2016; Ito et al., 2011; Wawron et al., 2013). This

suggests that infections are probably as a result of colonisation directly into the quarter, potentially through the teat canal, rather than as a result of systemic dissemination.

Prototheca may come from environmental sources. It has been reported that *P. bovis* was present in faecal samples from cows and calves from one herd with a history of protothecal mastitis, and largely absent from another herd with no history of protothecal mastitis, suggesting an environmental reservoir (Kurumisawa et al., 2018). However, *Prototheca bovis* has been isolated from the environment of farms without confirmed *Prototheca* infections, and infections have occurred on farms where *Prototheca* could not be isolated from the environment – including from faecal samples (Jagielski et al., 2019c, 2019b; Huilca-Ibarra et al., 2022).

There appears to be transmission between cows, as subsequent cases are much more likely after the first (Todd et al., 2018) and several outbreaks have occurred (Costa et al., 1996; Wawron et al., 2013; Ricchi et al., 2013; Gao et al., 2012). This could be through environmental contamination, by calves excreting *Prototheca* into the environment following ingestion of contaminated milk, or direct transmission via *Prototheca* cells in milk contaminating milking equipment (Shave et al., 2021; Milanov et al., 2016; Kurumisawa et al., 2018). It has been proposed that ingestion of *Prototheca*, particularly from contaminated dairy products, may pose an infection risk to humans. However, this seems unlikely, given the different species that tend to infect cattle and humans, and that most human infections are restricted to the skin.

Little is known about the mechanisms by which *Prototheca* species are able to initiate and maintain an infection, or about the host response against infection by algae. Several efforts have been made to remedy this lack of knowledge.

Firstly, pairwise comparisons have been made between species of *Prototheca* using either sequence data or mammalian tissue culture, usually between *P. bovis* to *P. ciferrii*. *P. bovis* has been observed to invade and kill phagocytes, while *P. ciferrii* can do neither (Shahid et al., 2017b; Zhao et al., 2021; Shahid et al., 2017a, 2020b), and *P. bovis* is more inflammatory than *P. ciferrii* (Deng et al., 2016). *P. wickerhamii* has also been observed to kill macrophages (Guo et al., 2023).

This approach suffers from a lack of fundamental knowledge about algae in general and *Prototheca* specifically. Limited genome assemblies and annotations make it difficult to understand the consequences of differential expression of genes. It is still unclear what is responsible for the different responses between *Prototheca* species. The limited scope of comparisons makes it difficult to know what is or is not an important difference for pathology.

Additionally, there is the risk of spurious comparisons. *P. ciferrii* was considered non-pathogenic by some authors performing comparisons, so it was being used as a non-pathogenic control, but there are now reports of infection by *P. ciferrii* (Hirose et al., 2018; Ahrholdt et al., 2012; Silveira et al., 2018). Despite apparent differences in infectivity, these

comparisons may be insufficient to identify fundamental factors for pathology in *Prototheca.*

Secondly, there have been attempts to identify virulence factors, using genome sequences, and antigens that are relevant for other pathogens, as they might provide insight into the pathology of *Prototheca* (Irrgang et al., 2015b, 2015a; Bakuła et al., 2021). However, it is not clear that the structures or sequences relevant for pathology in fungi and bacteria will have similar roles in *Prototheca* due to the evolutionary distance between green algae and these taxa (Burki, 2014). Additionally, the possible lack of classical virulence factors in some eukaryotic pathogens, like *Aspergillus*, where pathology may be as a result of complex combinations of virulence attributes allowing it to persist in weakened host environments, may make identification of virulence factors from sequence data challenging in *Prototheca* (Cramer and Perfect, 2009).

Finally, careful observation of infections has provided some interesting insights. *Prototheca* infections are associated with recruitment and proliferation of lymphocytes in cattle (Bisutti et al., 2023; Pegolo et al., 2022) as well as macrophages and polymorphonuclear cells in mice (Shahid et al., 2020b). However, it is not clear that adaptive immunity is required to control *Prototheca* infections, as humans with acquired immunodeficiency syndrome as a result of HIV infection are not obviously more susceptible to infection or disseminated disease (Lass-Flörl and Mayr, 2007).

P. bovis, *P. ciferrii*, and *P. wickerhamii* have all been found in phagocytes and the extracellular space (Santos et al., 2023; Ito et al., 2011; Whipple et al., 2020). It is not known

which cells are important for clearance or control, but some have speculated that neutrophils are important (Phair et al., 1981; Lass-Flörl and Mayr, 2007). In cattle infections, algae have been found in macrophages, but not in neutrophils, epithelial cells or myoepithelial cells, which may suggest *Prototheca* are using an intracellular niche to hide from neutrophils (Cheville et al., 1984). It has also been noted that human serum can kill *P. wickerhamii* cells, though it is unknown if serum from other species can kill *P. wickerhamii* or if serum can kill other species of *Prototheca* (Phair et al., 1981). The intracellular niche may therefore provide protection against humoral killing by antibodies and complement.

Ultimately, these channels of investigation are still in their infancy. Currently, we know very little about the pathology of any pathogenic *Prototheca* species.

An assumption that might hamper investigation is the assumption that disease progression is similar for all protothecosis, which may also be extrapolated into assuming that all *Prototheca* species use similar mechanisms to infect their hosts (Todd et al., 2018). This is reasonable, as all *Prototheca* species are closely related, and no obvious clinical differences between infections of different species have been reported. However, it is known that *Prototheca* species have host preferences, as *P. wickerhamii* is responsible for most human infections and *P. bovis* is responsible for most cattle infections (Toyotome and Matsui, 2022; Tashakkori et al., 2022; Jagielski et al., 2019b; Cullen et al., 2023; Lass-Flörl and Mayr, 2007). Additionally, *P. wickerhamii* and *P. bovis* are quite distantly related to each other within the lineage, despite seeming to be the main pathogenic species in the lineage (Jagielski et al., 2019a). Thus, it would not be unreasonable for unknown but meaningful differences to exist in the pathology in different *Prototheca* species.

Investigation of Accidental Pathology in *Prototheca*

In this thesis I investigate the basis of pathology in *Prototheca*. However, it is difficult to investigate the mechanisms by which *Prototheca* infect their hosts, due a lack of understanding of the fundamental biology across the genus.

It is not known what broad phenotypes distinguish between pathogenic species from non-pathogenic species, which makes targeted investigation into their underlying mechanisms difficult. Unbiased investigation, like genome-wide association studies (GWAS), require more data than is currently available. Additionally, there are many things "known" about the genus that are extrapolations from limited studies on individual or a small selection of strains (Suzuki, 2006; Sud and Feingold, 1979; Atkinson et al., 1972).

Additionally, as *Prototheca* appears to have several pathogenic lineages, these lineages may possess different pathogenic mechanisms. It is important that current and future comparisons, for example between pathogenic and non-pathogenic species, compare closely related species and do not falsely group together more distantly related pathogenic lineages. A better understanding of the phylogenetics of the AHP lineage may also improve understanding of how frequently the switch between pathogenicity and non-pathogenicity may have happened.

Thus, I will be capturing variation from a wide range of the AHP lineage using multiple isolates from species where possible. This will allow future work to target the mechanisms of variation that appears relevant to pathology.

This broad investigation into the virulence traits of *Prototheca* is guided by the overall hypothesis that pathogenic *Prototheca* species are accidental pathogens. As such, effort is made to investigate the phenotypic diversity of *Prototheca* species under conditions that are relevant to their existence in the environment as well as the host, and will not be limited to pathogenic species.

As discussed earlier, an argument for an accidental origin of pathogenicity requires separation of pathogenic and non-pathogenic species or strains by some environmental factor, or factors. Acquisition of this factor, or these factors, as a result of environmentally derived selection pressure should be both necessary and sufficient to enable pathogenicity in pathogenic strains. It will be difficult to prove that acquisition of this factor is not associated with a host, but it must be at least plausible – otherwise acquisition of pathogenicity could be explained as a coevolutionary process rather than an accidental one.

2) Core Methods

Many methods used in this thesis are unique to individual chapters, and are therefore addressed within methods sections of each chapter. However, some methods are shared across at least two chapters. These are therefore detailed below.

Chemicals

All chemicals were sourced from Sigma Aldrich, unless specified otherwise.

Media

A variety of media, described below, were used to grow algal cells. When solid media was needed, 20 g/l of agar was included to the compositions described below.

Prototheca Isolation Medium (PIM; 10 g/l potassium hydrogen phthalate, 0.9 g/l NaOH (Honeywell), 0.1 g/l $MgSO_4 \bullet 7H_2O$ (Merck), 0.2 g/l KH_2PO_4, 0.3 g/l NH_4Cl, 10.0 g/l glucose, 1 mg/l thiamine hydrochloride) (Pore, 1973). 5-fluorocytosine (5FC) is an optional component of PIM that is included to kill fungal species that are also able to grow on PIM when isolating *Prototheca* from the environment. As *Prototheca* strains were generally not being isolated from the environment, 5FC was not included, unless specified otherwise.

Bold Basal Medium with 3-fold Nitrogen and Vitamins (3N-BBM+V; 750 mg/l $NaNO_3$, 25 mg/l $CaCl_2 \bullet 2H_2O$, 75 mg/l $MgSO_4 \bullet 7H_2O$, 75 mg/l $K_2HPO_4 \bullet 3H_2O$, 175 mg/l KH_2PO_4, 25 mg/l NaCl, 4.5 mg/l Na_2EDTA, 580 ng/l $FeCl_3 \bullet 6H_2O$, 250 ng/l $MnCl_2 \bullet 4H_2O$, 30 ng/l $ZnCl_2$, $CoCl_2 \bullet 6H_2O$

ng/l 12 , 25 ng/l $Na_2MoO_4 \bullet 2H_2O$, 1 mg/l thiamine hydrochloride (vitamin B_1), 1.2 mg/l cyanocobalamin (vitamin B_{12})).

Modified Protease Peptone Medium (Na-PP; 1 g/l bacto peptone (SLS), 20 mg/l K_2HPO_4, 20 mg/l $MgSO_4 \bullet 7H_2O$, 170 mg/l $NaNO_3$).

Additionally, Sabouraud dextrose (SAB; 1% peptone, 2% dextrose; Merck Millipore) and YPD (1% yeast extract, 2% peptone, 2% dextrose) were purchased as complete media.

All media were sterilised by autoclaving except for PIM broth following the addition of 5FC and 3N-BBM+V broth, which were sterile filtered with a 0.2 µm filter (Fisherbrand).

Cultures and Strains

Strains were acquired from a variety of collaborators and AHP strains were assigned internal strain identifiers summarised in Appendix 1. *Prototheca* strains HP1 to HP26 were received from Uwe Rösler. *Prototheca* strains HP27 to HP54 were received from Tomasz Jagielski. *Auxenochlorella* strains HA1 to HA5 were purchased from the Culture Collection of Algae and Protozoa (CCAP). *Auxenochlorella* strains HA6 and HA7 were received from Thomas Pröschold. *Helicosporidium* strains HH1 to HH3 were purchased from the American Type Culture Collection (ATCC). *Chlorella vulgaris* CCAP 211/11B was received from Marianne Barnard and a sample of *Ulva*, identified as *Ulva intestinalis*, was received from Juliet Coates.

Working stocks of *Prototheca* and *Auxenochlorella* isolates were normally maintained on PIM agar plates which were grown at 25°C. These plates were moved to 4°C once colonies were sufficiently grown. This could take as little as two days for fast growing strains, or as long as 14 days for slow growing strains.

Colonies from these plates were used to inoculate broth precultures for experiments, usually in SAB broth, and were grown at 25°C for 48 hours unless stated otherwise. On occasion, stock plates kept at 25°C were used to inoculate precultures, particularly for strains that grew slowly. However, this had no noticeable impact on the resulting precultures. Due to variable availability of equipment, precultures were either grown in volumes of 6 ml, in 50 ml glass flasks, shaking at 150 rpm, or in volumes of 2 ml, in 14 ml plastic culture tubes (Greiner Bio-One), rotating at 20 rpm.

Incubators had no light source and so were usually quite dark, which may have affected *Auxenochlorella* strains. Due to slow growth, and a seemingly greater reliance on light, cultures of HA6 were grown by streaking onto PIM agar and left on the bench to grow (at approximately 20°C). Cultures reliably took 10+ days to grow. Where HA6 was required for experiments, cells were scraped off this agar and resuspended in Dulbecco's Phosphate Buffered Saline (PBS).

Strains of non-AHP, photosynthetic algae were grown in media according to their catalogue page at the CCAP (i.e. either 3N-BBM+V or Na-PP, with both solid and liquid media). No suitable equipment to provide artificial lighting was available, so strains were grown on a windowsill to expose them to natural sunlight.

Strains of *Prototheca* and *Auxenochlorella* were kept at -70°C for long term storage in Microbank 2D vials (Pro-Lab). Strains were removed from frozen storage by streaking onto a plate of PIM agar and incubated as described above.

Bioinformatics

On several occasions, organelle genomes were extracted from whole genome sequences, some of which were assembled as part of the work in this thesis and some of which were publicly available. To identify mitochondrial contigs or scaffolds, a BLASTn search (v2.12.0) (Camacho et al., 2009; Altschul et al., 1997, 1990) was performed on whole genome assemblies using default parameters and published *cytb* sequences (MF163459 and MZ604428.1) as queries. To identify plastid scaffolds, the same was done using sequences taken from published plastid genomes (NC_054192, bases 1000-2000; MF197536.1, bases 1000-1700; KJ001761.1, bases 1000-2000) as queries. Additionally, some sequences taken from annotations generated in GeSeq (v2.03) (Tillich et al., 2017) were used as queries for taxa that did not produce good alignments with the queries listed above.

A scaffold was taken to be an organelle genome if the query sequence was present on only one scaffold and the scaffold was of approximately the correct length. These scaffolds were extracted from overall assemblies using Samtools (v1.9, using htslib v1.9) (Danecek et al., 2021). The presence of other organelle-specific genes was confirmed using GeSeq using references from *A. prototheco*ides, *P. bovis*, *P. wickerhamii*, *C. vulgaris*, and *C. sorokiniana*, as appropriate.

Analysis

Data processing and statistical analyses in all chapters was undertaken with R scripts, unless specified otherwise. Data processing for some chapters was automated using Bash. Scripts are available on GitHub (https://github.com/tindersprite/PrototthecaThesis).

3) Identification of Algal Strains

Introduction

Prototheca infections have previously been reported in the United Kingdom (UK), in humans and dogs (Gaur et al., 2010; Gibb et al., 1991; Macartney et al., 1988). However, there is currently no published record of bovine protothecosis in the UK, despite cattle being the most frequent host of *Prototheca* infection world-wide. Bovine protothecosis typically presents as mastitis, with *Prototheca* cells present in the milk of infected cattle.

Broadly, methods to identify bovine protothecosis can be split into two categories. PCR-based methods involve extracting DNA from a milk sample and then using a PCR reaction (usually quantitative PCR (qPCR)) to detect the presence of *Prototheca*. qPCR methods typically detect the genus, but cannot identify species (ThermoScientific, 2015). Culture-based methods rely on culturing live *Prototheca* cells, usually on agar, and then identifying them by other methods (PCR, carbon assimilation, matrix-assisted laser desorption/ionization-time of flight, DNA melting, etc) (Onozaki et al., 2013; Ricchi et al., 2011; Fidelis et al., 2021; Huilca-Ibarra et al., 2022; Pore et al., 1983; Vasco-Julio et al., 2023). Currently, the most reliable methods for identifying *Prototheca* species are based on the cytochrome B (*cytb* or *cob*) gene, either through PCR-RFLP, multiplex PCR, or sequencing.

Additionally, the taxonomy of *Prototheca* has been extensively revised in recent years, and multiple methods to identify *Prototheca* species have been proposed. Many of the strains in

our collection were collected prior to the taxonomic revision of *Prototheca* in 2019 (Jagielski et al., 2019a), and were therefore assigned to taxa that are now considered invalid.

In this chapter I describe my work to identify *Prototheca* strains from milk samples.
In this chapter I also confirm the identity of strains in our existing collection and highlight some hitherto unreported problems of non-specificity.

Methods

Isolation of *Prototheca* Strains from Milk Samples

Frozen milk samples were provided by the National Milk Laboratories (NML) in Wolverhampton, UK. These samples were preserved by the addition of bronopol before being sent to the NML, where they were subsequently frozen at -20°C.

The isolation of live *Prototheca* cells from these milk samples was attempted using a method similar to protocols described before (Iskra et al., 2020; Jagielski et al., 2019b; Osumi et al., 2008). Samples were completely defrosted and shaken to mix them. 500 µl of milk was added to 2 ml of PIM broth (Pore, 1973), not containing 5FC, and incubated for 48 hours at 25°C. 100 µl of this culture was then streaked onto PIM agar, containing 5FC, and incubated at 25°C for 7 days. Colonies visually resembling *Prototheca* were to be subcultured and identified to species level by PCR-RFLP, as described below.

Isolation of *Prototheca* DNA from Milk Samples

Following failure to isolate living strains from milk, *Prototheca* were detected in milk by extracting DNA directly from milk, to be used as a template for PCR. A DNA extraction

method was developed based on the 'lytic' method described by Quigley *et al.* in 2012 (Quigley et al., 2012). 500 µl of milk was mixed with 500 µl of Milk Extraction Buffer 2X (20 mM ethylenediaminetetraacetic acid (EDTA; Pierce), 100 mM Tris base, 1.4 M NaCl, 2% cetyltrimethylammonium bromide (CTAB), 250 µg/ml Proteinase K) and 100-200 mg of 0.5 mm glass beads. This mixture was incubated at 55°C for at least 1 hour to chemically lyse *Prototheca* cells. 1 ml of a phenol:chloroform:isoamyl alcohol (25:24:1) mixture was added to the samples, which were then agitated twice in a Precellys Tissue Lyser at 6500 Hz for 20 seconds each to mechanically lyse *Prototheca* cells. DNA was then extracted with a phenol-chloroform method, involving two phenol washes and one chloroform wash (Sambrook, 2006). DNA was precipitated with 1 volume of 100% isopropanol and then dissolved in elution buffer (EB; 10 mM Tris base, controlled to pH 8) to be used for identification via PCR-RFLP and sequencing.

This method was validated by inoculating PCR-negative milk with cells of *P. bovis* with densities of multiples of ten ranging from 10^1 to 10^6 cells per ml.

Identification of *Prototheca* to species level

The *Prototheca* genus was redefined in 2019 based on the sequence of *cytb*, in part due to the convenience of distinguishing species using RFLP (Jagielski et al., 2019a). Thus algal strains were identified to species level by this established PCR-RFLP protocol, with some then taken for more detailed identification by sequencing. To provide the template for PCR, in most instances, cells from an agar culture were suspended in 10 µl 0.02 M NaOH and snap frozen in liquid nitrogen, similar to a method used previously (Wang et al., 1993). The template was quickly transferred to the PCR reaction vessel following snap freezing. On

some occasions, purified DNA, extracted with a phenol-chloroform method (as described above or in the methods section of Chapter 6) Genome Sequencing and Assembly of *Prototheca* (and Related Species)), was used. On other occasions, stock material, kept at -70°C, was mixed with an equal volume of 0.02 M NaOH and snap frozen in liquid nitrogen.

For PCR-RFLP and sequencing, the *cytb* gene was amplified using the established cytB_F1 (GYGTWGAACAYATTATGAGAG) and cytB_R2 (WACCCATAARAARTACCATTCWGG) primers (Jagielski et al., 2018b) and Red Taq DNA Polymerase Master Mix (VWR Life Science). The PCR conditions were as follows: 10 min of initial denaturation at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 50°C, and 30 s at 72°C, with a final extension period of 5 min at 72°C.

PCR amplified sequences were used to identify species by RFLP as described previously (Jagielski et al., 2019a). Briefly, amplified sequences were digested by the enzymes HpyCH4IV (TaiI) and RsaI and separated in an agarose gel to reveal patterns of bands that are mostly unique to species of *Prototheca*. This digest is insufficient to resolve all species, with separate digestion of the amplified sequence with MboI, MfeI or BfaI being necessary to distinguish *P. miyajii* from *P. tumulicola*, *P. cookei* from *P. pringsheimii*, or *P. cutis* from *P. paracutis* respectively. All restriction enzymes were purchased from New England Biolabs. Agarose gels were typically separated for 60 minutes at 70 V in gels that were 3% (w/v) agarose (Appleton woods).

Where the species identification from RFLP contradicted existing records for a given strain, PCR amplified sequences were sequenced using the cytB_F1 primer. Amplicons were purified using the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel) kit, according to

the manufacturer's instructions, for sequencing. Sequencing services were provided by Genewiz (Azenta Life Sciences). Species-level identification of *Prototheca* strains was done using *Prototheca*-ID, an online tool (Dziurzyński et al., 2021). A similar identification process was attempted for *Auxenochlorella* strains, using nucleotide BLAST of the nucleotide collection (nr/nt) restricted to Trebouxiophyceae. Visualisations of sequence data were produced in Jalview v2.11.2.6 (Waterhouse et al., 2009).

Bioinformatics for PCR

Algal genomes that might produce PCR product from the *cytb* PCR protocol provided above were identified via a BLASTn search using default parameters restricted to various taxa related to *Prototheca*, listed in Appendix 4, using the sequence of the reverse primer as a query. The reverse primer sequence was chosen as it is the longer of the two primer sequences. Additionally, sequences for organelle genomes of related genera were searched for manually, using the genus name in the GenBank search bar. All identified genomes were downloaded.

A BLASTn search using default parameters was performed on all downloaded genomes using the reverse complement of the reverse primer as a query sequence, with ambiguous bases replaced with N. The location with the lowest e-value that also contained the appropriate end of the primer sequence (3' for the forward primer, 5' for the reverse primer) was selected as the most likely binding site. Where two sites matched equally well, both sites were selected. Samtools was used to extract a 1.5 kbp region from this site in the direction that was thought to contain the forward primer binding site (Danecek et al., 2021).

These sequences were aligned in Jalview using the Clustal algorithm with default settings (Waterhouse et al., 2009).

Results

Live *Prototheca* could not be isolated from milk samples

The NML provides milk testing for all dairies in England and Wales, as well as some dairies in Scotland (Figure 3-1A). Fifteen samples of milk were provided by the NML, having been collected between September of 2021 and February of 2022 (Appendix 2), for our experiments. Records from the NML revealed that eight of these samples tested positive for *Prototheca* species, according to the Thermo Scientific PathoProof Complete-16 kit (a qPCR method). However, this kit could not identify *Prototheca* to species level. Some geographical data were provided along with sample data, indicating these each of these eight samples came from different regions within the island of Great Britain (Figure 3-1B).

To identify *Prototheca* from NML milk samples to species level, isolation of live cells was attempted, as described in the methods section for this chapter, for two randomly chosen samples. This attempt was made in April of 2022, using samples collected in September and November of 2021. Live cells could not be detected - following seven days of incubation at 25°C, no colonies were visible on the agar, nor was any growth apparent in the broth. This was readily explained by the addition of bronopol, an antimicrobial, to the milk prior to submission to the NML. Thus, molecular methods for detecting *Prototheca* were used.

Prototheca DNA could be purified from milk samples

Typical methods for preparing *Prototheca* DNA for PCR (suspending material in 0.02 M NaOH and breaking cells by snap freezing in liquid nitrogen or vortexing with 0.5 mm glass beads) were insufficient to detect *Prototheca* DNA in milk. PCR reactions with samples treated in these ways produced no bands. Thus, to detect *Prototheca* DNA in culture negative milk, a method based on phenol-chloroform DNA extraction was developed.

Milk was thawed and DNA was extracted using Milk Extraction Buffer 2X and phenol:chloroform:isoamyl alcohol, as described in the methods section of this chapter. The presence of algal DNA was tested for using the PCR reaction for *cytb* described in the methods section of this chapter. DNA from somatic cow cells in the milk and other bacterial pathogens that were reported as being present in the milk was likely purified along with algal DNA. However, cow and bacterial DNA are known to not be detected by the PCR reaction for *cytb* (Falcaro et al., 2021).

The DNA extraction method was validated by detecting *Prototheca* in milk inoculated with known densities of *Prototheca* cells. Milk inoculated with different densities of *Prototheca* cells produced bands of differing intensity, allowing for a rough quantification of cell number in the original sample (Figure 3-2A). DNA could also be detected from milk that had been inoculated one month prior and subsequently kept at -20°C (Appendix 3).

Using this method, DNA could be recovered from milk, allowing detection of *Prototheca* in culture negative milk (Figure 3-2B). Two out of the eight samples (samples number 6 and 7) that tested positive by the NML were confirmed as positive. These two samples were from

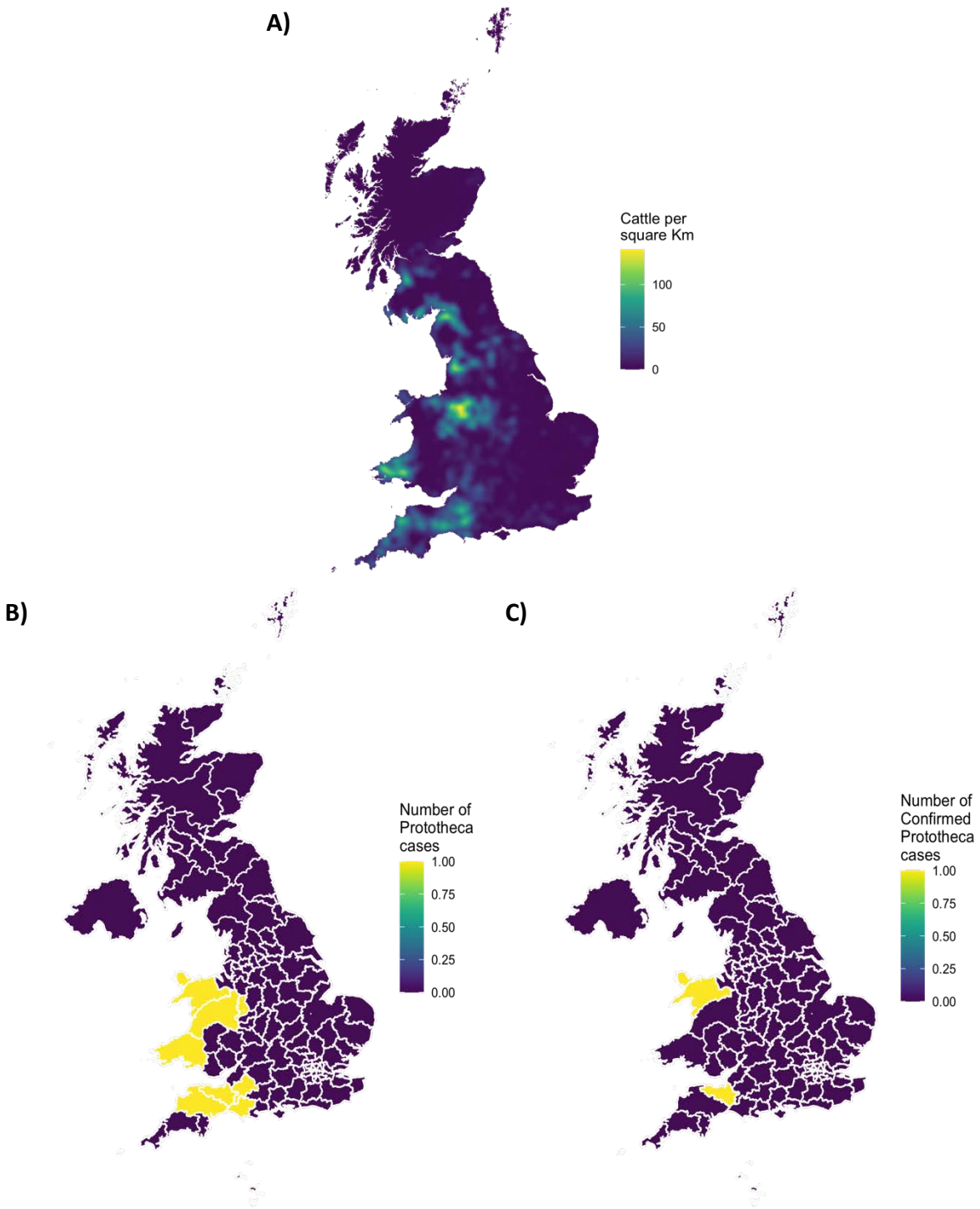


Figure 3-1. Chloropleth maps showing: A) The predicted population density of dairy cattle across Great Britain based on data from July 2021 (provided by the Livestock Demographics Data Groups from the Animal & Plant Health Agency). B) The number of NML milk samples from each region determined to contain Prototheca *by qPCR, as performed by the NML from September of 2021 to February of 2022. C) The number of NML milk samples from each region which could be confirmed to contain DNA* Prototheca *in this chapter. For maps B and C, regions on the map correspond to regions of the country with distinct area code within post codes.*

geographically distant farms (Figure 3-1C). This DNA was identified as being from *P. ciferrii* and *P. bovis* by PCR-RFLP (Figure 3-2C).

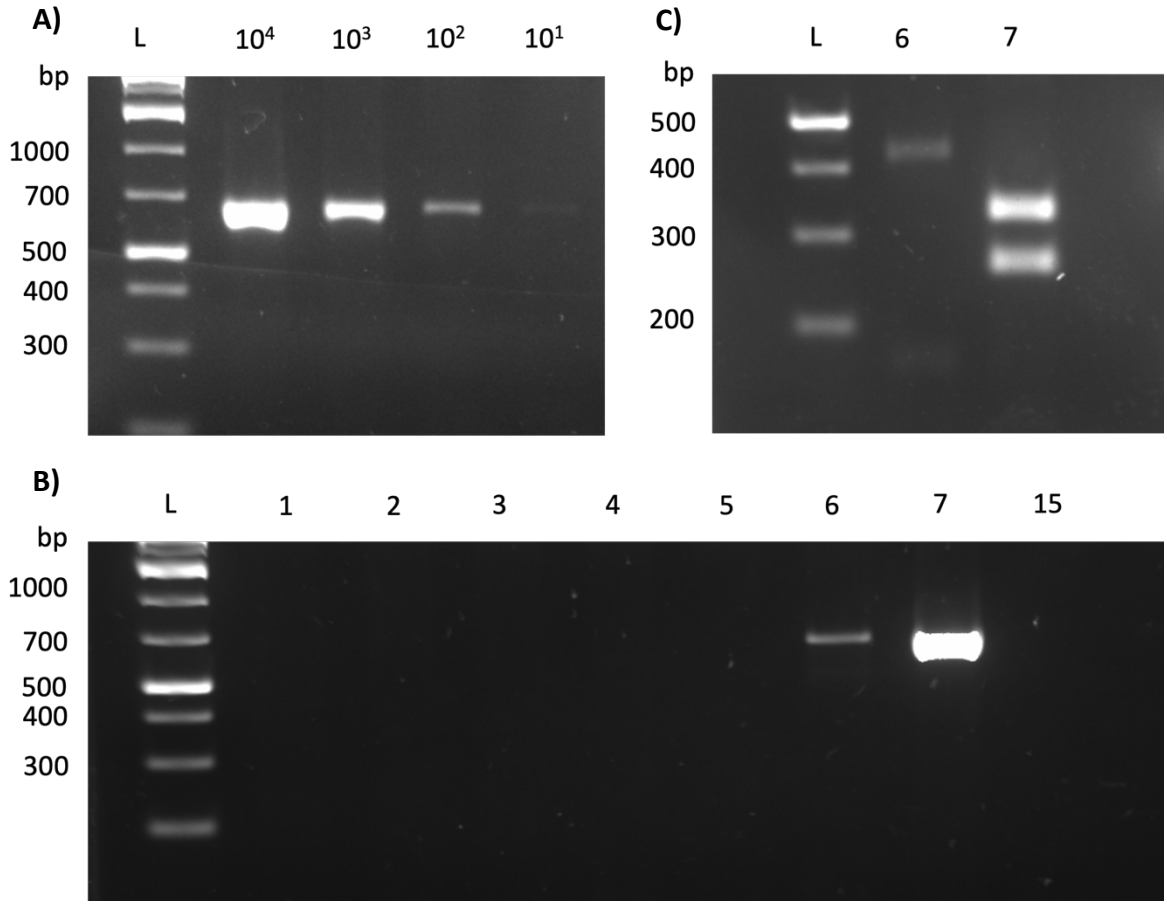


Figure 3-2. PCR products of the cytb *gene using DNA purified from milk as template. A) Milk samples that tested negative for* Prototheca *and were subsequently inoculated with* P. bovis *cells. B) Milk samples that tested positive for* Prototheca *when tested by the NML. C) PCR-RFLP of DNA purified from NML samples from which* Prototheca *DNA could be isolated by phenol-chloroform extraction. Numbers at the top of each lane in gel A indicate the cell density of* P. bovis*, in cells per ml, each sample was inoculated to. Numbers at the top each lane in gels B and C are sample numbers. In all gels, the ladder (L) is ThermoFisher GeneRuler 1 kb Plus DNA ladder.*

Following detection of *Prototheca* DNA, an attempt was made to isolate living cells from these milk samples. This attempt was made in October of 2022, using samples collected in September and October of 2021. No living cells could be recovered.

Prototheca strains were identified to species level

Having identified to species level the two milk samples from which algal DNA could be purified, the identity of the Host and Pathogen Interaction (HAPI) lab's existing collection of *Prototheca* was also investigated. Thus, all 53 strains in our collection were also identified to species level by PCR-RFLP.

48 of the 53 strains produced RFLP banding patterns consistent with their prior identification (Figure 3-3). However, five strains produced RFLP bands that were inconsistent with their assumed identity: HP25, HP26, HP30, HP33, and HP51 (Figure 3-3). Sequencing confirmed that HP33 was an isolate of *P. ciferrii* that happens to have the banding pattern for *P. bovis*, a known phenomenon (Jagielski et al., 2019a). The remaining four appear to be cases of swapped strains, though it is unknown where or when the swap may have occurred. These classifications were used in all subsequent work in this thesis, and a table containing them for reference is provided in Appendix 1.

Two strains of *P. wickerhamii* (HP6 and HP49) consistently produce an additional band at approximately 110 bp. However, a simulated digestion on the *cytb* sequence derived from a whole genome assembly of HP6, performed in Chapter 6) Genome Sequencing and Assembly of *Prototheca* (and Related Species) failed to show an additional band when compared to a similar sequence from HP5 and HP1.

CytB PCR-RFLP can be performed on two of the three genera of the AHP lineage

The *cytb* PCR reaction described by Jagielski et al (Jagielski et al., 2018b, 2019a) has been reported as being specific to *Prototheca* species (Falcaro et al., 2021). However, the genus

Prototheca is paraphyletic, including the genera *Auxenochlorella* and *Helicosporidium*, as covered in the overall introduction. It therefore seemed likely that the *cytb* PCR protocol would work on all members of the AHP lineage, not just *Prototheca*.

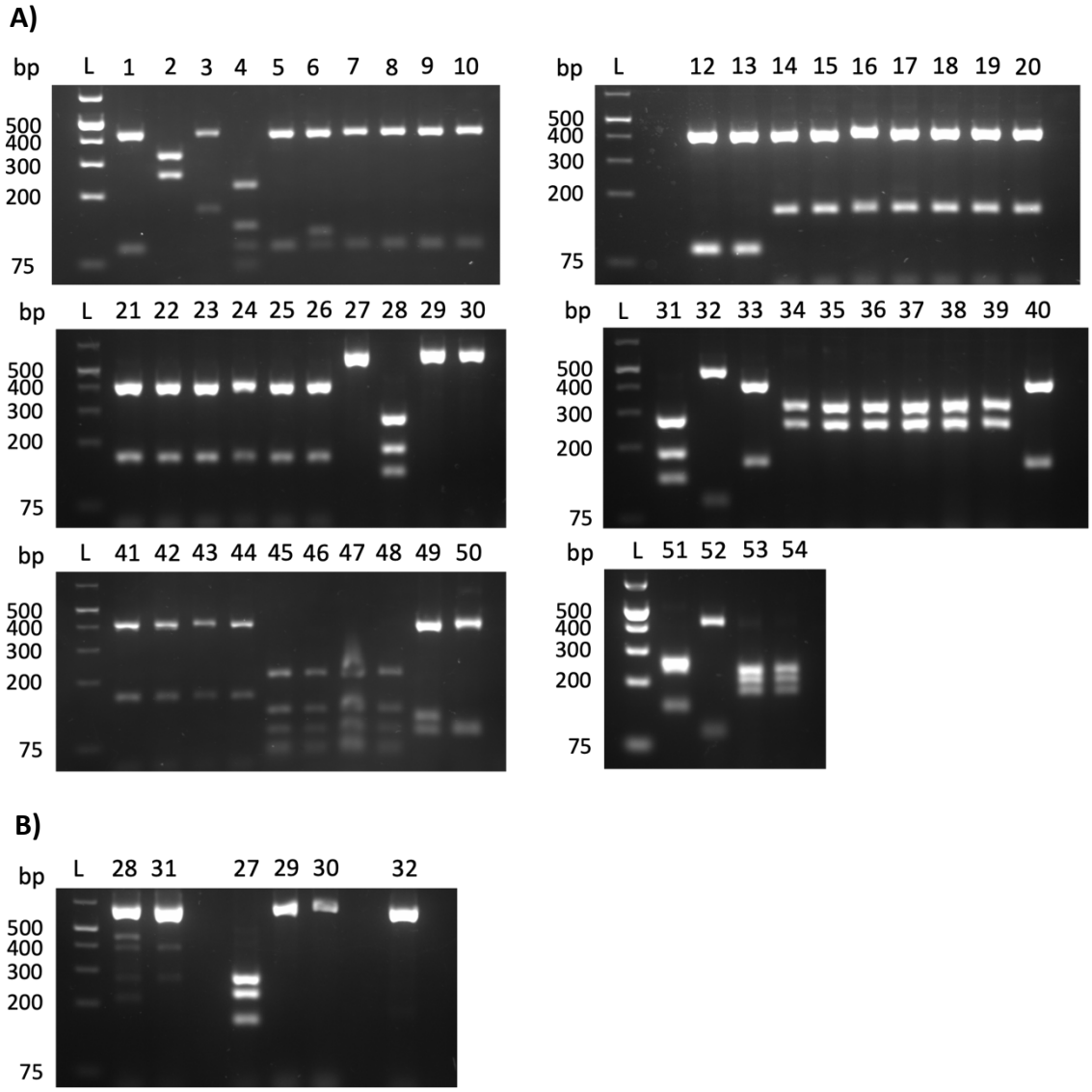


Figure 3-3. PCR-RFLP of Prototheca *strains possessed by the HAPI lab. Wells are numbered with the equivalent strain number, i.e. well 1 contains HP1, well 10 contains HP10, etc. A) PCR-RFLP of all strains digested with RsaI and TaiI. Notably, two strains of* P. wickerhamii *(HP6 and HP49) produce an additional band. B) Digests to distinguish between previously indistinguishable species. HP28 and HP31 were digested with BfaI; HP27, HP29 and HP30 were digested with MfeI; HP32 was digested with MboI. In all gels, the ladder (L) is ThermoFisher GeneRuler 1 kb Plus DNA Ladder.*

To test this, seven strains of *Auxenochlorella* and three strains of *Helicosporidium* were acquired. However, one strain of *Auxenochlorella* could not grow on PIM (Figure 3-4). This strain was subsequently identified as being a strain of *Leptochlorella* (pers. comm. Thomas

Pröschold). The six remaining strains of *Auxenochlorella* represent four species within the genus (Appendix 1). One strain, HA6, is visually different from the other *Auxenochlorella* strains, growing more slowly and appearing more dependent on light to grow, but all others seem phenotypically consistent with each other.

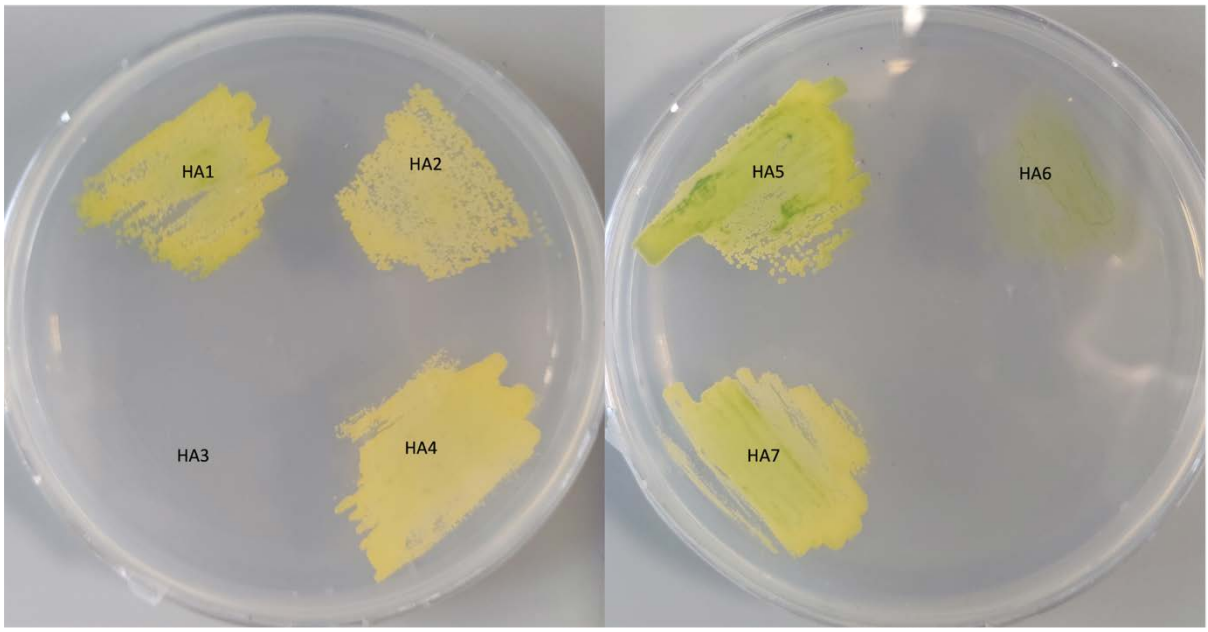


Figure 3-4. Agar plates containing strains that were originally identified as Auxenochlorella*. Plates are PIM agar, and were grown for 10 days at approximately 25°C in natural sunlight. Cultures are labelled with internal strain identifiers.*

The *cytb* PCR reaction was attempted on these six *Auxenochlorella* strains and three *Helicosporidium* strains. Live cells, snap frozen in liquid nitrogen, were used to provide the DNA template for *Auxenochlorella* reactions, but live *Helicosporidium* cells could not be produced for two of the three strains in our possession, and recovery of live cells from the third strain (HH3) was unreliable. For all three strains of *Helicosporidium*, frozen stock material was used to provide the DNA template, which is usually sufficient to produce a band for *Prototheca*, and live cells were also used for HH3.

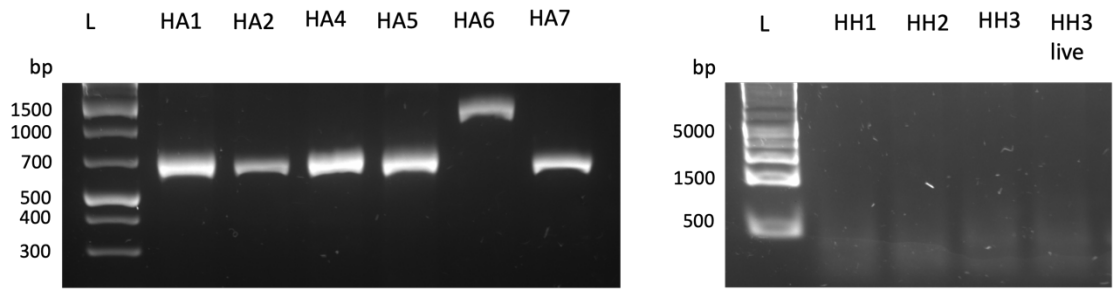


Figure 3-5. Colony PCR products following amplification of the cytb gene using material from Auxenochlorella *(left) and* Helicosporidium *(right). Wells are labelled with the appropriate internal strain identifier. Live cultures of* Auxenochlorella *were used to provide a template for all reactions. Live cultures were only used to provide a template for one* Helicosporidium *culture (labelled HH3 live); frozen stocks of cells were used as template for the three other reactions. In both gels, the ladder (L) is ThermoFisher GeneRuler 1 kb Plus DNA Ladder. Both gels were 3% TAE separated at 70 v for 60 minutes (*Auxenochlorella*) or 30 minutes (*Helicosporidium*).*

All strains of *Auxenochlorella* produced bands, with HA6 producing a larger band than the others (Figure 3-5). None of the *Helicosporidium* strains produced a band (Figure 3-5). Subsequently, RFLP using all four combinations of enzymes used for identification of *Prototheca* species (RsaI + TaiI; BfaI; MfeI; MboI) was performed on the PCR product from the *Auxenochlorella* strains. These digests did not provide any additional resolution beyond that of the PCR alone (Figure 3-6).

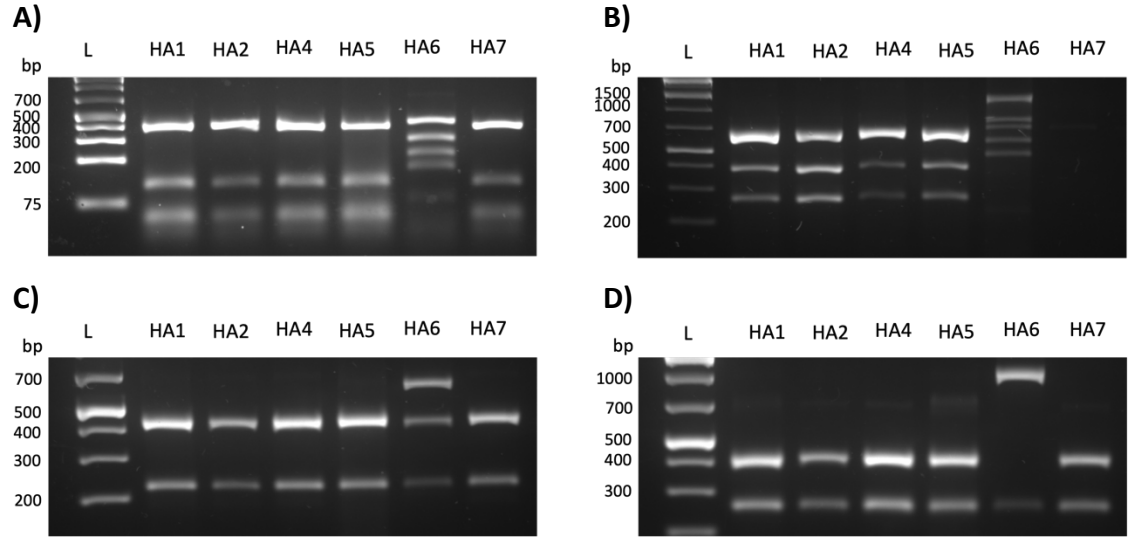


Figure 3-6. PCR-RFLP of Auxenochlorella *strains possessed by the HAPI lab. Wells are labelled with the appropriate internal strain identifier. A) PCR-RFLP of all strains digested with RsaI and TaiI. B) PCR-RFLP of all strains digested with BfaI. Amplification of HA7 failed for unknown reasons, but separate digestion showed the same pattern as HA1, HA2, HA4 and HA5. C) PCR-RFLP of all strains digested with MboI. D) PCR-RFLP of all strains digested with MfeI. In all gels, the ladder (L) is ThermoFisher GeneRuler 1 kb Plus DNA Ladder.*

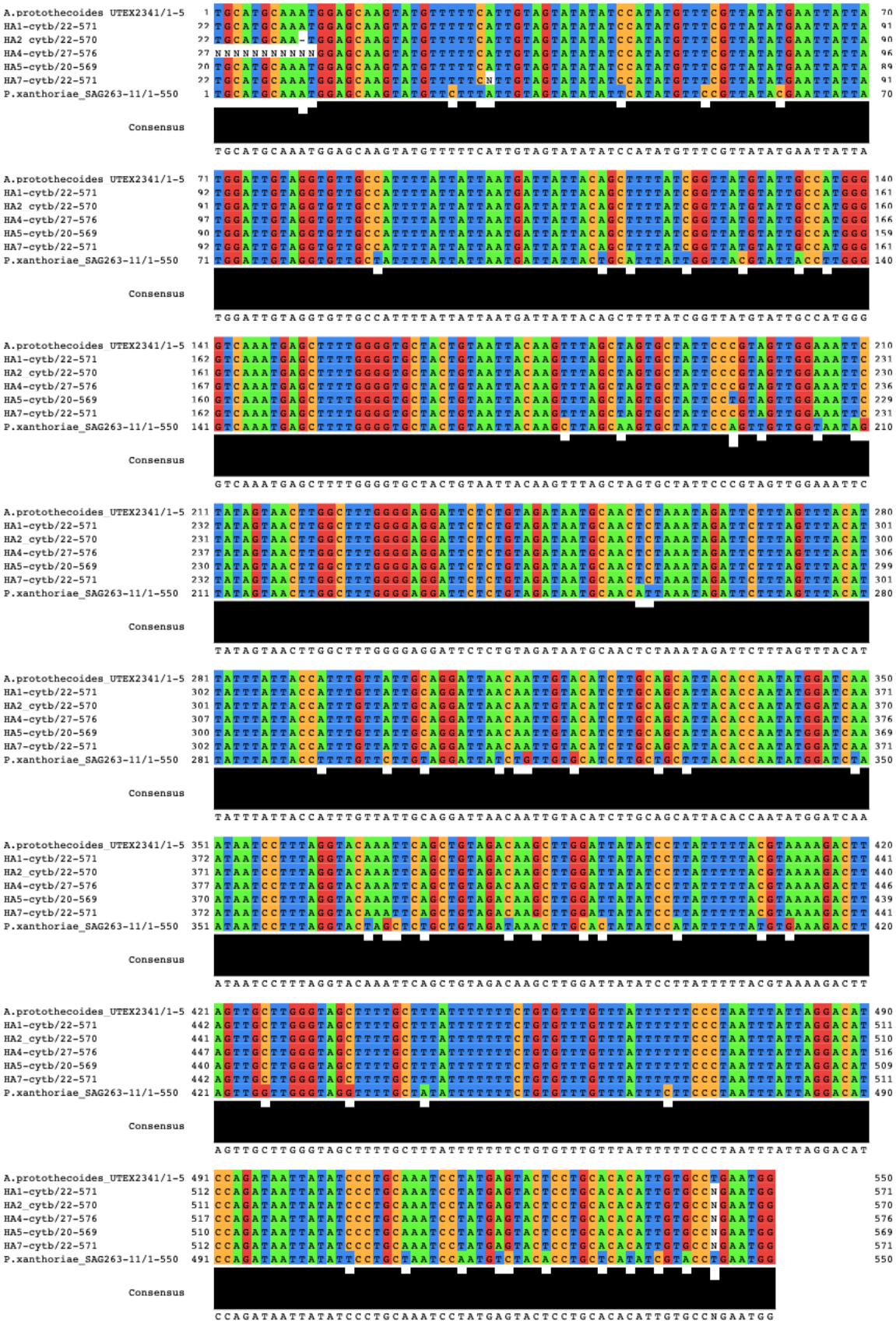


Figure 3-7. Alignment of cytb sequences from Auxenochlorella *strains in possession of the HAPI lab. Sequences have been aligned to publicly available sequences for* Auxenochlorella protothecoides *UTEX 2341 and* Prototheca xanthoriae *SAG 263-11. Alignment produced in Jalview using the Clustal algorithm with default settings.*

The PCR products from five of the six *Auxenochlorella* strains were sequenced to see if *cytb* could be used to identify species of *Auxenochlorella*. Amplification of *cytb* from HA6 was unreliable, and sufficient material for sequencing could not be produced in time. A BLAST search found all sequences to align well (≥99.5% sequence identity) with available sequences for *A. prototheco*ides mitochondrial DNA (Figure 3-7). All sequences were also found to be approximately 90% identical to sequences of *P. xanthoriae* (strain SAG 263-11, which was initially classified as *P. wickerhamii*).

These sequences are included in the *cytb* phylogeny in Chapter 7) Phylogenetics of *Prototheca* and Related Organisms, but they do not appear to have sufficient information to distinguish between *Auxenochlorella* species (e.g. *A. prototheco*ides from *A. symbiontica*). The *cytb* gene from HA6 was not sequenced, but this sequence was not necessary to distinguish HA6 from the other three species, as the length of the amplicon is sufficient for this.

The lack of PCR product using *Helicosporidium* as a template could have indicated that our *Helicosporidium* strains had degraded, or that *Helicosporidium* genuinely does not possess a suitable target sequence. A BLAST search of the *Helicosporidium* sp. Sj mitochondrial genome (accession GQ339576.1) using the *cytb* primers as queries found a region matching the reverse primer, but no real match for the forward primer. A search for both primers in the mitochondrial genomes of *P. wickerhamii*, *P. bovis*, and *A. prototheco*ides found matches for both primers. However, an alignment of mitochondrial DNA with primer sequences found a good alignment with all four mitochondrial DNA sequences for the forward and reverse primers (Figure 3-8). Thus, *Helicosporidium* likely fails to produce an

amplicon in this PCR reaction due to two point mutations at the binding site of the forward primer 3' end.

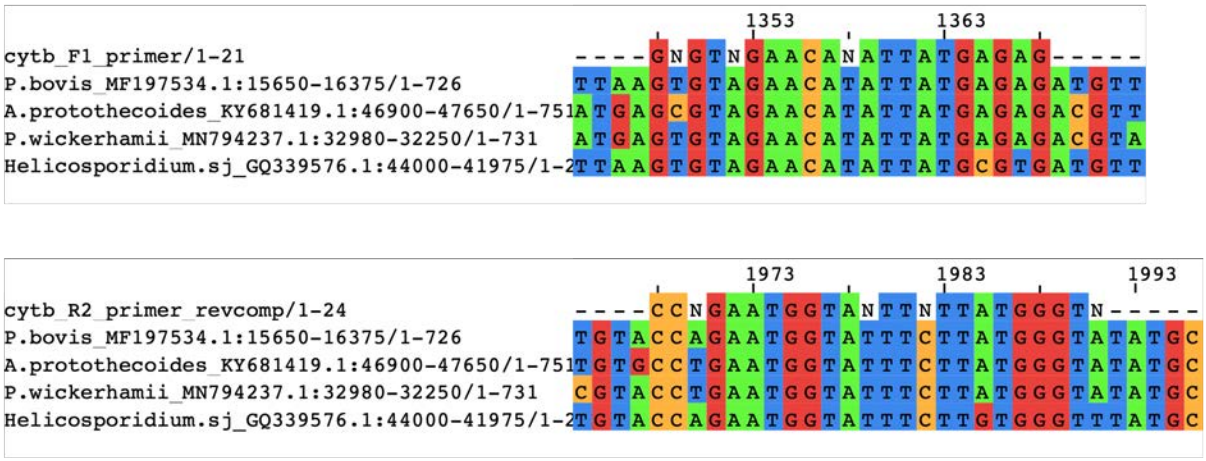


Figure 3-8. Alignment of AHP DNA with cytb primers, coloured by nucleotide. Top) Alignment with the sequence of the forward primer. Bottom) Alignment with the sequence of the reverse complement of the reverse primer. IUPAC ambiguity codes were removed from both primer sequences, replacing them with N. Alignment produced in Jalview using the Clustal algorithm with default settings.

CytB PCR protocol is not specific to the AHP lineage

To see whether the primers for *cytb* would detect non-AHP algae, the sequences of the cytB_F1 and cytB_R2 primers were aligned against publicly available algal sequence data. Ideally, such a search would include sequences from different genera, families, orders, and classes.

A BLAST search using the sequence of the cytB_R2 primer, with IUPAC ambiguity codes replaced with a consensus nucleotide from the alignment in Figure 3-8, returned several mitochondrial sequences from members of the Chlorellaceae family. Additionally, mitochondrial genomes from all genera in the family Chlorellaceae were searched for. A list of genera searched is available in Appendix 4. No additional mitochondrial genomes were found for members of the Chlorellaceae family.

Table 3-1. Taxonomic classifications of each organism for which mitochondrial DNA was found with a match to the cytb *reverse primer sequence. Strain information is included if the GenBank entry for the data contained strain information, "Unknown" if not.*

Class	Order	Family	Binomial name	Strain	Accession
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Chlorella variabilis*	NC64A	NC_025413.1
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Chlorella vulgaris*	NJ-7	MK948101.1
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Micractinium conductrix*	Unknown	KY629619.1
Trebouxiophyceae	Chlorellales	*incertae sedis*	*Picochlorum* sp.	BPE23	CM046021.1
Trebouxiophyceae	Chlorellales OR Watanabeales	Oocystaceae OR Watanabeaceae	*Jaagichlorella roystonensis*	Unknown	MN934958.1
Trebouxiophyceae	Chlorellales OR Watanabeales	Oocystaceae OR Watanabeaceae	*Jaagichlorella hainangensis*	Unknown	MN966687.1
Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Trebouxia* sp.	TR9	MH917293.1
Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Trebouxia* sp.	A1-2	MN642622.1
Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Lobosphaera incisa*	SAG 2468	NC_027060.1
Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Symbiochloris* sp.	SG-2018	NC_066038.1
Trebouxiophyceae	Trebouxiales	Botryococcaceae	*Botryococcus braunii*	Unknown	KR057902.1
Trebouxiophyceae	Unknown (Kato et al., 2023)	Unknown	*Medakamo hakoo*	M-hakoo 311	LC604817.1
Trebouxiophyceae	*incertae sedis*	Coccomyxaceae	*Coccomyxa subellipsoidea*	C-169	HQ874522.1
Chlorophyceae	Sphaeropleales	Sphaeropleaceae	*Atractomorpha echinate* (*cytb*)	UTEX 2309	KJ845686.1
Chlorophyceae	Sphaeropleales	Pseudomuriellaceae	*Pseudomuriella schumacherensis*	SAG 2137	KJ806273.1
Chlorophyceae	Chlamydomonadales	Chlamydomonadaceae	*Chlamydomonas reinhardtii*	CCAP 11/32 CW15+	U03843.1
Chlorophyceae	Chlamydomonadales	Chlamydomonadaceae	*Chlamydomonas moewusii*	Unknown	NC_001872.1

Ulvophyceae	Ulvales	Ulvaceae	*Ulva intestinalis*	CNS000531	MZ571476.1
Ulvophyceae	Ulvales	Ulvaceae	*Ulva* sp.	TM708	NC_053254.1
Chloropicophyceae	Chloropicales	Chloropicaceae	*Chloropicon* sp.	RCC4434	MK086009.1

At the time of writing, there are five families within the order Chlorellales: Chlorellaceae, Oocystaceae, Phyllosiphonaceae, Rhopalosolenaceae and Leptosiraceae (Schoch et al., 2020). No significant BLAST hits were found in a BLAST search using the sequence of cytB_R2 restricted to Leptosiraceae, and Rhopalosolenaceae and Phyllosiphonaceae were not recognised as valid taxa to restrict a BLAST search to. A BLAST search restricted to Oocystaceae found matches, but none in mitochondrial DNA. As with the Chlorellaceae family, genomes, mitochondrial or otherwise, from all genera in the other families of the order Chlorellales. A list of genera searched is available in Appendix 4. Only two mitochondrial genomes could be found, both for *Jaagichlorella* species. However, *Jaagichlorella* was recently reclassified into a new order: Watanabeales (Li et al., 2021). It is unclear whether these strains represent members of the Oocystaceae family as listed in algaebase (Schoch et al., 2020), or the Watanabeales family – if they are the latter, then only *Picochlorum* mitochondrial genomes were found within the Chlorellales order that were not also part of the Chlorellaceae family.

Some mitochondrial DNA from organisms from other orders within the Trebouxiophyceae class and from other classes were included. The list of genomes included, along with their taxonomic ranks, are available in Table 3-1.

A number of matches for cytB_R2 were found in chloroplast genomes of other microalgae and nuclear genomes of two plants. Thus, it was possible that these primers might detect chloroplast and nuclear DNA as well as mitochondrial DNA. However, this was not explored further.

Mitochondrial sequences were aligned to the *cytb* primers in small groups of related organisms, to prevent spurious alignments as a result of distantly related sequences. Good alignments for both primers were found in other members of the Chlorellaceae family, which the AHP lineage is part of (Figure 3-9A-D). Generally good alignment was seen in other members of the Chlorellales order, but the sequence of *Pichochlorum* had the same mutations in the region equivalent to the 3' end of the forward primer that appears to prevent amplification of *cytb* in *Helicosporidium* (Figure 3-9E&F). Members of the Trebouxiales order align well with the sequence of the reverse primer, but less well with the forward primer (Figure 3-9G&H). Finally, some reasonable matches with the forward primer and good matches with the reverse primer are seen in sequences from different classes (Figure 3-9I&J)

However, some sequences aligned poorly. Two sequences, from *J. roystonensis* and *Trebouxia* sp. TR9, aligned poorly with the forward primer and with other sequences from the same genus. This could indicate errors with identification of these strains.

Similarly, two regions from both *P. schumacherensis* and *Chlamydomonas reinhardtii (Chlam. reinhardtii)* aligned equally well with the reverse primer, though these matches were still poor, so both were included in the alignment. Both sequences from both taxa were included, but none of them aligned well with either primer or other sequences in the alignment. It seems likely that these regions are not homologous with the region containing *cytb* in *Prototheca*.

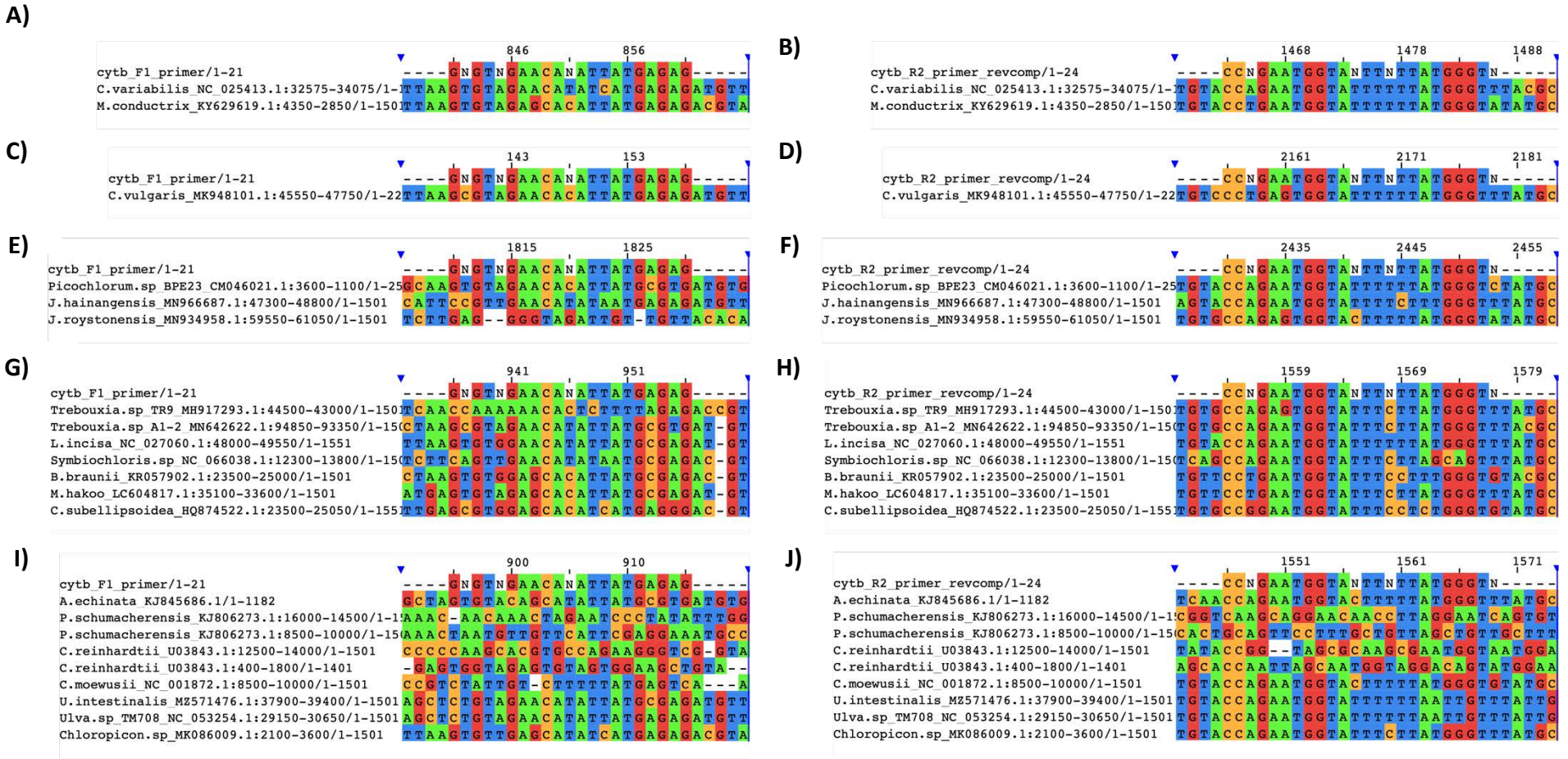


Figure 3-9. Alignment of non-AHP DNA with cytb primers, coloured by nucleotide. IUPAC ambiguity codes were removed from both primer sequences, replacing them with N. Sequences have been grouped to keep related taxa together, according to their relatedness to Prototheca *and to each other, but are otherwise arranged arbitrarily. A&B) Alignments of sequences from other genera within the Chlorellaceae family. C&D) Alignments of sequences from* C. vulgaris*, which appears to have an insertion which interferes with aligning against other Chlorellaceae mitochondrial genomes. E&F) Alignments of sequences from organisms which might be a different family within the Chlorellales order. G&H) Alignments of sequences from other orders within the Trebouxiophyceae class. I&J) Alignments of sequences from other classes.*

It can be challenging to tell, purely from sequence data, whether a PCR reaction will produce an amplicon. Thus, algal samples were acquired from collaborators and the CCAP to test the PCR protocol directly. The samples acquired are summarised in Table 3-2. The CCAP do not supply many of the taxa for which mitochondrial genome sequences were available, thus strains that would provide a similarly broad taxonomical range were acquired.

Table 3-2. Taxonomic classifications of each sample used as a template to determine the extent of the non-specificity of the cytb PCR reaction. Strain information is included where possible.

Class	Order	Family	Genus	Strain Identifier
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Prototheca*	Numerous
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Chlorella*	CCAP 211/11B
Trebouxiophyceae	Chlorellales	Chlorellaceae	*Keratococcus*	CCAP 223/1
Trebouxiophyceae	Chlorellales (Schoch et al., 2020) OR Unclear (Neustupa et al., 2013)	Chlorellaceae (Schoch et al., 2020) OR Unclear (Neustupa et al., 2013)	*Leptochlorella*	CCAP 211/54
Trebouxiophyceae	Chlorellales	Oocystaceae	*Crucigeniella*	CCAP 218/1
Trebouxiophyceae	*incertae sedis*	Coccomyxaceae	*Coccomyxa*	CCAP 211/97
Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Trebouxia*	CCAP 219/5A
Chlorophyceae	Chlamydomonadales	Chlamydomonadaceae	*Chlamydomonas*	CCAP 11/32A
Ulvophyceae	Ulvales	Ulvaceae	*Ulva*	NA

The PCR reaction was performed using cells snap frozen in a 0.2 M solution of NaOH. Amplicons were observed for all three samples that are clearly members of the Chlorellaceae family, as well as some more distantly related taxa (Figure 3-10). The most distantly related strain that could be detected by the PCR reaction was *Coccomyxa*. However, CCAP 211/54 (also known as HA3) produced a band unreliably – no bands were produced in later reactions (Figure 3-10A), despite being reliably produced in initial reactions (Figure 3-10B).

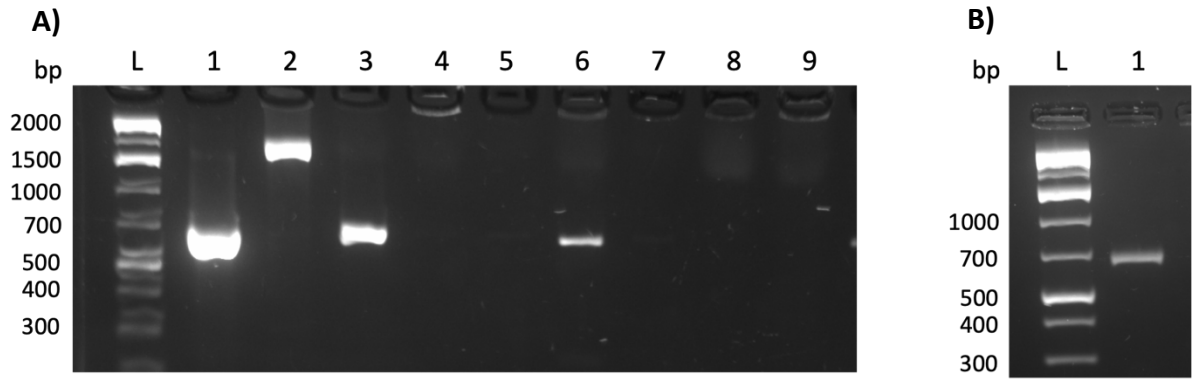


Figure 3-10. PCR products following amplification of the cytb gene from a wide range of algal taxa, using snap frozen cells as a template. A) Lanes contain 1) Prototheca wickerhamii*; 2)* Chlorella vulgaris *CCAP 211/11B; 3)* Keratococcus bicaudatus *CCAP 223/1; 4)* Leptochlorella *sp. CCAP 211/54; 5)* Crucigeniella apiculata *CCAP 218/1; 6)* Coccomyxa galuniae *CCAP 211/97; 7)* Trebouxia decolorans *CCAP 219/5A; 8)* Chlamydomonas reinhardtii *CCAP 11/32A; 9)* Ulva intestinalis*. Reaction was performed in May 2023. B) Lane 1 contains* Leptochlorella *sp. CCAP 211/54. Reaction was performed in June 2022. The ladder for both gels (L) is ThermoFisher GeneRuler 1 kb Plus DNA Ladder. Gel was 3% TAE run at 70 v for 60 minutes.*

The amplicon for CCAP 211/54, when an amplicon was produced, was sequenced. This amplicon had no significant similarity with any Trebouxiophyceae sequences using a search for highly similar sequences (megablast). With a search for slightly similar sequences (BLASTn), the best match is with *Trebouxia* sp. A1-2 mitochondrial DNA (percent identity 80.55%). There are matches with *Chlorella* species at about 79.1% identity. Thus, we can be confident that this strain is not an isolate of *Auxenochlorella*, though the amplicon sequence is insufficient to identify it as *Leptochlorella.*

Discussion

In this chapter, I have designed a protocol to purify algal DNA from milk samples and used this protocol to identify *Prototheca* in milk samples to species level, for the first time in the UK. I have also identified non-specificity in the previously published *cytb* PCR protocol.

The presence of *Prototheca* in UK milk

The presence of *Prototheca* could only be confirmed for two of the eight samples provided by the NML. Even these samples were culture negative, necessitating PCR-based methods to detect *Prototheca* in them. Milk samples from the NML were preserved with an antimicrobial (bronopol) before being sent to the NML, and were thawed and refrozen an unknown number of times. This was likely sufficient to kill any *Prototheca* cells in the samples before an attempt to recover them was made. However, bronopol appears to not completely inhibit the detection of *Prototheca* by molecular methods, as demonstrated here and elsewhere (Bauman et al., 2018; Magro et al., 2023).

The source of this discrepancy remains unclear. qPCR methods have similar lower levels of detection as seen for the method observed here (Bacova et al., 2021), both of which are lower than densities that are often seen in bulk milk (Libisch et al., 2022). Possible explanations for false negatives include degradation of the sample over time or during freeze/thaw cycles, or as a result of reactive oxygen species that may be produced by bronopol (Shepherd et al., 1988). Possible explanations for false positives include contamination by other algae, which is discussed below. However, given the number of samples that the NML processes daily, the worst-case-scenario of six false positives over six months would still indicate a low false positive rate overall.

It would be useful for future work to determine which of these explanations contribute to the discrepancy, as they would likely have implications for international surveillance of *Prototheca* and other mastitis pathogens. The extent of degradation of DNA following freezing and thawing, different methods of storage, or bronopol preservation on the time

scales seen here could all be investigated. Additionally, sequencing steps could be introduced when *Prototheca* is detected in milk samples by qPCR, which would be a useful addition to a national milk survey, or samples could be verified by a collaborator, as has occurred here.

All eight of the samples that tested positive for *Prototheca* by the NML are from the West of Great Britain. Only two of these samples could be confirmed to contain *Prototheca* DNA, one of which came from Wales and the other from Southwest England. However, it is difficult to extrapolate from these cases what conditions might predispose *Prototheca* infection/presence, both because there are relatively few cases and because these regions contain the highest density of dairy cattle in Great Britain. Additionally, the NML does not receive samples from Scotland or Northern Ireland.

To the best of my knowledge, the two milk samples from the NML from which *Prototheca* DNA could be isolated represent the first publicly available evidence of *Prototheca* infection of cattle in the UK, and the first time *Prototheca* infections have been identified to species level. However, it is not surprising that *Prototheca* is present in the UK as cases of cattle infected with *Prototheca* have been detected on all continents except for Antarctica (Asfour and El-Metwally, 2010; Shave et al., 2021). I identified my two positive strains as *P. bovis* and *P. ciferrii*. The presence of *P. bovis* is unsurprising, as it is the cause of most *Prototheca* infections in cattle (Toyotome and Matsui, 2022; Tashakkori et al., 2022; Jagielski et al., 2019b).

These findings are technically not a true confirmation of infection, as bulk tank milk can easily get contaminated by environmental microbes (Zecconi et al., 2020). *Prototheca*, especially *P. ciferrii*, is often present in farm environments, while *P. ciferrii* is only rarely associated with pathology (Osumi et al., 2008; Jagielski et al., 2019b; Ahrholdt et al., 2012; Bozzo et al., 2014). Thus, it is not unreasonable to suspect contamination, unless future studies identify a more prominent role for *P. ciferrii* in bovine protothecosis in the UK compared to other countries.

It is unlikely that these results are representative of the UK-wide burden of *Prototheca* presence or infection. Primarily because only two samples could be confirmed to contain *Prototheca*. Additionally, because bias may have been introduced in the selection of samples that made it to us – milk samples are only tested by qPCR for pathogens at the request of the farmers, not as a randomised selection, and only samples which had been tested this way were provided to us. Thirdly, both of these samples are of bulk milk (milk from multiple individuals that had been pooled), so it is unclear how many individual cows are suffering a *Prototheca* infection.

There has been some speculation about coinfection being an important factor for *Prototheca* infection. All bulk milk samples received from the NML tested positive for multiple pathogens (by the NML), including *Staphylococcus* and *Streptococcus*, but it is impossible to say whether these were from the same individuals as provided the *Prototheca* DNA.

Now that it appears that *Prototheca* are affecting British cattle, it would be beneficial to investigate the nation-wide burden of *Prototheca* infection. The difficulties encountered in confirming these cases of bovine *Prototheca* infections suggests that surveillance for *Prototheca* in the UK could be significantly improved.

An improved method of purifying algal DNA from milk

Relatively few national studies of *Prototheca* incidence have been done, and they tend to produce inconsistent results. Several methods of diagnosis and identification of *Prototheca* exist, which I group into culture- and PCR-based methods, both of which have benefits and drawbacks as they are currently used.

Culture-based methods have the advantage of producing viable strains, which can be tested further (e.g. for antimicrobial resistance), as well as selecting against contaminants that might produce false positives. However, it seems likely that the choice of growth conditions and media could impact the detection of *Prototheca*. Krukowski *et al.* (2020) detected *Prototheca* in 1.2% of Polish mastitis cases, which included both clinical and subclinical mastitis, using culture based methods (Krukowski et al., 2020). Jagielski *et al.* (2019) also used culture based methods in their national study of Poland, isolating *Prototheca* from 38.1% of clinical mastitis cases and 11.4% of subclinical mastitis cases (Jagielski et al., 2019b). However, Jagielski *et al.* (2019) used a variety of media including PIM to select for *Prototheca*, to avoid algal colonies being overgrown by fungal or bacterial colonies, while Krukowski *et al.* (2020) only used sheep blood and Sabouraud dextrose agar to identify *Prototheca*. Thus, I suspect the true incidence of *Prototheca* infections is closer to Jagielski's

estimate of 11-38% of cases than Krukowski's estimate of 1.2%. Additionally, culture-based methods are slow, requiring several days for colonies to grow.

PCR-based methods have the advantage of detecting *Prototheca* in samples that contain no viable cells (Bacova et al., 2021), as well as being significantly faster than culture-based methods, but the choice of lysis method can significantly impact which environmental DNA is detected (Deiner et al., 2015). The qPCR kits that I am aware of, including the one used by the NML, rely on lysis buffers, usually using proteinase K and sometimes thermal shock, to release DNA from *Prototheca* cells (Bauman et al., 2018; ThermoScientific, 2015; Zecconi et al., 2020). National studies using these methods tend to find *Prototheca* in 6-11% of herds. However, *Prototheca* cells are extremely robust, due to the layer of sporopollenin in the cell wall of *Prototheca* which is discussed in more detail in Chapter 6) Genome Sequencing and Assembly of *Prototheca* (and Related Species), and may not reliably be disrupted by either proteinase K or thermal shock. Other suggested methods use kits to extract DNA for subsequent detection of *Prototheca*, which may more reliably extract *Prototheca* DNA, but these methods tend to require a large volume of milk to extract material from (Bacova et al., 2021). Another potential disadvantage, as discussed below, is that PCR reactions designed to detect every member of the *Prototheca* genus can often detect other algal genera.

The detection method developed for and described in this chapter addresses the problem of lysis for PCR-based identification. It is as sensitive as the most sensitive qPCR methods (Bacova et al., 2021), which might be improved further (Odumeru et al., 2001), and requires much less material than methods that involve using kits for extraction (Liao and Liu, 2020;

Bacova et al., 2021). Additionally, it is a molecular-based method, meaning it is not dependent on viable cells being present in the milk. Purified DNA may also be easier to sequence, if a sample tests positive, which would address the problem of specificity.

The main disadvantage of this method is that it uses phenol, a hazardous chemical. Additionally, the equipment used here required significant manual handling, though bead beaters/tissue lyser machines that can handle more samples or may be loaded robotically may be available. Both of these may make this method difficult to scale up. This trade-off between sensitivity and specificity against scale has been observed elsewhere, with the 'lytic' method described by Quigley *et al.* providing the highest DNA yield, but requiring the most labour (Quigley et al., 2012).

Non-specificity of genus-wide PCR reactions

The PCR reaction to amplify *cytb* has been identified here as not specific to the AHP algae. The *cytb* protocol detected all algal samples in the same family as *Prototheca*, as well as some more distantly related taxa. Another PCR reaction capable of detecting all *Prototheca* species has been reported to be able to detect algae in different classes (Bacova et al., 2021). This non-specificity across multiple reactions suggests a significant risk of PCR reactions designed to detect the whole *Prototheca* genus detecting other algal genera.

Contamination may arise from airborne algae which are not removed by routine disinfection protocols, as they may not have been tested for algae, or environmental DNA in sources of water. Non-AHP amplicons, at least of *cytb*, are usually the same length as the amplicons of

Prototheca (with the surprising exceptions of HA6 and *Chlorella vulgaris*), meaning they are indistinguishable from *Prototheca* without sequencing.

It is not possible to accurately predict just from relatedness which algae will be detected, as *Coccomyxa* produced a band in the *cytb* reaction and *Crucigeniella* did not despite *Crucigeniella* being more closely related to *Prototheca*. Additionally, *Helicosporidium* is inside the AHP lineage, but is not detected by the *cytb* PCR protocol. Thus, qPCR reactions designed to detect the *Prototheca* genus should be checked for non-specific amplification from a variety of algal taxa, and safeguards to detect false positives from contaminants should be developed.

However, it is not clear that contamination is necessary to account for the discrepancy between which samples were found to contain *Prototheca* in this chapter and by the NML. The *cytb* PCR reaction produced amplicons using material of CCAP 211/54 as a template several months after its arrival, but was eventually unable to produce a band. This sample was kept at -20°C for the intervening time between the last unsuccessful amplification and the first unsuccessful amplification. This suggests that DNA can degrade in samples kept at -20°C, on the time scales relevant to my verification of the detection of *Prototheca* in NML samples.

Identification of strains in collection

It is disappointing, though not necessarily surprising, that *cytb* appears to not have sufficient diversity to distinguish between *Auxenochlorella* species, making it a poor candidate for species identification of the whole AHP lineage. For similar reasons, *cytb* is likely to not be

appropriate for resolving relationships between *Auxenochlorella* species. The use of other genes for this purpose is explored in Chapter 7) Phylogenetics of *Prototheca* and Related Organisms.

It is noteworthy that a reasonably large proportion (7.5%) of the strains in our collection had been misidentified. Whether this is reflective of wider collections remains unknown. It is perhaps indicative of the need to routinely use molecular identification methods to validate *Prototheca* strains.

4) Survival of Abiotic Stresses

Introduction

The accidental virulence hypothesis is premised upon environmental selection pressures inadvertently selecting for traits that confer pathogenicity. Thus, pathogenic and non-pathogenic species should be distinguishable by tolerance of at least one environmental factor.

Additionally, there is currently no clear way to identify whether species of *Prototheca* are pathogenic in the absence of case reports. For fungal taxa, pathogenic species can often be distinguished from non-pathogenic species based on growth under various abiotic challenges, as virulence traits usually provide some environmental benefit (Bhabhra and Askew, 2005; Cramer and Perfect, 2009). Thus, investigation of the tolerance of environmental factors may allow pathogenic species of *Prototheca* to be identified in the absence of case reports or a suitable animal model.

As for which stresses to investigate, tolerance of high temperatures and tolerance of low pH are promising starting points.

The Case for Thermotolerance

Mammalian hosts, as endotherms, elevate their body temperature above the temperature of their environment. The core temperature of a healthy human can range from 36°C to 37°C (Yousef et al., 2022), while the core temperature of a healthy cow can range from 38°C to 40°C (Normal Rectal Temperatures, n.d.; Suarez-Trujillo et al., 2022; Lees et al., 2019). Of

course, the temperature of more extreme parts of the body, like skin, is lower than the core temperature; human arm and leg skin tends to be in the range of 29-33°C (Lee et al., 2019). However, local inflammation from localised infection can raise local skin temperature by up to 5°C (Chanmugam et al., 2017). Fevers can raise body temperatures systemically, reaching over 39.5°C or 42°C in humans and cows respectively (Merck&Co, 2020; Eley, 2011).

Thermotolerance has been suggested as a necessary attribute for fungal infection in general (Robert and Casadevall, 2009; Köhler et al., 2015). Growth at body temperature is considered one of the main virulence factors for *Cryptococcus* (Kosmidis and Denning, 2017) and greater *in vitro* growth at 37°C is associated with greater *in vivo* virulence in *Aspergillus* (Cramer and Perfect, 2009; Paisley et al., 2005).

While environmental temperatures are usually below mammalian internal temperatures, they can be much higher in certain locations, providing opportunities to develop thermotolerance in the environment. Decomposing organic matter in compost piles can reach 65°C (Fernandes et al., 1994), and growth in compost is thought to contribute to the thermotolerance of *Aspergillus* (Bhabhra and Askew, 2005). Water heated in geysers or hot spring systems can reach local boiling temperatures (i.e. 70-100°C) (Jones et al., 2003).

If thermotolerance is necessary and sufficient for pathology, clinical strains or pathogenic species should all have higher growth rates at higher temperatures. If there is a threshold temperature above which clinical strains or pathogenic species grow and below which non-clinical strains or non-pathogenic species die, we might also expect this threshold to be

around the core temperature of their respective host (i.e. 37°C for human strains, 38.5°C for cattle strains).

The Case for Acid Tolerance

A connection between endosymbiosis and pathology has been proposed for other organisms, with both hypothesised to arise from the other in some contexts or as two strategies on the same continuum (Drew et al., 2021; Stat et al., 2008). In this case, it is the transition from photosynthetic endosymbiont to parasite that is of interest to us.

Isolates of both *Chlorella* and *Auxenochlorella* are able to form endosymbiotic relationships with a variety of protist lineages as well as some animal phyla (Huss et al., 1994; Kodama and Fujishima, 2005). Well studied hosts include *Hydra viridis* and *Paramecium bursaria.* Acid tolerance, specifically the ability to grow below pH 3.5, is linked to endosymbiosis for those relatives (Huss et al., 1994). For context, while soil pH can reach lower than pH 2, any soil with a pH of less than 4.5 is considered extremely acidic (Soil Survey Division Staff, 2017).

There appears to be no recognition process between a potential host and symbiont, at least for *Chlorella*; the potential symbiont is ingested and the potential host attempts to digest it (Kodama et al., 2007). If an ingested algal cell survives, it becomes an endosymbiont. An endosymbiont's ability to survive digestion may be widely applicable, as *Chlorogonium* symbionts from one Opisthokont protist were able to infect a protist from the SAR lineage (Kawai et al., 2018). Once established as an endosymbiont, the host must control the algal cells, which appears to be done actively via control of access to nutrients (Muscatine and

McNeil, 1989; Hamada et al., 2018; Lowe et al., 2016). Failure of a host to control the internalised algae is often fatal for the host cell. If interactions with host phagocytic cells is analogous to interactions with potential symbiotic hosts, ancestral capacity to establish endosymbiosis may contribute to virulence in *Prototheca* in terms of survival and damage.

It may seem unlikely that *Prototheca* would maintain the machinery to establish endosymbiosis, as it cannot photosynthesise and therefore cannot contribute to such a relationship. However, it is not clear that the relationship between *Chlorella* symbiont and its host is as mutualistic as was first assumed. At least one host appears to exploit its algal symbionts, reducing their fitness under every tested condition (Lowe et al., 2016). The only benefit to the algal cells appeared to be that they sometimes escape their host (Lowe et al., 2016). Thus, it is possible that endosymbiosis evolved in algae as an anti-predator defence. This might explain the persistence of endosymbiotic machinery in *Prototheca*, even after losing photosynthesis.

If endosymbiotic machinery is relevant for pathology and acid tolerance is predictive of that machinery, it may be that acid tolerance is predictive of pathology. The *H. viridis* phagolysosome acidifies to around pH 3.5-4 (Huss et al., 1994) and the *P. bursaria* reaches pH 2.4-3, thus being able to survive these low pHs is required for prospective endosymbionts. Immune phagolysosomes appear to reach a range of pH levels, from pH 3.5 (Murphy et al., 2012b) to pH 5 (Uribe-Quero and Rosales, 2017; Erwig and Gow, 2016; Kho et al., 2021); the pH tolerance threshold to identify pathogens should therefore exist in this range.

Possible nuances

Both acid tolerance and thermotolerance have been investigated in the past. Thermotolerance in particular is a common stress to be investigated, with authors often looking at growth at 37°C or higher (Masuda et al., 2016; Kano et al., 2022; Jagielski et al., 2022, 2019a). A lack of thermotolerance has also been used to rule out pathology in *P. paracutis* (Kunthiphun et al., 2019).

However, these results are usually reported categorically. For example, it is common to "+" to indicate growth and "-" to indicate the absence of growth, with possible additional symbols like "+/-", "W" or "weak" to indicate weak growth, or "variable" or "V" to indicate species where different strains could behave differently (Jagielski et al., 2019a; Kano et al., 2022; Marques et al., 2010; Nagatsuka et al., 2017; Jagielski et al., 2022). This makes identification of differences, which might become predictive thresholds, between pathogenic and non-pathogenic species more difficult.

It is also not clear whether all individuals of a given pathogenic species are equally pathogenic or if, as in other pathogens, there are specific lineages within each pathogenic species that are more virulent (Riley, 2014; Beale et al., 2015). It has been noted that *Prototheca* are abundant in the environments of their host organisms (i.e. humans and cattle) (Jagielski et al., 2022; Pore et al., 1983; Jagielski et al., 2019c; Osumi et al., 2008). Despite this abundance, *Prototheca* infections are rare. In fact, it has been noted that the environmental presence of *Prototheca* is surprisingly poorly associated with the rate of infections in several cattle herds, even when the environmental isolates are from pathogenic species (Jagielski et al., 2019b).

This could be explained in numerous ways, but one possible explanation is the existence of as-yet unknown pathogenic lineages existing within the pathogenic species which are responsible for most cases of infection. There is reason to suspect there may be pathogenic lineages within species of *Prototheca*: some strains of *C. vulgaris*, *C. sorokiniana* and *C. kessleri* strains can infect *Paramecium* while others cannot (Kodama et al., 2007), meaning infectivity of *Chlorella* is strain specific. Different isolates of *P. wickerhamii* were observed to cause different amounts of cytotoxicity in macrophages, and presented obviously different colony morphology, indicating that different strains of the same *Prototheca* species can have different pathogenic capacities, though these isolates were all clinical isolates (Guo et al., 2023).

In this chapter, I have therefore chosen to investigate these two abiotic stresses in pathogenic and non-pathogenic species of *Prototheca*. In addition, to investigate the possibility of pathogenic and non-pathogenic lineages within species, I have included clinical and environmental strains of pathogenic species.

Methods

Optical Density Growth Curves

Precultures, grown as described in the core methods chapter, were normalised to an optical density at 600 nm (OD_{600}) of 1, using a BioPhotometer Plus (Eppendorf), which was then used to inoculate 500 µl experimental cultures of SAB broth in a 48-well plate to an OD_{600} of 0.05. An example layout for a 48-well plate exposed to temperature stress is provided in Figure 4-1.

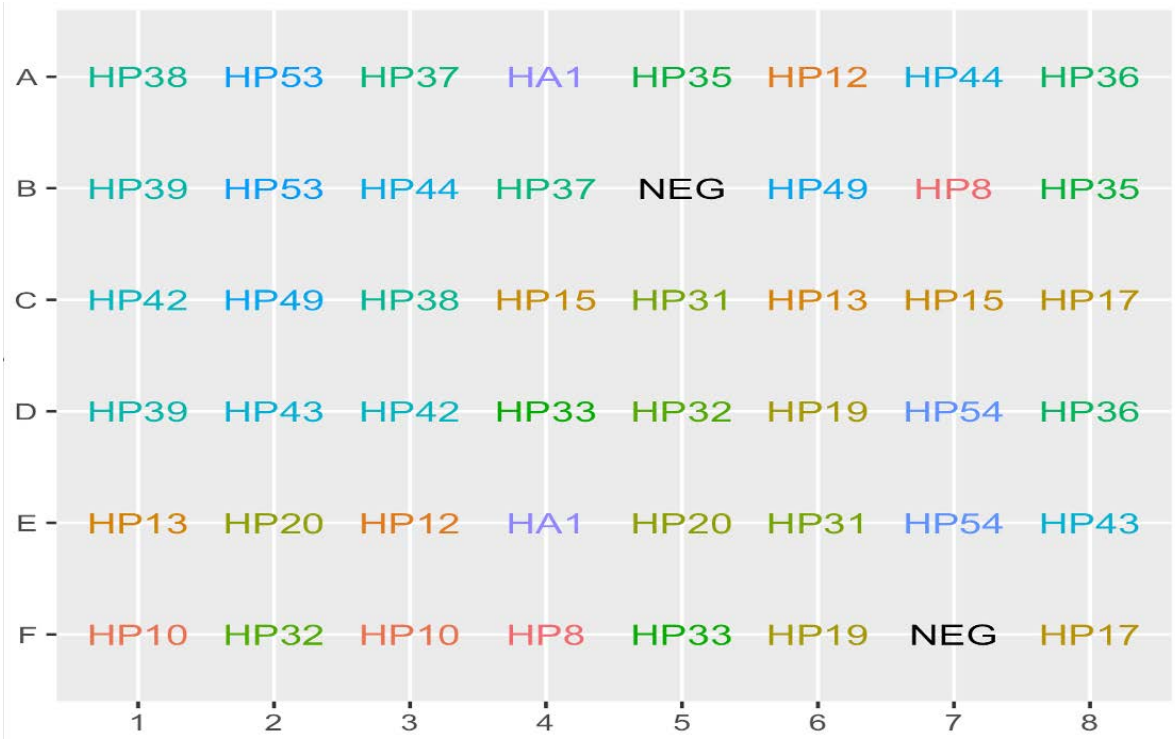


Figure 4-1. An example layout for a 48-well plate containing randomly distributed cultures exposed to temperature stress. Strains are coloured arbitrarily.

48-well plates containing experimental cultures were incubated within a FLUOstar Omega Microplate Reader (BMG Labtech; henceforth referred to as "plate reader") for at least 60 hours, with optical density (OD) measurements taken every 30 minutes at both 600 nm and 750 nm. 600 nm is a standard wavelength for measuring the OD of bacteria (Ratzke and Gore, 2018), and has been used for *Prototheca* in the past (Kwiecinski, 2015; Ely et al., 2019). However, there have been reports that a poor choice of wavelength can lead to poor estimates of cell density in photosynthetic algae, which can produce different amounts of pigments under different conditions (Griffiths et al., 2011). 750 nm has been recommended for photosynthetic microalgae, as the wavelength that has the least error (of the wavelengths tested, which did not include 600 nm) (Griffiths et al., 2011). OD_{750} has been used for *Chlorella* (Treves et al., 2013; Gerken et al., 2013). As these experiments were to include *Auxenochlorella* species, which produce photosynthetic pigments, it seemed

prudent to measure OD at 750 nm as well as 600 nm. The specific growth conditions of experimental cultures depended on the stress factor being examined.

To assess the effect of temperature stress, 48-well plates were incubated at 30°C, 33°C, 37°C, or 42°C in the plate reader, or were left at room temperature, which ranged from 25°C to 28°C at the time of the experiments.

To assess the effect of acid stress, broth was controlled to pH 2, 3, or 4 by the addition of concentrated HCl (Thermofisher), as measured by an 8100 pH and Temperature Meter (ETI), before autoclaving, to be used in experimental cultures. 48-well plates were incubated at room temperature for all strains, unless otherwise specified. The pH of a subset of wells was tested after growth using pH Test Strips 0.0-14.0.

For both temperature stress and acid stress, most cultures were grown in SAB. However, to check whether medium choice impacted survival, some cultures were grown in YPD instead.

The number of replicates performed for strains of different species was variable. The primary goal was to perform comparisons between species; it was therefore considered important to capture variation from a greater number of strains, where possible, to better represent the species as a whole. For species with three or more strains in our collection, growth was usually measured for two replicates per strain. Additionally, strains were chosen to include both environmental and clinical isolates where possible. Where there were only one or two strains of a given species in our collection, growth was measured for three replicates per strain.

To assess survival after incubation, 10 µl volumes of culture was taken from wells and spotted onto PIM agar and incubated at 25°C. Images were taken of these plates after approximately 2, 4 and 7 days. Images of agar plates and 48-well plates were taken with a ChemiDoc MP Imaging System (Bio-Rad).

Growth curves of optical density measurements were plotted using the loess method. Prediction intervals were calculated using the standard deviation of the data, only using data from eight-hour intervals.

Colony Forming Unit Growth Curves

Colonies from stock plates were used to inoculate SAB broth precultures, which were grown at 25°C for 72 hours in volumes of 2 ml, in 14 ml tubes, rotating at 20 rpm. Precultures were used to inoculate experimental cultures to a cell density of 10^6 cells per ml (using cell counts from a Corning Cell Counter (Corning)).

Experimental cultures were incubated at 25°C or 37°C in volumes of 2 ml, in 14 ml tubes, without shaking or rotation. After 0, 24, 48, 72, and 96 hours, experimental culture was taken and serially diluted 1:10, 1:100, 1:1,000, and 1:10,000. 15 µl of diluted culture was streaked onto PIM agar and incubated at 25°C for up to 10 days. Colony forming units (CFUs) were counted for each dilution and averaged to calculate cells per ml of the original culture. Where possible, colony counts over 250 and below 10 were excluded (i.e. where there were colony counts within this range from other dilutions) before calculating the mean and cells per ml, to remove inaccurate counts.

Quantitative Comparisons

To quantitatively compare growth, as measured by optical density, OD data was fitted to a logistic model using the growthcurver R package (Sprouffske and Wagner, 2016), allowing the calculation of summary statistics like carrying capacity, considered to be the height of the plateau in the stationary phase of growth, and growth rate.

Two-way analysis of variance (ANOVA) was used to detect significant differences in these statistics between strains of a given species. Strains that were significantly different from at least half of the rest of their species were removed from subsequent analysis, unless specified otherwise.

Two-way ANOVA was also used to detect whether temperature or pH interacted with species to affect carrying capacity and growth rates. To identify which species were affected by the applied stress, t tests were used to compare carrying capacity and growth rates at relevant thresholds (i.e. 30°C and 37°C, pH 3 and pH 4) within each species. Due to failure to fit models for some replicates, and subsequent inability to produce summary statistics, OD measurements taken at 0 and 60 hours from each well were compared via paired t tests to detect any growth at 37°C.

Quantitative comparisons of CFU counts were made using t tests. CFUs following exposure to 37°C and 25°C were compared for each time point and strain separately.

The Benjamini-Hochberg procedure was used to correct for multiple comparisons for all quantitative comparisons.

Results

Strains generally grow consistently within a species under temperature stress

To determine whether tolerance of high temperatures is associated with pathology, 45 strains were grown under temperature stress and had their growth measured as OD_{750} and OD_{600}. OD measurements were used to draw growth curves with the loess method.

A species was considered "pathogenic" if at least one isolate was derived from an infection or a credible case report of infection by that species could be found. Otherwise, species were considered non-pathogenic.

At 37°C, pathogenic and environmental species show visual differences in their growth (Figure 4-2). All environmental species appear to grow much less well at 37°C compared to 30°C. Impaired growth of some pathogenic species, e.g. *P. wickerhamii* and *P. ciferrii*, at 37°C relative to 30°C is also seen, but not to the same extent as for environmental species. All strains grew at "room temperature", an uncontrolled temperature that often ranged from 25°C to 28°C (Figure 4-2). All strains grew at 30°C, but *P. xanthoriae* and *P. moriformis* strains appear to grow less well than at room temperature.

Before growth at 37°C could be investigated quantitatively, it had to be verified that strains within a species grew similarly. The number of replicates per condition per strain is available in Appendix 5.

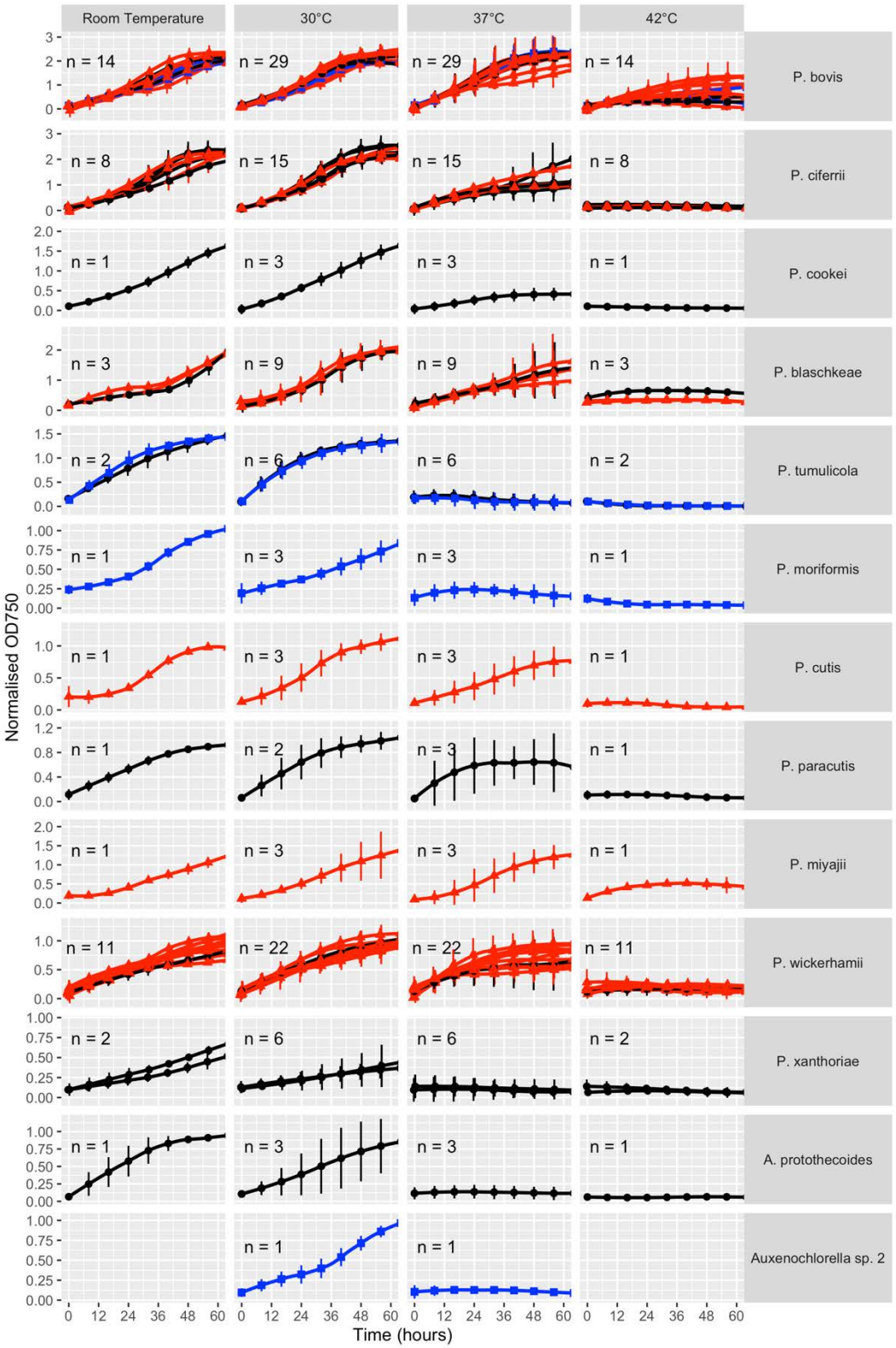


Figure 4-2. Growth curves of Prototheca *and* Auxenochlorella *strains under temperature stress, as measured by OD750. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. The variable number of replicates at each temperature is indicated in the top left of each graph and is broken down further in Appendix 5. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. A strain's origin is unknown if, when given the strain, we were not given information regarding the origin of the strain, or if there is evidence that the strain was swapped. Bars show the 95% prediction interval (mean ± 2 standard deviations) for a given strain at 8-hour intervals, slightly offset for improved visibility.*

Quantitative comparisons between strains and between species were performed using values calculated using the growthcurver R package, which fits measurements to the logistic equation to derive biologically meaningful statistics. However, not all growth curves could be fitted to a form of the logistic equation. Some could be fitted, but produced spurious statistics, including unreasonably high carrying capacities or growth rates. Examples of growth curves with extreme statistics are provided (Figure 4-3). Many of these errors were more likely under conditions where a strain failed to grow or grew poorly.

Three statistics were chosen to remove outlier models: carrying capacity in excess of 30, growth rate in excess of 1.5, and residual standard deviation (Sigma) in excess of 0.2. Models that produced spuriously high growth rates typically had very low carrying capacities, and models with spuriously high carrying capacities tended to have very low growth rates (Appendix 6).

Following removal of outliers, most models could be used. The number of models that failed each threshold is summarised in Table 4-1, and is shown in detail in Appendix 7. Only one model from each of our two strains of *P. tumulicola* was not rejected, which did provide enough data to make statistical comparison possible but reduced the power of this comparison. The number of replicates for each species where at least one model was not rejected is provided in Appendix 8. Visually, there is little difference between OD growth curves using measurements taken at 600 nm or 750 nm (Appendix 9) and therefore statistical analyses were performed only on OD_{750} data.

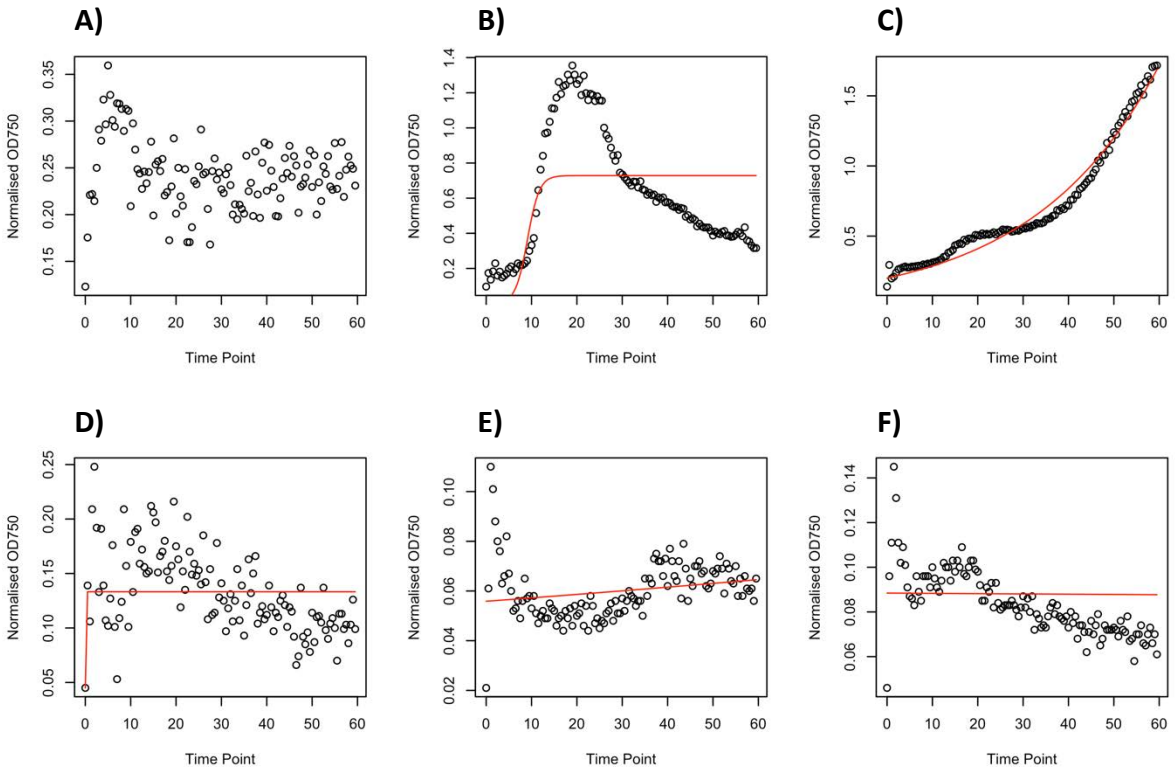


Figure 4-3. Examples of models with extreme statistics. Each graph represents data from one well from a single experiment. Each point is a measurement of OD750 taken at a given time point. Where the data could be fitted to a logistic model, the model has been plotted in red. Each graph demonstrates a different statistic indicating a poor fit. A) Complete failure to fit a model. B) Poor fit of model to data, with a sigma of 0.3. C) Unreasonably high carrying capacity, at 6.83×10^{7} OD units. D) Unreasonably high growth rate, at 40.7 OD units per hour. E) Excessively high generation time, at 285 hours. F) Negative inflection time, at 1419 hours before the start of the experiment.

Table 4-1. Summary of the number of wells exposed to temperature stress that failed to be fitted to a model, as well as the reason for failure. A more detailed breakdown is available in Appendix 7.

Statistic	Threshold	Number of Failures	Percent Failed
Carrying Capacity	30	21	3.34
Growth Rate	1.5	30	4.78
Sigma	0.2	2	0.32
Failure to Fit a Model	NA	12	1.91
All	All	65	10.35

Differences in carrying capacities and growth rates between strains within a species could be investigated for five species. Two-way ANOVA revealed significant differences in carrying capacity only for strains of *P. bovis* ($F = 6.465$, $p = 1.25\times10^{-5}$). This difference could entirely

be accounted for by HP18, which was significantly different from all 13 other strains (p values ranging from 4.96×10^{-7} to 1.40×10^{-4}) and was removed from subsequent analysis.

Two-way ANOVA revealed significant differences in growth rate between strains for *P. bovis* ($F = 3.712$, $p = 1.46\times10^{-3}$), *P. ciferrii* ($F = 36.76$, $p = 3.79\times10^{-9}$), *P. tumulicola* ($F = 15.2$, $p = 1.76\times10^{-2}$), and *P. wickerhamii* ($F = 95.416$, $p = 9.15\times10^{-13}$). However, no strain could be identified as significantly different from the rest of the species for three of these species. For *P. wickerhamii* and *P. bovis*, this was because very few pairwise differences were significant. For *P. tumulicola*, this was because we are only in possession of two strains, and thus are unable to tell which is more representative of the species. Six of the seven strains of *P. ciferrii* used were found to be significantly different from at least four other strains. However, removing these strains would remove almost all of the data for this species, so none were removed from further analysis.

Pathogenic species generally grow better than non-pathogenic species at 37°C

Two-way ANOVA revealed a significant effect of temperature ($F = 61.86$, $p < 2\times10^{-16}$) and species ($F = 37.367$, $p < 2\times10^{-16}$) on carrying capacity, as well as a significant interaction between the two ($F = 6.181$, $p = 1.26\times10^{-13}$). Two-way ANOVA also revealed a significant effect of temperature ($F = 28.2$, $p = 2.24\times10^{-15}$) and species ($F = 12.35$, $p = 1.17\times10^{-15}$) on growth rate, as well as a significant interaction between the two ($F = 13.26$, $p < 2\times10^{-16}$).

To determine which species had their growth affected by physiological temperatures, carrying capacities and growth rates at 30°C and 37°C were compared. Significant

differences were found in carrying capacities for five of the eight species for which comparisons were possible (Table 4-2).

Table 4-2. Comparisons of carrying capacities of Prototheca *species grown at 30°C and at 37°C. Rows containing pathogenic species are highlighted in pink. Comparisons for some non-pathogenic species are missing due to models being rejected.*

Species	Mean Carrying Capacity at 30°C	Mean Carrying Capacity at 37°C	Difference between means	Proportional difference	Adjusted p value	Significance
P. bovis	2.28	2.32	-0.0360	-0.0158	6.84×10^{-1}	NS
P. ciferrii	2.47	1.38	1.09	0.44	5.69×10^{-4}	***
P. cookei	1.89	0.412	1.48	0.783	3.64×10^{-2}	*
P. blaschkeae	2.52	1.53	0.987	0.392	3.64×10^{-2}	*
P. tumulicola	1.30	0.211	1.09	0.838	6.88×10^{-6}	***
P. moriformis	1.21	0.230	0.977	0.81	1.53×10^{-1}	NS
P. cutis	1.08	1.16	-0.0747	-0.0689	8.37×10^{-1}	NS
P. paracutis	1.01	0.786	0.220	0.219	2.76×10^{-1}	NS
P. miyajii	1.50	1.26	0.244	0.162	6.73×10^{-1}	NS
P. wickerhamii	1.03	0.747	0.283	0.275	6.71×10^{-5}	***
P. xanthoriae	0.652	0.112	0.54	0.828	7.50×10^{-2}	NS
*A. prototheco*ides	1.5	0.128	1.37	0.915	1.84×10^{-1}	NS

Carrying capacity was not significantly affected for *P. bovis*, *P. miyajii*, and *P. cutis*, all of which are pathogenic species. For the remaining pathogenic species, the magnitude of the difference was relatively small, at less than 50% of the carrying capacity at 30°C. Four non-pathogenic species were found to not have a significant difference between carrying capacity at 30°C and 37°C: *P. paracutis, P. moriformis, P. xanthoriae*, and *A. prototheco*ides. However, the percentage decrease of carrying capacities for *P. moriformi , P. xanthoriae*, and *A. prototheco*ides was greater than or equal to 81%, which is comparable to *P. cookei* and *P. tumulicola*, which are found to have significantly different carrying capacities. The impact of 37°C on the carrying capacity of *P. paracutis* genuinely appears limited, with a

proportional decrease of only 22% (less than both *P. wickerhamii* and *P. blaschkeae*). This was surprising, as discussed below.

When growth rate at 30°C was compared to growth rate at 37°C for each species, significant differences were only observed for *P. wickerhamii* (Table 4-3), though this difference was small. Approximately tenfold differences in growth rates were observed for *P. tumulicola* and *P. moriformis*, with the higher rates seen at 37°C.

Table 4-3. Comparisons of growth rates of Prototheca *species grown at 30°C and at 37°C. Rows containing pathogenic species are highlighted in pink. Comparisons for some non-pathogenic species are missing due to models being rejected.*

Species	Mean Growth Rate at 30°C	Mean Growth Rate at 37°C	Difference between means	Proportional difference	Adjusted p value	Significance
P. bovis	0.110	0.107	0.00240	0.0218	7.49×10^{-1}	NS
P. ciferrii	0.106	0.107	-0.00116	-0.0109	9.22×10^{-1}	NS
P. cookei	0.0778	0.136	-0.0580	-0.749	3.14×10^{-1}	NS
P. blaschkeae	0.0869	0.0767	0.0102	0.117	6.08×10^{-1}	NS
P. tumulicola	0.120	1.11	-0.991	-8.26	6.08×10^{-1}	NS
P. moriformis	0.0390	0.906	-0.867	-22.3	2.20×10^{-1}	NS
P. cutis	0.0965	0.068	0.0285	0.0295	4.00×10^{-1}	NS
P. paracutis	0.128	0.0946	0.0338	0.264	6.08×10^{-1}	NS
P. miyajii	0.0922	0.124	-0.0314	-0.341	6.08×10^{-1}	NS
P. wickerhamii	0.0749	0.150	-0.0752	1.00	1.76×10^{-3}	**
P. xanthoriae	0.0429	0.089	-0.0461	1.07	6.65×10^{-1}	NS
A. prototheccoides	0.0754	0.146	-0.0705	0.936	7.02×10^{-1}	NS

Due to the difficulties fitting models, a more direct method of measuring growth at 37°C was attempted. OD_{750} at 0 hours was compared directly to OD_{750} at 60 hours for all wells, not just those that produced acceptable models. Significant differences were observed in OD between 0 hours and 60 hours incubation at 37°C for all species except for *P. xanthoriae* (Table 4-4), suggesting that all species grew at 37°C. However, these differences were

smaller for non-pathogenic species (upper 95% confidence limit <0.5) than for pathogenic species (lower 95% confidence limit > 0.55), except for *P. paracutis*.

In general, it appears that pathogenic species grow better at higher temperatures, but it is difficult to find statistics to clearly separate these two groups.

Table 4-4. Growth of Prototheca *and* Auxenochlorella *species grown at 37°C, as determined by the difference in OD750 following 0 hours and 60 hours growth at 37°C. Rows containing pathogenic species are highlighted in pink.*

Species	Growth at 37°C	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Adjusted p value	Significance
P. bovis	2.15	2.23	2.07	3.23×10^{-47}	***
P. ciferrii	1.19	1.35	1.02	1.71×10^{-14}	***
P. cookei	0.344	0.437	0.252	1.67×10^{-4}	***
P. blaschkeae	1.22	1.44	0.994	2.98×10^{-9}	***
P. tumulicola	0.0404	0.0797	0.00115	4.85×10^{-2}	*
P. moriformis	0.104	0.198	0.0103	4.18×10^{-2}	*
P. cutis	0.731	0.846	0.615	9.18×10^{-6}	***
P. paracutis	0.531	0.824	0.238	5.87×10^{-3}	**
P. miyajii	1.17	1.35	0.984	9.18×10^{-6}	***
P. wickerhamii	0.635	0.691	0.578	9.05×10^{-25}	***
P. xanthoriae	0.00979	0.0377	-0.0181	4.63×10^{-1}	NS
*A. prototheco*ides	0.0794	0.110	0.0484	1.15×10^{-3}	**

Visual assessment of growth curves (Figure 4-2) appear to indicate that *P. paracutis*, *P. moriformis* and *P. cookei* grow initially at 37°C, but plateau earlier than pathogenic species. It was considered that OD growth curves would be unable to detect the difference between strains that had died or persisted at low ODs, or grew very slowly. Additionally, *P. moriformis* appears to decline after a peak at 24 hours. It was considered that OD growth curves may struggle to capture the difference between strains that could grow initially, but

subsequently died, and strains that grew consistently but poorly. To distinguish between some of these alternative explanations, it was necessary to check which strains were still alive, following exposure to temperature stress.

Most species survive exposure to 37°C for 3 days

Following growth curve experiments, 10 µl of culture from 48-well plates was spotted onto PIM agar and incubated at 25°C to test for survival. Images showing a subset of strains, representing every species in our collection (except for the species represented by HA6) plus HP18 and the other strain of *P. xanthoriae* (as these strains were found to behave differently from others of their species), are provided Figure 4-4.

All strains survive exposure to 25°C. This was expected, as all precultures were grown at 25°C, meaning that a strain that could not grow at 25°C would not be included in the growth curves.

P. xanthoriae is the only species that does not reliably survive exposure to 30°C. On some occasions, colonies did form following exposure to 30°C, but these colonies were sparse compared to other strains. However, there appear to be subtle differences between the two strains of *P. xanthoriae* tested, as HP54 appears to survive higher temperatures to a greater extent than HP53.

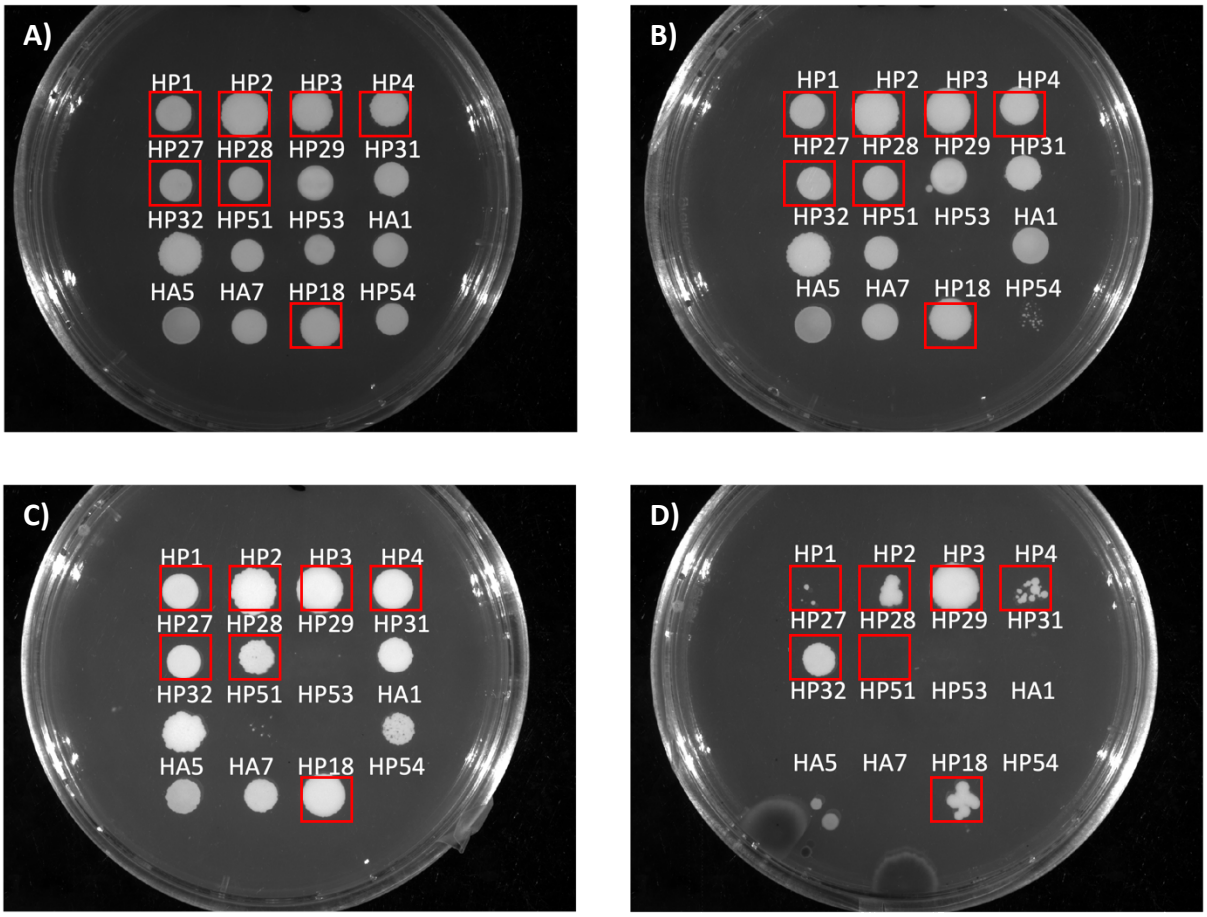


Figure 4-4. Agar plates showing survival of Prototheca *and* Auxenochlorella *strains following approximately 72 hours of exposure to temperature stress. Not all strains tested are shown, instead one strain from each species has been chosen. 14 species are represented by these 16 strains: HP1* – P. wickerhamii*; HP2* – P. ciferrii*; HP3 and HP18* – P. bovis*; HP4* – P. blaschkeae*; HP27* – P. miyajii*; HP28* – P. cutis*; HP29* – P. tumulicola*; HP31* – P. paracutis*; HP32* – P. cookei*; HP51* – P. moriformis*; HP53 and HP54* – P. xanthoriae*; HA1* – A. protothecoides*; HA5* – A. symbiontica*; HA7* – Auxenochlorella *sp. 2. HP18 and HP54 were included as they do not behave the same as the other strains of their species. The temperature stresses explored were: 25°C (A), 30°C (B), 37°C (C), and 42°C (D). Pathogenic species are indicated by a red box.*

In addition to *P. xanthoriae*, two other species do not survive 37°C: *P. moriformis* and *P. tumulicola*. These are all non-pathogenic species. However, *P. paracutis* and *P. cookei* clearly survive exposure to 37°C, despite the significantly lower plateau observed for *P. cookei* in the OD growth curves. Additionally, *Auxenochlorella* species tested survived exposure to 37°C inconsistently.

The only species to reliably survive exposure to 42°C are *P. bovis* and *P. miyajii*. *P. blaschkeae* often produces a few colonies following exposure to 42°C, but these colonies are

sparse enough to suggest that the culture is dying even if it does not fully die after 60 hours of exposure.

Qualitatively, it was occasionally observed that strains of *P. wickerhamii* can struggle at 37°C, occasionally producing more colonies from culture grown in wells at the edge of the 48-well plate. However, this trend was not always observed – for example, HP1 growth is robust in Figure 4-4, despite the well not being at the edge of the 48-well plate.

Most non-pathogenic species die during exposure to 37°C

Spotting culture after three days to determine survival, while efficient, was crude and binary. It was possible that cultures were dying at 37°C, but not sufficiently to observe in the time frame allowed. There was also no ability to observe short-term growth and subsequent death, as might be suggested by OD growth curves. Therefore, to investigate the death of strains at 37°C further, growth curves using CFUs were also made for non-pathogenic species.

Following exposure to 37°C, five of six non-pathogenic species show a decline in CFU, compared to exposure to 25°C (Figure 4-5). *P. paracutis* shows no significant difference at any time point, neither growing at 25°C nor dying at 37°C. CFUs for *P. cookei* declined by approximately half within 24 hours before stabilising for the rest of the exposure while *A. protothecoides* CFUs declined progressively over the course of four days without quite reaching zero. All other species reached zero CFUs as a result of incubation at 37°C, with *P. moriformis* and *P. tumulicola* reaching zero CFUs after 72 hours of exposure and *P.*

xanthoriae reaching zero CFUs after just 24 hours. All of these species grew when incubated at 25°C.

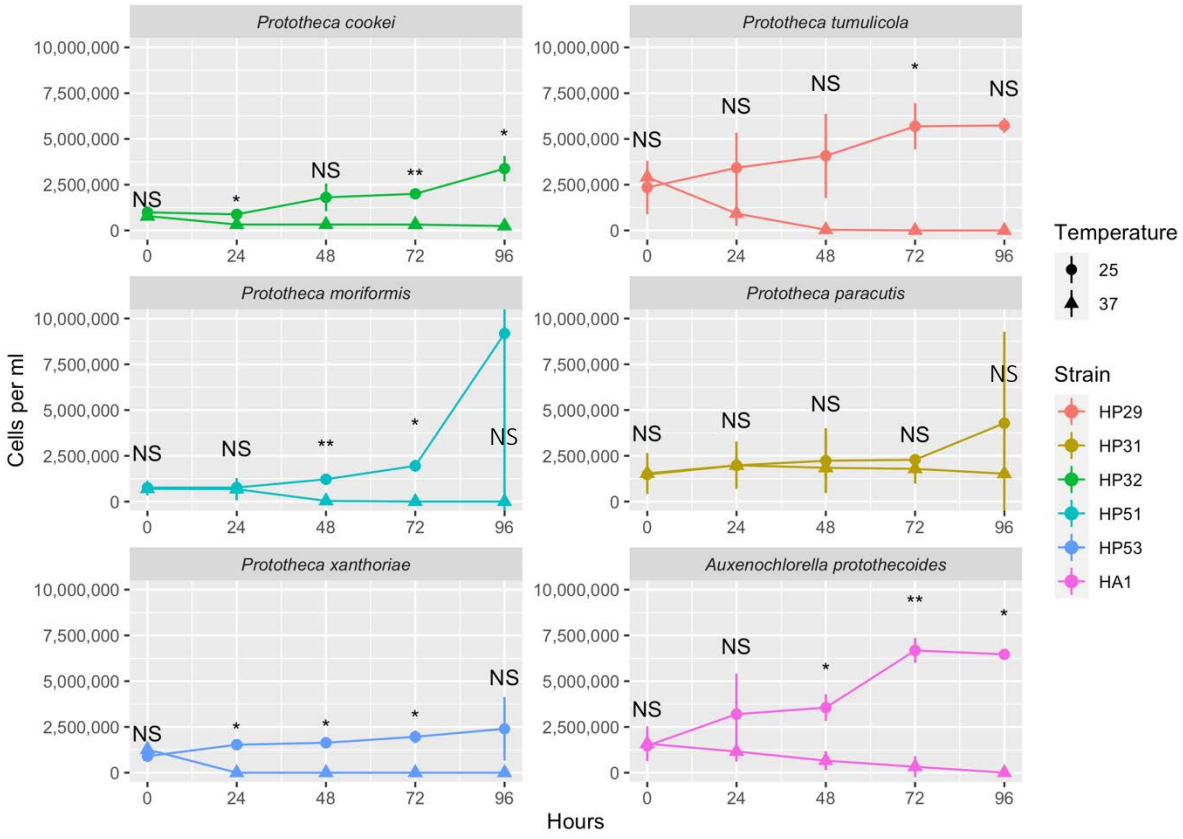


Figure 4-5. Growth curves of non-pathogenic Prototheca *and* Auxenochlorella *strains under temperature stress, as measured by CFUs. Each tile shows the growth of one strain of one species at both 25°C and 37°C, as indicated. Bars show the 95% confidence interval (mean ± 2 standard error) for a given strain and temperature. Stars are used to indicate significance differences between growth at 37°C and 25°C at a given time point (NS = P > 0.05 (i.e. not significant), * = P ≤ 0.05, ** = P ≤ 0.01, *** = P ≤ 0.001).*

All non-pathogenic AHP species investigated are therefore dying during exposure to 37°C, though to varying extents, except for *P. paracutis*.

Most species grow at skin temperature

Having identified that the human core body temperature of 37°C is a barrier to growth for non-pathogenic species, we wondered whether the human skin temperature would provide a similar barrier, as most *Prototheca* infections are cutaneous.

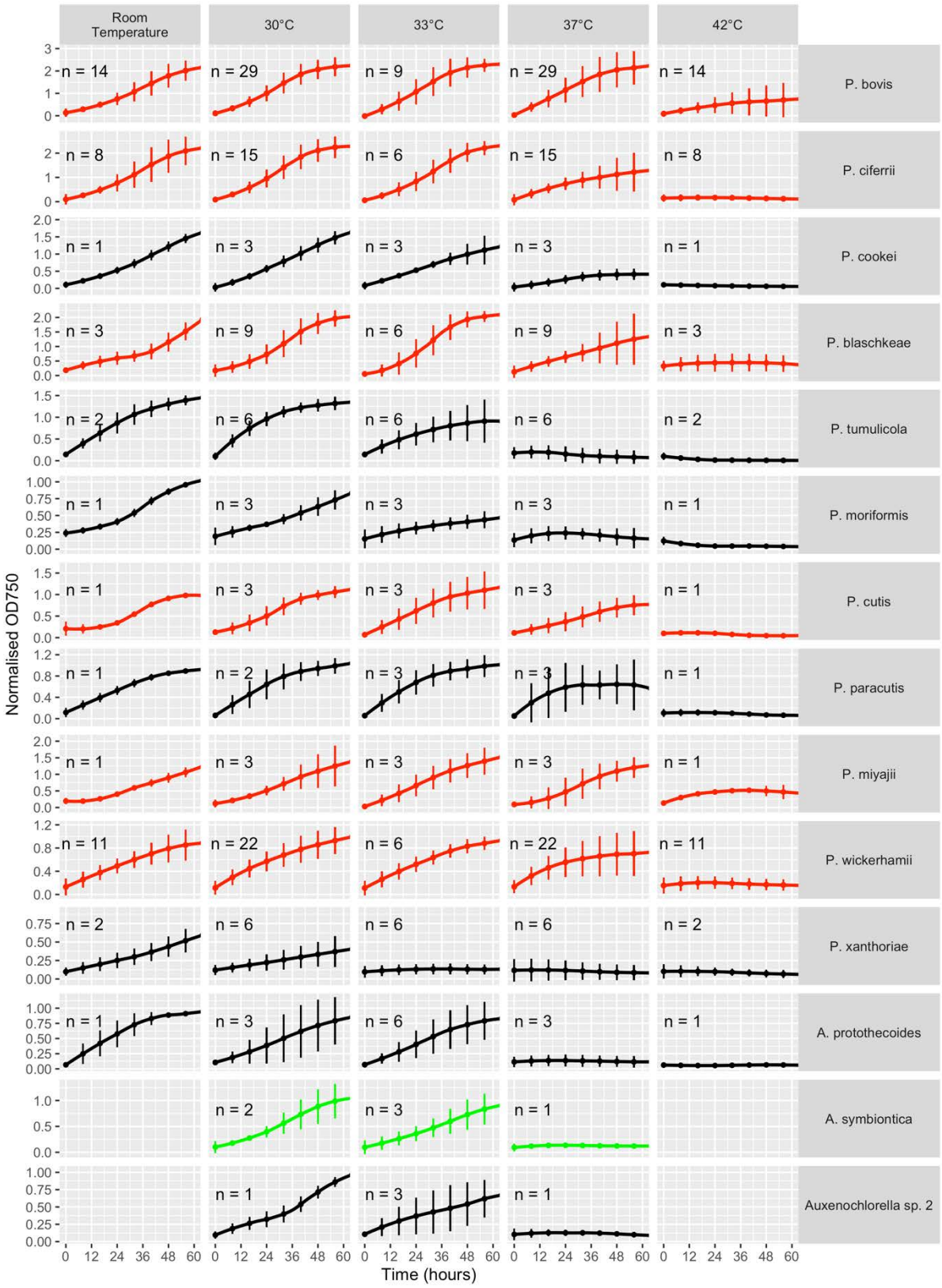


Figure 4-6. Growth curves of Prototheca *and* Auxenochlorella *species under temperature stress, presenting the same data as Figure 4-2 with the addition of data showing growth at 33°C, as measured by OD750. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents compiled data from every strain of a given species. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. The variable number of replicates at each temperature is indicated in the top left of each graph. Bars show the 95% prediction interval (mean ± 2 standard deviations) for a given strain at 8-hour intervals, slightly offset for improved visibility.*

When incubated at 33°C, the upper normal temperature range for human skin temperature, all *Prototheca* species except for *P. xanthoriae* grew (Figure 4-6). Some species which were observed to die at 37°C (*P. moriformis*, *P. tumulicola*, *P. cookei*) have visibly affected growth, but are still able to grow.

Following exposure to 33°C, all species produce colonies 7 days after being spotted onto agar, except for *P. xanthoriae*. Colonies for HP54 did emerge following 10 days of incubation.

All strains of *Auxenochlorella* tested retained at least some viable cells following exposure to 33°C. All strains except for HA6 produced dense colonies on agar. However, this was likely attributable to differences in the way HA6 grows, compared to our other *Auxenochlorella* strains - broth inoculated with the experimental culture of HA6 grew, indicating the presence of viable cells.

HP51, our only strain of *P. moriformis*, produced denser colonies from wells at the edge of the plate in all three replicates of exposure to 33°C, which is reminiscent of *P. wickerhamii* at 37°C. 33°C may therefore be close to the upper temperature limit for this strain.

Growth under temperature stress was not noticeably impacted by culture medium

To assess whether susceptibility to temperature was affected by culture media, growth curves were created using OD data from cultures grown in YPD, another media frequently used to grow *Prototheca*. A single experiment was performed at 30°C, 37°C and 42°C, using a subset of strains. The trends observed appear broadly similar to Figure 4-2 for all strains,

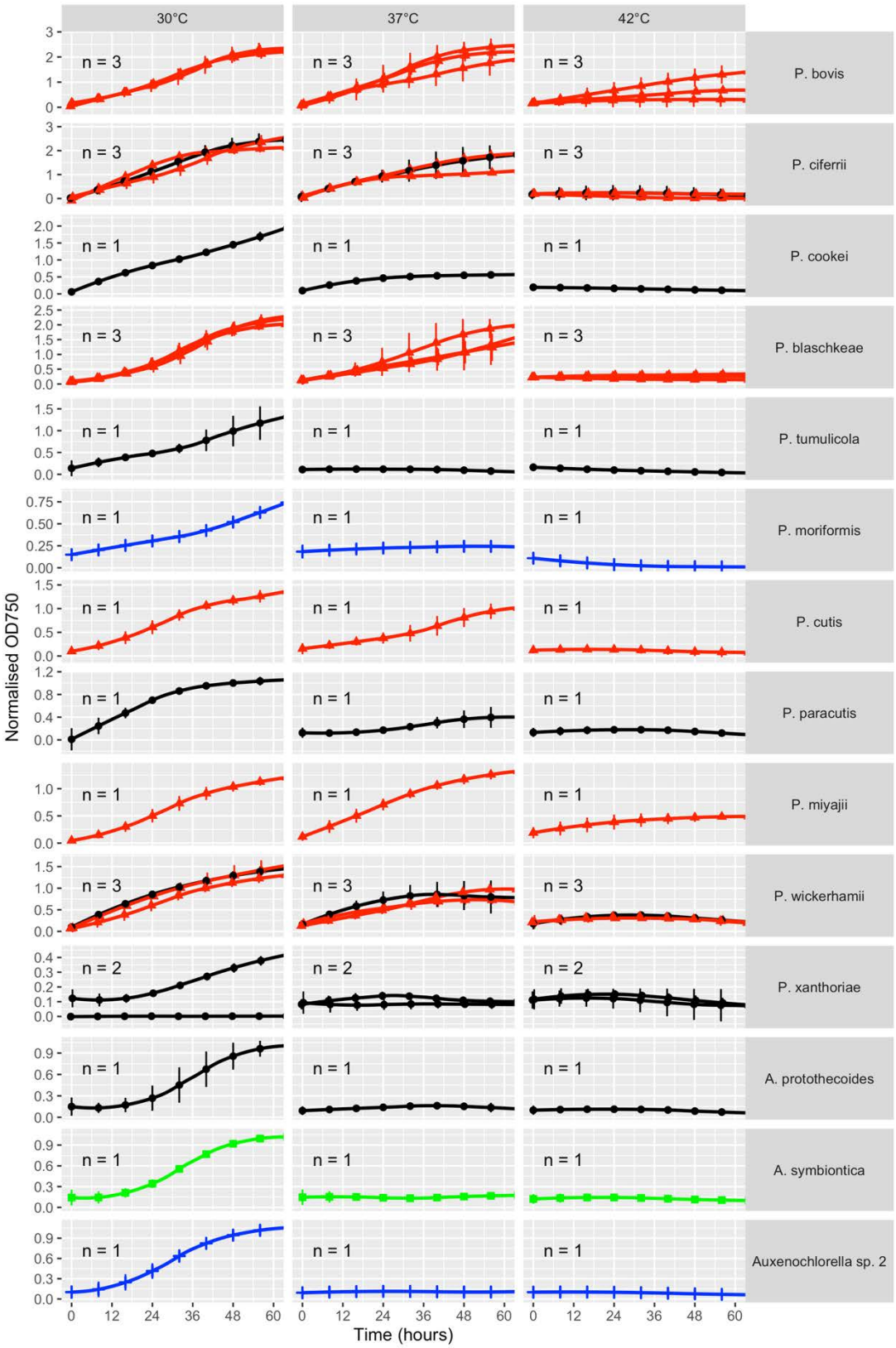


Figure 4-7. Growth curves of Prototheca and Auxenochlorella strains under temperature stress in YPD medium, as measured by OD750. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. Only one replicate per strain per temperature was measured. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. Bars show the 95% prediction interval (mean ± 2 standard deviations of technical replicates) for a given strain at 8-hour intervals, slightly offset for improved visibility.

except for HP53 (*P. xanthoriae*), which did not grow at 30°C (Figure 4-7). However, there was a problem with this preculture, not seen in subsequent YPD cultures. This failure to grow is, therefore, likely not representative of this strain's ability to grow at 30°C. *P. paracutis* may also grow less well at 30°C and 37°C in YPD compared to SAB.

As these curves generally resembled the OD growth curves in SAB, quantitative analyses and further replicates were not performed.

Strains generally grow consistently within a species under acid stress

To determine whether tolerance of low pH is associated with pathology, 34 strains were grown under pH stress at 25°C and had their growth measured as OD_{750} and OD_{600}. OD measurements were used to draw growth curves with the loess method.

An attempt was made to include HA6, which represented another unknown species of *Auxenochlorella*, in the acid stress experiments. However, the strain appears to grow too slowly to be detected in the time frame allowed, and appears to be more significantly impacted by the lack of light within the plate reader than all other *Auxenochlorella* strains attempted. Thus, data from HA6 was excluded from the analysis.

All strains grew at pH 4 (Figure 4-8), which represents an acidification of the base SAB media (approximately pH 5.5). Most strains grow at pH 3, with the only possible exception being *P. miyajii*. Several species, especially *P. blaschkeae* and *A. symbiontica*, appear to grow significantly less well at pH 3, though do manage some growth. Most species struggle to

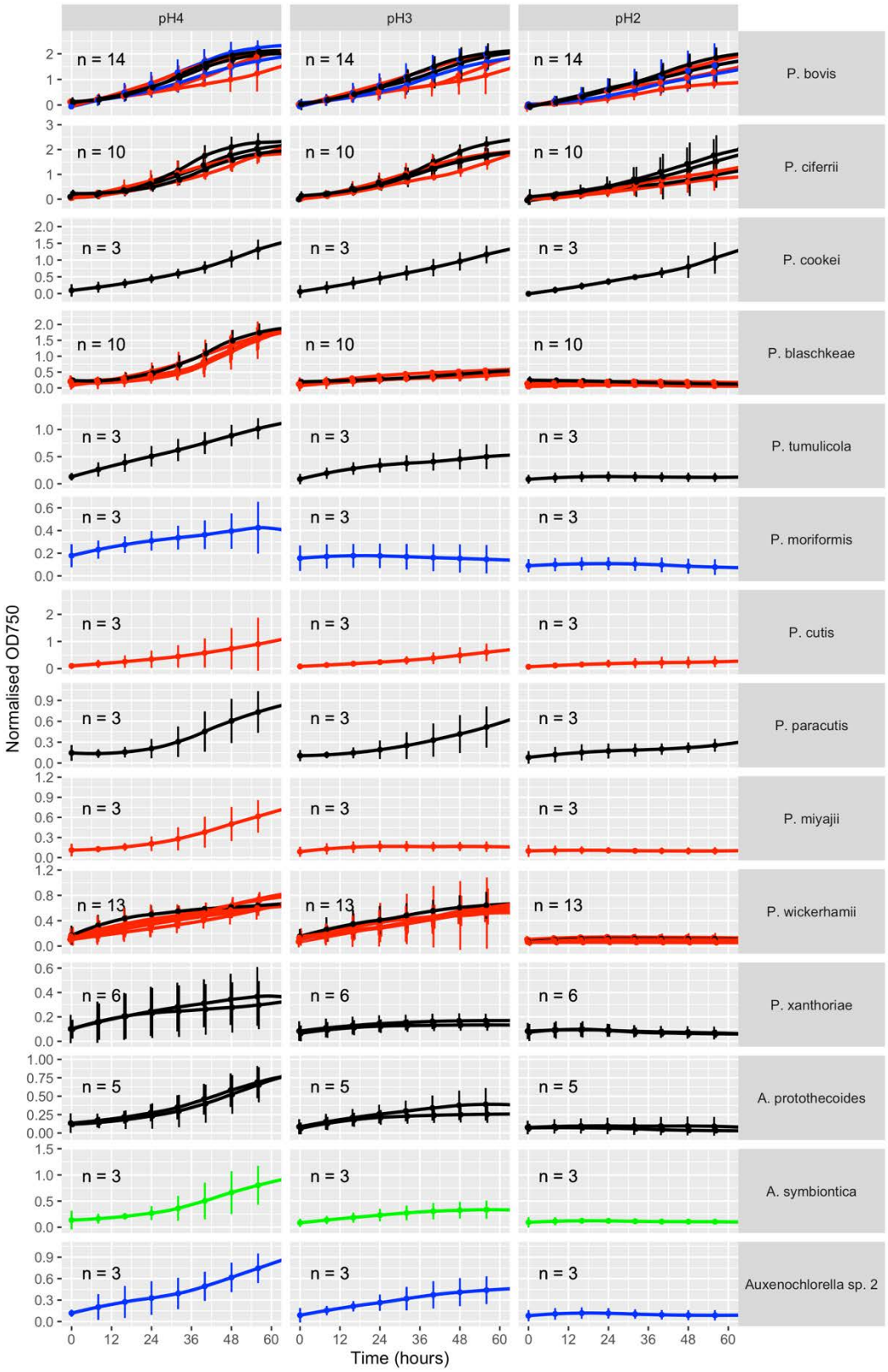


Figure 4-8. Growth curves of Prototheca *and* Auxenochlorella *strains under acid stress, as measured by OD750. Each tile shows the growth of strains of a given species at a particular pH. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. The variable number of replicates at each pH is indicated in the top left of each graph and is broken down further in Appendix 5. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. Bars show the 95% prediction interval (mean ± 2 standard deviations) for a given strain at 8-hour intervals, slightly offset for improved visibility.*

grow at pH 2, with the exception of the three species that were, until recently, known as *P. zopfii*: *P. bovis*, *P. ciferrii*, and *P. cookei*.

Before these relationships could be investigated quantitatively, it had to be verified that strains within a species grew similarly. As with temperature stress, the number of replicates per condition per strain is available in Appendix 5.

The thresholds that were used to identify outlier models under temperature stress were also used to identify outliers under acid stress.

Table 4-5. Summary of the number of wells exposed to pH stress that failed to be fitted to a model, as well as the reason for failure. A more detailed breakdown is available in Appendix 10.

Statistic	Threshold	Number of Failures	Percent Failed
Carrying Capacity	30	46	9.35
Growth Rate	1.5	29	5.89
Sigma	0.2	0	0
Failure to Fit a Model	NA	38	7.72
All	All	113	22.97

Despite the removal of a greater proportion of models, no species had fewer than two biological replicates under any condition. As with temperature stress, a detailed breakdown of the failures (Appendix 10) as well as the number of models not rejected for each species is provided (Appendix 11). As with temperature stress, growth curves generated from OD_{600} measurements are similar to curves generated from OD_{750} measurements (

Appendix 12), so statistical analyses were only performed on OD_{750} data.

Differences in carrying capacities and growth rates between strains within a species could be investigated for six species. As with temperature stress, significant differences in carrying capacity between strains within a species were only found for *P. bovis* ($F = 10.104$, $p = 7.99\times10^{-5}$). Following pairwise comparisons between strains, the carrying capacities of HP18 and HP14 were found to be different from all but one other strain (and were different from each other). Both strains were removed from subsequent analysis. Significant differences of growth rate within species were only found for *P. xanthoriae* ($F = 6.683$, $p = 2.95\times10^{-2}$). However, as there are only two strains of *P. xanthoriae*, we cannot be sure which is more representative of the species, so neither was removed.

Pathogenic species do not grow better than non-pathogenic species at low pH

Two-way ANOVA revealed a significant effect of pH ($F = 27.255$, $p = 1.59\times10^{-11}$) and species ($F = 5.022$, $p < 2\times10^{-16}$) on carrying capacity, as well as a significant interaction between the two ($F = 2.866$, $p = 2.62\times10^{-5}$). Two-way ANOVA also revealed a significant effect of pH ($F = 58.191$, $p < 2\times10^{-16}$) and species ($F = 2.804$, $p = 1.19\times10^{-3}$) on growth rate, as well as a significant interaction between the two ($F = 3.675$, $p = 1.63\times10^{-7}$).

As ability to grow at pH below 3.5 is predictive of endosymbiotic capacity of *Chlorella* strains, it was assumed that the relevant differences would be between pH 4 and pH 3. Thus, statistical comparisons between these two conditions were performed. However, when carrying capacity at pH 3 was compared to carrying capacity at pH 4 for each species, no significant differences were identified (Table 4-6). This was despite a greater than 90% decrease for *P. blaschkeae* and *P. miyajii*, and a greater than 75% decrease for *A.*

prototheccoides, *A. symbiontica*, and *P. tumulicola*. In all five of these cases, the lack of significance appears to be the result of a large standard error for carrying capacity at pH 4, to such an extent that four of these five species also struggle to distinguish their carrying capacity at pH 4 from zero.

Table 4-6. Comparisons of carrying capacities of Prototheca *and* Auxenochlorella *species grown at pH 4 and at pH 3. Rows containing pathogenic species are highlighted in pink.*

Species	Mean Carrying Capacity at pH 4	Mean Carrying Capacity at pH 3	Difference between means	Proportional difference	Adjusted p value	Significance
P. bovis	2.26	2.39	-0.129	-0.0572	3.33×10^{-1}	NS
P. ciferrii	3.13	2.40	0.736	0.235	3.33×10^{-1}	NS
P. cookei	5.31	1.43	3.89	0.732	3.33×10^{-1}	NS
P. blaschkeae	7.01	0.506	6.50	0.928	1.15×10^{-1}	NS
P. tumulicola	2.30	0.508	1.79	0.779	3.33×10^{-1}	NS
P. moriformis	0.619	0.180	0.440	0.71	3.33×10^{-1}	NS
P. cutis	4.21	2.02	2.19	0.52	5.01×10^{-1}	NS
P. paracutis	1.56	0.956	0.606	0.388	3.33×10^{-1}	NS
P. miyajii	1.88	0.174	1.70	0.907	3.33×10^{-1}	NS
P. wickerhamii	0.847	0.920	-0.073	-0.0862	7.48×10^{-1}	NS
P. xanthoriae	0.530	0.158	0.372	0.701	3.33×10^{-1}	NS
A. prototheccoides	2.90	0.387	2.51	0.867	3.33×10^{-1}	NS
A. symbiontica	1.70	0.382	1.31	0.774	3.33×10^{-1}	NS
Auxenochlorella sp. 2	1.67	0.481	1.19	0.711	3.33×10^{-1}	NS

Comparisons of growth rates at pH 4 and pH 3 also showed no significant differences (Table 4-7). As with carrying capacity, large proportional differences were seen, but it appears that higher growth rates are calculated for pH 3 than pH 4 for these species, despite expectations that pH 3 would be the more stressful condition.

Table 4-7. Comparisons of growth rates of Prototheca *and* Auxenochlorella *species grown at pH 4 and at pH 3. Rows containing pathogenic species are highlighted in pink.*

Species	Mean Growth Rate at pH 4	Mean Growth Rate at pH 3	Difference between means	Proportional difference	Adjusted p value	Significance
P. bovis	0.102	0.0866	0.0157	0.153	3.96×10^{-1}	NS
P. ciferrii	0.0924	0.0836	0.00879	0.0951	6.26×10^{-1}	NS
P. cookei	0.0509	0.0779	-0.0270	-0.531	3.96×10^{-1}	NS
P. blaschkeae	0.0670	0.0884	-0.0215	-0.320	3.98×10^{-1}	NS
P. tumulicola	0.0534	0.0806	-0.0273	-0.512	3.96×10^{-1}	NS
P. moriformis	0.0458	0.212	-0.166	-3.62	3.96×10^{-1}	NS
P. cutis	0.0646	0.0623	0.00231	0.0358	9.36×10^{-1}	NS
P. paracutis	0.0622	0.0468	0.0153	0.246	5.76×10^{-1}	NS
P. miyajii	0.0501	0.190	-0.140	-2.78	3.96×10^{-1}	NS
P. wickerhamii	0.0869	0.102	-0.0149	-0.172	6.65×10^{-1}	NS
P. xanthoriae	0.0928	0.126	-0.0331	-0.356	5.76×10^{-1}	NS
A. protothecoides	0.0456	0.100	-0.0548	-1.20	3.96×10^{-1}	NS
A. symbiontica	0.0573	0.0679	-0.0105	-0.183	6.58×10^{-1}	NS
Auxenochlorella sp. 2	0.0359	0.0587	-0.0228	-0.634	3.96×10^{-1}	NS

As with temperature stress, OD_{750} at 0 hours was compared directly to OD_{750} at 60 hours for all wells that contained strains growing at pH 3. All species showed significant differences between these time points (Table 4-8).

There is therefore no evidence for an association between growth at low pH and pathology.

Table 4-8. Growth of Prototheca *and* Auxenochlorella *species grown at pH 3, as determined by the difference in OD750 following 0 hours and 60 hours growth at 37°C Rows containing pathogenic species are highlighted in pink.*

Species	Growth at pH 3	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Adjusted p value	Significance
P. bovis	1.82	1.95	1.69	1.45×10^{-20}	***
P. ciferrii	1.94	2.03	1.85	2.58×10^{-16}	***
P. cookei	1.47	1.64	1.30	5.34×10^{-6}	***
P. blaschkeae	1.68	1.77	1.59	3.05×10^{-13}	***
P. tumulicola	1.05	1.18	0.914	6.01×10^{-4}	***
P. moriformis	0.361	0.523	0.199	1.82×10^{-2}	*
P. cutis	0.936	1.55	0.318	4.07×10^{-4}	***
P. paracutis	0.733	0.887	0.579	4.07×10^{-4}	***
P. miyajii	0.638	0.77	0.507	4.07×10^{-4}	***
P. wickerhamii	0.621	0.659	0.584	2.42×10^{-15}	***
P. xanthoriae	0.299	0.387	0.211	7.19×10^{-7}	***
*A. prototheco*ides	0.720	0.797	0.643	2.29×10^{-5}	***
A. symbiontica	0.848	1.03	0.668	5.99×10^{-4}	***
Auxenochlorella sp. 2	0.802	0.929	0.674	2.21×10^{-4}	***

Most *Prototheca* and *Auxenochlorella* strains survive exposure to extremely low pH

Following growth curve experiments, 10 µl of culture was spotted onto PIM agar to test for survival. Images showing a subset of strains, with the same selection of strains as was provided to illustrate survival of temperature stress, are provided (Figure 4-9).

Even though growth is not seen for most species at pH 2, and difficulties with modelling growth at pH 3 and 4 suggest these stresses are having an impact on growth, most species survive extreme pH.

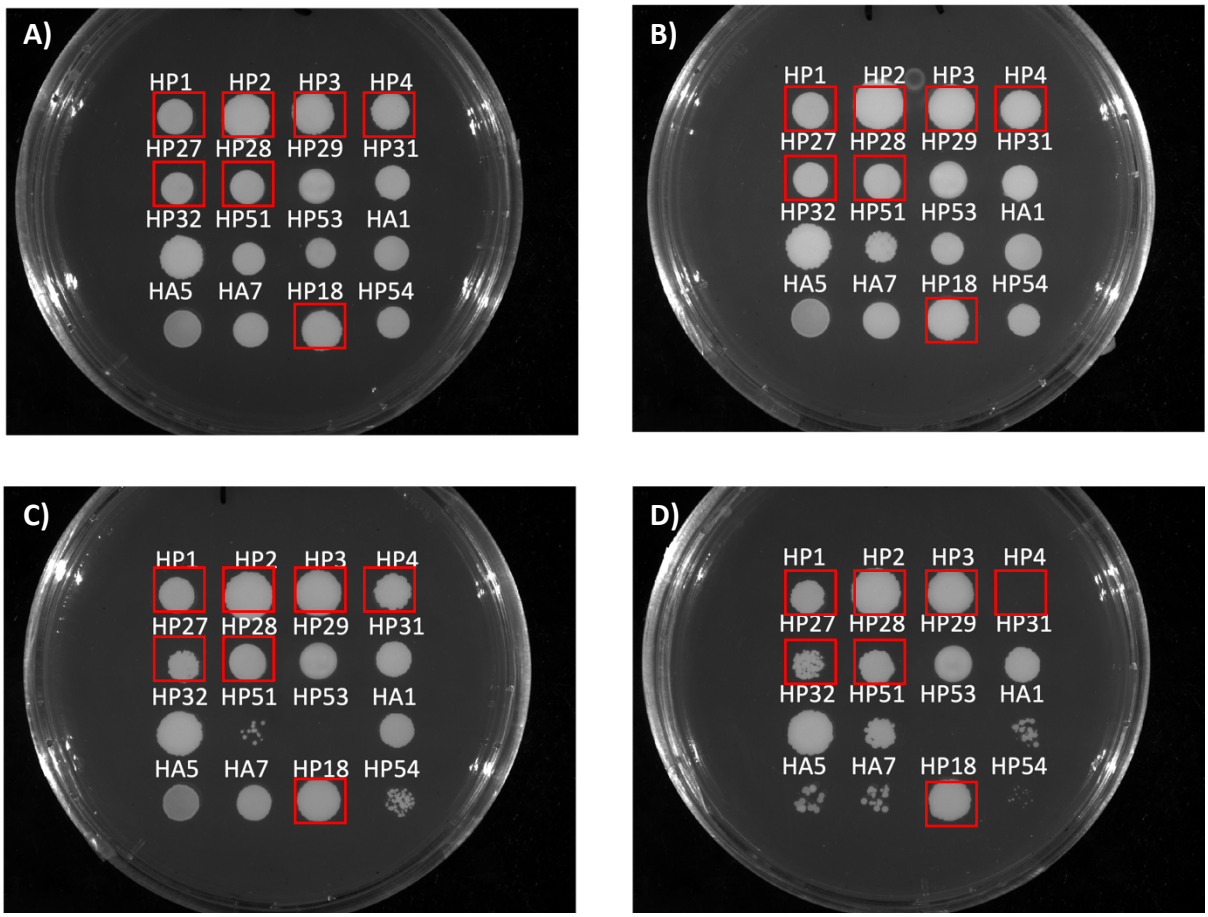


Figure 4-9. Agar plates showing survival of Prototheca *and* Auxenochlorella *strains following approximately 72 hours of exposure to acid stress. Not all strains tested are shown, instead one strain from each species has been chosen. 14 species are represented by these 16 strains: HP1* – P. wickerhamii*; HP2* – P. ciferrii*; HP3 and HP18* – P. bovis*; HP4* – P. blaschkeae*; HP27* – P. miyajii*; HP28* – P. cutis*; HP29* – P. tumulicola*; HP31* – P. paracutis*; HP32* – P. cookei*; HP51* – P. moriformis*; HP53 and HP54* – P. xanthoriae*; HA1* – A. prototheocoides*; HA5* – A. symbiontica*; HA7* – Auxenochlorella *sp. 2. HP18 and HP54 were included as they do not behave the same as the other strains of their species. The pH levels explored were: uncontrolled SAB at pH 5.5 (A), pH 4 (B), pH 3 (C), and pH 2 (D). Pathogenic species are indicated by a red box.*

There was greater visual variation in the colonies that grew after exposure to acid stress than for temperature stress. A dense, smooth circle at the point of spotting was considered to indicate good survival. Sometimes the dense circle had gaps, either on its edge or internally, but the colonies were too close to distinguish, which was still considered to be survival (though less robust than a solid circle). A complete absence of colonies, or a few individual colonies, was considered to indicate death, to varying degrees.

P. moriformis and *P. xanthoriae* are the only species to be consistently strongly impacted by pH 3, though both often manage to produce a few colonies (*P. xanthoriae*: 4/7 with some

colonies (3 of these instances being HP54), the remainder having none; *P. moriformis*: 3/5 with some colonies, the remainder having none). The circle of *P. miyajii* was always less dense following exposure to pH 3 than pH 4, but colonies were always produced. The circle of *A. protothecoides* and *A. symbiontica* species is affected in a similar way approximately 40% of the time, but is otherwise as dense at pH 3 as at pH 4.

At pH 2, *P. blaschkeae* (8/10) and *P. xanthoriae* (6/7) generally produce no colonies, and on the rare occasion that they do there are very few colonies – Figure 4-9 shows HP54 colonies at pH 2, but was the only plate (out of seven) to do so. All *Auxenochlorella* species are affected by pH 2, sometimes producing no colonies and sometimes producing a few. *A. symbiontica* appears to be the least affected, being the only species to produce colonies too dense to count following exposure to pH 2 (1/5), but it also produced no colonies on one occasion. *A. protothecoides* produced no colonies 60% of the time, while HA7 produced no colonies once.

P. miyajii always produces colonies following exposure to pH 2, but usually very few – Figure 4-9 is an outlier in this case. *P. wickerhamii* also always produces colonies, but occasionally the spot is sparse enough for individual colonies to be distinguished at pH 2. When this happens, the sparse wells are usually away from the edge of the 48-well plate. All other species survive all levels of pH stress investigated.

Peculiarly, cultures of *P. moriformis* (HP51) exposed to pH 2 appear to produce more colonies than cultures grown exposed to pH 3, when the culture media was SAB. This was seen in every experiment, despite the use of different batches of media at pH 2 and 3 for

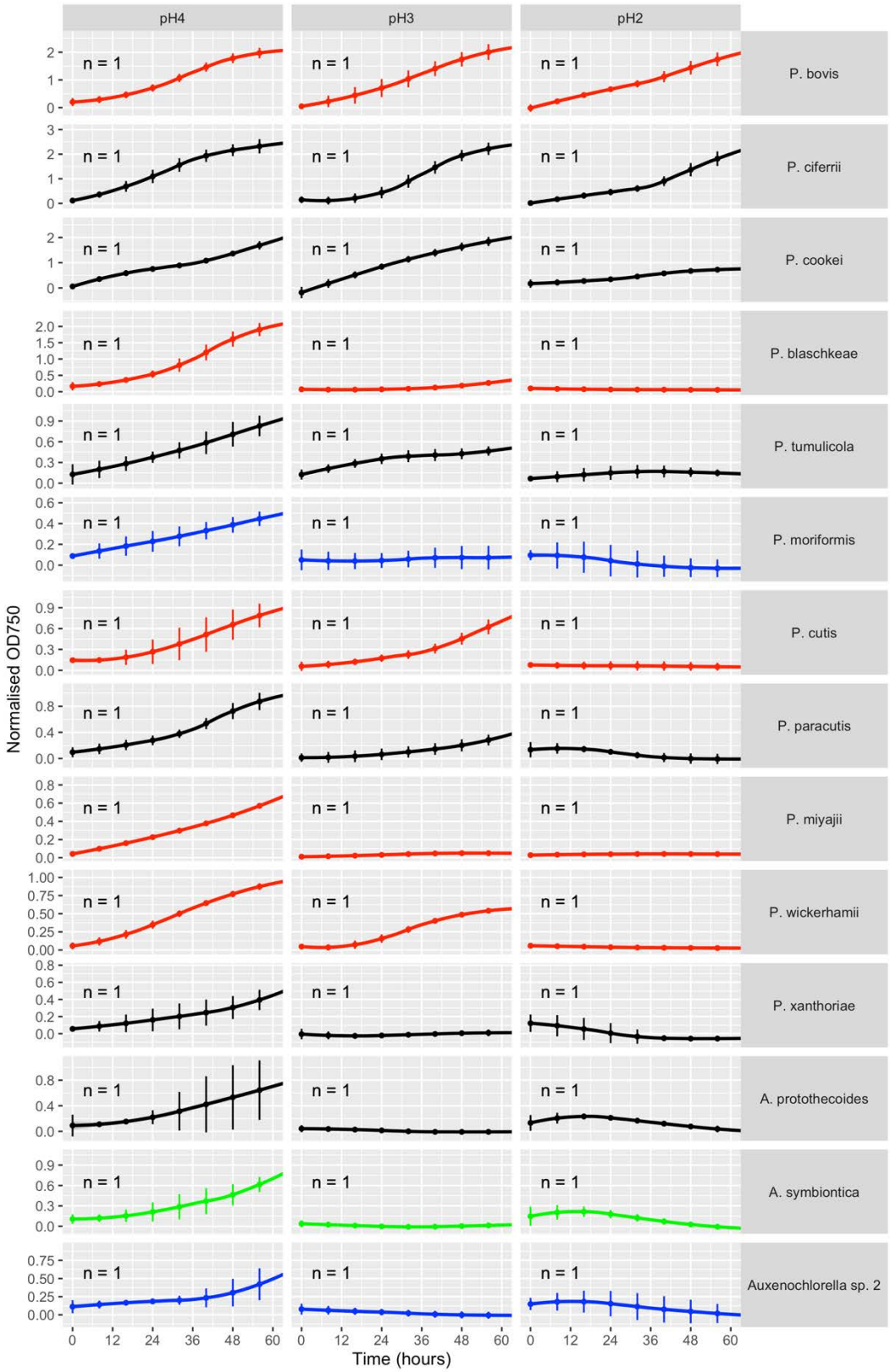


Figure 4-10. Growth curves of Prototheca and Auxenochlorella strains under acid stress in YPD medium, as measured by OD750. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. Only one replicate per strain per temperature was measured. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. Bars show the 95% prediction interval (mean ± 2 standard deviations of technical replicates) for a given strain at 8-hour intervals, slightly offset for improved visibility.

some replicates. This was not seen in the culture grown in YPD, in which *P. moriformis* produced colonies at all three pH levels, but produced fewest when exposed to pH 2. Survival was not noticeably different for any other species in YPD compared to SAB.

Growth under acid stress may be impacted by culture medium

It has been noted before that certain species of *Prototheca* are sensitive to the presence of some compounds at different pH levels (Kurtzman et al., 2011). We therefore decided to check whether acid tolerance was affected by media choice. As with temperature stress, growth curves were created using OD data from cultures grown in YPD. *Auxenochlorella* species and *P. paracutis* appear to grow slightly better in SAB compared to YPD at pH 3, but otherwise the difference between media appears minimal (Figure 4-10).

OD plateau in 48-well plates is not reflective of growth plateau in larger cultures

In the growth curves presented earlier in this chapter, different *Prototheca* species reach plateaus of different heights, with species related to *P. bovis* typically reaching higher ODs. To normalise experimental cultures, preculture ODs were measured and differences were also observed in typical ODs reached for each species (Figure 4-11). However, the heights of the growth curves plateaus do not appear correlated with the preculture ODs, despite comparable times and media composition. Precultures of our single *P. moriformis* strain and all *P. blaschkeae* strains grow poorly, relative to other *Prototheca* species, but their growth curves are not obviously different (when grown at room temperature) from other species. *P. wickerhamii* strain precultures grow well, but their growth curves plateau at much lower densities than *P. bovis* or *P. ciferrii*.

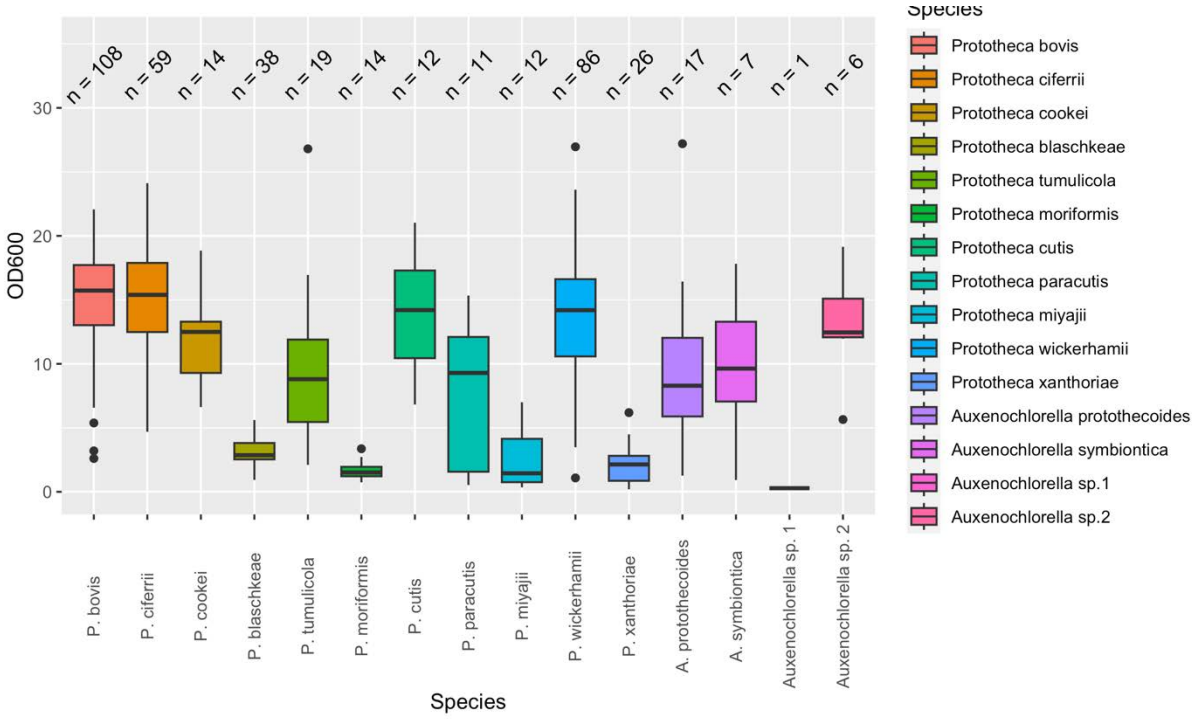


Figure 4-11. Prototheca *and* Auxenochlorella *preculture* OD_{600}*. At time of measurement, cultures had been grown for 48 hours at 25°C in SAB.*

This difference in plateau probably reflects differences in how strains grew in the wells. During growth curve experiments, all species consistently formed clumps in the middle of wells (Figure 4-12A). The visual size of the clump after incubation is a good indicator of the height of the plateau for the resulting growth curve. Additionally, some species form biofilm in SAB, which appears to cloud the media around their central clump and likely contributes to the higher plateau. This pattern of growth was consistent for all AHP strains grown in the plate reader. However, this growth pattern is not seen in 48-well plates grown outside of the plate reader (Figure 4-12B), suggesting that this is an artefact of the measurement. However, this is unlikely to affect the comparisons made earlier, as direct comparisons are never made between species.

B)

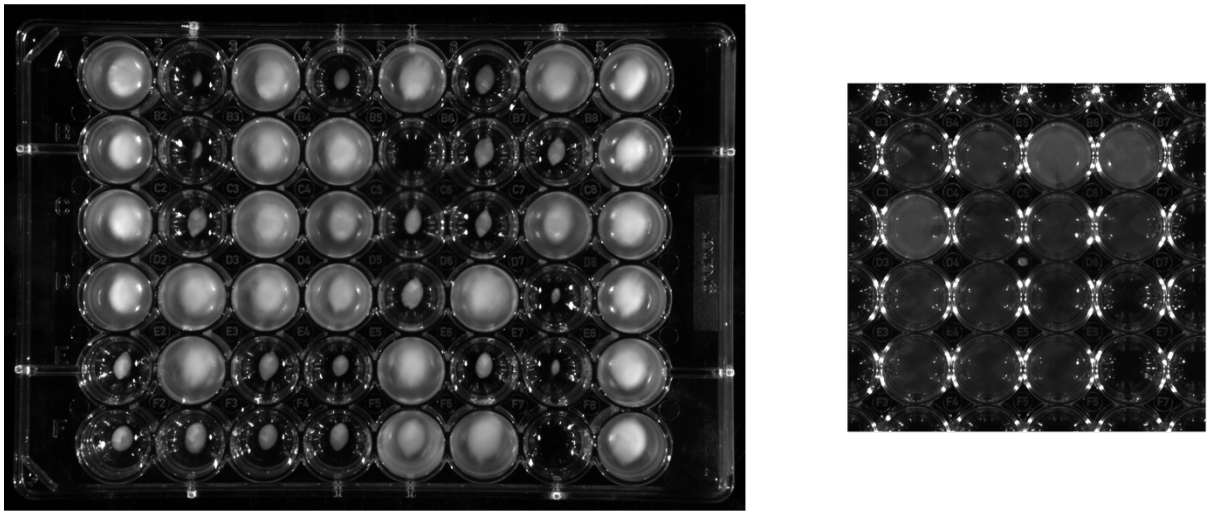


Figure 4-12. Prototheca *and* Auxenochlorella *cultures grown in 48-well plates. A) A 48-well plate from experiments to assess the effect of temperature stress on growth, grown at 30°C in the plate reader, containing a random selection of strains. Cultures noticeably clump in the centre of the wells, and some wells are noticeably cloudy, likely as a result of biofilm production. B) A selection of strains representing every species in our collection, grown at 30°C in a static incubator. Cultures are evenly spread throughout the well, while appearing less dense than in A.*

4 day incubations may be insufficient to see colony growth of *Prototheca* following stress

Following growth curves, culture from wells were spotted onto PIM agar, incubated at 25°C, and inspected for growth. Typically, plates were inspected on day 2, day 4, and day 7 following incubation, though these exact intervals were not always possible (e.g. an image was occasionally taken on day 8 instead of day 7).

It was noticed that some strains under some stressful conditions took longer than others to form visible colonies. For one experiment, agar plates were carefully inspected for signs of colony formation. Some colonies could be seen by eye after four days that were completely invisible after two days, and some colonies could be seen by eye after eight days that were very difficult to observe after four days (Figure 4-13).

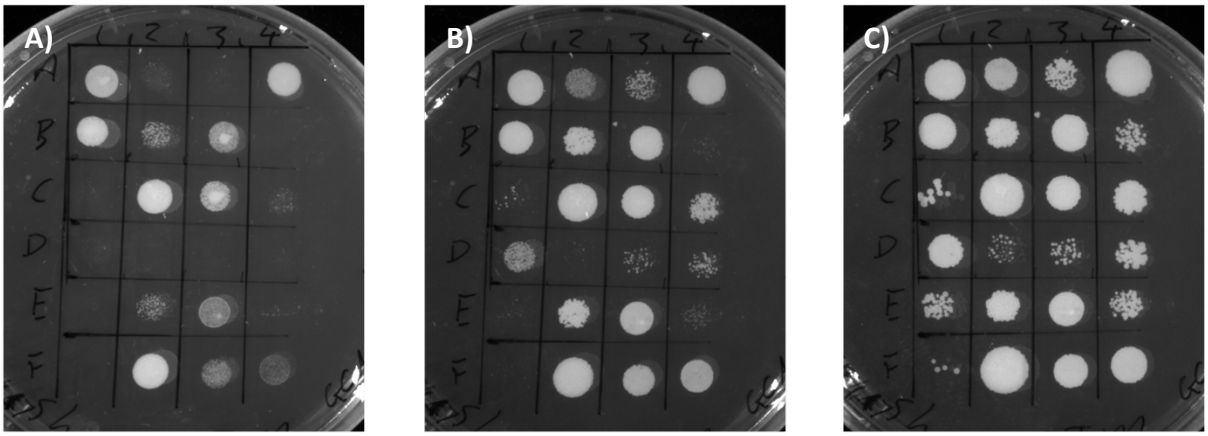


Figure 4-13. Images of an agar plate spotted with cultures exposed to acid stress to check survival after a growth curve experiment. The same plate was imaged two (A), four (B), and eight (C) days after inoculation. Colonies are visible on day four that were not visible on day two, despite close examination of the plate on day two. Colonies are easily visible on day eight that were visible on day four, but were difficult to see and could easily have been missed.

Discussion

In this chapter I have investigated the ability of members of the AHP lineage to grow under temperature and acid stress at both the strain and species level, with the goal of identifying stress tolerance that facilitates pathology within the lineage.

Relevance of stress tolerance to pathology

Temperature stress

The growth of *Prototheca* strains at different temperatures has been investigated before (Masuda et al., 2016; Kano et al., 2022; Jagielski et al., 2022, 2019a). However, to the best of my knowledge, the work here is the first to include: human skin temperatures, which were likely to be relevant as human *Prototheca* infections are mostly cutaneous; *Auxenochlorella* species, which are environmental members of the AHP lineage; growth curves and quantitative results, instead of endpoint, categorical results; and an explicit separation of clinical from non-clinical strains from multiple species, which might identify pathogenic lineages within species.

This work has revealed a range of thermotolerance within the AHP lineage, with *Auxenochlorella* appearing to be more thermotolerant than some species of *Prototheca*. The thermotolerance of species in the AHP lineage can be split into five groups: those able to grow at temperatures higher than 37°C (*P. bovis* and *P. miyajii*); those able to grow unimpeded at 37°C (*P. cutis*); those that struggle to grow, but are not killed at 37°C (*P. wickerhamii*, *P. ciferrii* and *P. blaschkeae*); those unable to survive, or at least unable to grow, at 37°C (*Auxenochlorella*, *P. tumulicola*, *P. moriformis*, *P. cookei*); and those unable to survive at 30°C (*P. xanthoriae*).

Pathogenic species can generally be distinguished quantitatively from non-pathogenic species by carrying capacity at 37°C, with non-pathogenic species showing a large proportional decrease in carrying capacity at 37°C compared to 30°C. However, this was not true for *P. paracutis*.

P. paracutis has previously been reported as unable to grow at 37°C (Kunthiphun et al., 2019; Kano et al., 2022), but we were unable to classify *P. paracutis* into any of the groups mentioned above. This was due to non-significant differences in carrying capacity and growth rates from growth curves derived from OD data, and an inability to identify growth at 25°C or death at 37°C from CFU data.

That thermotolerance can distinguish pathogenic from non-pathogenic species suggests that thermotolerance is necessary for pathology. It remains to be seen if thermotolerance is sufficient to allow pathology in a non-pathogenic species.

Acid stress

The ability for *Prototheca* species to survive low pH has also been investigated before, though not to the same extent as high temperatures and usually focused on cow-associated species (Jánosi et al., 2001; Marques et al., 2010). This work represents the taxonomically broadest investigation into the acid tolerance of the AHP lineage to date and has revealed a range of acid tolerance within the AHP lineage. However, there was no clear association between acid tolerance and pathology at the species level. The known pathogenic species *P. wickerhamii* is unable to grow at pH 2 and *P. blaschkeae* appears severely impacted by exposure to pH 3, while *P. cookei*, for which there are no reported cases of infection, appears unaffected by pH 2.

Instead, growth under acidic conditions appears linked to relatedness. All *Auxenochlorella* strains tested show similar growth patterns at different pH, as do all three of the recent former-*zopfii* species (which does not include *P. blaschkeae*, which was also originally classified as *P. zopfii* but was reclassified at an earlier date (Roesler et al., 2006)). However, this was difficult to investigate quantitatively.

It has been reported that "*P. zopfii*" and *P. blaschkeae* could not grow, or grew poorly, at pH 1 or 3 (Marques et al., 2010). Our work did not reach pH 1 but did demonstrate that *P. blaschkeae* was susceptible to acid at pH 3. However, we do see that *P. blaschkeae* survives exposure to pH 3, which was not reported previously, and we could not confirm the susceptibility of any of the three species that could have been *P. zopfii* to pH 3. It seems likely that this discrepancy can be accounted for by the choice of acid – Marques *et al.* chose

acetic acid, which has antimicrobial activity that appears distinct from its acidity (Trzaska et al., 2015), to reach pH 3 while hydrochloric acid was used here, which does not have this activity.

The role of stress tolerance in pathology

If thermotolerance is to be considered a virulence trait in *Prototheca*, it would be reasonable to think that mutations could increase thermotolerance, which could then increase the virulence of any strains that possess them.

However, it is not clear that stronger tolerance to stress in algae results in enhanced virulence or greater infectivity. It is true that *P. bovis* appears to be one of the species most tolerant of high temperatures and is responsible for most cattle cases. However, *P. wickerhamii* is somewhat inhibited by 37°C, unlike *P. cutis* and *P. miyajii*, and makes up the majority of human cases.

This could indicate that endothermy acts as a barrier to pathology, which must be overcome by thermotolerance for a species to become pathogenic. But, once endothermy is overcome, thermotolerance may have no additional impact on virulence.

Stress tolerance within a species

While it appears that thermotolerance can separate species with known pathogens from species without known pathogens, growth at 37°C does not separate clinical from environmental isolates. HP18 is the only strain of *P. bovis* identified to grow differently from other strains of the species, but it is a clinical isolate. Strains of *P. ciferrii* are quite variable in

their growth rate, but this variability does not separate clinical from environmental isolates. Similarly, acid tolerance does not separate pathogenic from non-pathogenic strains within a species.

Tolerance to high temperatures and acid, therefore, does not suggest that some strains of a pathogenic species are non-pathogenic. Note that this does not rule out the possibility, only that tolerance of these stresses is insufficient to distinguish between pathogenic and non-pathogenic strains, if such a distinction exists.

Future work

Experimental infection and other strains

If thermotolerance is predictive of pathology, it may be able to identify pathogenic species before a case of naturally occurring infection can be found. To do this, strains without reported cases of infection that grow or survive 37°C should be used to experimentally infect a suitable host.

All species tested here that were clearly able to grow at 37°C are already known to be pathogenic. However, it is possible that some of the species tested here are cryptic pathogens. Of the species tested here, *P. paracutis* is most likely to be a cryptic pathogen, as did not die at 37°C. A lack of death is not the same as the presence of growth, but growth of *P. paracutis* was not measured by CFU counts at 25°C, suggesting that some other stress was inhibiting its growth during this experiment. It is therefore possible that *P. paracutis* does grow at 37°C, though this growth is likely to be slow. Other possible candidates for

cryptic pathogens include *P. cookei*, which formed a stable population at 37°C in this chapter after an initial decline.

Additionally, several species of *Prototheca* were not tested here, and new species could be identified that are capable of growing at 37°C. For example, *P. vistulensis*, a recently discovered species, was not tested here but has been reported to grow at 35°C (Jagielski et al., 2022).

The observation that *P. xanthoriae* does not reliably survive 30°C may also inform future efforts to identify novel *Prototheca* species. *P. tumulicola* and *P. stagnora* have also been reported to not grow at 30°C (Nagatsuka et al., 2017), though *P. tumulicola* was observed to grow at 30°C in this chapter. Currently, work to isolate environmental strains of *Prototheca* sometimes culture isolates at 30°C (Jagielski et al., 2022), which is high enough to inhibit these environmental strains.

Future stresses

Tolerance of high temperatures and low pH are only two of many possible stresses that could be relevant for pathology. One such additional stress may be low oxygen stress, as tissues often contain oxygen below atmospheric levels, which can reduce further local to an infection (Zeitouni et al., 2016).

A second stress may be reactive oxygen or nitrogen species. Phagolysosomes contain reactive oxygen species (ROS) and reactive nitrogen species (RNS), produced by the respiratory burst (Erwig and Gow, 2016).

Mechanisms

Further investigation might involve dissecting the mechanisms that allow some *Prototheca* species to grow at higher temperatures. It is possible that *P. miyajii* and *P. bovis* may use different mechanisms to tolerate 42°C. Additionally, the greater apparent survival of *P. moriformis* pH 3 compared to pH 2 may indicate secretion of some organic acid.

This may have future applications for biotech as well as infectious disease. Algae that can grow at higher temperatures, or that produce secreted products, may be useful for fermenters. And even if *Prototheca* are not suitable for bioreactors, these traits may be transferrable to other algae, like *Auxenochlorella* or *Chlorella* species that are already of interest for biotechnological applications (Gao et al., 2014; Lim et al., 2010).

5) Phagocytosis of *Prototheca* and Related Organisms

Introduction

The data regarding phagocytosis of P. bovis *and* P. wickerhamii *in both murine and human cells presented in this chapter are published in* (Haider et al., 2023)*, as indicated in figure legends. Some of this data was generated by Dr Mohammed Haider, as indicated in figure legends.*

A mammalian immune response begins with innate immunity, the activation of which involves the recognition of microorganisms via microbe- or pathogen-associated molecular patterns (M- or PAMPs). MAMPs/PAMPs are general classes of molecules that are associated with microorganisms, including peptidoglycan and lipopolysaccharide from bacteria, β-glucans from fungi, and a variety of nucleic acids (Okude et al., 2021; Hatinguais et al., 2020). Recognition of these classes of molecules directs the immune response, and can trigger behaviour like phagocytosis (Murphy et al., 2012b).

As algae are not typical pathogens, the literature regarding their interactions with the immune system is sparse. In one instance, *Chlamydomonas reinhardtii*, a green alga, was used as a vector to deliver drugs into the lungs, and did not trigger an inflammatory response while doing so (Zhang et al., 2022). More typically, studies focus on potential pharmacological applications of derivates of algae (Zhang et al., 2020; Tomori et al., 2019; Orsi et al., 2010; Wang et al., 2021), which is of limited use in inferring how whole and living algae will interact with the immune system.

Within the AHP lineage, work is beginning to be done to investigate the interactions between *Prototheca* and macrophages, tissue resident phagocytes that are both capable of killing many microorganisms and play an important role in triggering inflammation (Murphy et al., 2012b). These studies have found that *P. bovis* and *P. ciferrii* are able to induce pro-inflammatory responses (via mitochondrially derived reactive oxygen species or the NLRP3 inflammasome), but that *P. bovis* does so to a greater extent, and that *P. bovis* is able to kill host cells, while *P. ciferrii* does not (Shahid et al., 2017b; Zhao et al., 2021; Shahid et al., 2017a). There is some evidence that *P. bovis* is able to actively invade macrophages, while *P. ciferrii* cannot (Shahid et al., 2020b). It has also been observed that *P. wickerhamii* can cause cytotoxicity against bone marrow derived macrophages (BMDMs) (Guo et al., 2023). In a paper now accepted for publication, we were able to show that *P. wickerhamii* is generally not phagocytosed by murine macrophages and is phagocytosed significantly less than *P. bovis* by human macrophages (Haider et al., 2023). Some data from that paper are included in this chapter.

However, this work covers only three of the six known pathogenic species of *Prototheca*, which themselves are only a fraction of the total number of species within the AHP lineage. To the best of my knowledge, there have been no investigations into the interactions between non-pathogenic species and mammalian immunity, apart from those that identify *P. ciferrii* as a non-pathogenic species despite reports of infection (Yang et al., 2020; Hirose et al., 2018). Thus, our understanding of the interaction between AHP algae and macrophages is extremely limited in scope. In this chapter, I aimed to explore the interactions between macrophages and the AHP lineage as a whole.

Knowing that *P. bovis* and *P. ciferrii* are phagocytosed similarly and that *P. bovis* is phagocytosed differently to *P. wickerhamii*, it was anticipated that phagocytic behaviour of *Prototheca* species might group by lineage. There are potentially two pathogenic lineages within *Prototheca*: *P. wickerhamii*, *P. cutis*, and *P. miyajii* are often grouped together as "human-associated species", while *P. bovis*, *P. ciferrii* and *P. blaschkeae* are often grouped together as "cattle-associated species" (Shahid et al., 2020a; Jagielski et al., 2019a; Kano et al., 2022). It is worth noting that *P. wickerhamii* is sometimes phylogenetically separate from the other human-associated species (Masuda et al., 2016).

Both pathogenic lineages of *Prototheca* contain species that are not pathogenic. *P. cookei* is another species that was previously classified as *P. zopfii*, and was therefore included in the cattle-associated group, but has never been reported to be pathogenic. *P. paracutis*, upon its discovery, was identified as being non-pathogenic and very closely related to *P. cutis*, and was therefore included in the human-associated group (Kunthiphun et al., 2019).

However, there are additional species that cannot be categorised as either cattle- or human-associated species. A potential third lineage comprises of *P. moriformis*, *P. tumulicola* and *P. stagnora*. This group will be referred to as the "true environmental", as there have been no reports of these species causing infection. Finally, *Auxenochlorella* appears to be a monophyletic lineage within the AHP, and is considered a fourth group in this chapter (Darienko and Pröschold, 2015).

It is less clear which group *P. xanthoriae* belongs to. It is a species of *Prototheca* which has never been associated with mammalian pathology, and so could be grouped with the true

environmental species. The type-strain was initially misclassified as *P. wickerhamii*, suggesting that it could be part of the human-associated group. However, phylogenetic trees often cluster *P. xanthoriae* closer to *Auxenochlorella* species than to other *Prototheca* species (Jagielski et al., 2019a; Suzuki et al., 2018). For the purposes of this chapter, *P. xanthoriae* will be grouped with *Auxenochlorella*.

The relationships between these groups are expanded upon in Chapter 7) Phylogenetics of *Prototheca* and Related Organisms.

Pathogenic species from the two pathogenic groups were compared to assess whether the difference in phagocytic capacity observed between *P. wickerhamii* and *P. bovis* is consistent between both pathogenic lineages. Comparisons were also made between species within each group, to assess whether phagocytic behaviour is consistent within each lineage.

This investigation focuses on macrophages. Macrophages are a component of innate immunity, that are recruited in the early stage of an infection and control infection by phagocytosing and killing pathogens (Hirayama et al., 2018). During experimental murine infections, *P. bovis* has been seen to attract macrophages to the site of infection (Shahid et al., 2020b). *Prototheca* have been found within macrophages, as well as extracellularly (Santos et al., 2023; Cheville et al., 1984). It is therefore reasonable that recognition of *Prototheca* by macrophages is likely to be an important element in establishing an infection, though they are unlikely to be the only component of immunity relevant to *Prototheca* infection.

Methods

Algal cell preparation

Algal cells from the AHP lineage used to infect macrophages were grown in SAB broth for two days, as described in the core methods chapter, except for HA6, for which material was scraped off PIM agar having been grown at 20°C under natural sunlight for a variable amount of time. CCAP strains were collected and used directly from cultures that were sent to us.

Cells were washed two or three times, depending on the number of recoverable cells, in Dulbecco's PBS. To pellet, cells were spun at 2700 g at 20°C for 1 minute. Pellets were resuspended by pipetting or scraping from the wall of the Eppendorf tube, as appropriate. Vortexing was avoided, as some species demonstrated a propensity to stick to the plastic tube rather than stay in suspension. Cells were counted on a CytoSmart automated cell counter (CytoSMART Technologies, Corning, NJ, USA).

Tissue Culture

Macrophages that were infected were either J774.1 (the European Collection of Authenticated Cell Cultures) or primary human macrophages. J774.1 (henceforth: murine cells) cells were grown in Dulbecco's Modified Eagle's Medium with low glucose, supplemented with 10% live foetal bovine serum, 2 mM L-glutamine, and 100 U/ml penicillin-streptomycin. Prior to infection, murine cells seeded in 24-well plates at 1×10^5 cells per well and were activated in an equivalent medium without serum and with 150 ng/ml phorbol 12-myristate 13-acetate (PMA) for at least 1 hour. Primary human

macrophages (henceforth: human cells) from two donors were prepared by Samuel Lara-Reyna and grown in complete RPMI 1640 medium supplemented with 10% heat-inactivated human serum, as described elsewhere (Haider et al., 2023). Leukocyte cones used for this chapter were obtained with ethical approval from the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham (approval reference ERN15_0804c). Informed consent was obtained prior to blood donation. All samples were fully anonymised and destroyed after experimentation. Prior to infection, human cells were seeded in 24-well plates at 5×10^5 cells per well, but were not activated with PMA.

Phagocytosis Assay

Macrophages were initially infected at a multiplicity of infection (MOI) of 3:1 (i.e. 3 algal cells per macrophage) for all species of *Prototheca*. Due to difficulties in counting strains related to *P. bovis*, which often resulted in undercounting algal cell cultures used to infect macrophages, later experiments with these strains were infected with these strains to a MOI of 1:1.

Phagocytosis with murine cells was done in CO_2 independent medium (ThermoFisher Scientific), while phagocytosis with human cells was done in complete RPMI 1640 medium supplemented with 10% heat-inactivated human serum supplemented with 5% CO_2. All assays were stained with lysotracker red (LTR; ThermoFisher Scientific) at 100 nM to detect the maturation of the phagolysosome following phagocytosis.

Time lapse images were taken in either a Nikon Eclipse Ti Live microscope or Zeiss Axio Observe Live microscope using a 20X objective for 1 hour. Images were taken at variable intervals across different experiments, ranging from 30 to 60 seconds. Videos were analysed in Fiji (ImageJ).

Quantification was performed for at least 50 macrophages per replicate. The phagocytic index of a given replicate was calculated as the proportion of macrophages which had phagocytosed at least one algal cell. Additionally, the number of algal cells within the macrophage as well as the frame in which the first phagocytosis event was completed were recorded.

Statistical analysis

For statistical analysis, algal species were usually grouped either according to pathology or by lineage, as described in the introduction.

To identify whether differences in phagocytic index existed within a group, data from both murine and human cells were compared. Two-way ANOVA, accounting for algal species and host species, was performed to identify whether significant differences existed between species and whether there was an interaction between algal species and host species.

Pairwise differences were only identified in groups where species were found to have a significant effect by two-way ANOVA. To do this, the data were first restricted to an individual cell type. Then Tukey's honest significance test was used on a one-way ANOVA model, only including algal species. Species were then grouped into statistically similar

groups, with one group being considered significantly different from another if at least half of the pairwise comparisons were statistically significant.

Results

Pathogenic species are not phagocytosed consistently

The main statistic used to compare different species was the proportion of macrophages that phagocytosed at least one algal cell, referred to here as the "phagocytic index". This was chosen both because it allows comparisons between species where phagocytosis is rare, and due to technical limitations (described below) in the two other statistics recorded. Phagocytic indexes for all six pathogenic species across both cell types are presented in Figure 5-1. The number of replicates for all algal strains exposed to macrophages is available in Appendix 13. Two-way ANOVA, including algal species and immune cell type identified significant differences in phagocytic indexes between species ($F = 27.116$, $p = 4.05\times10^{-11}$) and an interaction between algal species and immune cell type ($F = 3.169$, $p = 1.83\times10^{-2}$), but no significant effect of cell type ($F = 1.785$, $p = 0.1902$).

In murine cells, significant differences were observed in 10 pairwise comparisons (Appendix 14), resulting in three statistically similar groups. Most cattle-associated species of *Prototheca* were significantly different from human-associated species, with *P. ciferrii* phagocytosed significantly more than most other species. However, within the cattle associated species the only significant pairwise difference is between *P. bovis* and *P. ciferrii*, indicating quite similar phagocytic behaviour across the cattle-associated species.

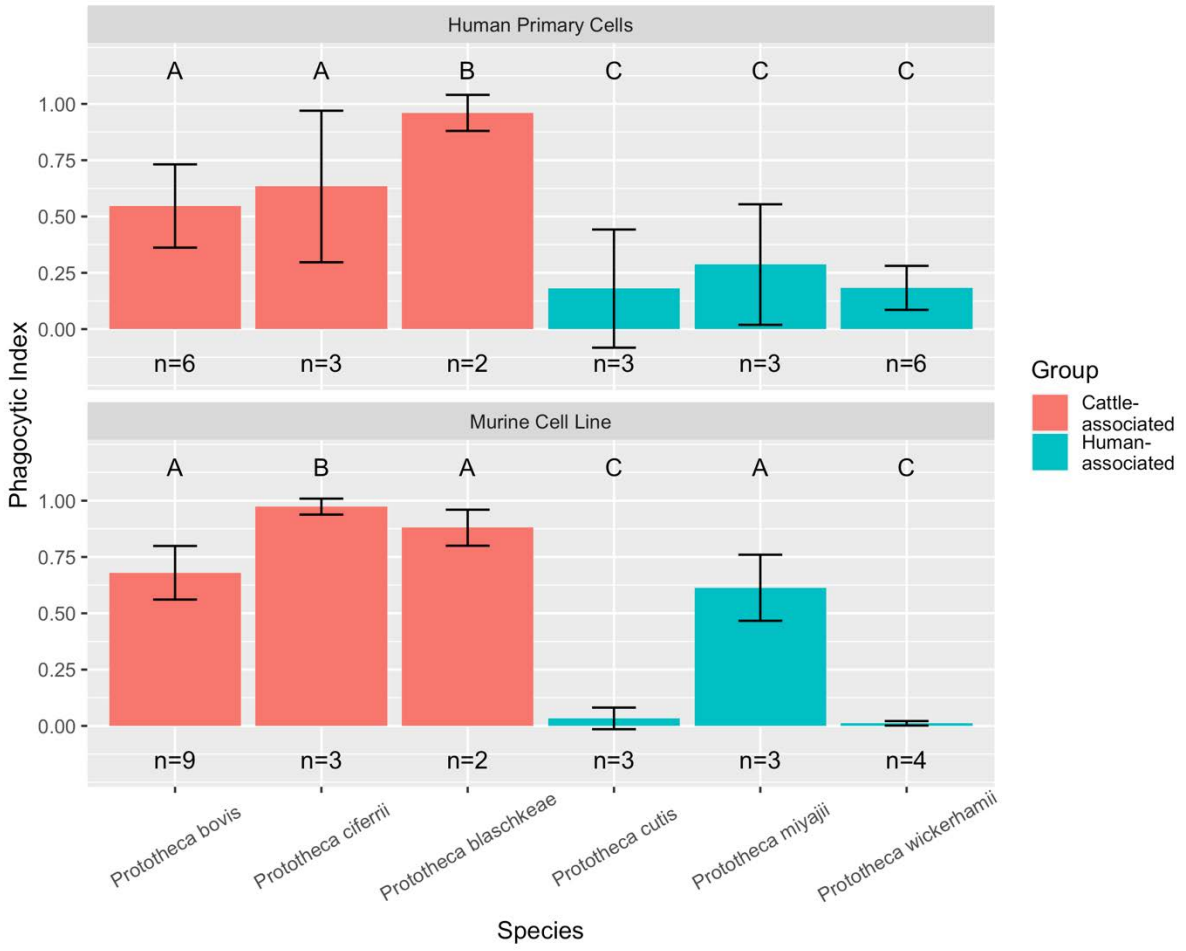


Figure 5-1. Phagocytic activity of macrophages in response to pathogenic species of Prototheca*. Algal cells were exposed to a murine macrophage cell line (J774.1 cells) and human primary macrophages for 1 hour. Bars are coloured according to host preference of the lineage, as defined in the introduction. Error bars are ± 2 standard error. Bars are collected into statistically similar groups indicated by letters based upon pairwise comparisons provided in Appendix 14. Six replicates worth of data of* P. bovis *and one replicate worth of data of* P. wickerhamii *exposed to murine cells were collected by Mohammed Haider. These replicates, as well as two replicates of* P. wickerhamii *exposed to J774.1 cells and all replicates of* P. bovis *and* P. wickerhamii *exposed to human primary cells have been published* (Haider et al., 2023).

These differences were more limited in human cells, where the only significant differences observed were *P. blaschkeae* against the three human pathogenic strains (*P. cutis*, *P. miyajii*, and *P. wickerhamii*; $p = 6.96\times10^{-3}$, 2.230×10^{-2}, and 2.63×10^{-3} respectively). This therefore resulted in three statistical groupings, with group A not significantly different from group B or C despite significant differences between B and C. However, it was noticed that macrophages derived from the second donor phagocytosed almost all AHP species more readily than macrophages from the first donor (Appendix 15), which is addressed in the discussion, and all replicates of *P. blaschkeae* were performed with these cells. In all

statistically significant comparisons, cattle-associated species had a greater phagocytic index than human-associated species.

Unexpectedly, significant differences were observed in the phagocytic index of human-associated species in murine cells, with *P. miyajii* having a much greater phagocytic index than the other two human associated species (Appendix 14). This trend might be present in human cells (Figure 5-1) but is not statistically significant.

Therefore, not all pathogenic species of *Prototheca* are phagocytosed in a similar manner. Additionally, while cow-associated pathogenic species generally behave consistently with each other, there is unexpected variation in the phagocytic capacity of the human-associated pathogenic species. Similar comparisons within each group were similarly investigated, to determine whether similar diversity existed within other groups and between pathogenic and non-pathogenic species.

Some AHP sub-lineages are phagocytosed consistently

Host phagocytes may exert a selective pressure on a pathogen, either by favouring less "phagocytosable" pathogens that are not internalised and killed by macrophages or by favouring more "phagocytosable" pathogens if an intracellular niche within macrophages is protective against other immune responses. Thus, elevated or diminished phagocytic indexes in pathogenic species could be interpreted as a result of selection pressures experienced during infection. However, it is difficult to draw these kinds of conclusions without comparing pathogenic and non-pathogenic species. It is possible that elevated or

diminished phagocytic indexes are a result of cell wall or plasma membrane modifications that occurred within a lineage independent of pathology.

To determine whether the elevated phagocytic index observed in cattle-associated species was likely to be an adaptation that occurred in pathogenic species as a result of immune selective pressures, phagocytic indexes were compared among the three pathogenic, cattle-associated species and one non-pathogenic, cattle-associated species (*P. cookei*).

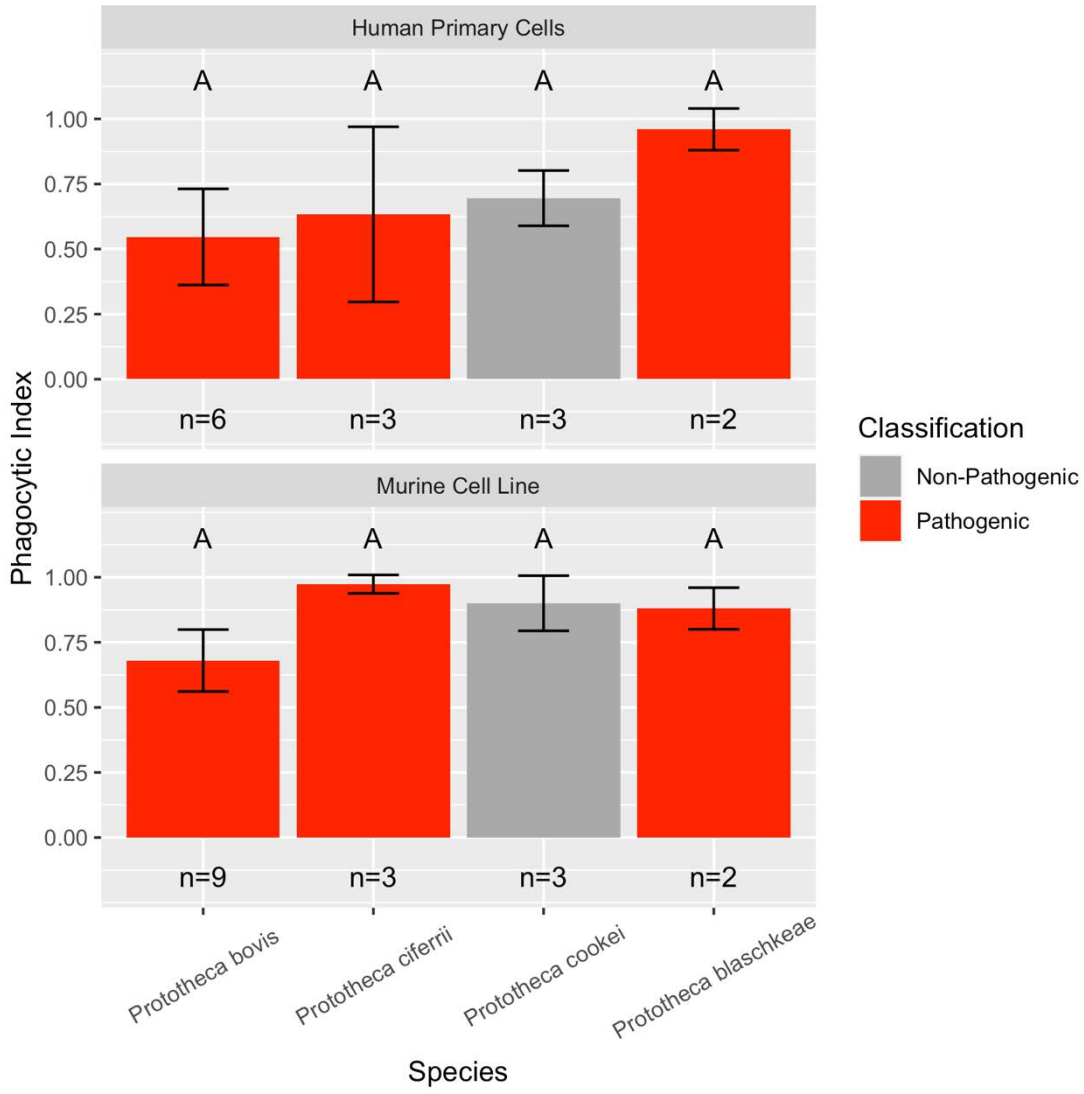


Figure 5-2. Phagocytic activity of macrophages in response to cattle-associated species of Prototheca. *Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to whether the species in question has at least one known instance of causing an infection. Error bars are ± 2 standard error.*

All bars are collected into the same statistically similar group indicated by letters based upon pairwise comparisons provided in Appendix 17. Six replicates worth of data of P. bovis *exposed to murine cells were collected by Mohammed Haider. These replicates, as well all replicates of* P. bovis *exposed to human cells have been published* (Haider et al., 2023).

All four species are readily phagocytosed by both murine and human macrophages (Figure 5-2, Figure 5-3, Appendix 16). Two-way ANOVA identified significant differences between algal species ($F = 3.867$, $p = 2.24\times10^{-2}$) and immune cell type ($F = 6.205$, $p = 2.04\times10^{-2}$), but no significant interaction between the two ($F = 1.187$, $p = 3.37\times10^{-1}$). However, the only significant pairwise difference between these strains was between *P. bovis* and *P. ciferrii* in murine cells (difference = 0.29 (95% CI = 0.0081-0.58), $p = 4.31\times10^{-2}$). Thus, no species in the cattle-associated group is significantly different from the others.

Additionally, there is therefore no significant difference between the non-pathogenic *P. cookei* and the three pathogenic cattle-associated species. This indicates that the elevated phagocytic index of this lineage is probably not an adaptation to selection pressures imposed by the immune system.

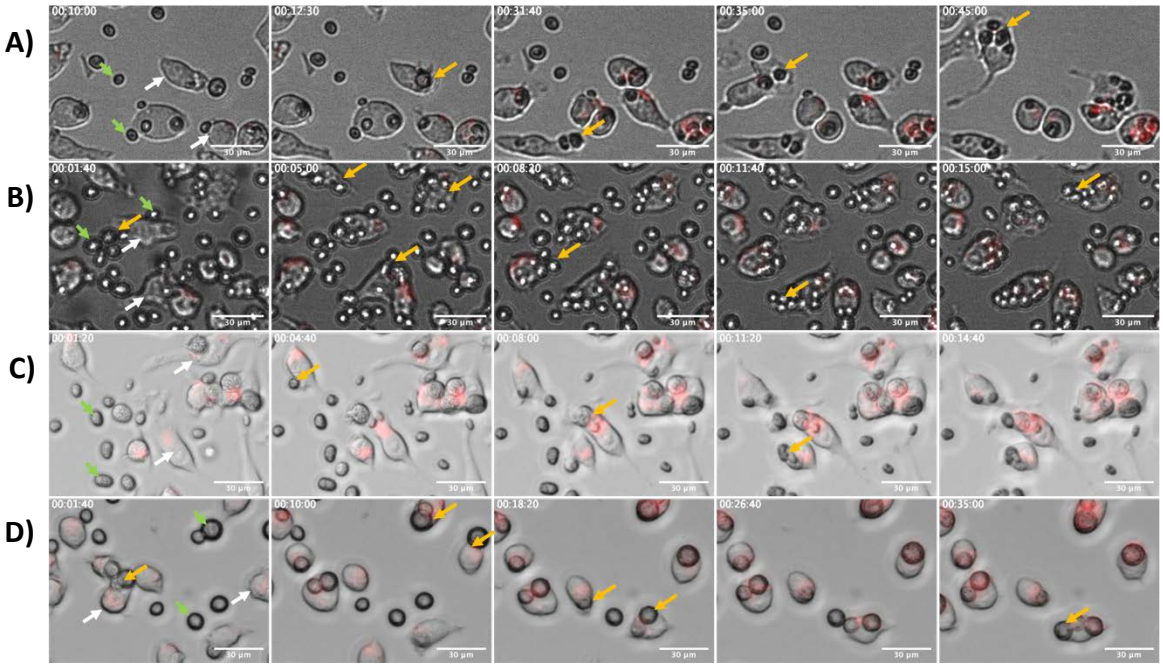


Figure 5-3. Phagocytosis of cattle-associated species of Prototheca *by murine cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; phagocytosis events are indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – primarily acidified phagolysosomes containing phagocytosed*

algal cells. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A) P. bovis*, HP3; B)* P. ciferrii*, HP2; C)* P. cookei*, HP32; D)* P. blaschkeae*, HP4.*

When *Prototheca* species that do not belong to either pathogenic lineage were exposed to macrophages, the phagocytic index was consistently low or negligible for both species in both murine and human cells (Figure 5-4). There was no significant difference between the species ($F = 4.308$, $p = 8.323\times10^{-2}$), but there was a significant difference between immune cell types ($F = 30.278$, $p = 1.51\times10^{-3}$) and a significant interaction between immune cell type and species ($F = 9.823$, $p = 2.022\times10^{-2}$). These significant differences appear to be a result of the phagocytic index of *P. tumulicola* in murine cells, which was different to all other conditions by virtue of being zero in all replicates. There is therefore little variation within this lineage either.

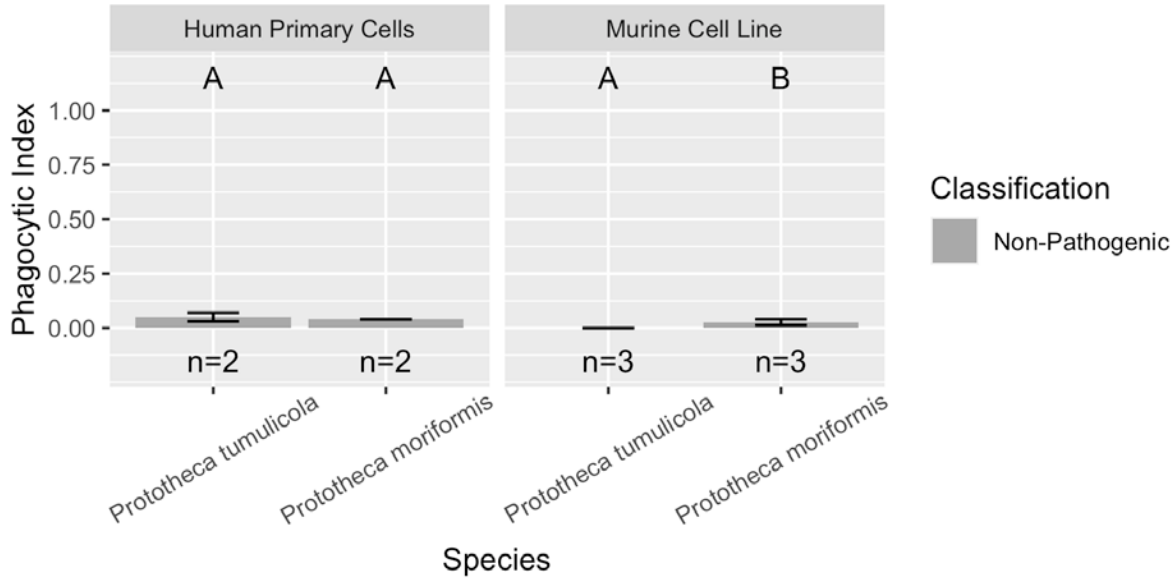


Figure 5-4. Phagocytic activity of macrophages in response to true environmental species of Prototheca*. Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to whether the species in question has at least one known instance of causing an infection. Error bars are ± 2 standard error. Bars are collected into statistically similar groups, indicated by letters, based upon pairwise comparisons provided in Appendix 18.*

However, during quantification it was noticed that murine macrophages appeared to move towards *P. moriformis* cells, despite a low phagocytic index (Figure 5-5). Macrophage recognition for chemotaxis and phagocytosis often use different receptors (Senoo et al.,

2016). This may suggest that some *Prototheca* species produce molecules that can be detected by and attract macrophages, even if they cannot stimulate phagocytosis.

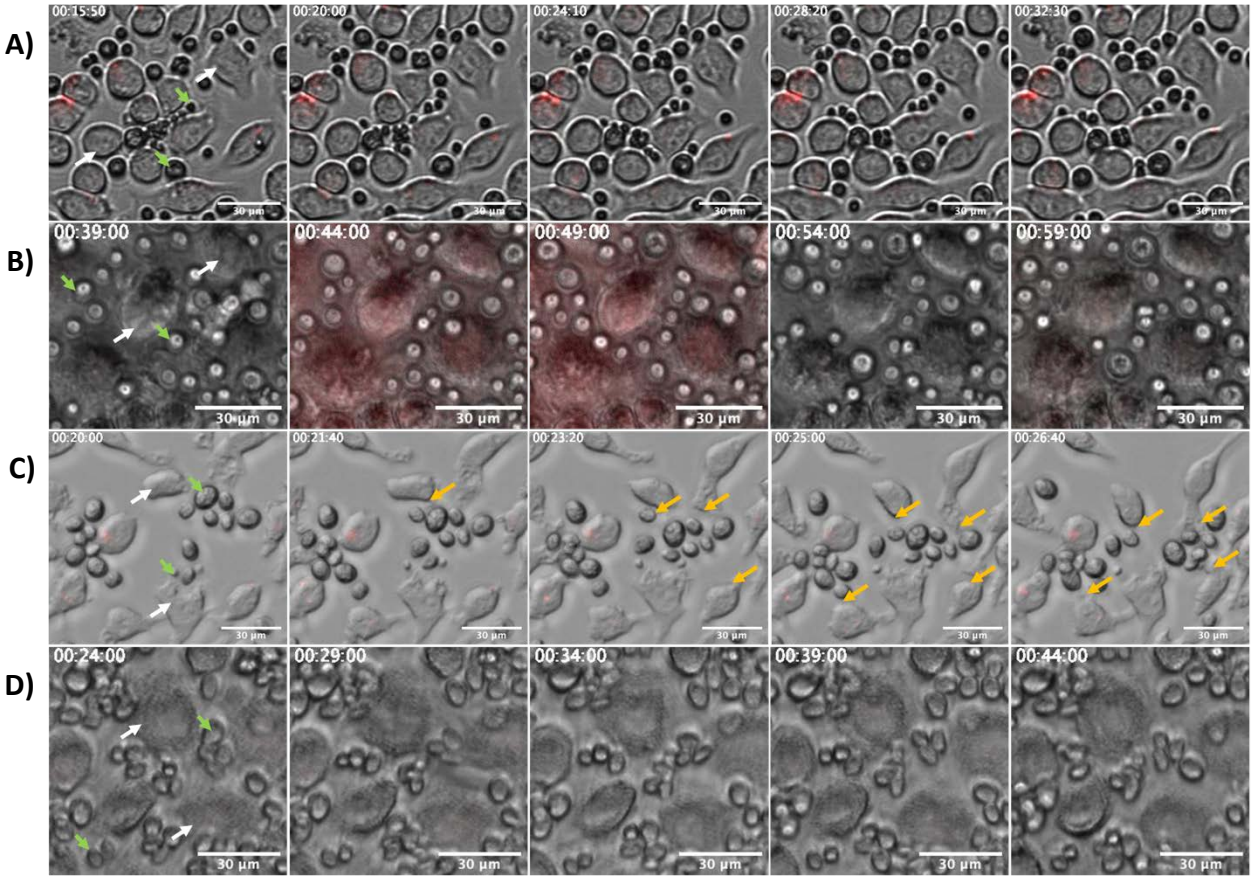


Figure 5-5. A lack of phagocytosis of true environmental species of Prototheca *by macrophages. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; suspected chemotaxis is indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – indicating a lack of acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from a single video, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. tumulicola*, HP29, in murine cells; B)* P. tumulicola*, HP29, in human cells from donor 2; C)* P. moriformis*, HP51, in murine cells; D)* P. moriformis*, HP51, in human cells from donor 1.*

Some AHP sub-lineages are not phagocytosed consistently

As with cattle-associated strains, the phagocytic indexes of the human-associated pathogenic strains were compared with the non-pathogenic *P. paracutis*.

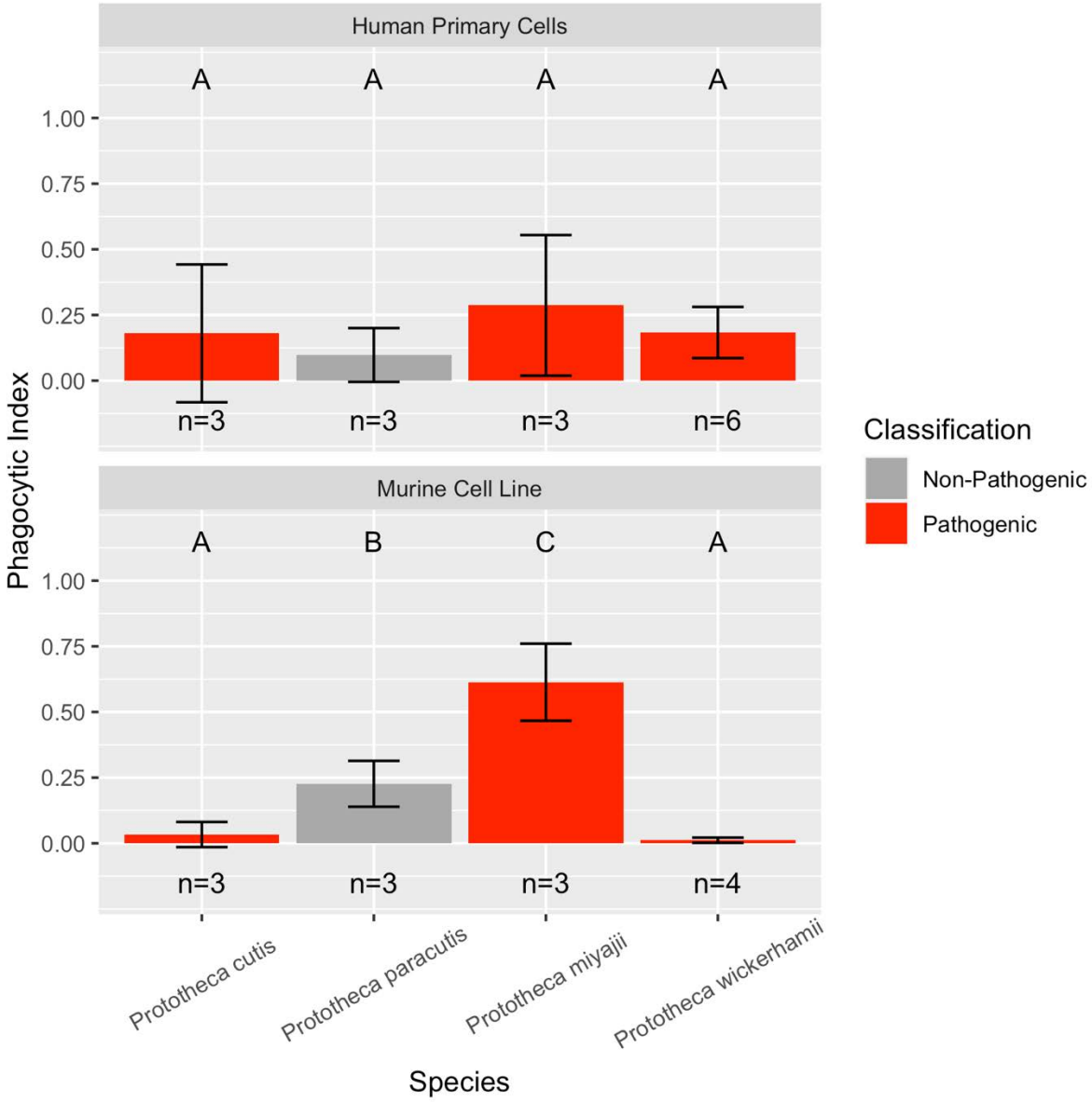


Figure 5-6. Phagocytic activity of macrophages in response to human-associated species of Prototheca*. Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to whether the species in question has at least one known instance of causing an infection. Error bars are ± 2 standard error. Bars are collected into statistically similar groups, indicated by letters, based upon pairwise comparisons provided in Appendix 19. One replicate worth of data of* P. wickerhamii *exposed to murine cells were collected by Mohammed Haider. This replicate, as well as two replicates of* P. wickerhamii *exposed to murine cells and all replicates of* P. wickerhamii *exposed to human primary cells have been published* (Haider et al., 2023).

The four human-associated species are generally phagocytosed less than the cattle-associated species and more than the true environmental species. However, they displayed much more variability between species (Figure 5-6). Two-way ANOVA revealed significant differences in phagocytic index between algal species ($F = 9.802$, $p = 3.47\times10^{-4}$) and a significant interaction between immune cell type and algal species ($F = 5.550$, $p = 6.15\times10^{-3}$).

As before, no significant differences were observed in human cells, but significant differences were observed between the phagocytic index of *P. miyajii* and all other algal species in murine cells (Appendix 19). Additionally, the phagocytic index of *P. paracutis* was significantly different from all other algal species in murine cells (Appendix 19).

The human-associated group therefore shows surprising variability in their phagocytic indexed though this difference appears to be lost in human macrophages. It is also therefore unclear what the ancestral state was within this lineage.

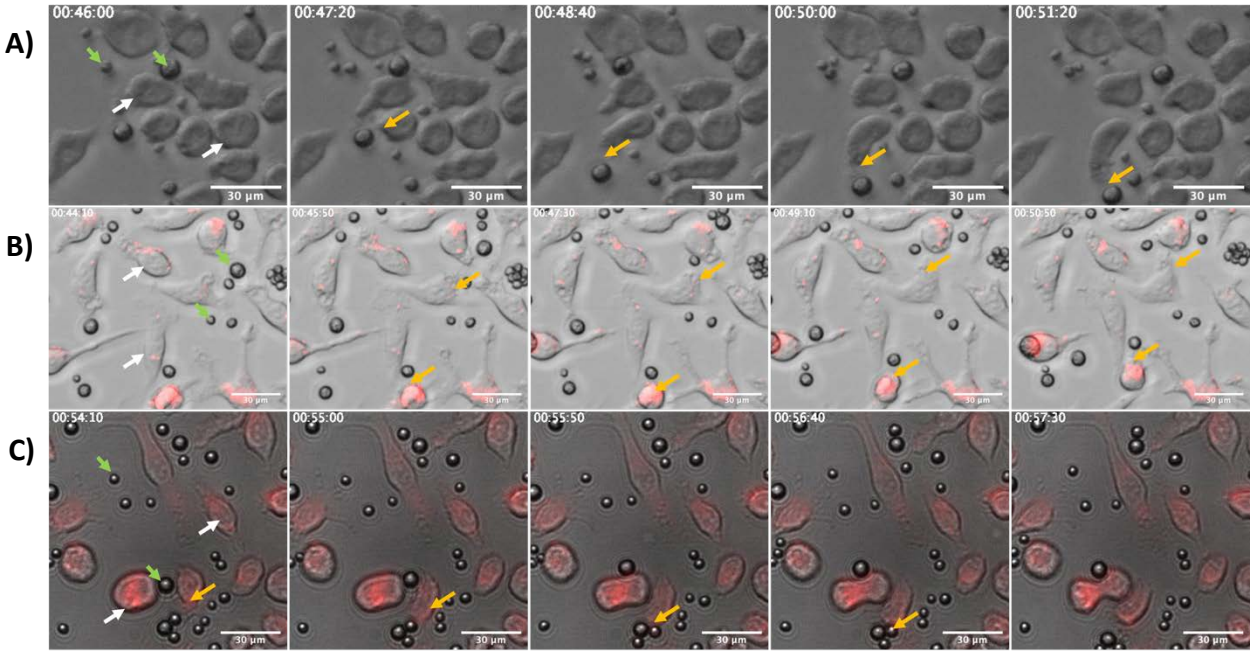


Figure 5-7. Chemotaxis of murine cells towards human-associated species of Prototheca*. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; suspected chemotaxis is indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – indicating a lack of acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from a single video, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. cutis*, HP28; B)* P. paracutis*, HP31; C)* P. wickerhamii*, HP52.*

As with *P. moriformis*, all human-associated species appeared to attract murine macrophages, even if they had a low phagocytic index (Figure 5-7). This was not observed in

human macrophages (Appendix 20), though chemotaxis was not clearly observed in human macrophages for any species.

It was also observed that, while murine macrophages do phagocytose *P. miyajii* cells, they appear to prefer to phagocytose larger cells (Figure 5-8). This apparent preference was not quantified or analysed statistically. Nor was it readily apparent in human cells.

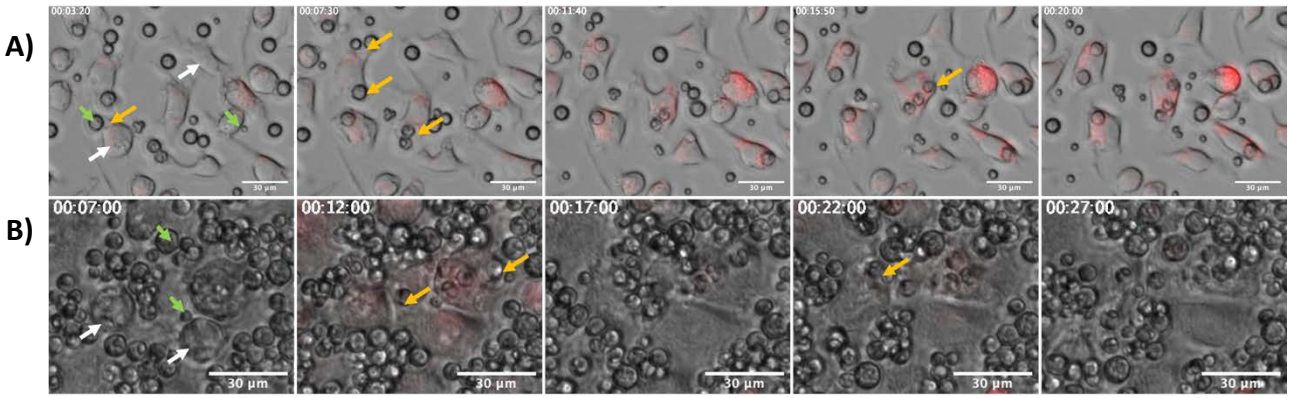


Figure 5-8. Phagocytosis of P. miyajii, *HP27, by macrophages. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; phagocytosis events are indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – which could not reliably identify acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from videos, at times indicated by the time stamp in the top left. Algal cells were exposed to: A) murine cells; B) human primary cells from donor 2.*

As the *Auxenochlorella* genus is closely related to the human-associated *Prototheca* species, it was considered that the phagocytic index of *Auxenochlorella* may shed light on the ancestral phagocytic capacity of human-associated *Prototheca* species.

However, these species also showed variation in their phagocytic index (Figure 5-9). Two-way ANOVA revealed that species has a significant effect on phagocytosis index ($F = 7.184$, $p = 2.19\times10^{-4}$). Cell type did not have a significant effect ($F = 1.971$, $p = 0.169$), nor was there a significant interaction between cell type and species ($F = 0.398$, $p = 0.808$). Pairwise comparisons revealed that *P. xanthoriae* is significantly different from all *Auxenochlorella* taxa except for *Auxenochlorella* sp. 1 in murine cells (Appendix 21), but that there were

otherwise no pairwise significant differences in murine or human cells. There is therefore some variability between species in the *Auxenochlorella* group.

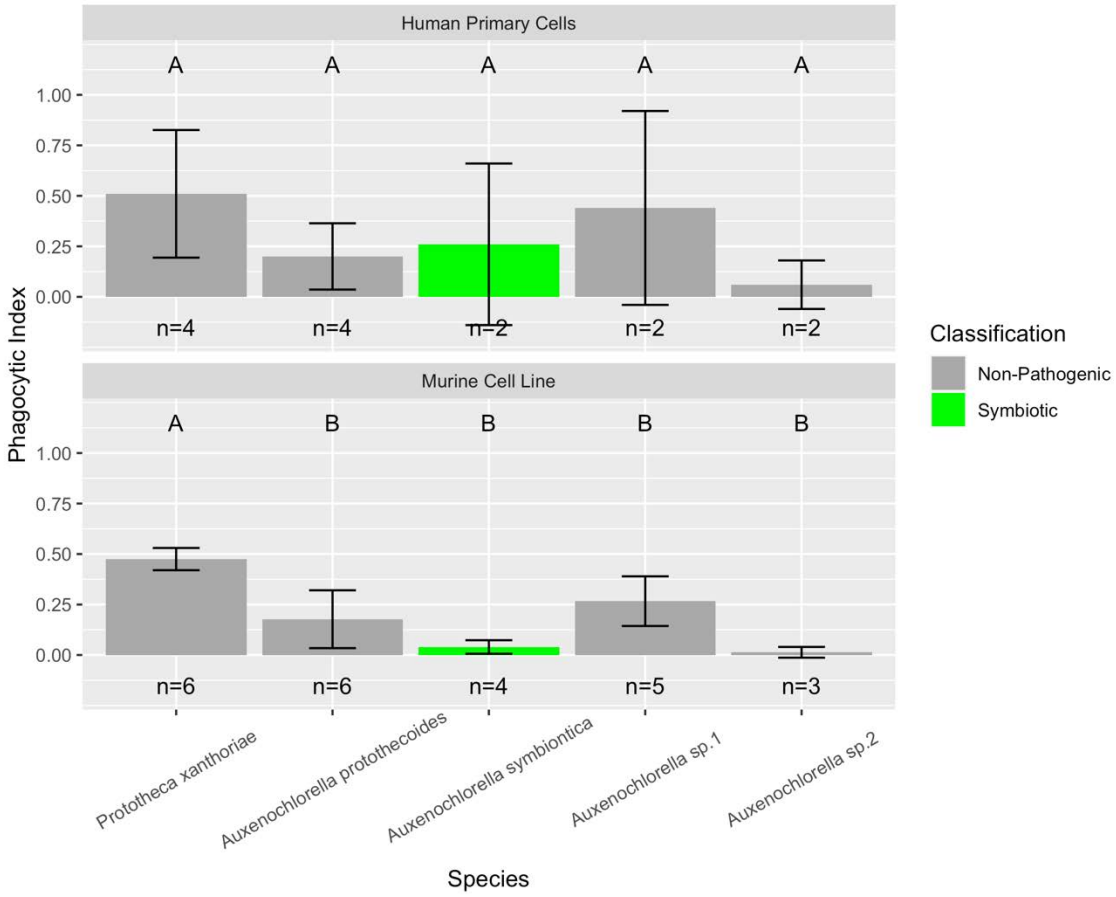


Figure 5-9. Phagocytic activity of macrophages in response to Auxenochlorella *and closely related species of* Prototheca*. Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to whether the species in question is known to be symbiotic. Error bars are ± 2 standard error. Bars are collected into statistically similar groups, indicated by letters, based upon pairwise comparisons provided in Appendix 21.*

Additionally, there was surprising variability of the phagocytic index within some species (Figure 5-10), but not others (Figure 5-11). Two strains were responsible for providing this variation: HA1 and HA6 (Figure 5-12).

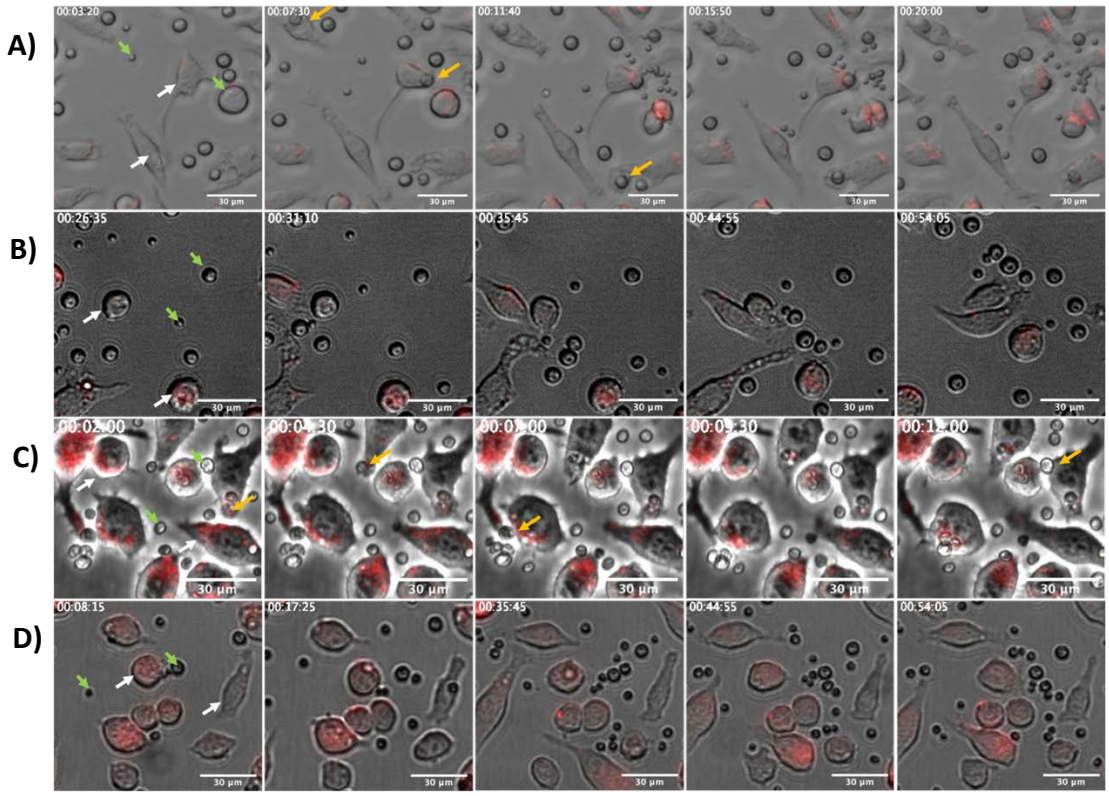


Figure 5-10. Inconsistent phagocytosis of strains of Auxenochlorella *by murine cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; phagocytosis events are indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – which could not reliably identify acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A&B)* A. protothecoides*, HA1; C&D)* Auxenochlorella *sp. 1, HA6. Phagocytosis is observed in rows A and C and is not observed in rows B and D.*

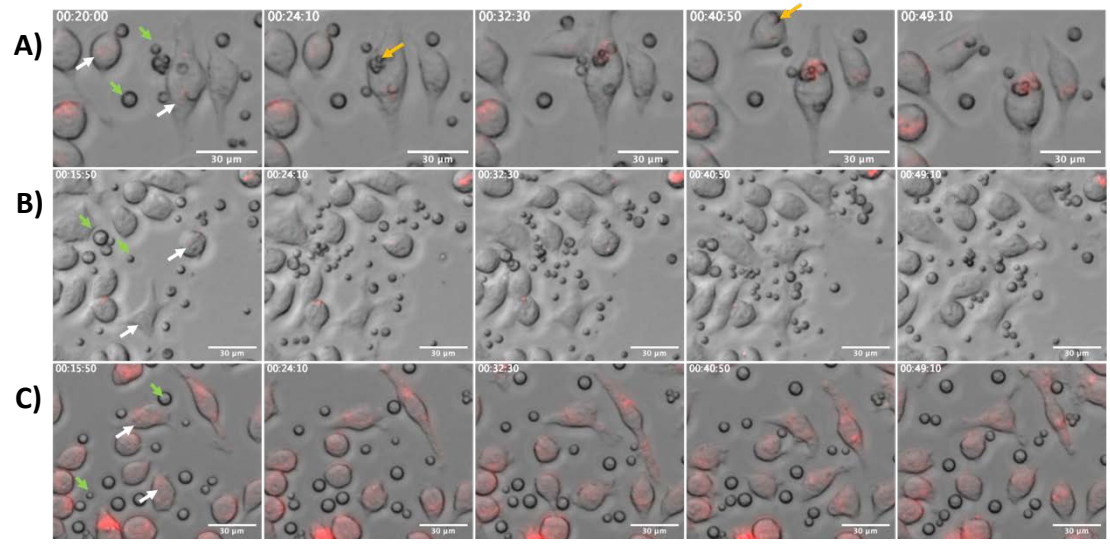


Figure 5-11. Consistent phagocytosis of strains of Auxenochlorella *and closely related species of* Prototheca *by murine cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; phagocytosis events are indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – which could not reliably identify acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. xanthoriae*, HP54; B)* A. symbiontica*, HA5; C)* Auxenochlorella *sp. 2, HA7.*

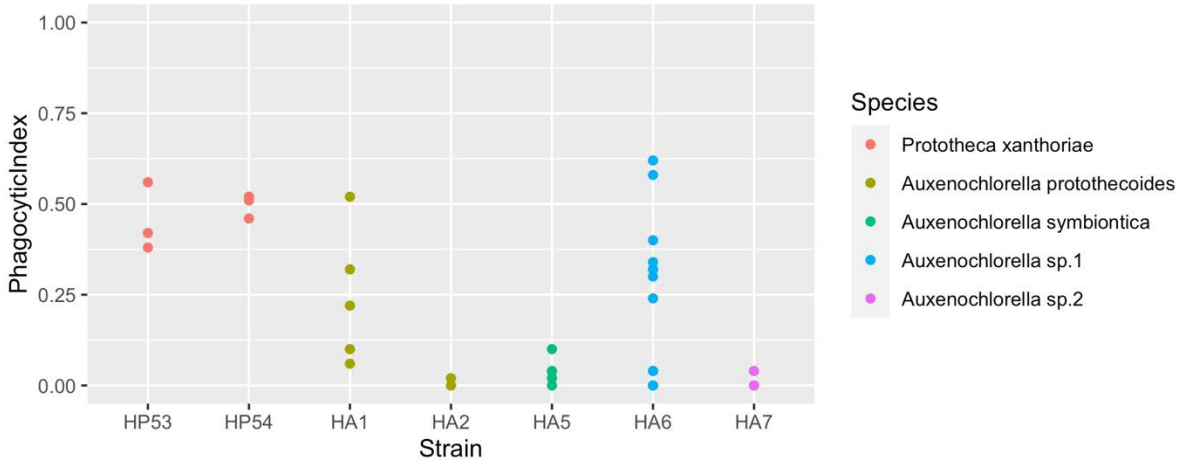


Figure 5-12. Phagocytic activity of murine cells in response to individual strains of Auxenochlorella *and closely related species of* Prototheca*. Each point represents the phagocytic index of a given strain from a single replicate. Points are coloured according to species.*

Due to the distinct growth requirements of HA6, the cultures of HA6 (representing *Auxenochlorella* sp. 1) provided to macrophages varied in age. Two younger cultures (16 days or younger) had phagocytic indexes less than 0.04, while older cultures (16 days or older) had phagocytic indexes greater than or equal to 0.24. Thus, it seems that phagocytosis is inducible by age for HA6, presumably as a result of cell wall remodelling that is known to happen as algal cultures age. It is not known whether this is true of other AHP species. There was no obvious cause of the variability in *A. protothecoides*.

Multiple metrics suggest cattle-associated species are phagocytosed more readily than other species

Three metrics are potentially of interest in identifying differences in dynamics between *Prototheca* and macrophages, as defined in the methods section: the phagocytic index, the proportion of macrophages that phagocytose at least one algal cell; the number of internalised algae; and the phagocytosis rate, which we consider to be the time at which the first phagocytosis event has been completed.

During quantification, the phagocytosis rate and number of algal cells within each macrophage were also recorded. Comparisons within each lineage group were made to determine whether there were significant differences, as was done for phagocytic index. Additionally, comparisons were made between groups, to determine whether different lineages were internalised differently.

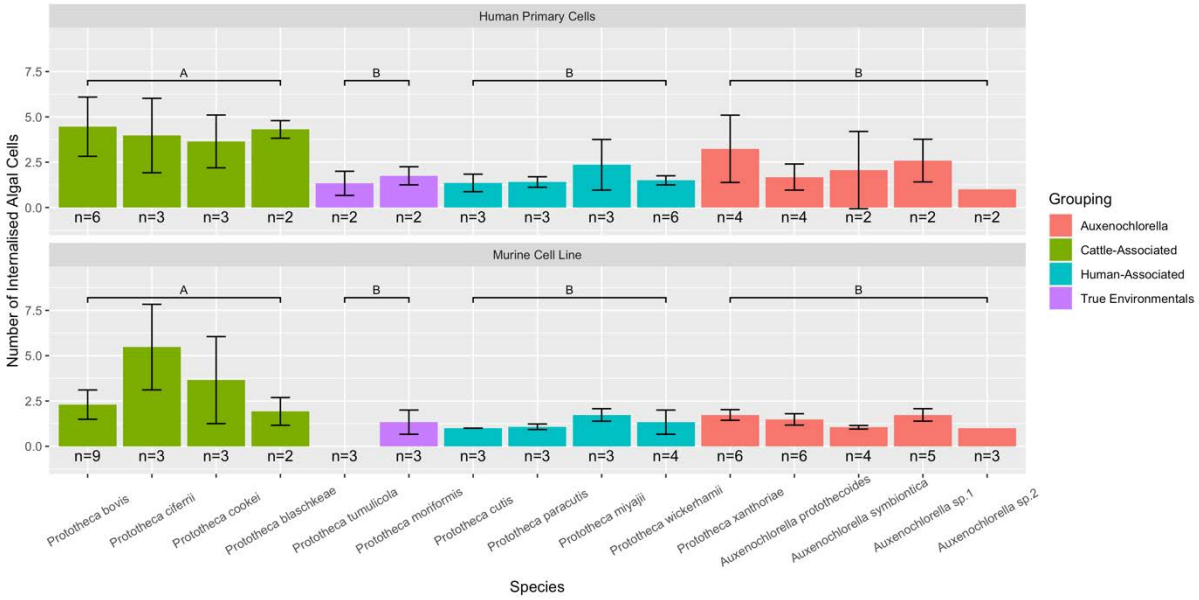


Figure 5-13. Number of algal cells phagocytosed per macrophages in response to species from the AHP lineage. Only data from macrophages that phagocytosed at least one algal cell is included. Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to the sub-lineage of each species, as defined in the introduction. Error bars are ± 2 standard error. Phagocytosis only occurred in one replicate for Auxenochlorella *sp. 2 in both cells, meaning standard error could not be calculated. Phagocytosis was not recorded for* P. tumulicola *in murine cells, thus no algae were internalised. Bars are collected into statistically similar groups, indicated by letters, based upon pairwise comparisons provided in Appendix 22. Six replicates worth of data of* P. bovis *and one replicate worth of data of* P. wickerhamii *exposed to murine cells were collected by Mohammed Haider.*

Two-way ANOVA only revealed significant differences in the number of internalised algae within a group for the human-associated species ($F = 3.288$, $p = 4.46\times10^{-2}$), but no significant difference was reported within any other group. However, within the human-associated species, there were no significant pairwise differences. Between groups, macrophages phagocytosed more algal cells from cattle-associated species than any other group in both

cell types, except for the true environmental species in murine cells (Figure 5-13, Appendix 22).

Two-way ANOVA revealed significant differences in phagocytosis rates within human-associated species ($F = 4.237$, $p = 1.975\times10^{-2}$) and the *Auxenochlorella* group ($F = 8.586$, $p = 1.67\times10^{-4}$). However, significant pairwise differences were only seen between species in the *Auxenochlorella* group, and only in murine cells, where *Auxenochlorella* sp. 1 was phagocytosed more quickly than all other *Auxenochlorella* species (Appendix 23). Between groups, it appears that cattle-associated species are generally phagocytosed more quickly than other groups in both cell types, though the difference is less in human cells (Figure 5-14, Appendix 24). Additionally, environmental species are phagocytosed more slowly than *Auxenochlorella* species, allowing these groups to be differentiated. However, there was no significant difference between environmental and human-associated *Prototheca* species, or *Auxenochlorella* and human-associated *Prototheca* species.

It therefore appears that cattle-associated species are phagocytosed more quickly, and that more algal cells are taken up per macrophage, than other lineages. But there is currently no indication that phagocytosis is different between cattle-associated species.

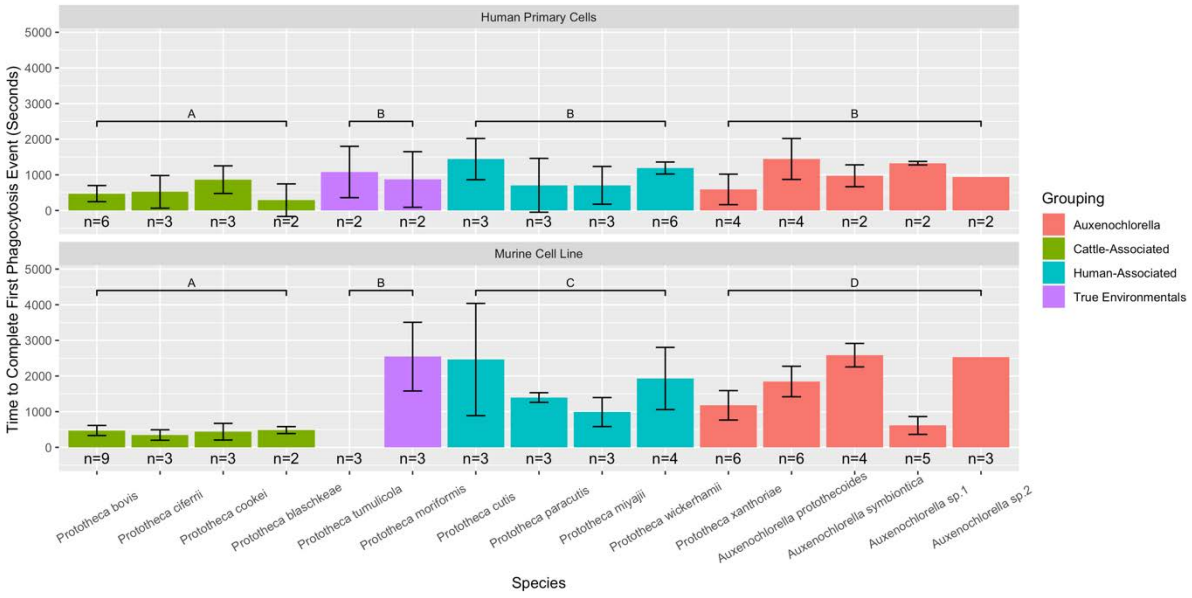


Figure 5-14. Rate of phagocytosis of macrophages in response to species from the AHP lineage. Algal cells were exposed to a murine macrophage cell line (J774.1) and human primary macrophages for 1 hour. Bars are coloured according to the sub-lineage of each species, as defined in the introduction. Error bars are ± 2 standard error. Phagocytosis only occurred in one replicate for Auxenochlorella *sp. 2 in both cells, meaning standard error could not be calculated. Phagocytosis was not recorded for P. tumulicola in murine cells, thus phagocytosis took no time. Bars are collected into statistically similar groups, indicated by letters, based upon pairwise comparisons provided in Appendix 24. Six replicates worth of data of* P. bovis *and one replicate worth of data of* P. wickerhamii *exposed to murine cells were collected by Mohammed Haider.*

There are, however, confounding factors that weaken confidence in these comparisons. Cattle-associated algal cell cultures were difficult to normalise, probably as a result of hydrophobicity of the cells. As such, these strains were present at a higher MOI, meaning more cells were present to be taken up. This may account for the greater number of internalised algae in this lineage. However, all cattle-associated species were similarly affected, meaning there is no suggestion that significant differences exist between cattle-associated species.

Additionally, due to technical issues, the amount of time taken to focus a video varied widely (from less than five minutes up to half an hour), weakening comparisons of phagocytosis rates. However, this would not have systematically affected one group more

than another, meaning this is likely to weaken the strength of comparisons but not unduly bias them.

These metrics are worth investigating more rigorously in the future to better understand the different dynamics between sub-lineages and species of the AHP lineage.

Phagocytosis of distantly related algae is limited but does occur

The variability of phagocytic indexes within the AHP lineage makes it unclear what phagocytic index would be considered an ancestral state for the lineage. Without this, it is difficult to infer what changes may or may not have been associated with pathology.

For this reason, algae from more distantly related taxa were provided to murine macrophages, to identify what a "normal" green algal phagocytic index might be. The relationships of these strains to *Prototheca* are provided in Table 3-2. There was sufficient material for two replicates of all strains except *Trebouxia decolorans*, which only had enough material for one replicate.

Generally, there was limited phagocytosis of non-AHP algae, with most replicates of most strains having a phagocytic index below 0.25 (Figure 5-15). *Chlamydomonas reinhardtii* was phagocytosed the least, with an index of 0.01, while *Coccomyxa galuniae* is reliably phagocytosed the most, with an index of 0.21. However, as with some *Auxenochlorella* strains, the phagocytic index of *Keratococcus bicaudatus* is quite variable (Appendix 25).

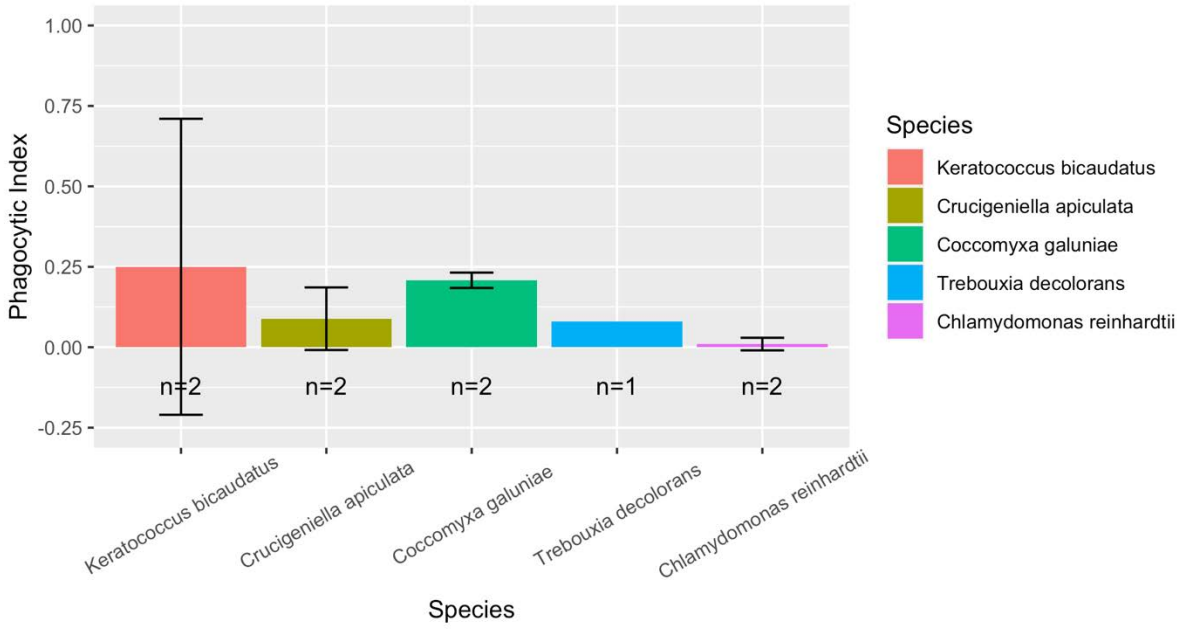


Figure 5-15. Phagocytic activity of macrophages in response to non-AHP green algae. Phagocytosis indexes are generally low, but phagocytosis does occur for non-AHP algae. Algal cells were exposed to a murine macrophage cell line (J774.1) for 1 hour. Bars are coloured according to algal species. Species are arranged in order of relatedness to Prototheca*, with the leftmost species being the closest related to* Prototheca*. The full taxonomy of these strains, and thus their relatedness to* Prototheca*, is available in Table 3-2. Error bars are ± 2 standard error.*

During quantification, particularly for the first replicate, there appeared to be limited chemotaxis of murine macrophages towards the algae. It is worth noting that four of the five cultures from the CCAP are known to contain bacteria, and while chemotaxis was observed in the second replicate for some algae, it is not clear that the macrophages were attracted to the algae and not these bacteria.

Immune cell death may have been observed with some species

The work undertaken in this chapter was not designed to observe macrophage death. However, during quantification, it was observed that individual macrophages would occasionally rapidly lose LTR staining, which was not observed in neighbouring macrophages. This was interpreted as being the result of macrophage death.

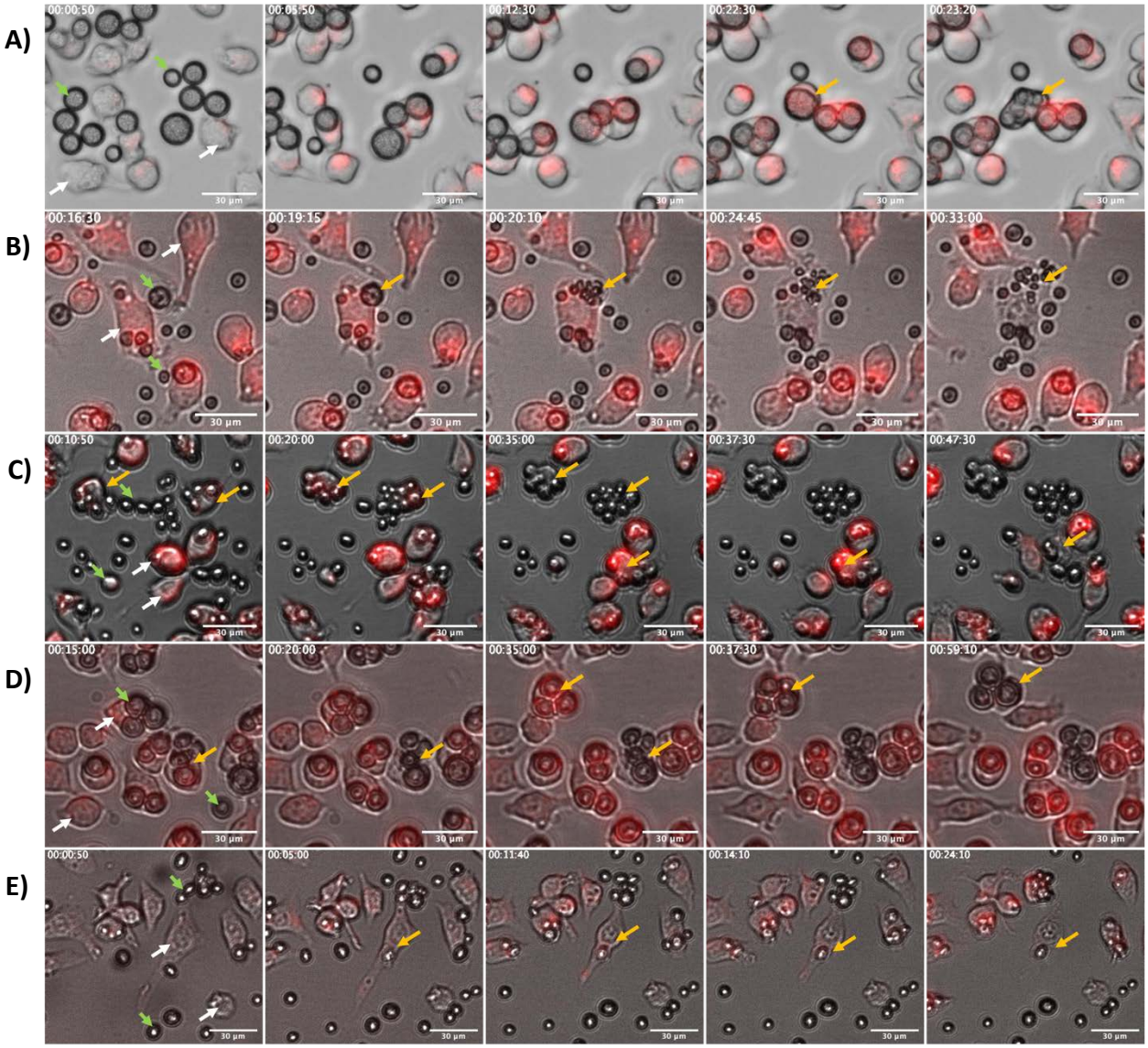


Figure 5-16. Apparent death of murine cells following phagocytosis of Prototheca *cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows. Loss of LTR staining across the whole cell was interpreted as death of the macrophage, which is indicated by yellow arrows. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row, as well as speculated cause of death, are: A)* P. blaschkeae*, HP4, lytic release; B)* P. miyajii*, HP27, lytic release; C)* P. bovis*, HP3, overfull macrophage; D)* P. blaschkeae*, HP4, overfull macrophage; E)* P. ciferrii*, HP2, no clear cause.*

Deaths were observed for four species: *P. blaschkeae*, *P. bovis*, *P. ciferrii*, and *P. miyajii* (Figure 5-16). Death was occasionally associated with the lytic release of daughter cells from a mature sporangium, which may indicate that some of these deaths are a result of mechanical force or a molecule released by this lysis. However, a few were not associated with lytic release of daughters but did occur in macrophages that were extremely full, while others were not for any clear reason.

There were also instances of algal division inside a macrophage that disrupted some LTR localisation, but did not cause the whole macrophage to lose the stain (Figure 5-17). This may not represent death of the macrophage, but may represent algal escape from the phagolysosome. It is unknown if these macrophages died following *Prototheca* escaping from the phagolysosome.

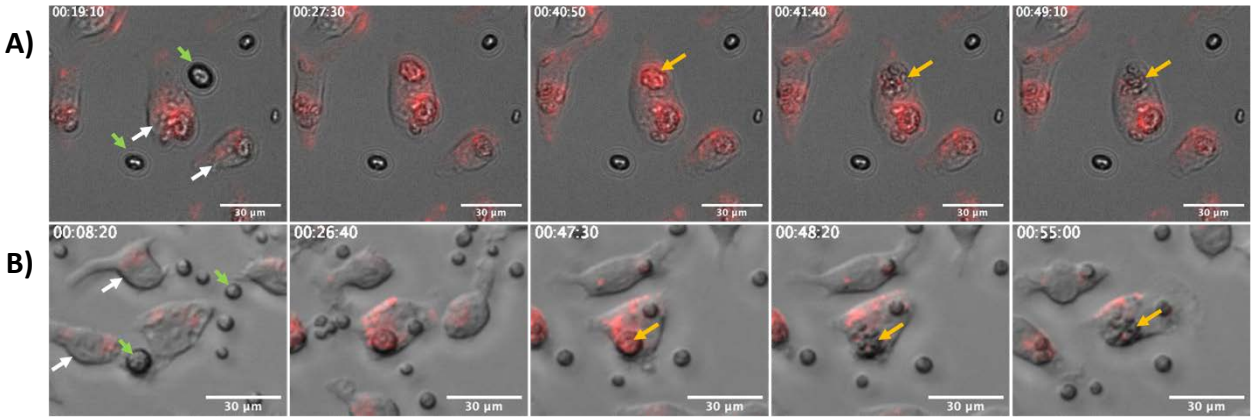


Figure 5-17. Apparent escape from murine phagolysosome following phagocytosis of Prototheca *cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows. Loss of LTR staining localised to an algal cell without loss of staining of the whole cell was interpreted as escape from the phagolysosome by the algal cell, which is indicated by yellow arrows. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. cookei*, HP32; B)* P. miyajii*, HP27.*

Discussion

In this chapter I have investigated the interaction between the AHP lineage and macrophages from two mammalian hosts. I have identified surprising variation between AHP species, as well as between host species.

Interpretation from the perspective of immune cells

To the best of my knowledge, this is the first investigation into interactions between *Prototheca* and the mammalian immune system that spans the genus, rather than focusing

on one or two species. Doing so appears to have revealed a surprising degree of complexity to these interactions between and within the various sub-lineages within the AHP lineage.

Firstly, it is interesting that there appear to be different trends in phagocytosis between cells derived from different host species. Several strains (i.e. *P. cutis*, *P. wickerhamii*, *A. symbiontica*) appear to have extremely low or negligible levels of phagocytosis in the murine cell line J774.1, but are phagocytosed more readily by human primary macrophages. This suggests some strains are recognised by some macrophages but not by others. Differences between murine and human immune responses to *Prototheca* have been observed (Shahid et al., 2020a).

From the data presented in this chapter, it is not clear whether this difference is due to differences in receptors (either as presence or absence of receptors, or subtle differences in how the same receptor in different animals might recognise pathogens) or due to changes that occurred in murine cells as part of being a cell line. However, in recently published work, we showed that induced BMDMs similarly struggled to phagocytose a strain of *P. wickerhamii* (Haider et al., 2023). Thus, it appears there are genuine differences in recognition between murine and human macrophages of *P. wickerhamii*, which may extend to the rest of the genus.

The possibility that the Immune cells of different host species have different capacities for recognising certain *Prototheca* species, particularly differences between mouse and human cells, has implications for establishing a mouse model to study *Prototheca* infection, as mice may not be capable of interacting with *Prototheca* infections as humans do. Additionally, it

may provide a possible explanation for the apparent host preference of pathogenic *Prototheca* species. It is currently unknown whether cattle macrophages resemble human or mouse cells, or possess their own pattern of phagocytosis. It is possible that mice are not representative of hosts for *Prototheca* infections as there are no recorded cases of infection of mice with *Prototheca*, only transient experimental association with rats (Pore and Shahan, 1988). Additionally, it is unknown whether this difference holds for other aspects of immunity (e.g. other immune cells, complement).

It Is Interesting that different host Individuals within the same species may have different capacities for recognising *Prototheca*, as observed here with human cells from the second donor having higher phagocytic indexes for most *Prototheca* species than cells from the first donor. A similar phenomenon was observed by Cox *et al.*, where the lymphocytes of a patient infected with *P. wickerhamii* did not respond to *Prototheca* cells, while those of three healthy individuals did (Cox et al., 1974). Similar inter-host variability has been reported for the pathogenic fungus *Cryptococcus* (Garelnabi et al., 2018). It remains to be seen whether this variability is relevant to infection.

It Is worth noting that the algal cells in this chapter are not opsonised. In our previous work we did not find that opsonising *P. wickerhamii* with human serum increased the phagocytic index substantially for murine macrophages (Haider et al., 2023), but it is likely that the trends observed here do not fully recreate phagocytic dynamics in the context of a real infection.

It was also noted that macrophages sometimes appear to be attracted towards algal cells that they were unable to phagocytose. This may indicate that *Prototheca* cells can attract immune cells even if those cells cannot recognise the algal cell for phagocytosis.

Interpretation from the perspective of AHP algae

Recognition and invasion

Phagocytosis events were observed for all species in human cells. Some species appear to have higher or lower phagocytosis indexes compared to other species, but there are limited statistically significant differences (the only significant pairwise difference was *P. blaschkeae* compared to *P. miyajii*). This may be due to the apparent differences between cells derived from different donors increasing the error for each *Prototheca* species.

However, there are clear differences between the recognition of different *Prototheca* species by murine macrophages. This suggests that there are differences between different *Prototheca* species that are relevant for their interaction with some mammalian immune systems. This could be another basis for host preference.

One possibility is that different species have different molecules on their surfaces, either on their plasma membrane or as components of their cell wall, only some of which can be recognised by macrophages for phagocytosis. This would likely account for the phagocytic index of *P. cookei* resembling its pathogenic close relatives *P. bovis* and *P. ciferrii*.

If so, it is unclear what these molecules are, or whether they are the same molecules in all phagocytosed *Prototheca* species. For example, it seems unlikely that *P. bovis*, *P. miyajii*,

and *P. xanthoriae*, all of which are phagocytosed readily but seem to be distantly related, are recognised by the same surface components, when closer relatives like *P. paracutis* and some *Auxenochlorella* species are phagocytosed at much lower levels. It is also unclear whether this molecule is available at all stages of the cell cycle, as it appeared as though larger cells of *P. miyajii* were phagocytosed more readily than small ones. Perhaps the cell walls of larger cells are modified to allow the lytic release of daughter cells, exposing different agonists for immune cell recognition.

A second possibility Is that some *Prototheca* species are actively invading macrophages rather than passively being phagocytosed. This may account for elevated phagocytic index of *P. miyajii* compared to its close relatives *P. paracutis* and *P. cutis*. However, invasion is unlikely to be sufficient to account for all of the differences observed in phagocytic indexes across *Prototheca* species. It has already been shown that *P. bovis* is able to invade mammalian cells while *P. ciferrii* is not (Shahid et al., 2020b), and *P. bovis* was not phagocytosed more than *P. ciferrii* using any metric in either cell type. Currently, only *P. bovis* has been suggested to actively invade mammalian cells.

A third possibility Is that some *Prototheca* species are inhibiting phagocytosis. This could account for the especially low phagocytic index of *P. wickerhamii* and *P. cutis* compared to *P. paracutis* and some *Auxenochlorella* species. The true environmental species, *P. tumulicola* and *P. moriformis*, unusually have very low phagocytic indexes in human macrophages as well as murine macrophages. I am sceptical that this is due to active mechanisms to prevent phagocytosis, as they are not able to survive physiological

temperatures (as demonstrated in Chapter 2). However, if they can be shown to actively inhibit phagocytosis, then this could be strong evidence of accidental pathology.

Of course, these possibilities are not mutually exclusive, and it is possible that all three are relevant in accounting for these trends of phagocytosis across the AHP lineage. Future work should elucidate the mechanisms of phagocytosis of *Prototheca*. It would be worth inhibiting actin polymerisation, as has been done before with cytochalasin D, to see if phagocytosis is dependent on actin in macrophages (Shahid et al., 2020b). It would also be worth providing macrophages with killed algae to see if phagocytosis increases, which could indicate that *Prototheca* have active mechanisms to prevent being phagocytosed, or decreases, which could indicate mechanisms for invasion. Heat killed *Candida albicans* and *Saccharomyces cerevisiae* are phagocytosed, which is blocked by cytochalasin D (Camilli et al., 2018). Ideally, this work should not extrapolate between species.

It Is also worth noting that strains of *P. wickerhamii* have been observed to cause different levels of cytotoxicity in BMDMs (Guo et al., 2023) and different strains of *P. ciferrii* have been reported to induce different cytokine responses in J774.1 cells (Yang et al., 2020). Only a limited repertoire of strains was explored here, with most species only represented by one strain, and so it may be that additional complexity remains within the phagocytic dynamics of each *Prototheca* species.

The possibility of facultative phagocytosis

Older cultures of HA6 were phagocytosed more than younger cultures, and there appeared to be a bias towards larger cells in the phagocytosis of *P. miyajii*. This raises the possibility of

changes in AHP species under certain conditions that facilitate phagocytosis. This could be because different cell wall components are present in the wall at different stages of the cell cycle, perhaps to facilitate release of daughter cells, or a culture in exponential phase compared to stationary phase. Differences in infectivity are also known for *Chlorella* symbionts, though the basis of that is also not known (Kodama and Fujishima, 2016).

Immune cell death

It is interesting that the four *Prototheca* species for which macrophage death was observed are the four pathogenic strains that are phagocytosed. No deaths were observed in videos of phagocytosis of *P. cookei*, which was readily phagocytosed but is not considered pathogenic, though *P. cookei* did appear able to escape the phagolysosome through division. The most straightforward explanation is that these pathogenic species have developed a virulence factor to kill macrophages, which *P. cookei* does not possess, either as a result of ancestral acquisition and subsequent loss in *P. cookei* or independent acquisition in *P. blaschkeae* and *P. bovis*/*P. ciferrii*. An alternative explanation would be that *P. cookei* has this killing capacity, but is unable to express it due to other factors with temperature stress. If a macrophage killing virulence factor exists, it would likely need to have arisen at least twice within the AHP lineage to be present in both *P. miyajii* and *P. bovis*.

It Is also possible that some amou”t of’this bias is an artefact. Macrophage death events were rare, and their detection was not an explicit part of experimental design. It is therefore not unlikely that the deaths reported here are an undercount. For example, *P. wickerhamii* has previously been reported to kill BMDMs, which was not observed here (Guo et al., 2023). It is also likely that the number of death events reflects the number of phagocytosis

events – fewer macrophages phagocytosed *P. wickerhamii* and *P. cutis*, making it less likely that these species would kill macrophages.

However, it seems unlikely for all of this bias to be artificial. Death events were not quantified, but they seemed more abundant for *P. bovis*, compared to the other three species. Previously, *P. bovis* has been observed to cause death in bovine mammary epithelial cells, while *P. ciferrii* was unable to do the same (Shahid et al., 2017b; Zhao et al., 2021). During real infections, *P. bovis* has also been observed to cause apoptosis in macrophages in cattle (Shahid et al., 2020b). The relevance of any of these behaviours to pathology is not known.

Ancestral state

With the variability in phagocytic index across the AHP lineage, it is difficult to identify whether the common ancestor of the AHP lineage would have possessed a negligible, low, or moderate phagocytic index. To acquire additional insight, murine macrophages were exposed to algal cells from different taxonomic groupings (at the level of genus, family, and class).

As the phagocytic index of these species varied between 0 and 0.25, it is difficult to exactly determine a consistent phagocytic index for green algae to use as an ancestral state. However, as these five non-AHP algae have low phagocytic indexes, it is likely that the enhanced capacity to be phagocytosed seen in the cattle-associated species, *P. xanthoriae*, and *P. miyajii* is the more derived state. Given that this occurs multiple times within the

lineage, it has presumably arisen multiple times, and may occur through different mechanisms.

It Is Interesting that the species most closely related to the AHP lineage (*Keratococcus bicaudatus*, which belongs to a different genus within the same family as *Prototheca*) has high variability in its phagocytic index across the two replicates. In particular, the replicate with the older culture has a markedly increased phagocytic index, which is similar to the trend observed in HA6. This may indicate facultative phagocytosis is a shared feature of the Chlorellaceae family.

During quantification, it appeared that macrophages were not attracted towards non-AHP algae. This contrasted with some *Prototheca* species, which appeared to attract macrophages even when they were not phagocytosed. This could indicate that the unknown substance that attracts macrophages towards many *Prototheca* species is not expressed in other algae, and may be unique to the AHP lineage.

6) Genome Sequencing and Assembly of *Prototheca* (and Related Species)

Introduction

The extraction and sequencing of genomic DNA from HP1, HP2, and HP3 was performed by Leanne Stones.

At the inception of this project there were no high quality *Prototheca* genomes available. As of 2023, that is not true, as three high quality genomes for *P. wickerhamii* exist (Guo et al., 2022; Bakuła et al., 2021). At the time of writing, only eleven nuclear genomes have been sequenced for *Prototheca*, eight of which are of clinical isolates (Guo et al., 2022; Maboni et al., 2021; Bakuła et al., 2021; Zeng et al., 2019; Suzuki et al., 2018; Severgnini et al., 2018). However, this is unlikely to be sufficient to perform genome-wide association studies to identify genes that contribute to pathogenicity or virulence.

Thus, in this chapter I initially aimed to assemble genomes from *Prototheca* species primarily to provide material to perform GWAS to find virulence-associated genes. Instead, I highlight pitfalls for genome assembly of the AHP lineage and provide material for the phylogenetics investigation in Chapter 7) Phylogenetics of *Prototheca* and Related Organisms.

Methods

DNA Extraction for Short Read Sequencing

Up to 6 ml of culture, grown as described in the core methods chapter, were collected by centrifuging at 3000 g for 1 minute at room temperature. All subsequent centrifugation was performed at 16000 g for five minutes at 4°C. Cells were resuspended in 400 µl of 0.02 M NaOH and transferred to 1.5 ml centrifuge tubes. Then, cells were flash frozen in liquid nitrogen and thawed for 15 minutes at 30°C in a heat block. Once thawed, cells were spun down and the supernatant was discarded. Typical pellets were approximately 100 µl.

An equal volume of 0.5 mm glass beads was added. The pellet and glass beads were resuspended in 400 µl of CTAB lysis buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris base pH 8.0, 20 mM EDTA), to which 400 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added. Tubes were vortexed four times for 40 seconds (160 seconds total) at high speed to lyse cells, and were kept on ice between vortexing. After lysis, the tubes were centrifuged and the aqueous phase was transferred to a new tube, to which an equal volume of phenol:chloroform:isoamyl alcohol was added. These tubes were vortexed for 40 seconds, centrifuged, and the aqueous phase was transferred to a new tube. This was repeated until no white interphase remained, at which point the aqueous phase was cleaned with an equal volume of chloroform:isoamyl alcohol (24:1), vortexing and centrifuging as described for phenol. DNA was precipitated from the aqueous phase by adding an equal volume of 100% room-temperature isopropanol, mixing by gentle pipetting and immediately centrifuging for 3 minutes. DNA pellets were washed in 70% ethanol, left to air dry, and resuspend in EB buffer (10 mM Tris-Cl, pH 8.5).

The length of the extracted DNA was checked by agarose gel electrophoresis, In a 0.75% TAE gel run for 60 minutes at 70 V. 260/230 and 260/280 ratios were checked with a NanoDrop 2000c spectrophotometer (ThermoFisher), and concentration was quantified using a Qubit3 fluorometer and the Qubit dsDNA high sensitivity kit (ThermoFisher). Paired-end Illumina sequencing was provided by MicrobesNG (https://www.microbesng.com; standard service).

Assembly of Short Reads

The computational work described in this chapter was performed using the University of Birmingham's BlueBEAR High Performance Computing (HPC) service, which provides a HPC service to the University's research community. See http://www.birmingham.ac.uk/bear for more details.

Sequencing reads were quality checked with FastQC (v0.11.9) (Andrews, 2010) and checked for contamination by FastQ Screen (v0.14.0) (Wingett and Andrews, 2018) and decontaMiner (v1.4) (Sangiovanni et al., 2019). The outputs of FastQC and FastQ Screen were summarised by MultiQC (v1.5) (Ewels et al., 2016). An attempt was made to check the ploidy of one strain using Jellyfish (v2.2.10) (Marçais and Kingsford, 2011) and SmudgePlot (v0.2.5) (Ranallo-Benavidez et al., 2020).

Genomes were assembled using SPAdes (v3.14.1) with default settings (Bankevich et al., 2012). The quality of assemblies was checked using QUAST (v5.0.2) (Gurevich et al., 2013) and BUSCO (v5.3.1) (Manni et al., 2021). QUAST and BUSCO quality statistics were also generated for a number of published genomes sourced from GenBank and the China

National GeneBank DataBase. BUSCO completeness checks were made using the chlorophyta_odb10 list of genes.

To identify if any chlorophyta_odb10 genes were consistently missing from the genomes of a species, comparisons were made between lists of BUSCO genes present in each genome from a given species. The minimal intersection was defined as the BUSCO genes that were identified as complete or duplicated in all genomes of a given species; the maximal union was defined as the BUSCO genes that were identified as complete, duplicated, or fragmented in any genome of a given species. These lists were compared to identify if some lineages were consistently missing genes.

Gene ontology (GO) enrichment analysis was performed on some of these lists of genes. BUSCO gene IDs were converted to UniProt accession codes using OrthoDB (v10.1) (Kriventseva et al., 2019). BUSCO gene IDs were converted into accession codes from *C. variabilis* were chosen rather than accession codes from an AHP species as no accessions for *Prototheca* were available and it was assumed that *Chlorella* genomes would be better annotated than *Auxenochlorella* genomes. GO enrichment analysis was performed on these *Chlorella* accession codes using the GO enrichment tool of PlantRegMap, using annotations from *C. variabilis* NC64A and default parameters (Tian et al., 2020).

Organelle genomes were identified and extracted from whole genome assemblies, as described in the core methods section.

High Molecular Weight DNA Extraction and Long Read Sequencing

Three separate methods were attempted to extract high molecular weight (HMW) DNA from *Prototheca*.

For all methods, cultures were grown in SAB and extracted DNA was initially visualised with gel electrophoresis to assess if fragments were larger than 20 kbp. For extractions with fragments larger than 20 kbp, pulsed-field gel electrophoresis (PFGE) using standard protocol 4 of a Pippin pulse (Sage Science) was used to visualise longer DNA. Ladders used in all gels were Lambda DNA Hind III Digest and Lambda DNA Mixed Digest. DNA was quality checked with a NanoDrop and quantified with a Qubit3, as described above.

Liquid Nitrogen Grinding

To extract DNA via grinding in liquid nitrogen, methods to extract nuclei from plant cells (Li et al., 2020) and to extract DNA from purified nuclei were combined. To extract nuclei, cultures were grown to volumes between 100 ml and 200 ml and then spun at 3000 g to collect the algal cells, which typically produced 1-3 g of material. Cells were washed in 0.02 M NaOH and ground using a pestle and mortar in liquid nitrogen. Ground cells were suspended in ice-cold nuclear isolation buffer (NIB; 10 mM Tris base, 80 mM KCl, 10 mM EDTA, 0.5 M sucrose, 1 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochloride, 0.5% (v/v) Triton X-100, 0.5% (v/v) 2-Mercaptoethanol) and incubated for at least fifteen minutes while agitating at 15 rpm. Nuclei were washed twice by centrifuging at 2400 g for 12 minutes at 4°C, discarding supernatant, and then resuspending in 10 ml NIB using a paint brush. DNA was extracted from nuclei with the Nanobind Plant Nuclei Big DNA Kit from Circulomics according to the product manual. Nuclei were lysed by

the combined action of proteinase K, a lysis buffer, and incubation at 55°C for 30 minutes, inverting to mix every five minutes. A silica disc was added to bind and protect the DNA before the addition of 1 volume of 100% isopropanol to precipitate the DNA. The DNA bound to the silica disc was washed and then eluted with elution buffer (from the kit).

CTAB Incubation

To extract DNA via chemical lysis of cells, 6 ml cultures of algal cells were spun at 3000 g and washed twice in deionised water. Cells were then incubated in 500 µl pre-warmed CTAB isolation buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris pH 8.0, 20 mM EDTA) at 55°C for 1 hour. DNA was then extracted with phenol:chloroform:isoamyl alcohol as described above, but tubes were rolled to mix the aqueous and organic phases instead of vortexing.

Enzymatic Digestion

A method to form protoplasts of *"Chlorella protothecoides*", a classification considered synonymous with *Auxenochlorella protothecoides*, was used and modified (Lu et al., 2012). 2 ml cultures of algae were spun at 3000 g, washed once in deionised water, and resuspended in 1 ml of protoplasting buffer (0.6 M D-mannitol, 25 mM Tris base pH 6). Algal cells were digested with combinations of 1% (w/v) snailase (Abbexa), 2% (w/v) cellulase R-10 (Onozuka) and 2% (v/v) glusulase (PerkinElmer). Digests were performed at 30°C or 37°C for 24 or 48 hours, agitated at 150 rpm.

To test for the presence of protoplasts, digested cells were diluted 1:100 in protoplasting buffer as a control and pure water or 2.5% sodium dodecyl sulphate (SDS) to lyse protoplasts. Cells were left for at least five minutes, to allow protoplasts diluted in water or

SDS to lyse, and then the number of cells was counted using a haemocytometer. The reduction of cells diluted in water or SDS relative to the cells diluted in protoplasting buffer was taken to be the number of protoplasts after treatment.

Results

Assemblies based on paired-end reads are poor quality

Genomic DNA, suitable for Illumina sequencing, was extracted from *Prototheca* cultures using the first method described in the methods section. Interestingly, during DNA extractions, cultures of *P. bovis* and related species typically had a larger interphase zone during extraction and a lower yield of DNA. Extractions from this lineage often completely failed. This likely reflects biological differences between the lineages, which will be addressed in the discussion. Sequencing statistics, i.e. coverage, number of reads, and length of insert, are available in Appendix 26. All sequencing reads were 250 bp long prior to trimming.

The ploidy of a genome, and possible heterozygosity as a result of ploidy, have Implications for genome assembly. A ploidy unaware assembly programme may struggle to resolve sequencing reads that appear to match to the same region, but mismatch by a single base. This may give rise to duplications in the assembly. Microalgae are usually haploid (Lemieux et al., 2019), but ploidy has not been explicitly checked in *Prototheca*.

Sequencing reads from HP5, a strain of *P. wickerhamii*, were used to test the ploidy of the strain using the tool smudgeplot. Additionally, sequencing reads from an *Escherichia coli*

strain (SRR17866859) were included, as it was not clear from the documentation whether smudgeplot could resolve haploid genomes.

The results from smudgeplot (Figure 6-1) suggest that both the *Prototheca* strain and *E. coli* strain are diploid. As *E. coli* are known to be haploid, it appears that smudgeplot cannot detect haploidy. Other methods to determine ploidy were not within the scope of this project (Krishan, 1975; Viruel et al., 2019; Vazač et al., 2018). For now, we have assumed that *Prototheca* are haploid.

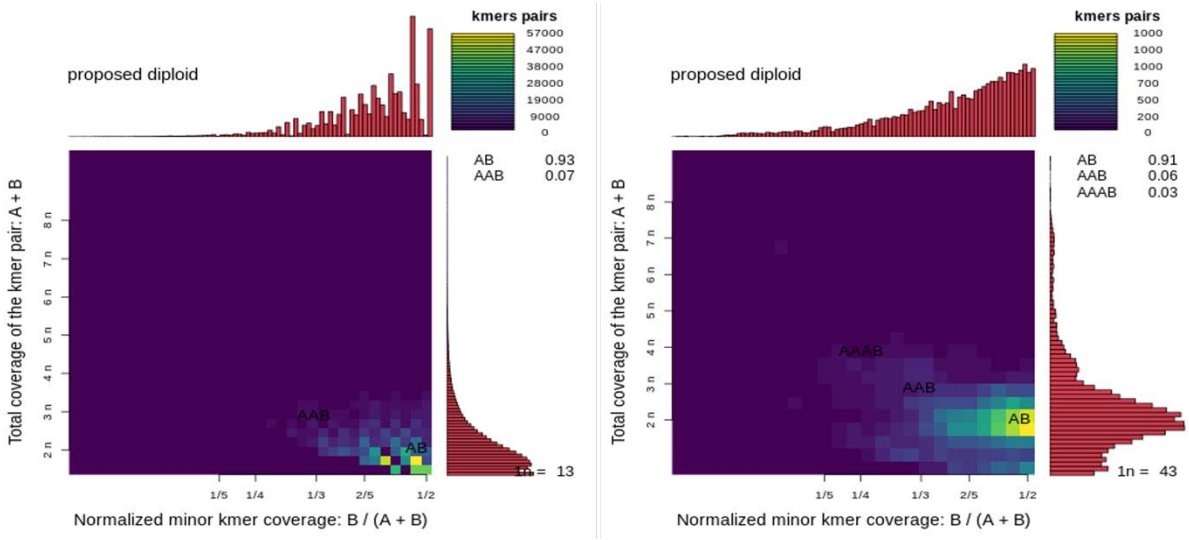


Figure 6-1. Investigating ploidy in Prototheca *with smudgeplot. Left, smudgeplot result for HP5, a strain of* P. wickerhamii *with an unknown ploidy. Right, smudgeplot result for* E. coli*, a known haploid. Despite* E. coli *being a known haploid, both genomes are proposed to be diploid, suggesting the tool cannot detect haploidy.*

14 AHP strains from the HAPI lab collection were sequenced and assembled into draft genomes. QUAST statistics and BUSCO completeness scores are provided in Table 6-1 and Figure 6-2 respectively.

Table 6-1. QUAST statistics for internal draft genomes.

Species	Strain (internal ID)	Assembly length (bp)	Number of Scaffolds	N50 (bp)	L50	Largest Scaffold (bp)
P. bovis	HP3	23,870,425	21,146	1,276	5,473	36,786
P. bovis	HP14	37,814,366	17,297	3,258	3,304	28,714
P. bovis	HP18	39,693,261	18,516	3,144	3,682	40,358
P. bovis	HP25	21,960,082	8,581	3,952	1,611	42,085
P. bovis	HP26	28,971,989	15,739	2,490	3,407	39,120
P. ciferrii	HP2	31,381,674	22,365	1,860	4,655	41,335
P. wickerhamii	HP1	26,847,773	6,598	8,755	863	52,264
P. wickerhamii	HP5	28,633,442	5,762	11,301	662	91,915
P. wickerhamii	HP6	17,566,260	3,638	15,076	294	187,195
P. xanthoriae	HP53	12,715,853	6,946	2,495	1,518	55,763
P. xanthoriae	HP54	22,321,916	6,547	5,073	1,293	55,763
A. protothecoides	HA1	22,584,531	5,209	9,228	594	118,361
A. symbiontica	HA5	25,807,475	9,628	5,951	1,180	105,887
Auxenochlorella sp. 2	HA7	21,043,448	1,782	20,445	320	84,707

To provide context, all available published genomes for *Prototheca* and some genomes representing *Auxenochlorella*, *Chlorella*, and *Micractinium* were scored in the same way. Assemblies reported by Zeng *et al.* (2019) could not be found, and thus were not included (Zeng et al., 2019). The QUAST statistics and BUSCO completeness scores of published genomes are provided in Table 6-2 and Figure 6-3 respectively.

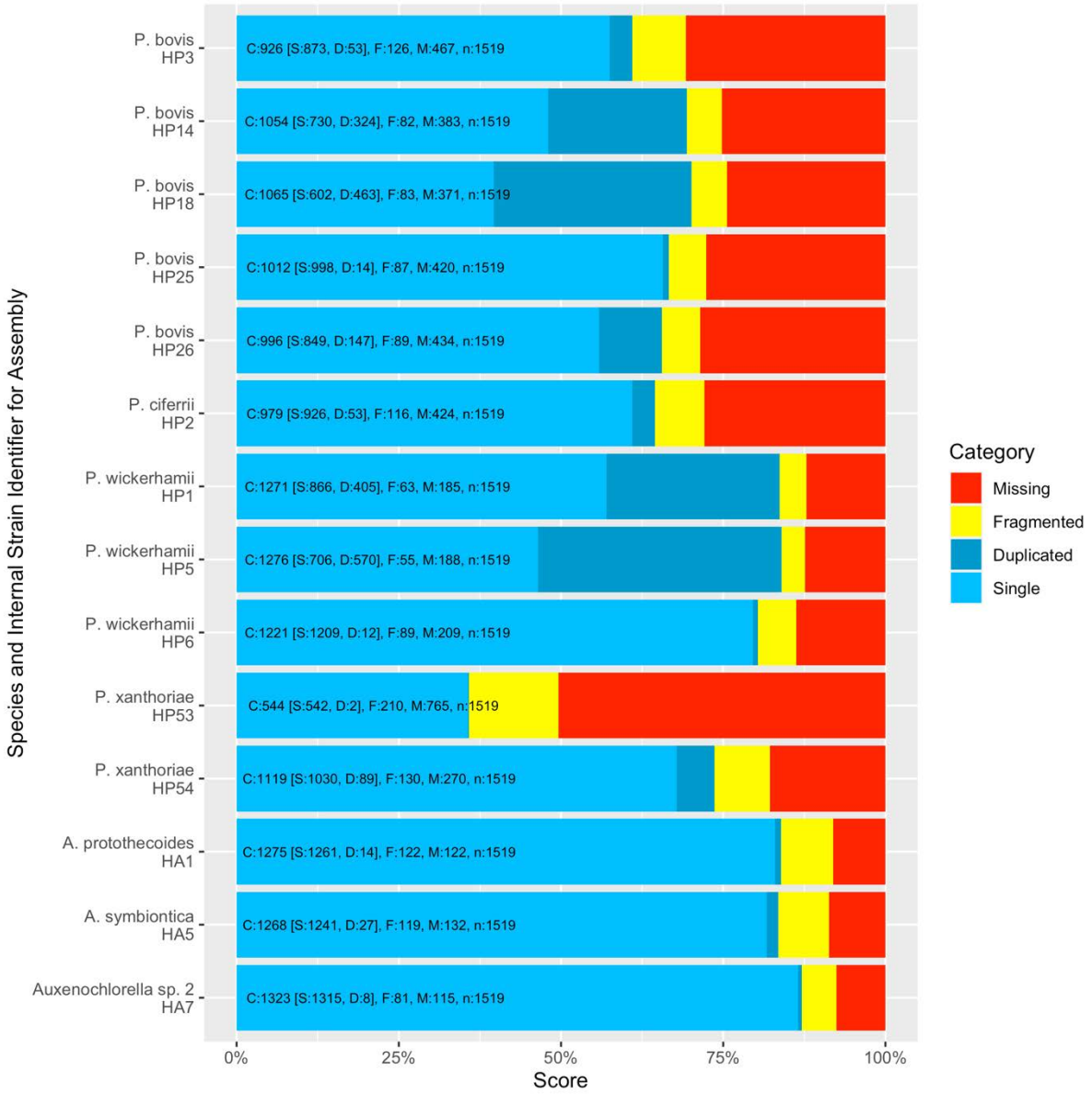


Figure 6-2 BUSCO completeness of internal draft genomes. Sequenced isolates are identified by species name and strain identifier. Bars show the proportion of genes from the chlorophyta_odb10 database found to be present with a single copy (light blue), present with multiple copies (dark blue), present but fragmented (yellow), and missing (red). Bars are labelled with the number of genes in each category, as well as the number of genes present as either single or multiple copies I, and the number of genes in the database (n).

The draft assemblies do not produce scaffolds that approach whole chromosomes (Table 6-1), as *Prototheca* chromosomes appear to have lengths around 3 Mbp (Suzuki, 2006). Additionally, draft genomes are significantly more fragmented than many published genomes, with more scaffolds, a shorter N50, and a shorter longest scaffold. Under these metrics, draft assemblies are only comparable to the genomes for *Helicosporidium* ATCC 90520 and *P. cutis* 20-25310 (Table 6-2).

Table 6-2. QUAST statistics for published genomes. Accession numbers refer to sequences held in GenBank, hosted by NCBI, except where marked by asterisks (*). Accession numbers marked by asterisks refer to sequences held in the China National GeneBank DataBase (CNGBdb).

Species/Strain	Assembly length (bp)	Number of Scaffolds	N50 (bp)	L50	Largest Scaffold (bp)	Accession	Reference
Helicosporidium ATCC 50920	12,373,820	5,666	3,036	1,246	24,678	GCA_ 000690575.1	(Pombert et al., 2014)
P. bovis SAG 2021	24,744,314	4553	7,940	913	57,068	GCA_ 003612995.1	(Severgnini et al., 2018)
P. ciferrii SAG 2063	26,448,487	6955	7,030	911	97,625	GCA_ 003613005.1	(Severgnini et al., 2018)
P. stagnora JCM 9641	16,896,228	27	1,107,247	6	1,567,793	GCA_ 002794665.1	(Suzuki et al., 2018)
P. cutis 20-25310	19,235,735	4,809	5,846	1,016	65,909	GCA_ 016906445.1	(Maboni et al., 2021)
P. cutis JCM 15793	20,029,045	46	1,409,608	6	2,465,851	GCA_ 002897115.2	(Suzuki et al., 2018)
P. wickerhamii ATCC 16529	16,700,629	19	1,578,614	5	2,447,261	GCA_ 016906385.1	(Bakuła et al., 2021)
P. wickerhamii S1	17,573,978	19	1,639,047	5	2,552,045	I0036161 *	(Guo et al., 2022)
P. wickerhamii S931	17,453,189	26	1,406,360	6	1,845,983	I0036160 *	(Guo et al., 2022)
C. protothecoides Sp0170	22,902,383	306	285,543	24	1,144,356	GCA_ 000733215.1	(Gao et al., 2014)
C. pyrenoidosa FACHB-9	56,992,954	1346	1,392,758	14	3,905,163	GCA_ 001430745.1	(Fan et al., 2015)
C. sorokiniana UTEX 1228	61,391,140	64	2,415,094	10	4,567,720	GCA_ 002939045.1	(Hovde et al., 2018)
C. sorokiniana UTEX 1230	58,534,920	20	4,091,730	7	5,120,617	GCA_ 003130725.1	(Hovde et al., 2018)
C. sorokiniana UTEX 1412	57,881,618	65	2,025,419	10	5,449,605	GCA_ 003116155.1	(Hovde et al., 2018)
C. sorokiniana UTEX 1602	59,566,223	159	2,592,956	9	4,192,086	GCA_ 002245835.2	(Arriola et al., 2018)
C. variabilis NC64A	46,159,515	414	1,469,606	12	3,119,887	GCA_ 000147415.1	(Blanc et al., 2010)
C. vulgaris 211/11P	40,437,440	44	2,825,136	6	5,422,624	GCA_ 023343905.1	(Cecchin et al., 2019)
M. conductrix SAG 241.80	61,018,900	300	1,210,495	13	5,958,903	GCA_ 002245815.2	(Arriola et al., 2018)

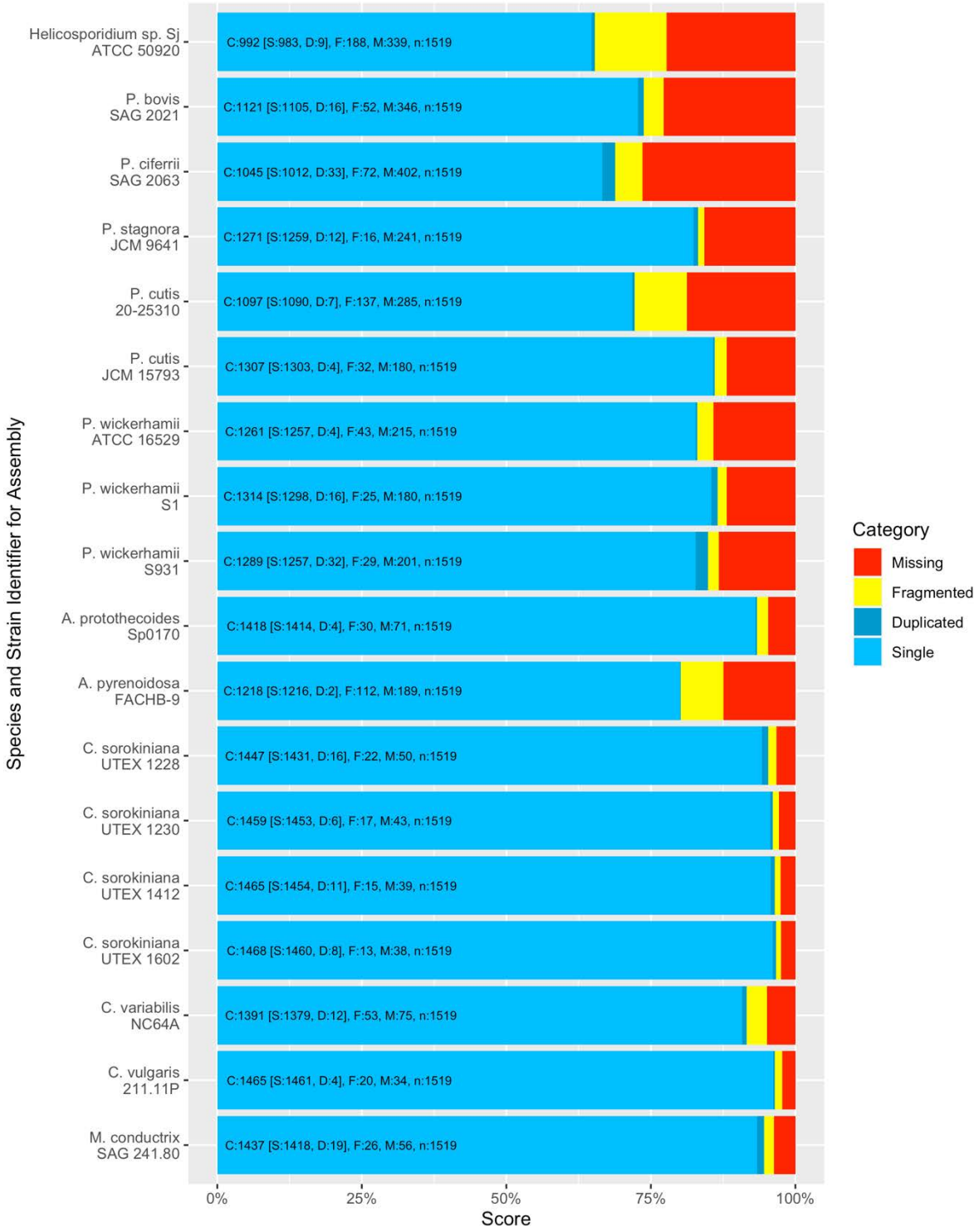


Figure 6-3 BUSCO completeness of published genomes. Sequenced isolates are identified by species name and strain identifier. Bars show the proportion of genes from the chlorophyta_odb10 database found to be present with a single copy (light blue), present with multiple copies (dark blue), present but fragmented (yellow), and missing (red). Bars are labelled with the number of genes in each category, as well as the number of genes present as either single or multiple copies I, and the number of genes in the database (n).

Draft assemblies also appear to have substantial duplication not seen in published genomes. Two of three draft *P. wickerhamii* genomes (HP1 and HP5, Table 6-1) are 50% larger than published ones (Table 6-2) and four of the draft assemblies have at least 25% of the BUSCO genes duplicated (Figure 6-2), while published genomes tend to be less than 2% duplicated (Figure 6-3). These statistics are likely connected; HP6 is the least fragmented draft genome from *P. wickerhamii* and HP25 is the least fragmented draft genome from *P. bovis*, and these genomes are the smallest draft genomes of their respective species and show the lowest levels of duplication. Fragmented genomes are known to result in inflated gene counts (Denton et al., 2014).

It Is potentially Interesting that assemblies of *P. bovis* and *P. ciferrii* genomes are typically more fragmented than other *Prototheca* genomes, in both the draft genomes presented here and published genomes (Table 6-1, Table 6-2). This may reflect genuine differences in genomes between cattle-associated species and other *Prototheca* species, e.g. presence of repetitive sequences.

AHP species do not contain all chlorophyte BUSCO genes

Despite greater levels of fragmentation, it appears that draft genomes have similar BUSCO completeness scores to published genomes. This presented an opportunity to assess the suitability of the chlorophyte BUSCO list of genes to assess *Prototheca* assemblies.
While BUSCO lists are manually curated to contain genes that should be present in all taxa within a lineage, sometimes some taxa do not contain the full list of genes. Notably, *Chlorella* have been identified as missing some of the genes in the core eukaryotic gene list (Hanschen and Starkenburg, 2020).

For species with more than one available genome, the lists of BUSCO genes present in each genome were combined as described in the methods section for this chapter. Initially, all that was considered was the number of genes present in each species.

Table 6-3. Number of BUSCO genes present across genomes of species with more than one available genome. The minimal intersection of BUSCO genes represents the number of complete genes present in every genome of a given species. The maximal union of BUSCO genes represents the number of complete or fragmented genes present in any genome of a given species. Species are arranged alphabetically. Photosynthetic species are marked with an asterisk (*).

Species	Minimal Intersection	Maximal Union
P. bovis	774	1230
P. ciferrii	856	1214
P. cutis	1084	1352
P. wickerhamii	985	1377
P. xanthoriae	528	1261
*A. protothecoides**	1266	1456
*C. sorokiniana**	1391	1500

No algal species investigated here contains all 1519 genes in the Chlorophyta list of BUSCO genes, even when combining all BUSCO genes present in any available genomes, though genomes of photosynthetic species contained more of these genes (Table 6-3). Across all 6 genomes of *P. wickerhamii*, the maximal union contained 1377 genes, meaning 142 genes were completely absent (i.e. were not identified, even as fragments, in any genome). Similarly, across all 6 genomes of *P. bovis*, 289 genes were completely absent. There was substantial overlap between these lists: 137 of the 289 genes in at least one of these species was lost in both *P. bovis* and *P. wickerhamii* genomes. *C. variabilis* UniProt accessions could be found for for 285 of these genes, and 248 of these had GO annotations.

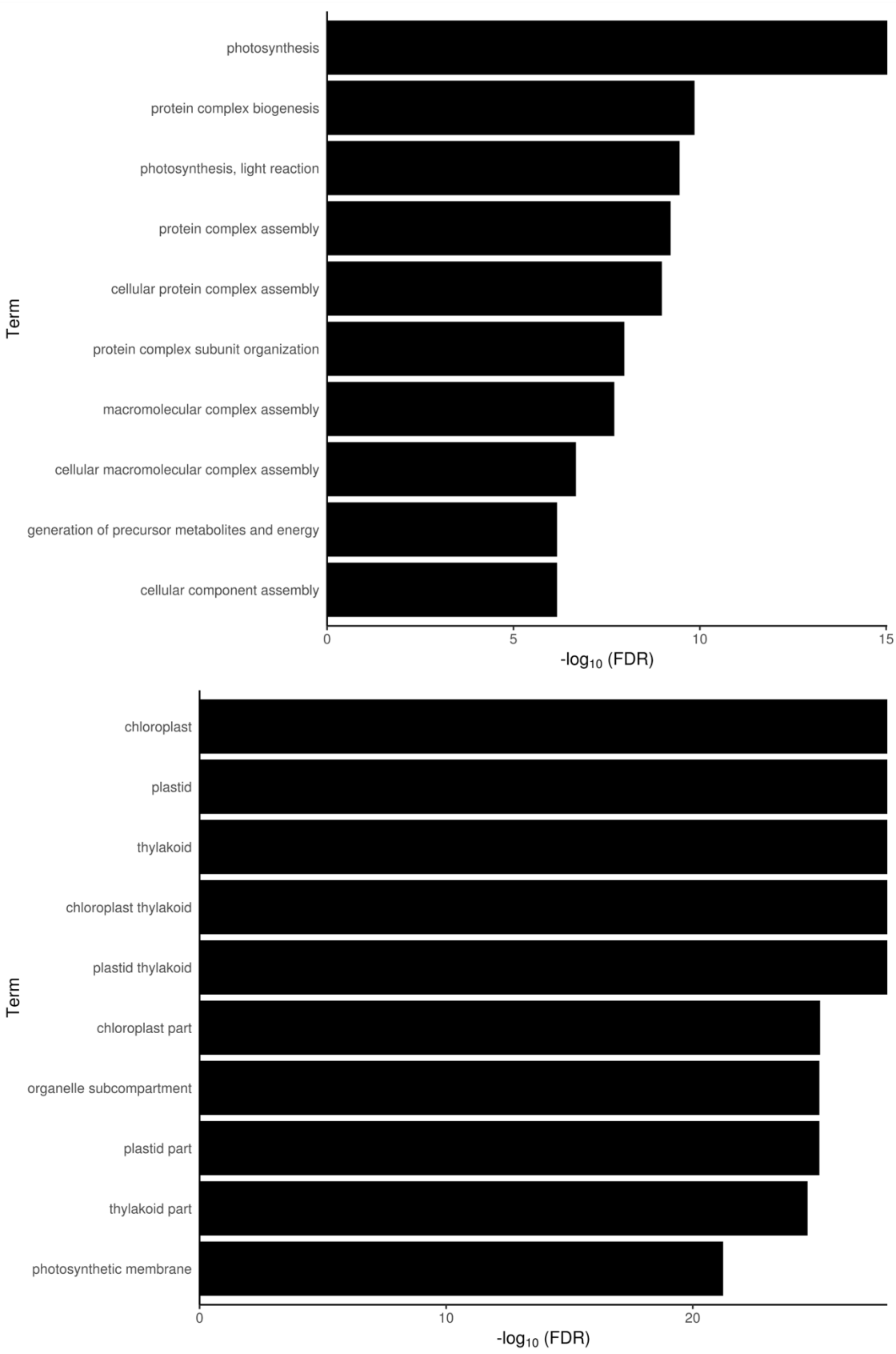


Figure 6-4. GO enrichment analysis for BUSCO genes lost in both P. bovis *and* P. wickerhamii *assemblies, showing biological process (top) and cellular component (bottom) GO terms.*

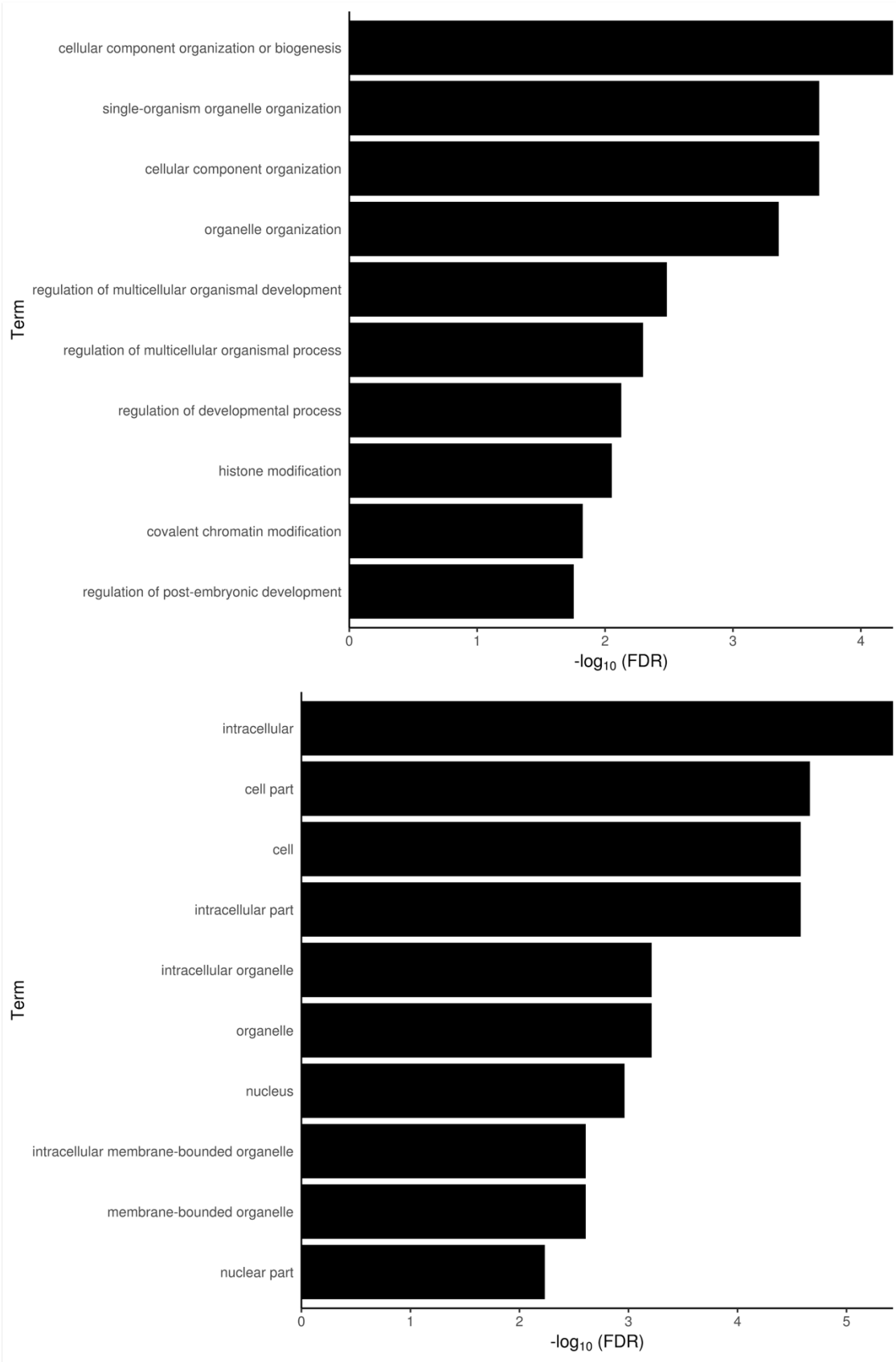


Figure 6-5. GO enrichment analysis for BUSCO genes lost in P. bovis *but not* P. wickerhamii *assemblies, showing biological process (top) and cellular component (bottom) GO terms.*

Unsurprisingly, the genes lost in both *P. bovis* and *P. wickerhamii* genomes were significantly enriched for GO terms relating to photosynthetic biological processes and chloroplast cellular components (Figure 6-4). It seems likely that these are missing as a result of not being present in the genomes of these species, rather than due to assembly errors.

However, the genes that are lost only in *P. bovis* appear to have more varied functions, with significantly enriched GO terms including nuclear cell components and processes relating to development and the organisation or modification of DNA (Figure 6-5). It is not clear what the implications of losing these genes would be, nor is it clear if these losses are as a result or indicative of adaptation to pathology. Finally, it is unclear whether the genes that are only lost in *P. bovis* are truly lost, or if some of them are a consequence of the greater fragmentation of these assemblies.

Notably, no individual *P. wickerhamii* (Figure 6-6) or *P. bovis* (Figure 6-7) genome contained all the genes present in the maximal union, though *P. wickerhamii* assemblies using long read sequencing technologies come close. It is not clear whether this represents individual variation between strains, or assembly errors for these genomes.

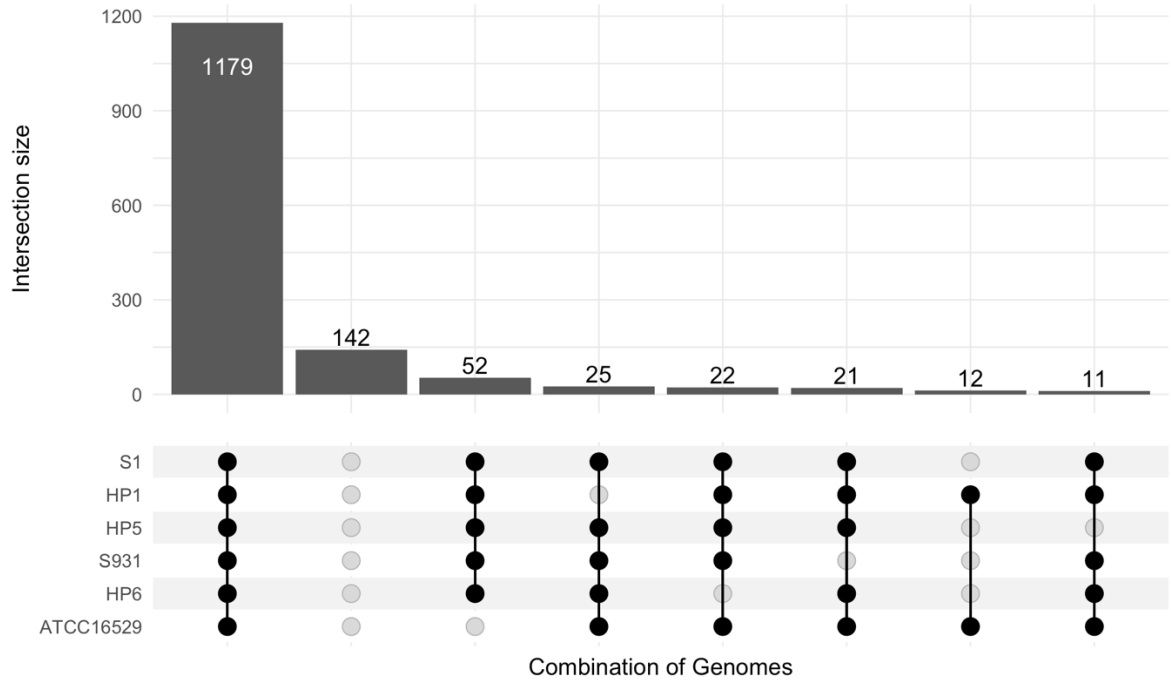


Figure 6-6. Number of BUSCO genes present across genomes of P. wickerhamii *strains. Combinations of genomes are indicated by shaded circles, with a black circle indicating a set of genes is present in a given genome and a grey circle indicating a set of genes is not present in a genome. Genes are counted as being present in a genome if they are complete or fragmented. Combinations of genomes with fewer than 10 genes are excluded.*

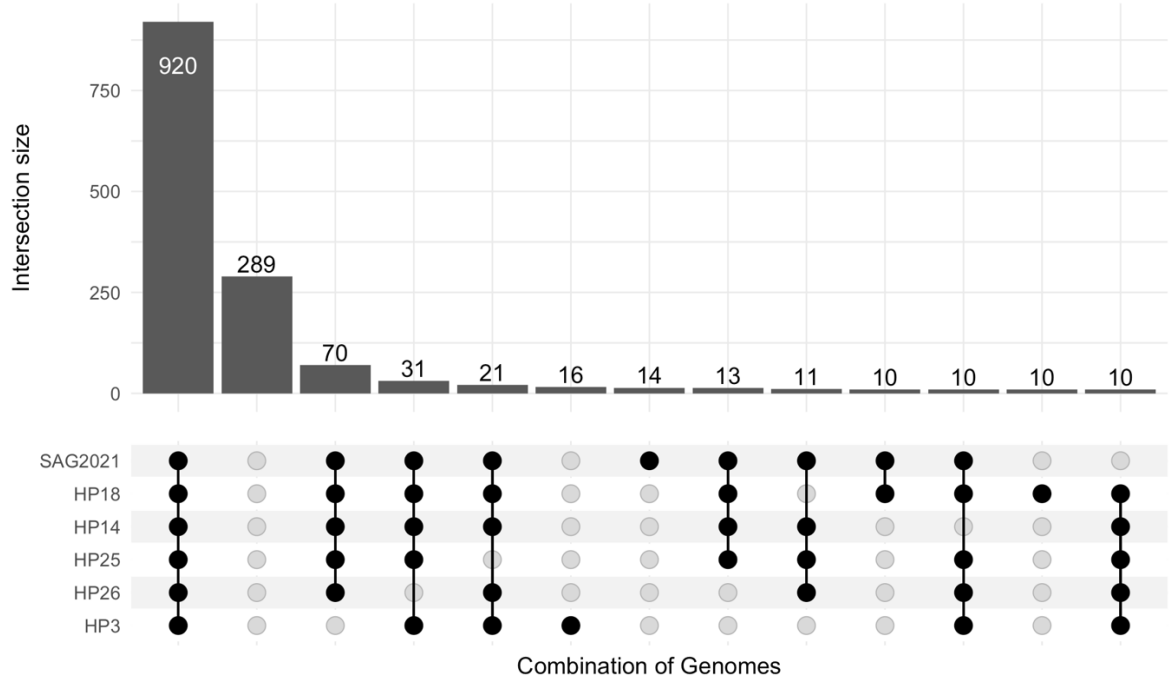


Figure 6-7. Number of BUSCO genes present across genomes of P. bovis *strains. Combinations of genomes are indicated by shaded circles, with a black circle indicating a set of genes is present in a given genome and a grey circle indicating a set of genes is not present in a genome. Genes are counted as being present in a genome if they are complete or fragmented. Combinations of genomes with fewer than 10 genes are excluded.*

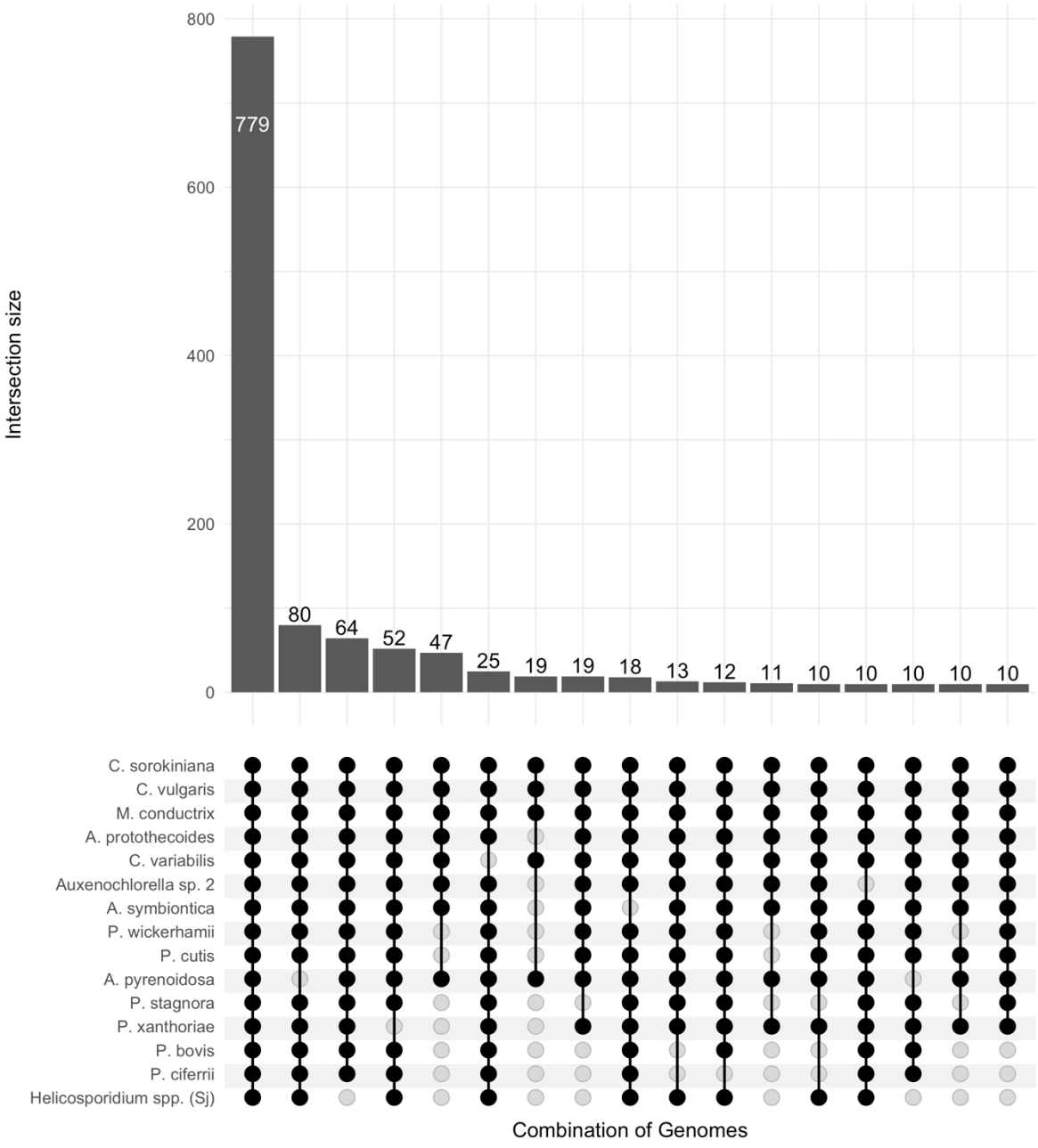


Figure 6-8. Number of BUSCO genes present across genomes of algal species. Combinations of genomes are indicated by shaded circles, with a black circle indicating a set of genes is present in a given genome and a grey circle indicating a set of genes is not present in a genome. Genes are counted as being present in a genome if they are complete or fragmented. Combinations of genomes with fewer than 10 genes are excluded.

The number of BUSCO genes absent from sub-lineages or individual species was also explored (Figure 6-8). With the currently available genomes, only 779 BUSCO genes are present in every species. However, a large number of genes are present in all but one species, and these species usually only have one genome sequence or (as in the case of *P. xanthoriae*) all genomes are poor quality. These likely represent assembly or sequencing

errors, rather than true losses of the genes in question. More interestingly, there are 47 genes present in all photosynthetic genomes present but none of the non-photosynthetic genomes. However, care should be taken with interpreting these. More genomes from *Prototheca* and *Helicosporidium* species with only one or two genomes available should be sequenced and included before biologically meaningful conclusions can be drawn, to avoid mistakes from assembly errors.

Complete organelle genomes are present in draft assemblies

It was also noticed that the largest contigs in many of the draft assemblies are approximately the same size or larger than available organelle genomes for these species. BLAST was used to identify which contigs might contain organelle-specific sequences, and contigs with approximately the correct length were extracted from assemblies. The identity of these contigs as organelle genomes was confirmed with GeSeq.

Both mitochondrial and plastid genomes could be identified as single contigs in every *Auxenochlorella* assembly, as well as in all assemblies of *P. wickerhamii* and *P. xanthoriae*. Additionally, complete plastid and mitochondrial genomes could be identified in four and two assemblies respectively of *P. bovis*. These genomes are used in the phylogenetic analyses in Chapter 7) Phylogenetics of *Prototheca* and Related Organisms.

Long Read Sequencing was Unsuccessful

The apparent duplication or absence of genes from draft assemblies was considered a significant impediment to GWAS and to reliably identifying loss or gain of genes, or expansion or contraction of gene families. As fragmented genomes are known to result in

inflated gene counts (Denton et al., 2014), less fragmented assemblies were desired. It was thought that the inclusion of long read sequence data would help reduce the fragmentation of draft genomes.

Long read sequencing is dependent on the extraction of HMW DNA. DNA can be considered HMW at around 20 kbp (Stark et al., 2020), but the term often refers to molecules that are 50 kbp or longer (NewEnglandBiolabs, n.d.; DNATechnologiesCore, n.d.). The intention was to use Oxford Nanopore sequencing for long reads, which has no upper limit for sequencing lengths. Therefore, we aimed to extract DNA of at least 50 kbp, but longer DNA would have been beneficial.

A method to extract HMW DNA from *Prototheca* exists, but requires grinding cells at 20 Hz (Jagielski et al., 2017b), something we were not able to do. Therefore, three alternative methods were explored for the extraction of HMW DNA from *Prototheca*: grinding in liquid nitrogen, incubation in a CTAB lysis buffer, and enzymatic digestion of algal cell walls to form protoplasts.

Grinding in liquid nitrogen was our first choice of technique as it is well established as a method of extracting HMW DNA from plants (Li et al., 2020) and algae (Stark et al., 2020; Jagielski et al., 2017b). Cells from several strains of *P. wickerhamii* and *P. bovis* were ground in liquid nitrogen to purify their nuclei and extract their DNA as described in the methods section of this chapter.

HMW DNA, in excess of 50 kbp, was successfully extracted from strains of *P. wickerhamii*, though the extraction was unreliable. However, HMW DNA could not be isolated from *P. bovis* (Figure 6-9). Due to the poorer quality of *P. bovis* and *P. ciferrii* genomes assembled from short reads relative to other species of *Prototheca*, as well as the poorer quality of published genomes of these species, extraction of HMW DNA from these species was considered the main priority.

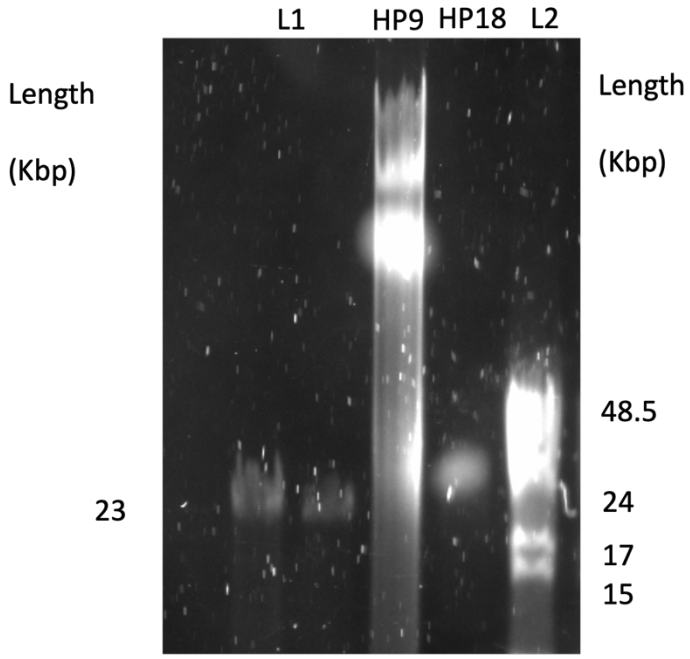


Figure 6-9. Pulsed-field gel electrophoresis of genomic DNA extracted from strains of Prototheca*, by grinding in liquid nitrogen. Lanes containing DNA are labelled with the internal strain identifier of the strain in question. Samples represent the highest molecular weight DNA extracted from their respective species:* P. wickerhamii *(HP9) and* P. bovis *(HP18). Two ladders are present: L1 is Lambda Hind III Digest; L2 is Lambda DNA Mixed Digest.*

Following the failure to extract HMW DNA of *P. bovis* by grinding in liquid nitrogen, cells from strains of *P. bovis* were incubated in CTAB lysis buffer before DNA was extracted with phenol, as described in the methods section for this chapter. No DNA could be recovered using this method.

Suspecting that the cell wall was the primary reason for cells failing to lyse, an attempt was made to digest the cell wall of *Prototheca* cells with combinations of enzyme cocktails

(snailase, cellulase R-10, and glusulase). The methods were taken from a paper in which protoplasts of *Chlorella prototheco*ides were produced (Lu et al., 2012). It was assumed that this strain of "*Chlorella prototheco*ides" was actually *A. prototheco*ides, as these species are equivalent (Guiry and Guiry, 2015; Shihira and Krauss, 1965; Kalina and Puncocharova, 1987). Thus, digestion was attempted on cells of *P. wickerhamii*, *P. bovis*, and *A. prototheco*ides.

The number of protoplasts was Inferred by the reduction In counted cells following dilution in pure water or 2.5% SDS, as described in the methods section of this chapter. There was no difference in cell counts between cells diluted in water or SDS compared to cells diluted in protoplasting buffer. Therefore, no protoplasts were formed following any digestion. *P. bovis* cells following digestion with all three enzymes for 24 hours at 37°C became "sticky" – forming a pellet that could not be disrupted with mild agitation – as did *A. prototheco*ides cells following digestion with snailase and cellulase for 24 hours and *P. bovis* cells following digestion with snailase and glusulase after 48 hours. Otherwise, digestion had no noticeable effect.

Extraction of HMW DNA was therefore unsuccessful, and long-read sequencing was not performed.

Discussion

In this chapter I have attempted to sequence the genomes of AHP species for use in phylogenetic analyses and identification of virulence traits through comparative genomics.

Assemblies from short reads

The initial intention for this chapter was to perform GWAS to compare pathogenic and non-pathogenic strains or species, to identify relevant virulence traits. However, the draft assemblies generated from Illumina reads appear too low quality to trust any differences found.

QUAST statistics reveal draft assemblies are very fragmented. All assemblies have at least two orders of magnitude more scaffolds than the expected number of chromosomes. Additionally, there appears to be duplication in some draft genomes. Two of three assemblies for *P. wickerhamii* (HP1 and HP5), two of five assemblies for *P. bovis* (HP14 and HP18) and the one assembly of *P. ciferrii* (HP2) are much longer than published genomes of their respective species (Bakuła et al., 2021; Guo et al., 2022; Severgnini et al., 2018). Additionally, much higher levels of BUSCO duplication are seen in draft assemblies than published genomes.

P. wickerhamii assemblies are higher quality than *P. bovis* or *P. ciferrii*, as they have fewer scaffolds which tend to be longer. This is likely partially due to difficulties in extracting DNA from *P. bovis* and *P. ciferrii*. Cultures of these species tended to produce much more biofilm, resulting in a larger interphase and potentially less DNA from the same number of cells. It has also been suggested that the durability of *Prototheca* species to some stresses is based on the robustness of their cell wall (Lassa et al., 2011). In Chapter 4) Survival of Abiotic Stresses, it was noticed that *P. bovis* is much more resistant to temperature and pH variation than *P. wickerhamii*. If durability of the cell wall is a component of this general stress tolerance and this durability makes nucleic acid extraction from these strains more

difficult, we may infer that the cell walls of *Prototheca* species within the lineage containing *P. bovis* are more durable than those outside it, even if it is not known how or why. The difficulty of isolating high-quality DNA from *P. bovis* and its relatives may account for the scarcity of genomes from these species, despite them appearing to be the most significant pathogens within *Prototheca*.

However, problematic lysis is unlikely to account for all of the fragmentation seen in *P. bovis* and *P. ciferrii* genomes. The published genomes of these species make use of mate pair sequencing to scaffold contigs, as does the genome of *P. cutis* JCM 15793, but these assemblies remain relatively fragmented, unlike *P. cutis* JCM 15793 (Table 6-2) (Suzuki et al., 2018; Severgnini et al., 2018). There may, therefore, be features of these genomes that make assembly problematic, like repetitive or duplicated regions of the genome.

It Is clearly difficult to get a *Prototheca* assembly purely based on Illumina reads. However, most of the genomes assembled from short reads are similarly complete, according to BUSCO. This may suggest that reference-guided assemblies, using Illumina data, may be sufficient to provide high-quality assemblies, once suitable references are available. It appears that *Prototheca* genomes are missing a number of chlorophyte BUSCO genes, unlike *Auxenochlorella* and *Chlorella*, and *P. bovis* and *P. ciferrii* are missing more than *P. wickerhamii*.

Difficulties extracting HMW DNA

Attempts were made to extract HMW DNA from *Prototheca*, especially *P. bovis*, to improve the quality of the draft assemblies. Several methods exist for isolating HMW DNA from

microalgae. However, it remains a technical challenge – DNA extraction from algae has been noted as being more difficult than from fungi or bacteria due to the complexity and durability of their cell walls (Jagielski et al., 2017b; Lloyd et al., 2021) and the cells of *A. protothecoides* and *Prototheca* cells have been noted as being particularly durable compared to other microalgae (He et al., 2016; Lloyd and Turner, 1968).

Of the three methods attempted here, only grinding in liquid nitrogen produced HMW DNA, but did so inconsistently and only for *P. wickerhamii*. Given that it was also difficult to extract DNA for short read sequencing, this likely reflects a difference in the cell walls of *P. bovis* and *P. wickerhamii*. It is also possible that the unicellularity of *Prototheca* species made grinding less effective than would be expected, as the methods chosen were originally designed for plants, not unicellular microalgae (Li et al., 2020).

The exact composition of the *Prototheca* cell wall is not known. Glucose, galactose and mannose are monosaccharides within the cell wall, but cellulose does not appear to be a significant polysaccharide within the *Prototheca* cell wall (Conte and Pore, 1973). Sporopollenin has been reported as a component of *Prototheca* cell walls, of both the *wickerhamii*- and *bovis*-related lineages (Ueno, 2009; Lassa et al., 2011) but not all species (Atkinson et al., 1972), and has been suggested as being a major contributor to their durability (Jánosi et al., 2001; Lassa et al., 2011; Jagielski et al., 2011). Sporopollenin is an extremely durable polymer, likely synthesised from fatty acids and phenolics, but its durability means its exact structure remains unknown (Ríos et al., 2013; Mackenzie et al., 2015). Sporopollenin is resistant to most acids, bases, solvents and detergents, but can be

removed by some strong oxidising solutions, like hydrogen peroxide (Edlund et al., 2017), or certain organic bases, including 2-aminoethanol (ethanolamine) (Southworth, 1974).

The attempts to form protoplasts may also have been hindered by the presence of sporopollenin. Others, using similar methods, have failed to form protoplasts because only part of the cell wall was susceptible to digestion (He et al., 2016; Atkinson et al., 1972). It is possible that chemical methods to remove sporopollenin could be included in future attempts to form AHP protoplasts – for example, the mouth secretions of some Gelechiidae moths (specifically of the genus *Deltophora*) appear to be able to dissolve sporopollenin (Luo et al., 2011). If successful, chemical methods to form *Prototheca* protoplasts would enable resolution of whole chromosomes using PFGE and facilitate long read sequencing.

7) Phylogenetics of *Prototheca* and Related Organisms

Introduction

When investigating the phylogenetic relationships of microalgae, many authors use the sequences of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) or ribosomal DNA (rDNA) sequences (Neustupa et al., 2013; Darienko et al., 2015; Pröschold et al., 2011; Darienko and Pröschold, 2015; Bock et al., 2011). RuBisCO is not suitable to investigate the relationships between *Prototheca* species and related taxa, as *Prototheca* species are no longer photosynthetic. Historically, the locus of choice to investigate and define *Prototheca* species was the nuclear encoded rDNA, which included part of the large-subunit (LSU or 28S), small-subunit (SSU or 18S) and internal transcribed spacer region (Jagielski et al., 2018b). More recently, the partial sequence of the mitochondrial *cytb* has been used (Jagielski et al., 2018b).

Broadly, phylogenetic relationships derived from these two loci are in agreement (Figure 1-4) (Shave et al., 2021). However, there are points of contention between phylogenies based on these loci. At a higher level, there is the question of which genera are necessary to include to identify a monophyletic lineage that contains all *Prototheca* species.

The AHP lineage is well supported by both loci, but analyses based on rDNA and *cytb* have suggested that *Jaagichlorella* and *Chlorella*, respectively, belong within the genus, and are closely related to *Auxenochlorella* (Plieger and Wolf, 2022; Jagielski et al., 2019a). The inclusion of *Jaagichlorella* in the AHP lineage is likely incorrect, as this GenBank entry is actually a misclassified isolate of *Auxenochlorella* (Darienko and Pröschold, 2015), but it

remains unclear whether *Chlorella* is a genuine member of the AHP lineage. It has been suggested that *Micractinium* may be encompassed within *Chlorella* (Masuda et al., 2016; Wu et al., 2019), while others suggest that *Micractinium* may be a sister genus to *Chlorella* (Luo et al., 2010; Pröschold et al., 2020). Depending on which genera are required to fully encompass all *Prototheca* species while remaining monophyletic, this lineage may be referred to as the AHP, CHAP or CHAMP lineage.

At the species level, some relationships within *Prototheca* are well-supported while others are poorly supported.

One such well-supported group includes species that were formerly identified as *P. zopfii*. This includes *P. blaschkeae*, which was distinguished from *P. zopfii* in 2006 (Roesler et al., 2006), and *P. bovis*, *P. ciferrii*, and *P. cookei* (Jagielski et al., 2019a). This group was been referred to as the cattle-associated species earlier in this thesis based on the pathogenic members of this group, as *P. bovis* and *P. blaschkeae* are the main pathogens of cattle and *P. ciferrii* is often found in cattle environments (Toyotome and Matsui, 2022; Tashakkori et al., 2022; Jagielski et al., 2019b; Osumi et al., 2008) , but "former-*zopfii*" species might be more accurate. Another well-supported group includes *P. cutis* and *P. miyajii*, both of which were originally identified as *P. wickerhamii* (Satoh et al., 2010; Masuda et al., 2016), as well as *P. paracutis*. It has also been well documented that these two clades are relatively distant from each other in the overall AHP lineage.

Less well supported is the association between *P. moriformis*, *P. stagnora*, and *P. tumulicola*, a group referred to as "true environmental" species earlier in this thesis as the grouping is

the only clade that contains only *Prototheca* while containing no pathogenic species. Analyses based on rDNA suggest these species form a clade in the overall AHP lineage, which is usually more closely associated with the former-*zopfii* species (Jagielski et al., 2019a; Shave et al., 2021; Jagielski et al., 2018b; Plieger and Wolf, 2022; Kunthiphun et al., 2019). However, analyses based on *cytb* sequences suggest that this group is paraphyletic, being basal to the overall AHP lineage (Jagielski et al., 2019a, 2022). There is also ambiguity regarding *P. xanthoriae*, which was also originally identified as *P. wickerhamii*, but is usually considered more closely related to *Auxenochlorella* species than other *Prototheca* species (Jagielski et al., 2019a; Suzuki et al., 2018).

The final uncertainty with *Prototheca* species is whether *P. wickerhamii* is more closely related to *Auxenochlorella* and *P. xanthoriae* or the clade containing *P. cutis*. In an earlier chapter, *P. wickerhamii* was grouped with the *P. cutis* clade to form the human-associated species, but this was based upon capacity to infect humans rather than strong phylogenetic support.

Finally, the relationships of *Helicosporidium* and *Auxenochlorella* within the AHP lineage are not clear. *Auxenochlorella* is usually strongly associated with the former-*wickerhamii* species but has not been investigated for monophyly, largely due to sequence data only being available for one species. By contrast, it currently appears that the *Helicosporidium* genus is monophyletic (Mancera et al., 2012), but *Helicosporidium* frequently moves within the AHP lineage with different analyses (Plieger and Wolf, 2022; Jagielski et al., 2018b; Suzuki et al., 2018; Ueno et al., 2005).

Here, I aim to Investigate these relationships between members of the AHP/CHAMP lineage, and determine whether the AHP lineage requires the inclusion of *Chlorella* and *Micractinium* to be a monophyletic clade.

Methods

Data Collection

The data used for phylogenetics were a combination of publicly available data were taken from GenBank and sequencing performed in earlier chapters. Publicly available sequences were acquired over time between March 2022 and July 2023. To investigate the possibility of the CHAMP lineage, genomes and genes were taken from species from each of the five genera (*Chlorella*, *Helicosporidium*, *Auxenochlorella*, *Micractinium*, and *Prototheca*).

Organelle genomes were obtained from all *Prototheca* and *Auxenochlorella* species for which whole genomes have been sequenced. Additionally, *cytb* sequences for *Prototheca* species without published whole genomes were obtained using BLASTn, as described below. Where multiple sequences were available, multiple sequences were included. However, not all sequences were included where many sequences were available for a given species, as the addition of these sequences typically does not improve support for higher-level relationships (Jinatham et al., 2021; Jagielski et al., 2022) while increasing the computational cost of analyses. Typically, two or three published sequences were included in addition to the sequences produced in this thesis.

Many species of *Chlorella* have been proposed and reclassified over time, as happened with *Prototheca* (Pröschold et al., 2011; Bock et al., 2011). This made it challenging to know which species are still considered valid, and should therefore be included. Up to three sequences for *Chlorella* species were acquired for each of the three "true" *Chlorella* species (*C. sorokiniana*, *C. variabilis*, and *C. vulgaris*) for which multiple organelle genomes were available (Bock et al., 2011). Two sequences from species defined since these revisions (*C. ohadii* and *C. dessicata*) were also included. Only one *Micractinium* genome was included, as it is not a contentious point within the literature regarding *Prototheca* and it was expected that the relationship between *Prototheca* and *Micractinum* would reflect the relationship between *Chlorella* and *Prototheca*.

Previous analyses have used *Chalmydomonas* sequences as an outgroup (Jinatham et al., 2021; Jagielski et al., 2022). However, as noted in Chapter 3) Identification of Algal Strains, *Chlamydomonas* is in a different class to *Prototheca* (Table 3-1) and had poor matches for the primers used to amplify *cytb* (Figure 3-9). It was thought that this evolutionary distance and poor match might impact alignments and produce spurious relationships. In the same analysis *Symbiochloris* and *Lobosphaera* were observed to have good alignment with the primers used to amplify *cytb*, and were more closely related to the AHP lineage without being part of it (as members of the same class but different orders). Thus, *Symbiochloris* and *Lobosphaera* were chosen as outgroups.

Organelle genomes were identified and extracted from whole genome assemblies, where necessary, as described in the core methods section.

Phylogenetics

To generate phylogenies based on individual genes, AHP sequences of *cytb* were obtained using BLASTn with default parameters, using sequences from *P. bovis* and *P. wickerhamii* as queries. Additionally, sequences for some *Prototheca* species were sought by manually searching GenBank using accession numbers or species names. Despite this, *cytb* sequences for some species, like *P. paracutis*, could not be found and therefore could not be included. Sequences of *cytb* derived from *Prototheca* were not successful in finding matches in some *Chlorella* or more distantly related genomes. In these cases, start and stop positions for *cytb* sequences were either derived from GeSeq annotations or alignments of *cytb* primers to organelle genomes, as described in Chapter 3) Identification of Algal Strains.

Other organelle sequences were obtained using the start and stop positions included in annotations produced by GeSeq from whole organelle sequences.

Phylogenetic analyses focused on organelle genes as organelle genomes can clearly be assembled from relatively low-quality data (i.e. paired-end illumina reads). Thus, sequences from additional taxa are likely to be acquired relatively easily. Additionally, matches to organelle genes were typically only found in one scaffold in assemblies – the scaffold containing the whole organelle genome. This improved confidence the genes selected were orthologs. Nuclear genes, including tubulin, were explored for analyses, but matching sequences were found in multiple scaffolds, reducing confidence that sequences were orthologs. However, it was not clear if these multiple matches were due to assembly errors or duplication in genomes.

Sequences were aligned using Clustal Omega (v1.2.3) (Sievers et al., 2011) in Geneious 2023.1.2 (https://www.geneious.com) with the default settings, unless specified otherwise. Maximum likelihood (ML) phylogenetic trees were generated using the RaxML plugin (v8.2.11) (Stamatakis, 2014) in Geneious, using a generalised-time reversible GAMMA substitution model and 500 bootstrap replicates.

Synteny

The arrangement of regions within organelle genomes was also explored to investigate relationships within the CHAMP lineage. To show meaningful structural differences, organelle genome sequences must be in the same orientation and start in the same place, as organelle genomes are circular. This was done manually, initially using sequences of *cytb* and *cysT* for mitochondrial and plastid genomes respectively followed by iterative manual refinement. Syntenic regions were visualised using the MAUVE (v1.1.3) (Darling et al., 2010) plugin in Geneious.

Results

The partial *cytb* sequence supports a monophyletic AHP lineage

Previous phylogenies based on the sequence of *cytb* have identified that *Chlorella* species are within the AHP lineage. However, they do so with a limited number of sequences from a limited number of *Chlorella* strains, and completely exclude *Micractinium* sequences (Jagielski et al., 2022, 2019a). It was thought that this limitation in sequence data might provide insufficient resolution to represent the relationships within the CHAMP lineage. Thus, sequence data from five species of *Chlorella* and one species of *Micractinium* was obtained for phylogenetics.

In searching available mitochondrial genome sequences from *C. vulgaris* for *cytb*, it was found that all three had matches for the *cytb* sequence from *C. variabilis*. However, it was not a single match. In all sequences, a match for the first half of the *cytb* sequence was separated from a match for the second half by 1400±1 bp (Figure 7-1). A compatible observation was made in Chapter 3) Identification of Algal Strains, where a PCR reaction to amplify *cytb* from a strain of *C. vulgaris* produced a band that was approximately 2 kbp instead of the expected 700 bp (Figure 3-10).

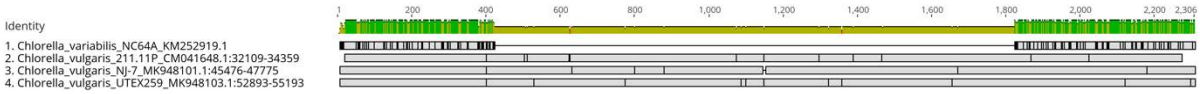


Figure 7-1. Alignment of the cytb *sequence from* C. vulgaris *aligned against* C. variabilis*. Sequences were initially aligned using the in-built Geneious Alignment tool in Geneious before being manually edited to show matches identified by BLAST. Single nucleotide polymorphisms are indicated in each sequence with a vertical black line. Locations with 100% nucleotide identity are indicated in green.*

This was interpreted as being caused by an insertion in the *C. vulgaris* sequence for *cytb*. This insert was manually removed before aligning, as it would interfere with the alignment algorithm. Inserts in the sequence of *cytb* were not observed in the sequences of other taxa. A taxonomically unrestricted BLAST search optimised for slightly similar sequences performed in August of 2023 found small but significant matches in mitochondrial genomes of other algae, but the full sequence was only found in genomes of *C. vulgaris*.

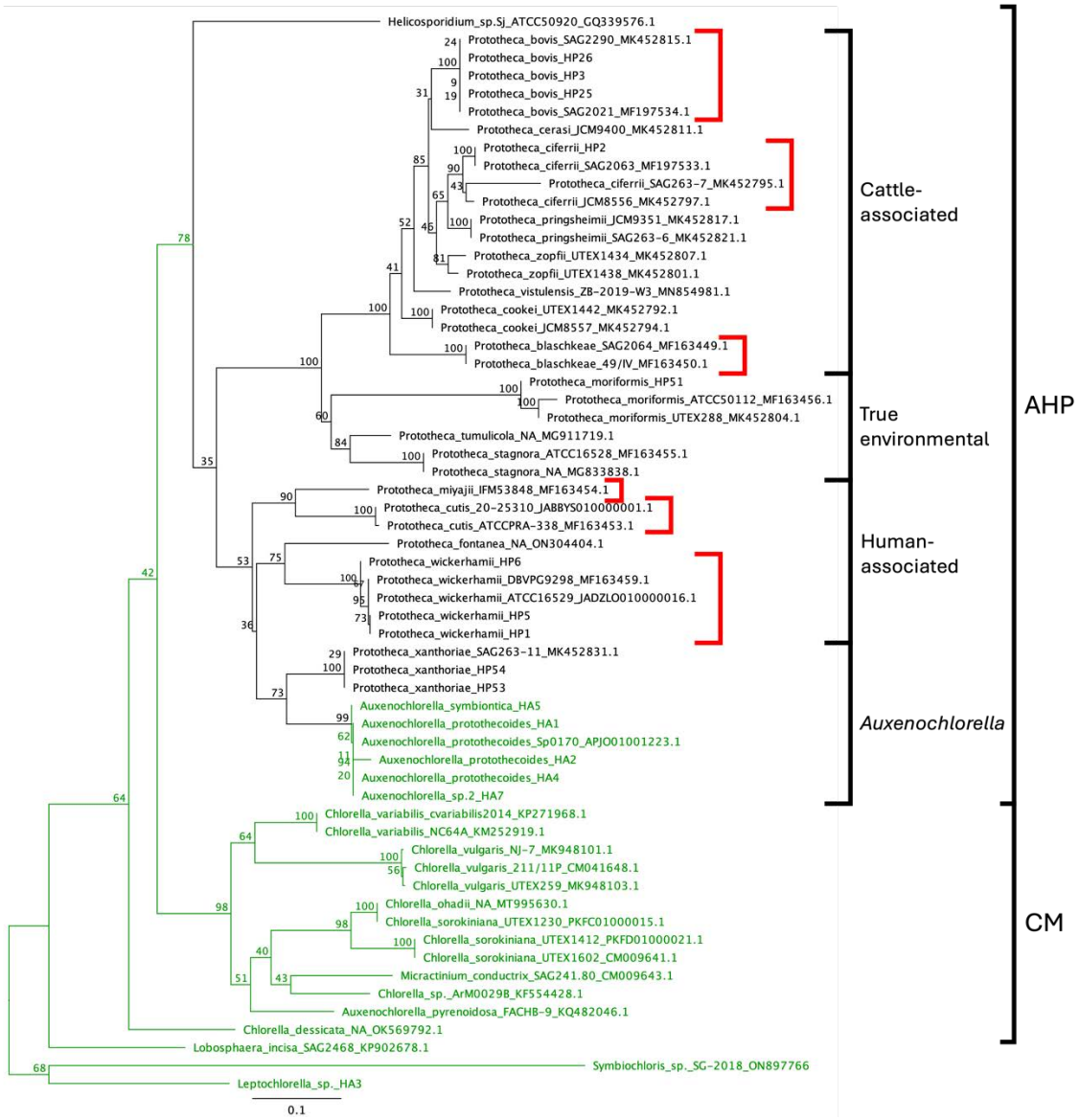


Figure 7-2. Maximum likelihood phylogeny based on partial cytb *gene sequences of CHAMP species. The alignment comprised of 60 sequences of 598 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 11;* Helicosporidium*, 1;* Auxenochlorella*, 7*; Micractinium*, 1;* Prototheca*, 37. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The separation of the AHP lineage from* Chlorella *and* Micractinium *is indicated with black brackets, as are the four AHP sub-lineages used in Chapter 5) Phagocytosis of Prototheca and Related Organisms. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to* Symbiochloris *sp. SG-2018 and* Leptochlorella *sp. HA3. Scale bar indicates nucleotide substitution rate.*

Additionally, previous phylogenies based on *cytb* use limited sequence data from *Auxenochlorella*, typically limited to a single sequence from *Auxenochlorella protothecoides*. In Chapter 6) Genome Sequencing and Assembly of *Prototheca* (and Related Species), genomes from two additional species, *Auxenochlorella symbiontica* and an unknown species

represented by HA7, were sequenced. Furthermore, the genome from the algal strain FACHB-9, which was originally identified as *Chlorella pyrenoidosa*, is publicly available (Fan et al., 2015). However, *C. pyrenoidosa* has been reported as being synonymous with *Auxenochlorella pyrenoidosa* (Calvo-Pérez Rodó and Molinari-Novoa, 2015). Thus, this strain was considered a member of *Auxenochlorella*. Sequences from four *Auxenochlorella* taxa could therefore be included.

Sequences for *cytb* could not be found for some species of *Prototheca*, and therefore could not be included.

With the addition of *Micractinium* and more *Chlorella* sequence data, a phylogeny based on *cytb* no longer identifies *Chlorella* as being within the AHP lineage (Figure 7-2). Additionally, the true environmental species are no longer considered basal to the AHP lineage, instead forming a sister lineage to the cattle-associated species, as is seen in analyses based on rDNA. Instead, *Helicosporidium* is identified as basal to the AHP lineage.

However, several relationships are not well supported in the phylogeny based on the sequence of *cytb*. Notably, there is low support for the node that suggests *P. wickerhamii* is more closely related to *Auxenochlorella* than *P. miyajii*. The node separating the clade containing the human-associated species and *Auxenochlorella* from the true environmental and cattle-associated species is also poorly supported. To attempt to investigate these relationships further, additional phylogenies were produced using other genes.

Other organelle gene sequences support a monophyletic AHP lineage

Five additional genes were chosen to investigate relationships between AHP species, all of which were encoded in organelle genomes. Organelle genes were chosen as it appears to be easier to sequence whole organelle genomes than nuclear genomes, thus these phylogenies are likely to be easily improved in the future.

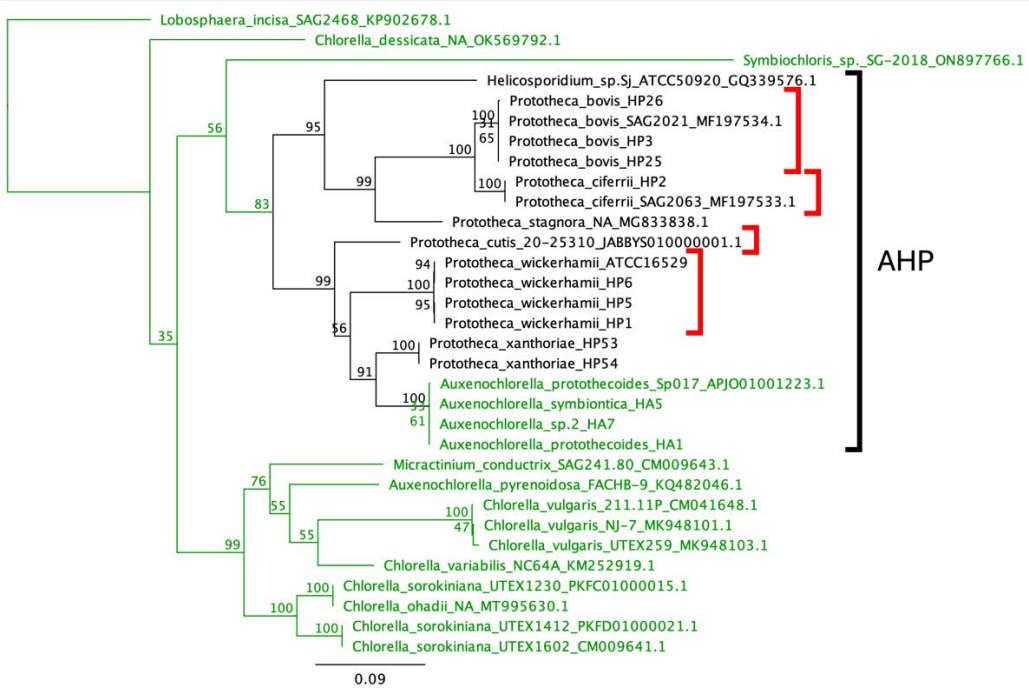


Figure 7-3. Maximum likelihood phylogeny based on mitochondrial atp1 *gene sequences of CHAMP species. The alignment comprised of 32 sequences of 1542 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 9;* Helicosporidium*, 1;* Auxenochlorella*, 5*; Micractinium*, 1;* Prototheca*, 14. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The AHP lineage is indicated with a black bracket. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to* Lobosphaera incisa *SAG2468. Scale bar indicates nucleotide substitution rate.*

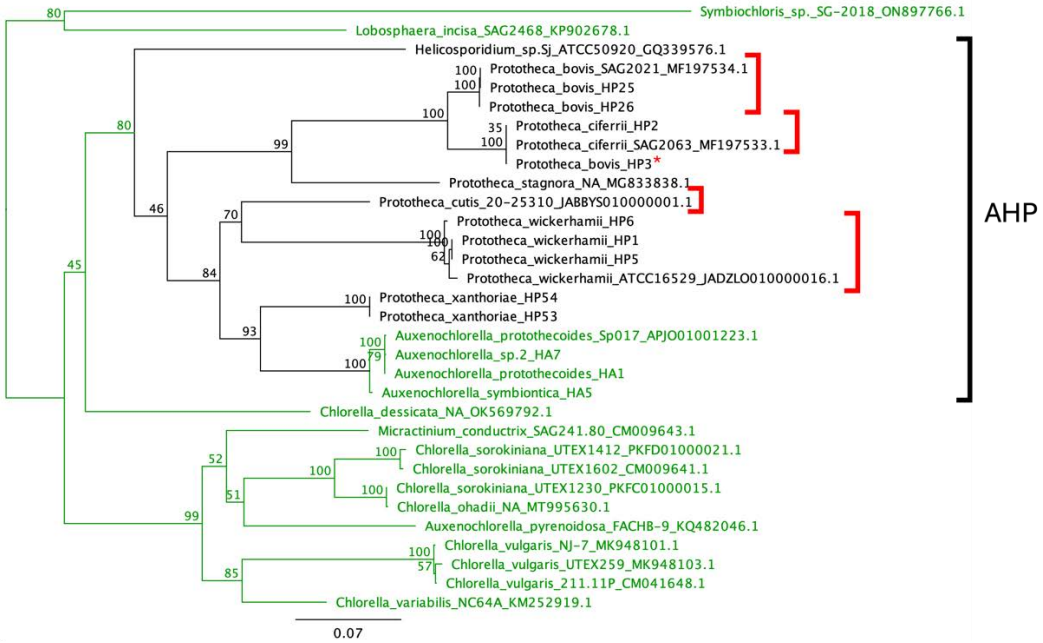


Figure 7-4. Maximum likelihood phylogeny based on mitochondrial cox1 *gene sequences of CHAMP species. The alignment comprised of 32 sequences of 1584 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 9;* Helicosporidium*, 1;* Auxenochlorella*, 5*; Micractinium*, 1;* Prototheca*, 14. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The sequence from* P. bovis *HP3 that does not group with other* P. bovis *species is indicated by a red star. The AHP lineage is indicated with a black bracket. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to* Lobosphaera incisa *SAG2468 and* Symbiochloris *sp. SG-2018. Scale bar indicates nucleotide substitution rate.*

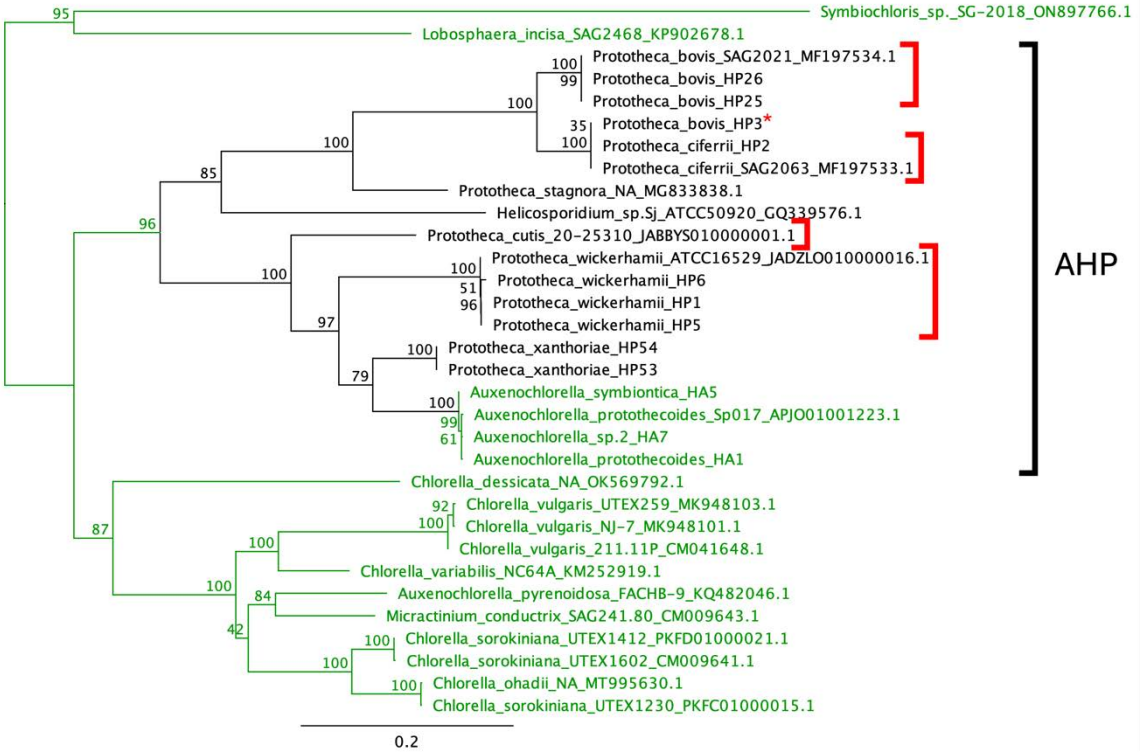


Figure 7-5. Maximum likelihood phylogeny based on mitochondrial nad5 *gene sequences of CHAMP species. The alignment comprised of 32 sequences of 2081 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 9;* Helicosporidium*, 1;* Auxenochlorella*, 5*; Micractinium*, 1;* Prototheca*, 14. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The*

sequence from P. bovis *HP3 that does not group with other* P. bovis *species is indicated by a red star. The AHP lineage is indicated with a black bracket. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to* Lobosphaera incisa *SAG2468 and* Symbiochloris *sp. SG-2018. Scale bar indicates nucleotide substitution rate.*

Phylogenies were generated from three genes chosen from mitochondrial genomes: *atp1* (Figure 7-3), *cox1* (Figure 7-4), and *nad5* (Figure 7-5). The sequence of *cox1* was fragmented in many genomes, though to a different extent in different taxa: *cox1* was present as 3 fragments in *P. wickerhamii*, *P. xanthoriae*, *P. stagnora*, *Auxenochlorella* species, and *M. conductrix*; 2 fragments in *P. cutis* and some *Chlorella* species; and 1 fragment in *P. bovis*, *P. ciferrii* and other *Chlorella* species. Fragments were manually combined, as was done for *cytb* in *C. vulgaris*. No other sequence was fragmented.

When using chloroplast genomes for phylogenetics, it is generally recommended to use the sequences of *nad2*, *nad5*, *atpA*, *atpE*, *psbC* and *psbD* (Mekvipad and Satjarak, 2019). However, all of these genes were absent in chloroplast genomes from at least two of the explored taxa and some were absent in all chloroplast genomes, according to GeSeq annotations (Table 7-1).

Table 7-1. Presence or absence of recommended genes for phylogenetic analysis in CHAMP chloroplast genomes.

Gene	*P. bovis* HP25	*P. wickerhamii* HP1	*P. xanthoriae* HP54	*prototheco*ides HA1	*C. variabilis* NC64A
nad2	Absent	Absent	Absent	Absent	Absent
nad5	Absent	Absent	Absent	Absent	Absent
atpA	Absent	Absent	Present	Present	Present
atpE	Absent	Absent	Present	Present	Present
psbC	Absent	Absent	Absent	Present	Present
psbD	Absent	Absent	Absent	Present	Present

Instead, phylogenies were created based on the sequences of *rpl2* (Figure 7-6) and *cysT* (Figure 7-7), as all chloroplast genomes tested contained these genes.

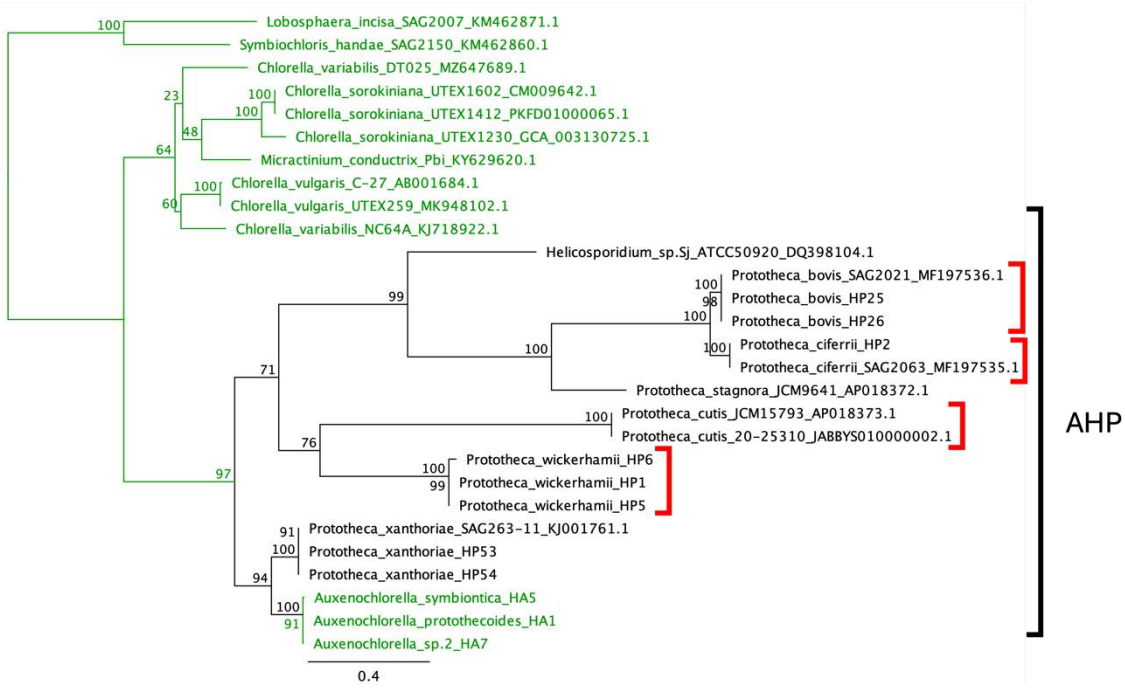


Figure 7-6. Maximum likelihood phylogeny based on chloroplast cysT *gene sequences of CHAMP species. The alignment comprised of 28 sequences of 1070 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 7;* Helicosporidium*, 1;* Auxenochlorella*, 3*; Micractinium*, 1;* Prototheca*, 14. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The AHP lineage is indicated with a black bracket. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to* Lobosphaera incisa *SAG2468 and* Symbiochloris *sp. SG-2018. Scale bar indicates nucleotide substitution rate.*

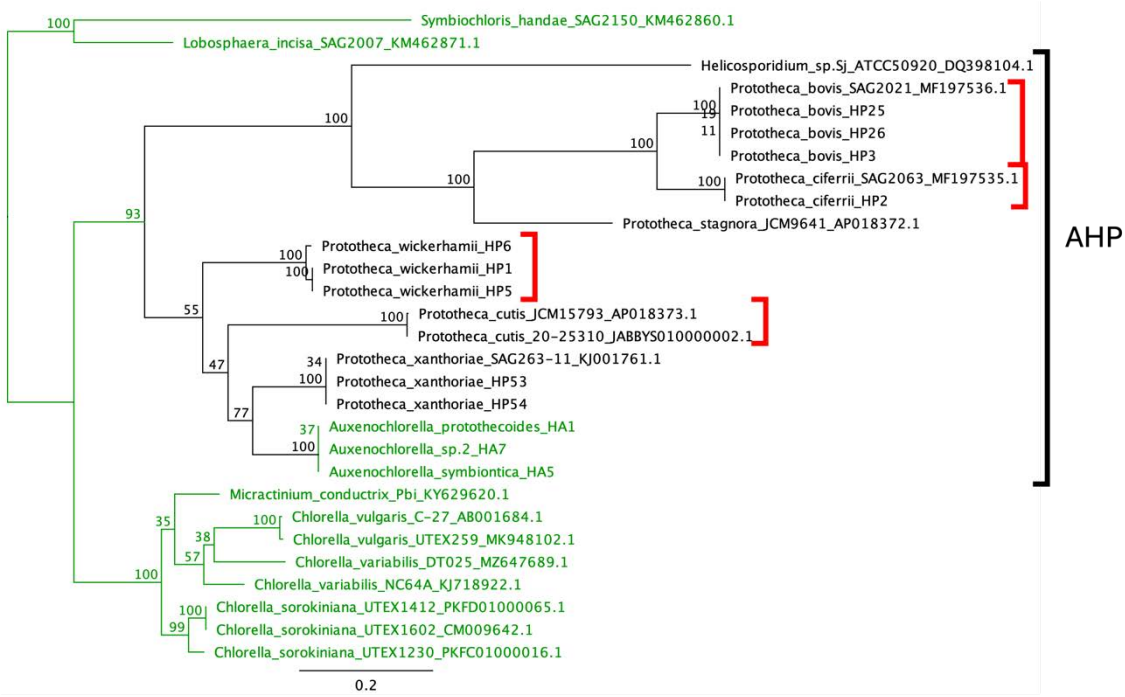


Figure 7-7. Maximum likelihood phylogeny based on chloroplast rpl2 *gene sequences of CHAMP species. The alignment comprised of 29 sequences of 930 nucleotides. CHAMP species are represented by the following number of genomes:* Chlorella*, 7;* Helicosporidium*, 1;* Auxenochlorella*, 3*; Micractinium*, 1;* Prototheca*, 15. The labels and branches of photosynthetic lineages are coloured green. Species capable of infecting mammals are indicated with red brackets. The AHP*

lineage is indicated with a black bracket. Numbers at the nodes are ML bootstrap values. Sequences are provided with genus and species names, strain ID (if possible) and GenBank accession number (if available), separated by underscores. The phylogeny was rooted to Lobosphaera incisa *SAG2468 and* Symbiochloris *sp. SG-2018. Scale bar indicates nucleotide substitution rate.*

The separation of the AHP lineage from *Chlorella* and *Micractinium* is observed with all five genes. The separation of the lineages containing *P. bovis* from the lineage containing *P. wickerhamii* is also observed with all five genes, though in the phylogeny based on *cysT* these two lineages are themselves sister to the lineage containing *Auxenochlorella* and *P. xanthoriae*. Sequences derived from *P. stagnora*, the only true environmental species for which organelle genomes were available, were included within the same lineage as *P. bovis* with all five genes with strong support.

Additionally, all genes except for *cox1* support the inclusion of *Helicosporidium* sp. Sj within a lineage that includes *P. bovis* and excludes *P. wickerhamii*, *Auxenochlorella* and all former-*wickerhamii* species.

Unfortunately, the relationships between *P. cutis*, *P. wickerhamii*, and *Auxenochlorella* remain unclear – the phylogenies created from *cox1* and *cysT* sequences suggest *P. wickerhamii* is more closely related to *P. cutis* while the phylogenies created from *atp1* and *nad5* sequences suggest *P. wickerhamii* is more closely related to *Auxenochlorella*.

In all phylogenies, sequences from the strain FACHB-9, originally identified as *C. pyrenoidosa* but taken to be *A. pyrenoidosa* in this chapter, grouped with sequences taken from *Chlorella* not *Auxenochlorella*. Thus, this isolate is presumably not a member of *Auxenochlorella*.

Organelle synteny supports two lineages within the AHP lineage

There is taxonomic information in the arrangements of genes in a genome, as well as the order of nucleotides within them. Thus, to not rely solely on phylogenies based on individual genes, the structures of organelle genomes were compared within the CHAMP lineage.

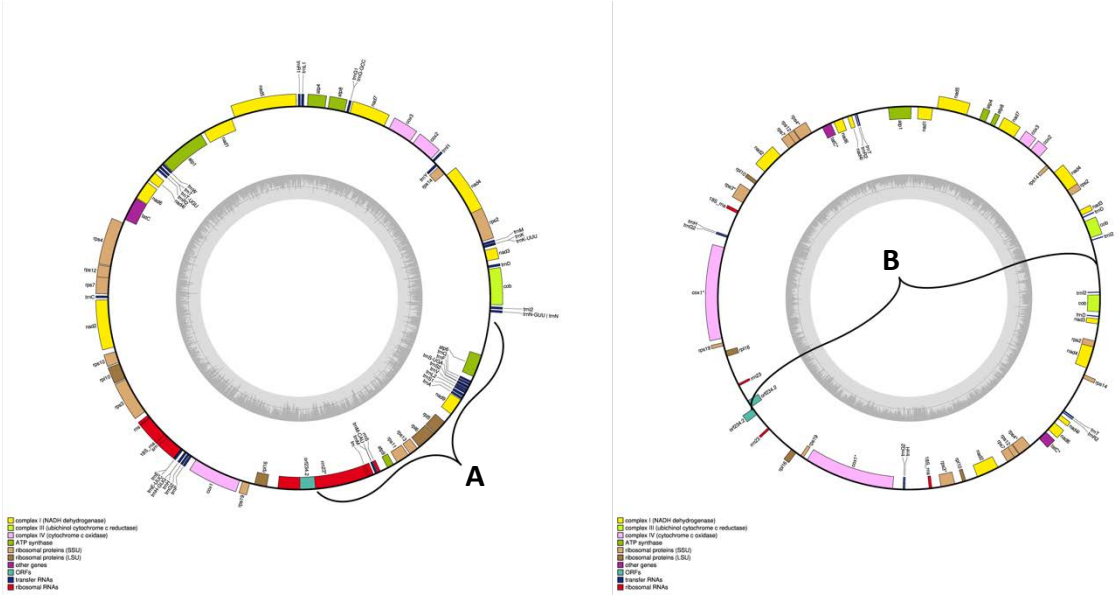


Figure 7-8. Circular plots of the mitochondrial genomes of P. bovis *HP2 (left) and* P. stagnora *(right). Genes encoded in the region labelled A are not present in the mitochondrial genome of* P. stagnora*, while genes encoded in the region labelled B appear to be duplicates in the mitochondrial genome of* P. stagnora*. In this instance, the duplicated region appears to be an inversion of the original sequence.*

Probable assembly errors were identified in the mitochondrial genomes of *P. stagnora*, and *M. conductrix* that were included in this analysis. The mitochondrial genome of *P. stagnora* appeared to have a duplication of one region and deletion of another relative to the mitochondrial genome of *P. bovis*, according to the annotations generated by GeSeq (Figure 7-8). Similarly, duplication of one region in the *Micractinium* genome was observed (Figure 7-9). Duplication in organelle genomes as a result of assembly error is not uncommon and these apparent duplications prevented whole genome alignment using MAUVE. Thus, the duplicated sections were manually removed before mitochondrial genomes were aligned using MAUVE.

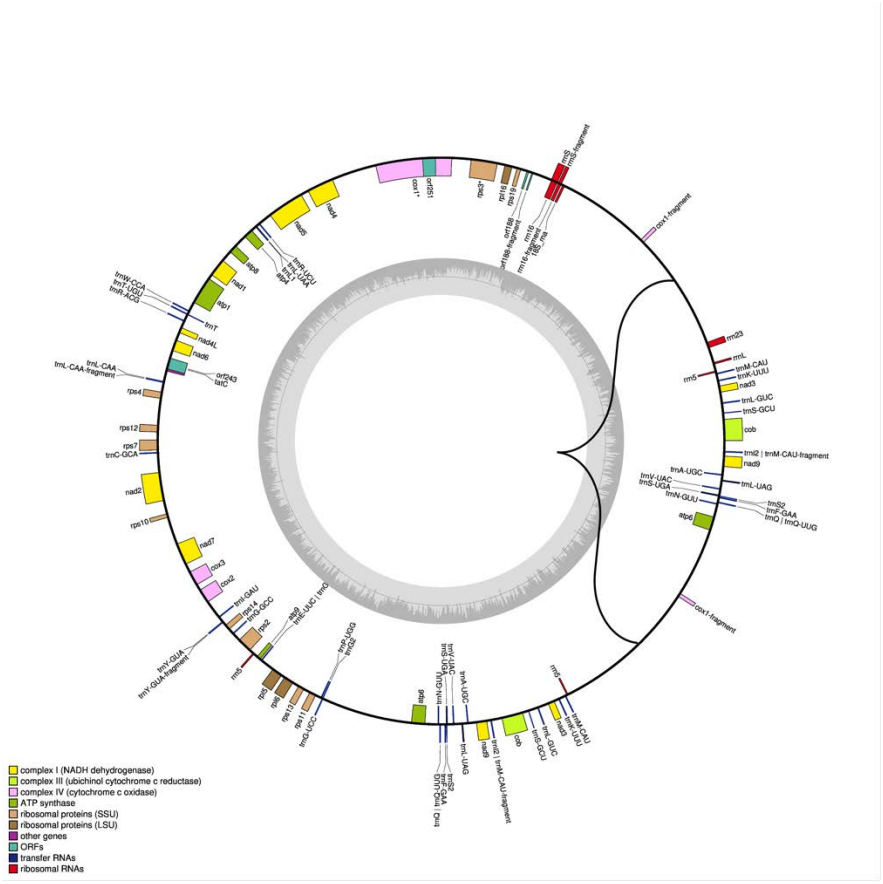


Figure 7-9. Circular plot of the mitochondrial genomes of M. conductrix *SAG 241.80. Genes encoded in the indicated region appear to be duplicated. In this instance, the duplicated region appears to be in the same orientation as the original sequence.*

Additionally, the mitochondrial genome appears to be present in the genome assembly for *A. pyrenoidosa* (FACHB-9) as part of a larger scaffold. The whole scaffold was aligned to the mitochondrial sequences derived from *C. sorokiniana* UTEX 1230 and the draft assembly of HA1, both of which aligned to the first 70 kbp of the scaffold in question. As this is approximately the length of the *Chlorella* mitochondrial genome, this was manually extracted from the scaffold.

The structures of mitochondrial genomes visually clustered into three main groups, which were generally similar within each group but showed differences between groups (Appendix 27). The differences between the mitochondrial genomes from *Auxenochlorella*, *P. cutis*, *P. wickerhamii*, and *P. xanthoriae* are limited, with a single structural rearrangement

(translocation or inversion) sufficient to match other genomes in the group (Figure 7-10).

The *Auxenochlorella* genomes are almost structurally identical.

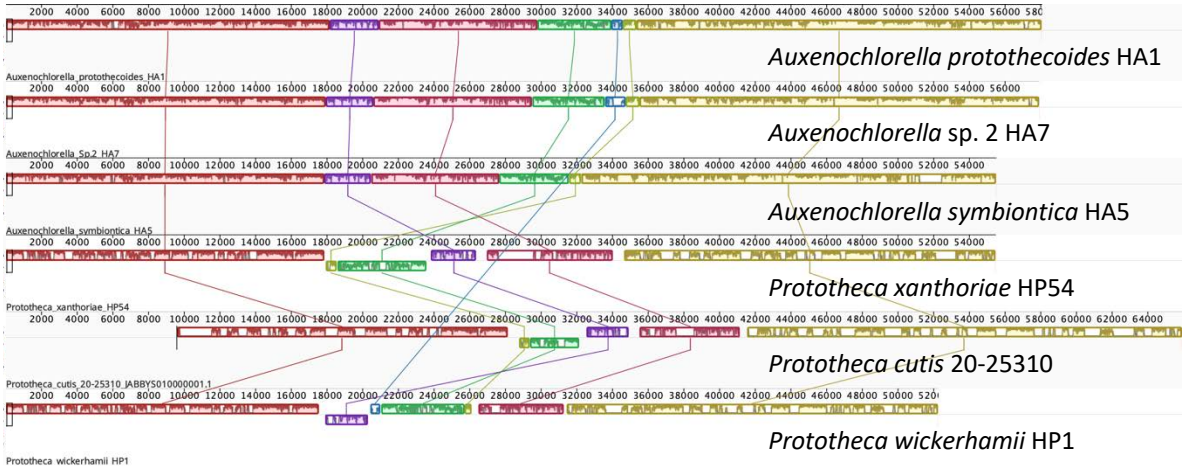


Figure 7-10. Arrangement of syntenic blocks of mitochondrial sequences from Auxenochlorella *and human-associated* Prototheca*, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.*

Structurally, the mitochondrial genomes of *P. bovis, P. ciferrii* and *P. stagnora* are virtually identical, apart from the deletion in *P. stagnora* that results in the loss of several blocks at the end of the genome which may be an artefact (Figure 7-11). The *Helicosporidium* mitochondrion is also quite similar to these three, with two large inversions sufficient to almost match the *P. bovis* mitochondrion.

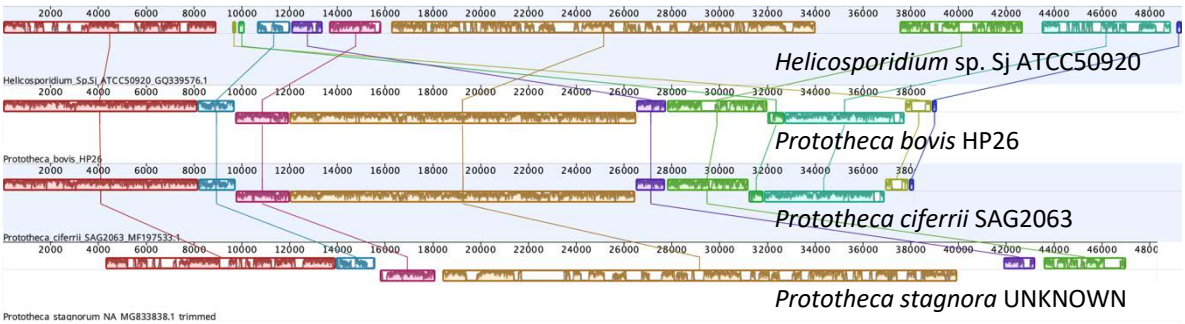


Figure 7-11. Arrangement of syntenic blocks of mitochondrial sequences from Helicosporidium*, and true environmental and cattle-associated species of* Prototheca*, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.*

Larger structural changes have occurred between these two groups, and between these groups and the *Chlorella* mitochondrial genomes (Figure 7-12). Additionally, there is significant structural variation within the *Chlorella* mitochondrial genomes, which include genomes from *A. pyrenoidosa* and *M. conductrix*.

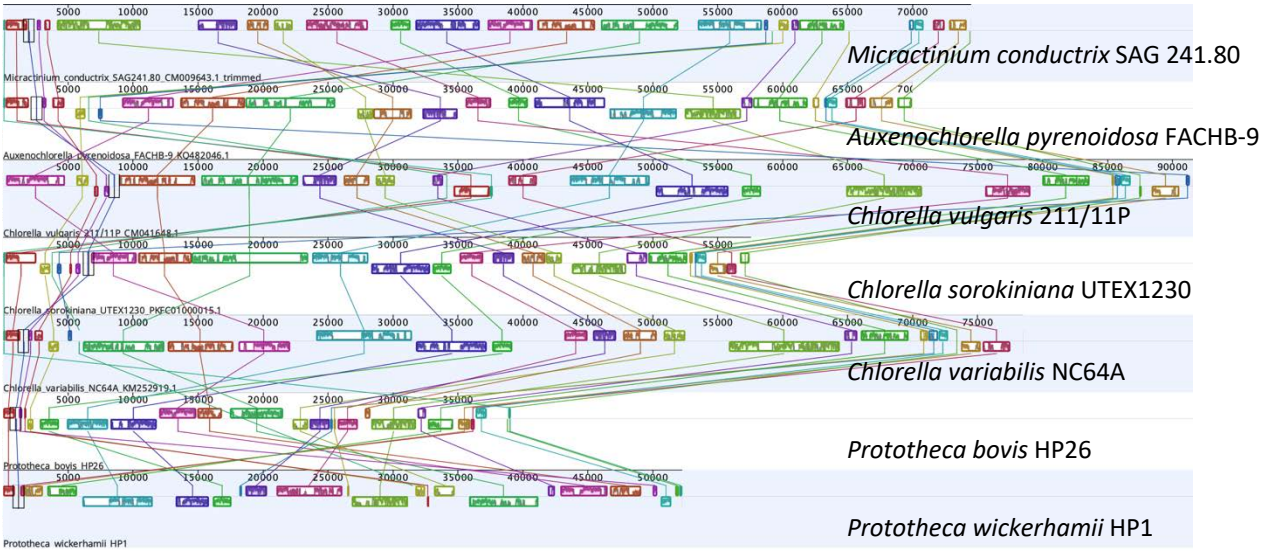


Figure 7-12. Arrangement of syntenic blocks of mitochondrial sequences from a selection of CHAMP species, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.

Fewer structural changes were observed in the chloroplast genomes of CHAMP species, but the same three groups that could be seen in the structures of mitochondrial genomes were still visible (Appendix 28). An intact chloroplast genome could not be found in the assembly for the strain FACHB-9. There are no structural changes in the chloroplast genomes from *Auxenochlorella*, *P. cutis*, *P. wickerhamii*, and *P. xanthoriae* apart from deletion of some regions (Figure 7-13).

The chloroplast genomes of *P. bovis*, *P. ciferrii*, *P. stagnora* and *Helicosporidium* show a similar pattern as their mitochondria. There are no structural rearrangements between

Prototheca genomes, but some differences exist between *Prototheca* and *Helicosporidium* genomes (Figure 7-14).

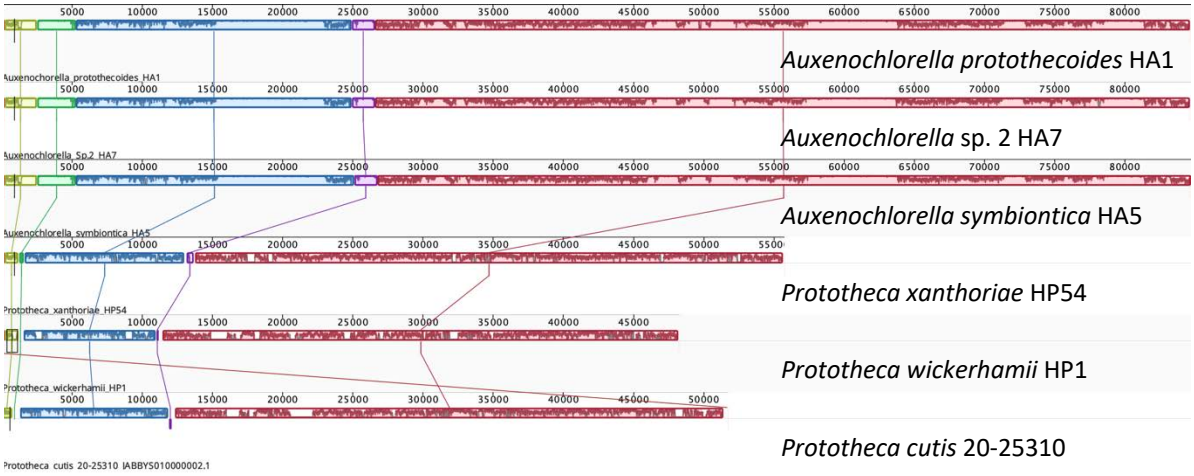


Figure 7-13. Arrangement of syntenic blocks of chloroplast sequences from Auxenochlorella *and human-associated* Prototheca*, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.*

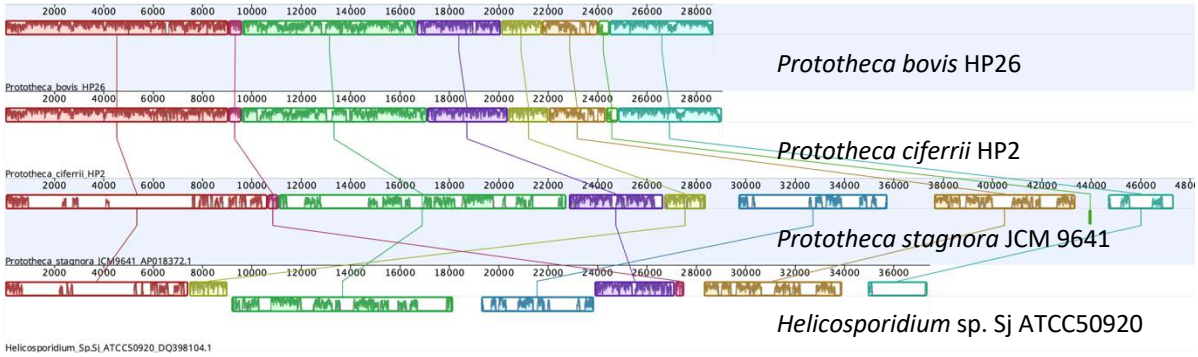


Figure 7-14. Arrangement of syntenic blocks of chloroplast sequences from Helicosporidium*, and true environmental and cattle-associated species of* Prototheca*, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.*

Finally, it is clear that large rearrangements have occurred between *P. wickerhamii* and *P. bovis*, as well as within *Chlorella* species (Figure 7-15). It is readily apparent that *Prototheca* have extremely reduced chloroplast genomes compared to *Chlorella*, and large blocks of the

P. bovis chloroplast genome do not find matches in *P. wickerhamii* and related genomes (though these regions do have matches in *P. ciferrii* and *P. stagnora* genomes).

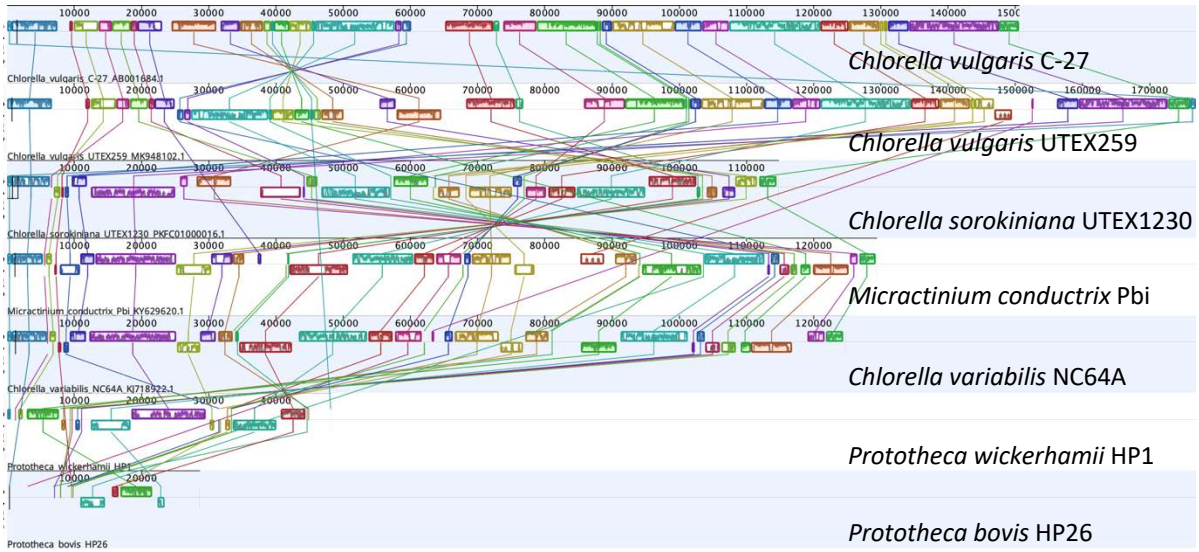


Figure 7-15. Arrangement of syntenic blocks of chloroplast sequences from a selection of CHAMP species, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.

Discussion

In this chapter I use sequence and structural data from genomes to investigate the phylogenetic relationships between the CHAMP species, using both publicly available sequence data and genomes assembled in the previous chapter.

The CHAMP lineage

With the addition of more *Chlorella* sequences, the partial *cytb* sequence no longer suggests that *Chlorella* or *Micractinium* are within the AHP lineage. This is supported by organelle structure. Thus, there is no indication that it is necessary to include *Chlorella* or *Micractinium* within the monophyletic lineage that includes all *Prototheca* species. The AHP lineage is therefore a monophyletic clade.

However, it is not possible to comment on the monophyly of the CHAMP lineage from the analysis in this chapter. The structure of organelle genomes suggests that the *Chlorella* species included in this investigation are quite distantly related to each other, relative to *Prototheca* species, as large changes have occurred in both mitochondrial and chloroplast genomes. Additionally, several genera (*Coccomyxa*, *Dicloster*, *Marvania*, *Nanochlorum*, *Parachlorella*, *Picochlorum*, *Pseudochloris*) have been suggested to be more closely related to *Chlorella* than *Prototheca* based on plastid gene sequences and the sequence of β-tubulin (Mancera et al., 2012; Bakuła et al., 2020). *Prototheca* have historically been considered to be equivalent to *Chlorella* that have lost the ability to photosynthesise (Conte and Pore, 1973; Lass-Flörl and Mayr, 2007), but future molecular analyses may suggest that other photosynthetic algal taxa are more closely related to *Prototheca* than *Chlorella*.

The AHP Lineage

Within the overall AHP lineage, there is consistent separation of species into two major sub-lineages, in phylogenetics and organelle genome structures. One sub-lineage contains *Auxenochlorella*, the human-associated species of *Prototheca* (*P. wickerhamii*, *P. cutis*) and species that were originally identified as *P. wickerhamii* (*P. xanthoriae*). The other sub-lineage contains the cattle-associated species of *Prototheca* (*P. bovis*, *P. ciferrii)*, the species that were originally identified as *P. zopfii*, and the "true environmental" *Prototheca* species (*P. moriformis*, *P. stagnora*). It appears that this second sub-lineage also includes *Helicosporidium*. In homage to the moniker for the overall AHP lineage, perhaps these two sub-lineages could be referred to as the AP and the HP lineages.

One of the historical suggestions to resolve the current paraphyly of *Prototheca* has been to allocate *Prototheca* species into different genera (Todd et al., 2018). At a minimum, the current *Prototheca* species within the AP and HP lineages should be split into different genera. Additionally, as both sub-lineages contain pathogenic species, comparisons between pathogenic and non-pathogenic species of *Prototheca* should probably compare species within each sub-lineage.

The AP Sub-Lineage

Most phylogenies investigating the AHP lineage only include one species of *Auxenochlorella*: *A. protothecoides*. In previous chapters, *cytb* and whole genomes were sequenced for two additional species. The sequence of *cytb* was noted to not contain significant variation between three of the four species considered *Auxenochlorella* in this chapter. This, along with organelle structures and phylogenies from other genes provide confidence that *Auxenochlorella* is monophyletic, unlike *Prototheca*.

The exception to this would be the strain FACHB-9, which was identified as *Chlorella pyrenoidosa* and considered synonymous with *Auxenochlorella pyrenoidosa* in this chapter. However, the organelle genomes are not structurally similar to other *Auxenochlorella* species, nor does it group with *Auxenochlorella* in any phylogeny. Thus, this strain likely actually belongs to the *Chlorella* genus.

The species that are currently classified as *Prototheca* in the AP sub-lineage form at least two distinct lineages within the AP sub-lineage. All evidence suggests that *P. xanthoriae* is closely related to *Auxenochlorella*, forming a clade within the AP sub-lineage. Due to a lack

of organelle genomes from *P. paracutis* and *P. miyajii*, the relationships between these strains and *P. cutis* could not be investigated. However, the clade containing these species has been well supported by published work on *cytb* and rDNA sequences.

The relationships between *P. Wickerhamii* and the other taxa within the AP sub-lineage remain unclear. There is equal support for it being included within the clades containing *P. cutis* or *Auxenochlorella*. However, *cytb* data suggests *P. wickerhamii* may form its own clade with *P. fontanea*, for which organelle genomes are not yet available.

If these clades were to form the boundaries of new genera, this sets an upper limit of four genera within the AP sub-lineage, assuming photosynthesis is considered an essential characteristic of *Auxenochlorella.* The relationships, as presented in this chapter also suggest that photosynthesis has been lost in *Prototheca* four times: in the HP sub-lineage, in the *P. cutis* clade, in the *P. wickerhamii* clade (assuming it is not within the *P. cutis* clade), and in *P. xanthoriae*. There is currently no obvious pattern of acquisition or loss of pathology.

The HP Sub-Lineage

The analysis of the HP sub-lineage in this chapter is limited by the amount of *Helicosporidium* sequence data available. *Helicosporidium* organelle genomes have some structural similarity to those of *P. bovis* and *P. stagnora*, but large structural revisions have clearly taken place. It is therefore possible that *Helicosporidium* is basal to the AHP lineage, at which point the lineage that contains *P. bovis* and *P. stagnora* could be referred to as

"true" *Prototheca*, as this lineage contains both of the original species that were identified as *Prototheca* (Kruger, 1894).

It Is also not possible to comment on whether *Helicosporidium* is monophyletic, as only one sequence was used. However, previous analysis based on the sequence of β-tubulin suggest that *Helicosporidium* is monophyletic (Mancera et al., 2012).

It is not strictly necessary to separate the *Prototheca* species within the HP sub-lineage, but there are two distinct clades that appear to form. One contains all species that were originally identified as *P. zopfii*, and the other contains the "true environmental" species.

There may be significant phenotypic differences between these clades. It has been noted that strains of *P. tumulicola* and one strain of *P. stagnora* did not require vitamins to be supplemented in their growth media (Nagatsuka et al., 2017). Additionally, phenotypic differences between these clades have been observed in earlier chapters of this thesis, particularly with regards to recognition by the mammalian immune system. It is possible that these differences are sufficient to warrant different genera. There is therefore an upper limit of three genera within the HP sub-lineage, making for an upper limit of seven genera in the overall AHP lineage, four of which contain confirmed pathogens.

8) Overall Discussion

Phenotypic differences within the genus

In this thesis I set out to investigate the basis of pathology in *Prototheca*. I have identified phenotypic differences across the genus in response to abiotic stress and in interactions with the mammalian immune system, as well as differences between lineages in extracting DNA and assembling genomes.

My investigation of many species within the lineage, rather than focusing on a single pairwise comparison, has provided some interesting insights. Notably, *P. wickerhamii* and *P. bovis* are very distinct, despite being considered the most pathogenic species within the genus. In fact, differences between all cattle-associated species and other members of the AHP lineage exist for several measures.

One of the most notable differences that has impacted my own research has been the fact that it is significantly more challenging to extract high-quality DNA from cattle-associated species than other *Prototheca*, leading to more fragmented and less complete genomes. I also demonstrated that cattle-associated species are much more readily phagocytosed, and are similarly tolerant to extremely acidic conditions. Other phenotypic differences between the lineages were noticed, including differences in biofilm formation and sinking speed (which may reflect native buoyancy) of algal cells, but were not explored further. These differences add support to the notion that the cattle-associated strains belong in a different genus to other *Prototheca* species, as has been proposed before (Todd et al., 2018).

Phylogenetic analyses based on organelle genomes suggest that currently known *Prototheca* species could be split into at least two new genera, or as many as seven.

The differences between cattle-associated species and other *Prototheca* species suggests structural differences between these two groups, particularly in their cell wall and plasma membrane. This could have implications for immunity. Different molecules may be present, leading to different interactions with the immune system. It may also indicate that pathogenic species employ different strategies during infection; *P. bovis* is consistently robust against stress, which may translate to surviving the extreme conditions imposed by an Immune response, unlike *P. wickerhamii*, which may rely more on immune evasion.

Differences in stress responses and immune responses could have implications for drug susceptibility or risk of dissemination, though these have not been explored here nor are they readily apparent from clinical literature. However, these differences should encourage future researchers in the field to not extrapolate features for the whole genus from studies of a single species.

There was no indication of significant diversity in abiotic stress tolerance between strains of the same species. For logistical reasons, it was not possible to test for intra-specific diversity in phagocytic rates for most species. We previously noticed intra-specific differences in phagocytosis rate, but not phagocytic index, in *P. bovis* (Haider et al., 2023). There is therefore currently no indication that there are infectious sub-lineages within species.

One weakness of this work is that traits were measured without the overall context of an infection. I cannot, therefore, determine whether a trait is necessary for infection, only that a trait is present or absent in pathogenic or non-pathogenic species. Experimental infection will be necessary to test some of these hypotheses. However, host species may be a relevant factor to be aware of during experimental infection experiments. Variation in phagocytic behaviour was observed between cells from different host species in Chapter 5) Phagocytosis of *Prototheca* and Related Organisms, and differences between human, cattle and mouse immune responses have been noted in the past (Shahid et al., 2020a; Bisutti et al., 2023; Pegolo et al., 2022; Shahid et al., 2020b).

Are *Prototheca* accidental pathogens?

The overall hypothesis of this thesis was that *Prototheca* species were accidental pathogens. If so, then it was predicted that: pathogenic *Prototheca* species could be separated from non-pathogenic species according to traits that could evolve in the environment, and that responses that would be considered indicative of specialised adaptation to pathogenic lifestyles would be present in non-pathogenic species.

This threshold has been met. Pathogenic *Prototheca* species are more tolerant of physiological temperatures than non-pathogenic species, and a range of phagocytic behaviours (negligible phagocytosis to abundant phagocytosis) were observed in both pathogenic and non-pathogenic species. *Prototheca* species therefore could be accidental pathogens.

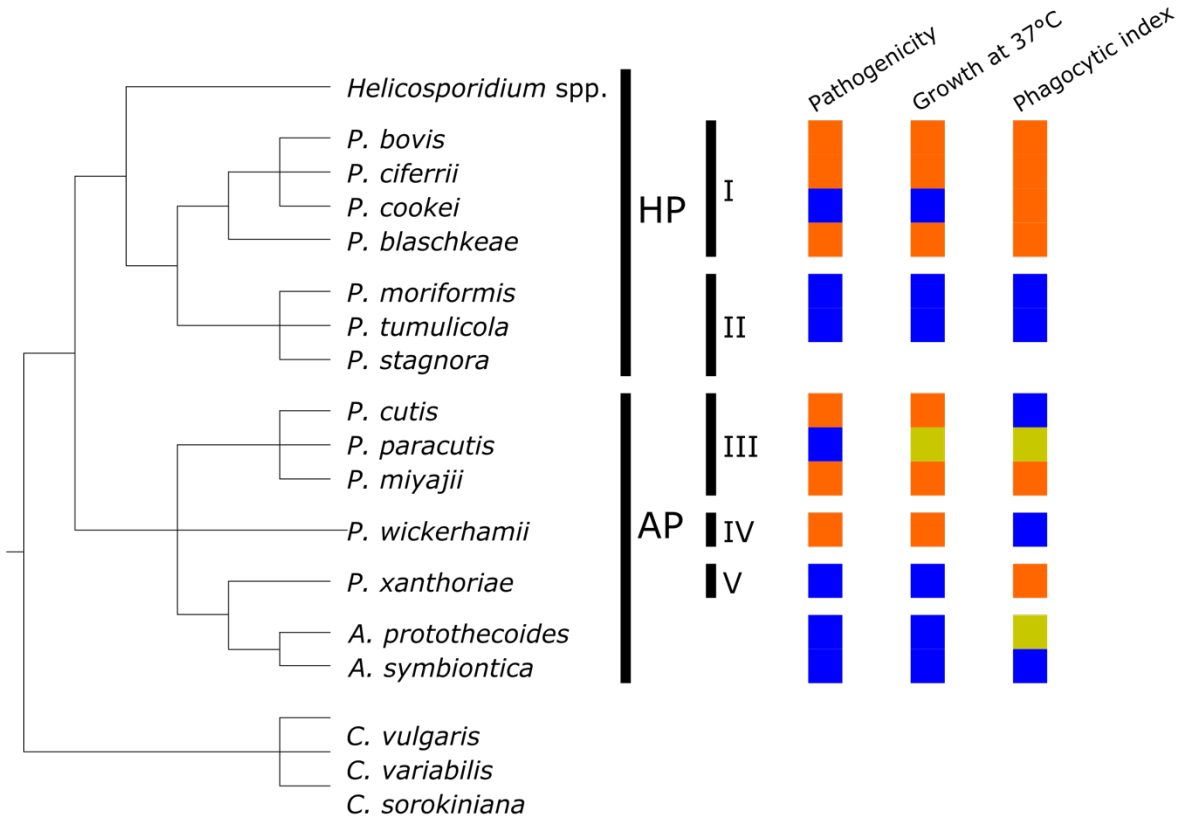


Figure 8-1. Summary of evidence for accidental pathology in Prototheca *species. A cladogram presents the higher-level relationships, derived from six organelle genes, between the species for which phenotyping data was collected. These relationships are broken into the two suggested main sub-lineages for the AHP lineage. The five possible new genus boundaries are indicated with roman numerals. Phenotype data are presented as coloured boxes: orange represents the presence of a qualitative attribute or a high level of a quantitative attribute; blue represents the absence of a qualitative attribute or a low level of a quantitative attribute; yellow represents ambiguity in the presence of a qualitative attribute or a middling level of a quantitative attribute. The distribution of the ability to grow at 37°C matches the distribution of pathogenicity within* Prototheca *and* Auxenochlorella*, fulfilling the first requirement to demonstrate accidental pathology. The distribution of phagocytic index, which could be interpreted as indicative of specific adaptation against immune stress, does not match the distribution of pathogenicity within* Prototheca *and* Auxenochlorella*, fulfilling the second requirement to demonstrate accidental pathology.*

However, the evidence presented in this thesis is not sufficient to demonstrate that *Prototheca* species are accidental pathogens. The following suggested work will form the basis of a strong argument for accidental pathology in *Prototheca.*

Requirement: Mechanisms

To make a strong argument for accidental pathology, broad phenotyping will be insufficient. It will be necessary, particularly for tolerance to immune stresses, to identify the mechanisms of resistance in pathogens, to determine whether they are present in non-

pathogenic species. Care should be taken when investigating these mechanisms to compare closely related species – investigating the basis for the high phagocytic index of *P. bovis* and *P. cookei* is more likely to be fruitful if *P. moriformis* is used as a comparison than if *P. wickerhamii* or *P. cutis* were used (which should in turn be compared to *P. miyajii*, *P. paracutis*, or *P. xanthoriae*).

Factors which separate pathogenic species from non-pathogenic, however, have no requirement to share a mechanism. The mechanism of the thermotolerance of *P. bovis* and *P. miyajii* being different would not be problematic for an accidental origin of pathology. In this case, it is the phenotype rather than the mechanism that is important. Care should still be taken to compare closely related species – differences in thermotolerance mechanisms between *P. bovis* and *P. xanthoriae* are likely to be starkest, but differences between *P. bovis* and *P. cookei* are more likely to be relevant for pathogenicity.

Most pathogenic fungi are opportunistic eukaryotic pathogens, much like *Prototheca*. Fungal opportunistic pathogens often possess "virulence traits" – traits that confer advantages in non-host environments that also confer advantages when infecting a host – rather than classical virulence factors (Cramer and Perfect, 2009). There is little reason to believe that this will not also be true of *Prototheca*. It may therefore be difficult to knock out genes to show that they are necessary for survival. Investigation of these mechanisms may use advanced imaging or comparison of strains with different phenotypes through "-omics" technologies (Shui et al., 2015; Barber et al., 2021; Gerstein et al., 2019). If there are virulence factors that can be knocked out without compromising the overall growth of the organism, more like classical virulence factors, transposon sequencing methods like TraDIS

(Langridge et al., 2009) or SATAY (Michel et al., 2017) may be beneficial in identifying them. Gene knockouts or inhibition of signalling in immune cells may be useful in identifying how they recognise and attempt to control *Prototheca* (Haider et al., 2023).

When investigating the mechanisms of attributes that appear relevant for pathology, it may be important to consider that different pathogenic species may have developed different mechanisms to underly the same phenotype. It would not be unexpected for the basis for thermotolerance in *P. bovis* to be different from the basis for thermotolerance in *P. miyajii*, as these two species are distantly related despite both surviving 42°C.

Requirement: Additional Conditions

Additionally, it is unlikely that the conditions explored here are the only ones that will determine pathogenicity for *Prototheca* species. Some *Chlorella* species can also grow at 37°C, but it remains unclear whether they are pathogenic (Hovde et al., 2018; Kim et al., 2019). Therefore, the range of factors investigated needs to be expanded. This should include additional stresses, of the sort explored here.

Additional factors, including tolerance of ROS, RNS (Erwig and Gow, 2016), and low oxygen stress (Zenewicz, 2017) can be explored using similar methods to the approach described here, to determine whether growth or survival of them separates pathogenic from non-pathogenic species. Additionally, it may be that susceptibility to these stresses changes when they are combined versus when they are experienced in isolation. Finally, the ability to survive in the environments of appropriate hosts may also be relevant – though this

appears unlikely to play a significant role with the disinfectants explored by other colleagues within the lab (Zhang, 2022; Shu, 2023) – as well as their environmental abundance.

It would also be of Interest to me”sure’tolerance of additional components of the immune response, including susceptibility to complement and polymorphonuclear (PMN) cells, to determine whether non-pathogenic and pathogenic species are similarly tolerant. It may also include resistance to specific enzymes, like lysozyme. It will also be important to know what cells are important for clearance, how they are recruited, and how different *Prototheca* species are recognised.

Characteristics other than survival under stress should also be explored. The growth phase of algal cultures affects the infectivity of *Chlorella* cells, as *Chlorella* cultures in logarithmic phase infect fewer *Paramecium* than *Chlorella* in stationary phase (Kodama and Fujishima, 2016). As age appears to impact the phagocytic index of HA6, there is reason to suspect this is relevant for the AHP lineage as well. The composition and thickness of the *Chlorella* cell wall changes depending on its growth conditions (Safi et al., 2014), which may alter how *Prototheca* interact with immunity if the same is true of AHP species. Additionally, sporopollenin has been considered important for *Prototheca* surviving immunity (Asfour and El-Metwally, 2010), but not all species possess sporopollenin (Atkinson et al., 1972).

Bioinformatics should also be used to supplement phenotyping; GWAS combined with *in vitro* and clinical phenotyping has been used to identify virulence traits in fungi (Gerstein et al., 2019), though care will need to be taken to ensure that appropriate comparisons are being made. Additionally, there may still be challenges in getting high enough quality

genomes from cattle-associated strains, given the difficulty they posed in Chapter 6) Genome Sequencing and Assembly of *Prototheca* (and Related Species). Reference-guided assembly may facilitate sequencing once high quality references become available for these species, assuming there is limited structural diversity in *Prototheca* genomes (Lischer and Shimizu, 2017).

When investigating new stresses, it is worth recalling that *P. wickerhamii*, despite being one of the more pathogenic species, is slightly inhibited by exposure to 37°C, unlike *P. bovis*. *P. miyajii* has a similar tolerance to high temperatures as *P. bovis*, as both can grow at 42°C, but causes significantly fewer infections. Thus, some traits may contribute quantitatively to pathogenicity, but others may contribute qualitatively. This is most likely when the trait allows overcoming of some barrier, like surviving physiological temperatures (Casadevall, pers. comm. May).

Primed for Pathology

Though the scope of this work is broader than many others looking for virulence traits in *Prototheca*, by including several non-pathogenic species of *Prototheca* as well as some strains of *Auxenochlorella*, there may still be key attributes that are outside the scope of this work. Pathology appears to have arisen multiple times within the lineage, and has been reported in a few cases for *Chlorella* (Hart et al., 2014; Jones et al., 1983), suggesting the whole AHP lineage likely has other traits that predispose, or at least facilitate, the development of pathology.

One such possible capacity is mixotrophy that allows photosynthetic organisms to grow exclusively with external sources of carbon, which would be particularly necessary for growth within an opaque mammalian host. As an obligate heterotroph, mechanisms to grow on external carbon sources must necessarily exist within all *Prototheca* species, and they appear present in most *Auxenochlorella* and at least some *Chlorella* species (Kim et al., 2019). While the concept of a true obligate autotroph appears controversial, there are at least some algae that are not capable of growing without light (Anderson et al., 2018).

Similarly, mammalian hosts sequester the essential minerals iron, manganese and zinc to inhibit the growth of pathogens, as part of nutritional immunity (Kehl-Fie et al., 2010). This presents a barrier to many pathogens, and must be overcome by countermeasures like siderophores (Citiulo et al., 2012; Becker and Skaar, 2014). *Auxenochlorella* and *Prototheca* are able to grow on PIM, a minimal isolation medium that contains few transition metals. In particular, it contains no iron, manganese, or zinc (Pore, 1973), unlike some other algal growth media which explicitly include these as trace elements (UTEX, n.d.; CCAP, n.d.). The adaptations that allow *Auxenochlorella* and *Prototheca* to grow on PIM may also allow them to bypass this aspect of nutritional immunity.

Host Contribution

It is also worth considering that pathology is a result of an interaction between host and microbe (Pirofski and Casadevall, 2012; Casadevall and Pirofski, 2001). It is possible that some differences in pathology are due to differences in hosts.

In humans, immunodeficient individuals are more susceptible to *Prototheca* infections. Similar deficiencies could be present in individuals of other host species, who would otherwise not be susceptible to *Prototheca* infections. However, this could also be on the level of whole host species. *Prototheca* infections are associated with recruitment and proliferation of lymphocytes rather than PMN or macrophages in cattle, which drive chronic inflammation (Bisutti et al., 2023; Pegolo et al., 2022). However, in mouse models, it appears that macrophages and PMN cells are recruited (Shahid et al., 2020b). If PMN cells are required, but lymphocytes are insufficient, for control and clearance of *Prototheca* then this difference in leukocyte recruitment could account for why cattle are the main host for *Prototheca* infections and there are no known mouse cases of infection. This may also account for the differences in presentation of infections between human, cattle, and canine (where infections are much more severe, usually resulting in death (Stenner et al., 2007)) hosts.

Spectrum of Pathology

It is possible that, while pathology was accidental in origin, not all *Prototheca* species remain accidental pathogens.

Regardless of the mechanism of entry, *Prototheca* species have a clear exit mechanism from cases of mastitis, as *Prototheca* cells are present in the milk. *Prototheca* infections in cattle therefore have a complete infection cycle. Cattle also appear able to infect each other (Todd et al., 2018; Milanov et al., 2016).

Due to the existence of a complete infection cycle, cattle present the best-known environment for *Prototheca* to evolve host-adapted virulence factors. As *P. bovis* is responsible for most cases of mastitis, we would expect *P. bovis* to be the most likely species to possess these host-adapted virulence factors. *P. bovis* has been reported to kill or invade phagocytes, which could be the product of these virulence factors (Shahid et al., 2020b). Thus, *P. bovis* may be an opportunistic rather than accidental pathogen.

Does accidental pathology need reworking?

Accidental pathology requires traits that evolved in the environment, i.e. not in the context of immunity, being sufficient to allow a pathogen to infect a host, by chance. This presents two possible areas of ambiguity: what distinguishes an environmental from an immune context, and how do we determine that the origin of the traits was truly by chance?

Some authors explicitly suggest that the relevant immune context for accidental pathology is mammalian (May et al., 2016) but most do not, instead providing examples like protist predation or global temperature rise that are unambiguously non-immune (Casadevall, 2008; Casadevall et al., 2019). However, there are contexts that are "ambiguously immune", which may be relevant for accidental pathology. Vertebrates generally share the same overall immune system principles, so it is unlikely that there will be significant differences between mammalian immunity and avian immunity (Sharma, 1991); thus a pathogen adapted to avian hosts would not be an accidental pathogen. Additionally, insects and nematodes are used as a model system for mammalian immunity, because they can be infected by human pathogens and elements of their immune systems are analogous to ours (Marsh and May, 2012; Ten et al., 2023; Murphy et al., 2012a); invertebrates may therefore

provide relevant selection pressures to acquire pathogenicity in mammals. Even more distantly, several microbial genera include pathogens of both plants and humans (*Aspergillus* (Pawar et al., 2008; Dolezal et al., 2014); *Fusarium* (Arie, 2019; Dignani and Anaissie, 2004); *Pseudomonas* (Silby et al., 2011)), and plant pathogens are known to occasionally infect humans (Dutta and Ray, 2023; Kim et al., 2020), suggesting pathogenicity in plants may confer benefits to infecting mammals. There is even some scope for the evolution of traits that would appear to counter mammalian adaptive immunity, depending on the mechanisms of the memory-like responses that have been observed in insects (Cooper and Eleftherianos, 2017).

A solution to this ambiguity arises through the removal of abstraction – some aspects of insect or even plant immunity will resemble mammalian immunity and others will not. Thus, the details and mechanisms of traits should determine what is and is not a relevant immune context. If a trait is required for mammalian infection by a plant pathogen but plays no role in plant pathology, then it should not be considered to have emerged in an immune context. For example, *Agrobacterium* and *Burkholderia* use different mechanisms to cause disease in human and plant hosts, which might allow the human-relevant mechanisms to be considered accidental, while *Pseudomonas* uses the same toxins mouse and plant infections, which should not be considered accidental (Kim et al., 2020).

Relatedly, it may also be challenging to identify whether a virulence trait originated “by chance” (i.e. as a result of non-immune selection pressures). In this thesis, I have attempted to address this by investigating non-pathogenic relatives of pathogenic species, but this

approach may be confounded by ancestral pathogens and subsequent loss of pathology in modern non-pathogenic species.

One possible solution to an uncertain origin of virulence traits is in the observation made in the introduction in the definition of an accidental pathogen, that an accidental pathogen may derive no fitness benefit from infection. Under this framework, accidental pathology could be interpreted as “currently accidental” rather than “of an accidental origin”, though this would not account for recent associations between a pathogen and a new host. If a pathogen derives fitness benefits from infection, then it will develop traits to enhance traits to facilitate infection, becoming an opportunistic pathogen rather than an accidental pathogen.

It Is worth noting that the fitness benefit of Infection may Iie dispersal Into new locations, rather than just increasing in abundance in one location or transmitting to new hosts. Additionally, it is worth considering that host death might be a valid escape mechanism, especially for microorganisms that are normally saprophytic and tolerant of anoxic conditions.

This framing then provides a third axis of pathology, in addition to the two described in the overall introduction. The requirement for a microorganism to infect a host as part of its life cycle determines whether it is an obligate or facultative pathogen; the ability of a microorganism to cause infection in a healthy host determines whether it is a primary or secondary pathogen; and whether the microorganism experiences fitness benefits from infection determines whether it is an accidental or adapted pathogen. By necessity, there

would be no obligate accidental pathogens, but all other combinations could exist, though some would be much less stable (and therefore less common) than others.

Virulence traits should also be identified as functional outside of an immune context. An absence of benefit to infection, and a plausible fitness benefit in an environmental context for all traits sufficient for pathogenicity would be sufficient to conclusively identify that a pathogen is accidentally pathogenic.

9) List of References

Abdelhameed, K.G. (2016) Detection of *Prototheca zopfii* in Raw Milk and Cheese with Special Reference to Their Antibiogram. *Journal of Food Safety*, 36 (2): 214–219. doi:10.1111/jfs.12233.

Aboal, M. and Werner, O. (2011) Morphology, fine structure, life cycle and phylogenetic analysis of *Phyllosiphon arisari*, a siphonous parasitic green alga. *European Journal of Phycology*, 46 (3): 181–192. doi:10.1080/09670262.2011.590902.

Abu Ali, B.M., Alzayer, H., Alwazzeh, M.J., et al. (2021) *Prototheca zopfii* as a Cause of Pneumonia and Disseminated Infection in Febrile Neutropenia: A Case Report and Literature Review. *American Journal of Case Reports*, 22: e933694. doi:10.12659/AJCR.933694.

Adiba, S., Nizak, C., van Baalen, M., et al. (2010) From Grazing Resistance to Pathogenesis: The Coincidental Evolution of Virulence Factors. *PLoS ONE*, 5 (8): e11882. doi:10.1371/journal.pone.0011882.

Ahn, A., Choe, Y.J., Chang, J., et al. (2017) Chronic Eosinophilic Meningoencephalitis by *Prototheca wickerhamii* in an Immunocompetent Boy. *Pediatric Infectious Disease Journal*, 36 (7): 687–689. doi:10.1097/INF.0000000000001552.

Ahrholdt, J., Murugaiyan, J., Straubinger, R.K., et al. (2012) Epidemiological analysis of worldwide bovine, canine and human clinical *Prototheca* isolates by PCR genotyping and MALDI-TOF mass spectrometry proteomic phenotyping. *Medical Mycology*, 50 (3): 234–243. doi:10.3109/13693786.2011.597445.

Altschul, S.F., Gish, W., Miller, W., et al. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3): 403–410. doi:10.1016/S0022-2836(05)80360-2.

Altschul, S.F., Madden, T.L., Schäffer, A.A., et al. (1997) Gapped BLAST and PSI-BLAST: a new

generation of protein database search programs. *Nucleic Acids Research*, 25 (17): 3389–3402. doi:10.1093/NAR/25.17.3389.

Anderson, R., Charvet, S. and Hansen, P.J. (2018) Mixotrophy in Chlorophytes and Haptophytes—Effect of Irradiance, Macronutrient, Micronutrient and Vitamin Limitation. *Frontiers in Microbiology*, 9: 1704. doi:10.3389/FMICB.2018.01704/BIBTEX.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data.* Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

dos Anjos, C., Sellera, F.P., Gargano, R.G., et al. (2019) Algicidal effect of blue light on pathogenic *Prototheca* species. *Photodiagnosis and Photodynamic Therapy*, 26 (February): 210–213. doi:10.1016/j.pdpdt.2019.04.009.

Arie, T. (2019) Fusarium diseases of cultivated plants, control, diagnosis, and molecular and genetic studies. *Journal of Pesticide Science*, 44 (4): 275–281. doi:10.1584/JPESTICS.J19-03.

Arnold, P. and Ahearn, D.G. (1972) The Systematics of the Genus *Prototheca* with a Description of a New Species *P. filamenta*. *Mycologia*, 64 (2): 265–275. doi:10.1080/00275514.1972.12019261.

Arriola, M.B., Velmurugan, N., Zhang, Y., et al. (2018) Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga. *The Plant Journal*, 93 (3): 566–586. doi:10.1111/tpj.13789.

Asfour, H.A.E. and El-Metwally, A.E. (2010) Microbiological and histopathological investigations on *Prototheca* mastitis in dairy animals. *Global Veterinaria*, 4 (4): 322–330. Available at: http://www.idosi.org/gv/gv4(4)10/2.pdf.

Ashford, B.K., Ciferri, R. and Dalmau, L.M. (1930) A new species of *Prototheca* and a variety of the same isolated from the human intestine. *Arch Protistenk*, 70: 619–938.

Asiag, N., Lapid, R., Aizenberg, Z., et al. (2022) Spinal cord protothecosis causing paraparesis

in a dog. *Journal of veterinary diagnostic investigation*, 34 (4): 684–688. doi:10.1177/10406387221093048.

Atkinson, A.W., Gunning, B.E.S. and John, P.C.L. (1972) Sporopollenin in the cell wall of *Chlorella* and other algae: Ultrastructure, chemistry, and incorporation of 14C-acetate, studied in synchronous cultures. *Planta*, 107 (1): 1–32. doi:10.1007/BF00398011.

Azaman, S.N.A., Wong, D.C.J., Tan, S.W., et al. (2020) De novo transcriptome analysis of *Chlorella sorokiniana*: effect of glucose assimilation, and moderate light intensity. *Scientific Reports*, 10: 17331. doi:10.1038/s41598-020-74410-4.

Azua-Bustos, A., González-Silva, C., Arenas-Fajardo, C., et al. (2012) Extreme environments as potential drivers of convergent evolution by exaptation: The Atacama Desert Coastal Range case. *Frontiers in Microbiology*, 3: 426. doi:10.3389/FMICB.2012.00426/BIBTEX.

Bacova, R., Kralik, P., Kucharovicova, I., et al. (2021) A novel TaqMan qPCR assay for rapid detection and quantification of pro-inflammatory microalgae *Prototheca* spp. in milk samples. *Medical Mycology*, 59 (8): 784–792. doi:10.1093/mmy/myaa120.

Baker, D.M., Freeman, C.J., Wong, J.C.Y., et al. (2018) Climate change promotes parasitism in a coral symbiosis. *The ISME Journal 2018 12:3*, 12 (3): 921–930. doi:10.1038/s41396-018-0046-8.

Baker, G.C., Beebee, T.J.C. and Ragan, M.A. (1999) *Prototheca richardsi,* a pathogen of anuran larvae, is related to a clade of protistan parasites near the animal-fungal divergence. *Microbiology*, 145: 1777–1784. doi:10.1099/13500872-145-7-1777.

Bakuła, Z., Gromadka, R., Gawor, J., et al. (2020) Sequencing and Analysis of the Complete Organellar Genomes of *Prototheca wickerhamii*. *Frontiers in Plant Science*, 11: 1296. doi:10.3389/fpls.2020.01296.

Bakuła, Z., Siedlecki, P., Gromadka, R., et al. (2021) A first insight into the genome of

Prototheca wickerhamii a major causative agent of human protothecosis. *BMC Genomics*, 22: 168. doi:10.1186/s12864-021-07491-8.

Balloux, F. and van Dorp, L. (2017) Q&A: What are pathogens, and what have they done to and for us? *BMC Biology*, 15: 91. doi:10.1186/s12915-017-0433-z.

Bankevich, A., Nurk, S., Antipov, D., et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19 (5): 455–477. doi:10.1089/cmb.2012.0021.

Barber, A.E., Sae-Ong, T., Kang, K., et al. (2021) *Aspergillus fumigatus* pan-genome analysis identifies genetic variants associated with human infection. *Nature Microbiology*, 6 (12): 1526–1536. doi:10.1038/s41564-021-00993-x.

Bauman, C.A., Barkema, H.W., Dubuc, J., et al. (2018) Canadian National Dairy Study: Herd-level milk quality. *Journal of Dairy Science*, 101 (3): 2679–2691. doi:10.3168/jds.2017-13336.

Baurain, D., Brinkmann, H., Petersen, J., et al. (2010) Phylogenomic Evidence for Separate Acquisition of Plastids in Cryptophytes, Haptophytes, and Stramenopiles. *Molecular Biology and Evolution*, 27 (7): 1698–1709. doi:10.1093/MOLBEV/MSQ059.

Beale, M.A., Sabiiti, W., Robertson, E.J., et al. (2015) Genotypic Diversity Is Associated with Clinical Outcome and Phenotype in Cryptococcal Meningitis across Southern Africa. *PLoS Neglected Tropical Diseases*, 9 (6): e0003847. doi:10.1371/JOURNAL.PNTD.0003847.

Becker, K.W. and Skaar, E.P. (2014) Metal limitation and toxicity at the interface between host and pathogen. *FEMS Microbiology Reviews*, 38 (6): 1235–1249. doi:10.1111/1574-6976.12087.

van Bezooijen, B.P.J. and Newling, D.W.W. (2002) Protothecosis of the Urinary Tract. *Journal of Urology*, 167 (1): 252–252. doi:10.1016/S0022-5347(05)65431-6.

Bhabhra, R. and Askew, D.S. (2005) Thermotolerance and virulence of *Aspergillus fumigatus*:

Role of the fungal nucleolus. *Medical Mycology*. 43 (SUPPL.1). doi:10.1080/13693780400029486.

Bisutti, V., Mach, N., Giannuzzi, D., et al. (2023) Transcriptome-wide mapping of milk somatic cells upon subclinical mastitis infection in dairy cattle. *Journal of animal science and biotechnology*, 14 (1): 93. doi:10.1186/S40104-023-00890-9.

Blanc, G., Duncan, G., Agarkova, I., et al. (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*, 22 (9): 2943–2955. doi:10.1105/tpc.110.076406.

Bliven, K.A. and Maurelli, A.T. (2012) Antivirulence genes: Insights into pathogen evolution through gene loss. *Infection and Immunity*, 80 (12): 4061–4070. doi:10.1128/IAI.00740-12.

Boamah, D.K., Zhou, G., Ensminger, A.W., et al. (2017) From Many Hosts, One Accidental Pathogen: The Diverse Protozoan Hosts of *Legionella*. *Frontiers in Cellular and Infection Microbiology*, 7 (NOV): 477. doi:10.3389/FCIMB.2017.00477.

Bock, C., Krienitz, L. and Pröschold, T. (2011) Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea*, 11 (2): 293–312. doi:10.5507/fot.2011.028.

Borza, T., Popescu, C.E. and Lee, R.W. (2005) Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryotic Cell*, 4 (2): 253–261. doi:10.1128/EC.4.2.253-261.2005.

Boyd, A.S., Langley, M. and King, L.E. (1995) Cutaneous manifestations of *Prototheca* infections. *Journal of the American Academy of Dermatology*, 32 (5 PART 1): 758–764. doi:10.1016/0190-9622(95)91456-0.

Bozzo, G., Bonerba, E., Di Pinto, A., et al. (2014) Occurrence of *Prototheca* spp. in cow milk samples. *New Microbiologica*, 37 (4): 459–464.

Brosius, J. (2019) Exaptation at the molecular genetic level. *Science China. Life sciences*, 62 (4): 437–452. doi:10.1007/S11427-018-9447-8.

Brown, N.F., Wickham, M.E., Coombes, B.K., et al. (2006) Crossing the Line: Selection and Evolution of Virulence Traits. *PLOS Pathogens*, 2 (5): e42. doi:10.1371/JOURNAL.PPAT.0020042.

Brown, S.P., Cornforth, D.M. and Mideo, N. (2012) Evolution of virulence in opportunistic pathogens: Generalism, plasticity, and control. *Trends in Microbiology*, 20 (7): 336–342. doi:10.1016/j.tim.2012.04.005.

Buendía, A., Patiño, E., Rijlaarsdam, M., et al. (1998) [Endocarditis due to an alga of the genus Prototheca sp. A saprophyte of water and of tree sap?]. *Archivos del Instituto de Cardiologia de Mexico*, 68 (4): 333–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9810371 (Accessed: 22 May 2020).

Burki, F. (2014) The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harbor Perspectives in Biology*, 6 (5). doi:10.1101/CSHPERSPECT.A016147.

Burki, F., Okamoto, N., Pombert, J.F., et al. (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proceedings of the Royal Society B: Biological Sciences*, 279 (1736): 2246–2254. doi:10.1098/RSPB.2011.2301.

Caisová and Lenka (2011) Molecular phylogeny and taxonomic revision of chaetophoralean algae (Chlorophyta). *Theses.Cz*. Available at: http://theses.cz/id/7phqgv/Caisova_PhD_Thesis_2011_shortened_version.pdf%5Cnpapers2://publication/uuid/495F4A2A-6CC6-47AD-A513-55DBB3ABD359.

Calvo-Pérez Rodó, J.D. and Molinari-Novoa, E.A. (2015) A Nomenclatural and Cultural Note on *Chlorella peruviana* G. Chacon and other species of the Genus *Chlorella* Beij. (Chlorellaceae, Chlorellales). *The Biologist*, 13 (1): 153–156.

Camacho, C., Coulouris, G., Avagyan, V., et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, 10: 421. doi:10.1186/1471-2105-10-421/FIGURES/4.

Camilli, G., Eren, E., Williams, D.L., et al. (2018) Impaired phagocytosis directs human monocyte activation in response to fungal derived β-glucan particles. *European journal of immunology*, 48 (5): 757. doi:10.1002/EJI.201747224.

Capra, E., Cremonesi, P., Cortimiglia, C., et al. (2014) Simultaneous identification by multiplex PCR of major *Prototheca* spp. isolated from bovine and buffalo intramammary infection and bulk tank. *Letters in Applied Microbiology*, 59 (6): 642–647. doi:10.1111/lam.12326.

Carey, W.P., Kaykova, Y., Bandres, J.C., et al. (1997) Cutaneous Protothecosis in a Patient with AIDS and a Severe Functional Neutrophil Defect: Successful Therapy with Amphotericin B. *Clinical Infectious Diseases*, 25 (5): 1265–1266. doi:10.1086/516974.

Casadevall, A. (2005) Fungal virulence, vertebrate endothermy, and dinosaur extinction: is there a connection? *Fungal Genetics and Biology*, 42 (2): 98–106. doi:10.1016/J.FGB.2004.11.008.

Casadevall, A. (2008) Evolution of intracellular pathogens. *Annual Review of Microbiology*, 62: 19–33. doi:10.1146/annurev.micro.61.080706.093305.

Casadevall, A., Kontoyiannis, D.P. and Robert, V. (2019) On the emergence of *Candida auris*: climate change, azoles, swamps, and birds. *mBio*, 10 (4). doi:10.1128/MBIO.01397-19.

Casadevall, A. and Pirofski, L. anne (2001) Host-Pathogen Interactions: The Attributes of Virulence. *The Journal of Infectious Diseases*, 184 (3): 337–344. doi:10.1086/322044.

Casadevall, A. and Pirofski, L.A. (2009) Virulence factors and their mechanisms of action: The view from a damage-response framework. *Journal of Water and Health*, 7 (SUPPL. 1): 2–18. doi:10.2166/wh.2009.036.

Casadevall, A. and Pirofski, L.A. (2014) Ditch the term pathogen. *Nature*. 516 (7530) pp. 165–166. doi:10.1038/516165a.

CCAP (n.d.) *Bold's Basal Medium (BB)*.

Cecchin, M., Marcolungo, L., Rossato, M., et al. (2019) *Chlorella vulgaris* genome assembly and annotation reveals the molecular basis for metabolic acclimation to high light conditions. *The Plant Journal*, 100 (6): 1289. doi:10.1111/TPJ.14508.

Chan, J.C., Jeffers, L.J., Gould, E.W., et al. (1990) Visceral Protothecosis Mimicking Sclerosing Cholangitis in an Immunocompetent Host: Succesful Antifungal Therapy. *Reviews of Infectious Diseases*, 12 (5): 802–807. doi:10.1093/CLINIDS/12.5.802.

Chanmugam, A., Langemo, D., Thomason, K., et al. (2017) Relative Temperature Maximum in Wound Infection and Inflammation as Compared with a Control Subject Using Long-Wave Infrared Thermography. *Advances in Skin and Wound Care*, 30 (9): 406–414. doi:10.1097/01.ASW.0000522161.13573.62.

Chen, F., Saab-chalhoub, M., Tao, J., et al. (2023) Disseminated Protothecosis Due to *Prototheca zopfii* and Literature Review. *American Journal of Dermatopathology*, 00 (00): 1–5.

Cheville, N.F., McDonald, J. and Richard, J. (1984) Ultrastructure of *Prototheca zopfii* in bovine granulomatous mastitis. *Veterinary pathology*, 21 (3): 341–348. doi:10.1177/030098588402100312.

Chodat, R. (1913) Monographies d'algues en culture pure. *Mat. Crypt. Suisse*, 4: 234–241.

Chrisman, C.J., Alvarez, M. and Casadevall, A. (2010) Phagocytosis of *Cryptococcus neoformans* by, and Nonlytic Exocytosis from, *Acanthamoeba castellanii*. *Applied and Environmental Microbiology*, 76 (18): 6056–6062. doi:10.1128/AEM.00812-10.

Citiulo, F., Jacobsen, I.D., Miramón, P., et al. (2012) *Candida albicans* Scavenges Host Zinc via

Pra1 during Endothelial Invasion. *PLOS Pathogens*, 8 (6): e1002777. doi:10.1371/JOURNAL.PPAT.1002777.

Cláudia Barros, A., Gonçalves, A.L. and Simões, M. (2020) "Chapter 6: Microalgal and cyanobacterial biofilms." In Simões, M., Borges, A. and Simões, L.C. (eds.) *Recent Trends in Biofilm Science and Technology*. pp. 127–156. doi:10.1016/B978-0-12-819497-3.00006-4.

Cock, J.M. and Coelho, S.M. (2011) Algal models in plant biology. *Journal of Experimental Botany*, 62 (8): 2425–2430. doi:10.1093/JXB/ERR117.

Conte, M. V. and Pore, R.S. (1973) Taxonomic implications of *Prototheca* and *Chlorella* cell wall polysaccharide characterization. *Archiv für Mikrobiologie*, 92 (3): 227–233. doi:10.1007/BF00411203.

Cooper, D. and Eleftherianos, I. (2017) Memory and specificity in the insect immune system: Current perspectives and future challenges. *Frontiers in Immunology*, 8: 539. doi:10.3389/FIMMU.2017.00539/BIBTEX.

Cordy, D.R. (1973) Chlorellosis in a Lamb. *Veterinary Pathology*, 10 (2): 171–176. doi:10.1177/030098587301000209.

Costa, E.O., Carciofi, A.C., Melville, P.A., et al. (1996) Prototheca sp. outbreak of bovine mastitis. *Journal of Veterinary Medicine, Series B*, 43 (6): 321–324. doi:10.1111/j.1439-0450.1996.tb00321.x.

Cox, G.E., Wilson, J.D. and Brown, P. (1974) Protothecosis: a Case of Disseminated Algal Infection. *The Lancet*, 304 (7877): 379–382. doi:10.1016/S0140-6736(74)91760-7.

Cramer, R.A. and Perfect, J.R. (2009) "Recent advances in understanding human opportunistic fungal pathogenesis mechanisms." In *Clinical Mycology*. 2nd ed. Elsevier Inc. pp. 15–31. doi:10.1016/B978-1-4160-5680-5.00002-5.

Crispens, C.J. and Marion, K. (1975) Algal infection in a corn snake (Elaphe guttata guttata).

Lab Anim Sci, 25 (6): 788–789.

Cullen, G.D., Yetmar, Z.A., Fida, M., et al. (2023) *Prototheca* Infection: A Descriptive Study. *Open Forum Infectious Diseases*, 10 (6). doi:10.1093/OFID/OFAD294.

Danecek, P., Bonfield, J.K., Liddle, J., et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, 10 (2): 1–4. doi:10.1093/GIGASCIENCE/GIAB008.

Darienko, T., Gustavs, L., Eggert, A., et al. (2015) Evaluating the species boundaries of green microalgae (Coccomyxa, Trebouxiophyceae, Chlorophyta) using integrative taxonomy and DNA barcoding with further implications for the species identification in environmental samples. *PLoS ONE*, 10 (6): e0127838. doi:10.1371/journal.pone.0127838.

Darienko, T., Gustavs, L. and Pröschold, T. (2016) Species concept and nomenclatural changes within the genera *Elliptochloris* and *Pseudochlorella* (Trebouxiophyceae) based on an integrative approach Verbruggen, H. (ed.). *Journal of Phycology*, 52 (6): 1125–1145. doi:10.1111/jpy.12481.

Darienko, T. and Pröschold, T. (2015) Genetic variability and taxonomic revision of the genus *Auxenochlorella* (Shihira et Krauss) Kalina et Puncocharova (Trebouxiophyceae, Chlorophyta) De Clerck, O. (ed.). *Journal of Phycology*, 51 (2): 394–400. doi:10.1111/jpy.12279.

Darienko, T. and Pröschold, T. (2019) The genus *Jaagichlorella* Reisigl (Trebouxiophyceae, Chlorophyta) and its close relatives: An evolutionary puzzle. *Phytotaxa*, 388 (1): 47–68. doi:10.11646/phytotaxa.388.1.2.

Darling, A.E., Mau, B. and Perna, N.T. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLOS ONE*, 5 (6): e11147. doi:10.1371/JOURNAL.PONE.0011147.

Davies, R.R., Spencer, H. and Wakelin, P.O. (1964) A case of human protothecosis.

Transactions of the Royal Society of Tropical Medicine and Hygiene, 58 (5): 448–450. doi:10.1016/0035-9203(64)90094-X.

Deiner, K., Walser, J.C., Mächler, E., et al. (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183: 53–63. doi:10.1016/J.BIOCON.2014.11.018.

Deng, Z., Shahid, M., Zhang, L., et al. (2016) An Investigation of the Innate Immune Response in Bovine Mammary Epithelial Cells Challenged by *Prototheca zopfii*. *Mycopathologia*, 181 (11–12): 823–832. doi:10.1007/s11046-016-0053-0.

Denton, J.F., Lugo-Martinez, J., Tucker, A.E., et al. (2014) Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies Guigo, R. (ed.). *PLoS Computational Biology*, 10 (12): e1003998. doi:10.1371/journal.pcbi.1003998.

Dignani, M.C. and Anaissie, E. (2004) Human fusariosis. *Clinical Microbiology and Infection*, 10 (Suppl. 1): 67–75. doi:10.1111/J.1470-9465.2004.00845.X.

DNATechnologiesCore (n.d.) *High Molecular Weight DNA Isolation (HMW-DNA) | DNA Technologies Core*. Available at: https://dnatech.genomecenter.ucdavis.edu/high-molecular-weight-dna-isolation-hmw-dna/ (Accessed: 17 May 2022).

Dolezal, A.L., Shu, X., OBrian, G.R., et al. (2014) *Aspergillus flavus* infection induces transcriptional and physical changes in developing maize kernels. *Frontiers in Microbiology*, 5 (JULY): 384. doi:10.3389/FMICB.2014.00384/ABSTRACT.

Drew, G.C., Stevens, E.J. and King, K.C. (2021) Microbial evolution and transitions along the parasite–mutualist continuum. *Nature Reviews Microbiology*, 19 (10): 623–638. doi:10.1038/s41579-021-00550-7.

Dutta, S. and Ray, U. (2023) Paratracheal abscess by plant fungus *Chondrostereum purpureum*- first case report of human infection. *Medical Mycology Case Reports*, 40: 30.

doi:10.1016/J.MMCR.2023.03.001.

Dziurzyński, M., Decewicz, P., Iskra, M., et al. (2021) Prototheca-ID: a web-based application for molecular identification of Prototheca species. *Database: The Journal of Biological Databases and Curation*, 2021 (2021): 1–4. doi:10.1093/DATABASE/BAAB073.

Edlund, A.F., Olsen, K., Mendoza, C., et al. (2017) Pollen wall degradation in the brassicaceae permits cell emergence after pollination. *American Journal of Botany*, 104 (8): 1266–1273. doi:10.3732/ajb.1700201.

El-Ani, A.S. (1967) Life cycle and variation of *Prototheca wickerhamii*. *Science*, 156 (3781): 1501–1503. doi:10.1126/science.156.3781.1501.

Eley, T. (2011) *National Animal Disease Information Service*. Available at: https://www.nadis.org.uk/disease-a-z/cattle/the-healthy-cow/ (Accessed: 6 September 2022).

Ely, V.L., Felipetto Cargnelutti, J., Segabinazzi Ries, A., et al. (2023) Prototheca bovis in goats: Experimental mastitis and treatment with polypyrrole. *Microbial Pathogenesis*, 174: 105950. doi:10.1016/J.MICPATH.2022.105950.

Ely, V.L., Gressler, L.T., Sutili, F.J., et al. (2019) Biofilm formation by Prototheca zopfii isolated from clinical and subclinical bovine mastitis in distinct growth conditions under different dyes. *Ciencia Rural*, 49 (2): e20180574. doi:10.1590/0103-8478cr20180574.

Ely, V.L., Pereira, D.I.B., Costa, M.M., et al. (2022) Activity of biogenic silver nanoparticles against isolates of Prototheca species from bovine mastitis. *Letters in Applied Microbiology*, 75 (1): 24–28. doi:10.1111/LAM.13696.

Erwig, L.P. and Gow, N.A.R. (2016) Interactions of fungal pathogens with phagocytes. *Nature Reviews Microbiology*, 14 (3): 163–176. doi:10.1038/nrmicro.2015.21.

Esteban, G.F., Fenchel, T. and Finlay, B.J. (2010) Mixotrophy in Ciliates. *Protist*, 161 (5): 621–

641. doi:10.1016/j.protis.2010.08.002.

Ewels, P., Magnusson, M., Lundin, S., et al. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32 (19): 3047–3048.

Falcaro, C., Furlanello, T., Binanti, D., et al. (2021) Molecular characterization of Prototheca in 11 symptomatic dogs. *Journal of Veterinary Diagnostic Investigation*, 33 (1): 156–161. doi:10.1177/1040638720976423.

Fan, J., Ning, K., Zeng, X., et al. (2015) Genomic Foundation of Starch-to-Lipid Switch in Oleaginous *Chlorella* spp. *Plant Physiology*, 169 (4): 2444. doi:10.1104/PP.15.01174.

Fernandes, L., Zhan, W., Patni, N.K., et al. (1994) Temperature distribution and variation in passively aerated static compost piles. *Bioresource Technology*, 48 (3): 257–263. doi:10.1016/0960-8524(94)90155-4.

Fernández, N.B., Taverna, C.G., Vivot, M., et al. (2019) First bloodstream infection due to *Prototheca zopfii* var. *hydrocarbonea* in an immunocompromised patient. *Medical Mycology Case Reports*, 24: 9–12. doi:10.1016/j.mmcr.2019.02.003.

Fidelis, C.E., Franke, M., de Abreu, L.C.R., et al. (2021) MALDI-TOF MS identification of *Prototheca* algae associated with bovine mastitis. *Journal of Veterinary Diagnostic Investigation*, 33 (6): 1168–1171. doi:10.1177/10406387211036459.

Fidelis, C.E., Leite, R. de F., Garcia, B.L.N., et al. (2022) Antimicrobial activities of polyhexamethylene biguanide against biofilm-producing *Prototheca bovis* causing bovine mastitis. *Journal of Dairy Science*, 0 (0). doi:10.3168/JDS.2022-22468.

Figueroa-Martinez, F., Nedelcu, A.M., Smith, D.R., et al. (2015) When the lights go out: The evolutionary fate of free-living colorless green algae. *New Phytologist*. 206 (3) pp. 972–982. doi:10.1111/nph.13279.

Figueroa-Martinez, F., Nedelcu, A.M., Smith, D.R., et al. (2017) The plastid genome of

Polytoma uvella is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to nonphotosynthetic lifestyles. *Plant Physiology*, 173 (2): 932–943. doi:10.1104/pp.16.01628.

Frese, V.K. and Gedek, B. (1968) Ein fall von protothecosis beim reh. *Berl. Munch. Tierarztl. Wochenschr*, 81: 171–178.

Fujimoto, M., Inaba, Y., Takahashi, T., et al. (2018) Image Gallery: Granulomatous dermatitis due to infection with the chlorophyllic green alga *Desmodesmus*. *British Journal of Dermatology*, 179 (4): e167. doi:10.1111/bjd.17007.

Galán, F., García-Martos, P., Palomo, M.J., et al. (1997) Onychoprototheсosis due to *Prototheca wickerhamii*. *Mycopathologia*, 137 (2): 75–77. doi:10.1023/A:1006893614334.

Gandham, N., Vyawahare, C., Chaudhaury, N., et al. (2015) Onychoprototheсosis: An uncommon presentation of protothecosis. *Indian Journal of Medical Microbiology*, 33 (3): 435–437. doi:10.4103/0255-0857.158583.

Gao, C., Wang, Y., Shen, Y., et al. (2014) Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics*, 15: 582. doi:10.1186/1471-2164-15-582.

Gao, J., Zhang, H. qi, He, J. zhong, et al. (2012) Characterization of *Prototheca zopfii* Associated with Outbreak of Bovine Clinical Mastitis in Herd of Beijing, China. *Mycopathologia*, 173 (4): 275–281. doi:10.1007/s11046-011-9510-y.

Garelnabi, M., Taylor-Smith, L.M., Bielska, E., et al. (2018) Quantifying donor-to-donor variation in macrophage responses to the human fungal pathogen *Cryptococcus neoformans*. *PLoS ONE*, 13 (3): e0194615. doi:10.1371/JOURNAL.PONE.0194615.

Gärtner, G. (1996) Soil Algae. *Methods in Soil Biology*, pp. 295–305. doi:10.1007/978-3-642-60966-4_19.

Gaur, S., Marrin, C. and Barnes, R.A. (2010) Disseminated protothecosis following traumatic Hickman line removal in a patient with leukaemia. *Medical Mycology*, 48 (2): 410–412. doi:10.3109/13693780903188698/2/TMMY_A_419042_O_F0002G.JPEG.

Gawryluk, R.M.R., Tikhonenkov, D. V., Hehenberger, E., et al. (2019) Non-photosynthetic predators are sister to red algae. *Nature*, 572 (7768): 240–243. doi:10.1038/S41586-019-1398-6.

Geisen, V., Mayer, C., Harrer, J., et al. (2020) [Disseminated protothecosis with ulcerative granulomatous colitis in a Rhodesian Ridgeback from Germany]. *Tierarztliche Praxis. Ausgabe K, Kleintiere/Heimtiere*, 48 (5): 369–375. doi:10.1055/A-1238-1554.

Gentles, J.C. and Bond, P.M. (1977) Protothecosis of atlantic salmon. *Medical Mycology*, 15 (2): 133–139. doi:10.1080/00362177785190211.

Gerken, H.G., Donohoe, B. and Knoshaug, E.P. (2013) Enzymatic cell wall degradation of *Chlorella vulgaris* and other microalgae for biofuels production. *Planta*, 237 (1): 239–253. doi:10.1007/s00425-012-1765-0.

Gerstein, A.C., Jackson, K.M., McDonald, T.R., et al. (2019) Identification of Pathogen Genomic Differences That Impact Human Immune Response and Disease during *Cryptococcus neoformans* Infection. *mBio*, 10 (4). doi:10.1128/MBIO.01440-19.

Gibb, A.P., Aggarwal, R. and Swainson, C.P. (1991) Successful treatment of *Prototheca* peritonitis complicating continuous ambulatory peritoneal dialysis. *Journal of Infection*, 22 (2): 183–185. doi:10.1016/0163-4453(91)91679-R.

Ginel, P.J., Pérez, J., Molleda, J.M., et al. (1997) Cutaneous protothecosis in a dog. *Veterinary Record*, 140 (25): 651–653. doi:10.1136/vr.140.25.651.

Godofredo, V.R., Enokihara, M.M.S. e. S., Tomimori, J., et al. (2020) Cutaneous protothecosis in kidney transplant recipient. *Anais Brasileiros de Dermatologia*, 95 (2): 210–213.

doi:10.1016/j.abd.2019.08.022.

Góes, L.D.M., Monteiro, V. da S. and Souza, A.T.O. de (2021) Case for diagnosis. Ulcer and papular lesions in a patient with diabetes mellitus. Protothecosis. *Anais Brasileiros de Dermatologia*, 96 (5): 613–616. doi:10.1016/J.ABD.2021.03.004.

Gregory, T.R. (2009) Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evolution: Education and Outreach*, 2 (2): 156–175. doi:10.1007/S12052-009-0128-1/FIGURES/3.

Griffiths, M.J., Garcin, C., van Hille, R.P., et al. (2011) Interference by pigment in the estimation of microalgal biomass concentration by optical density. *Journal of Microbiological Methods*, 85 (2): 119–123. doi:10.1016/j.mimet.2011.02.005.

Grzesiak, B., Głowacka, A., Krukowski, H., et al. (2016) The In Vitro Efficacy of Essential Oils and Antifungal Drugs Against *Prototheca zopfii*. *Mycopathologia*, 181 (7–8): 609–615. doi:10.1007/s11046-016-9994-6.

Grzesiak, B., Kołodziej, B., Głowacka, A., et al. (2018) The Effect of Some Natural Essential Oils Against Bovine Mastitis Caused by *Prototheca zopfii* Isolates *in vitro*. *Mycopathologia*, 183 (3). doi:10.1007/s11046-018-0246-9.

Guéguen, N., Le Moigne, D., Amato, A., et al. (2021) Lipid Droplets in Unicellular Photosynthetic Stramenopiles. *Frontiers in Plant Science*, 12: 639276. doi:10.3389/FPLS.2021.639276/BIBTEX.

Guiry, M.D. and Guiry, G.M. (2015) *AlgaeBase*.

Guo, J., Chen, J., Li, T., et al. (2023) Integration of transcriptomics, proteomics, and metabolomics data for the detection of the human pathogenic *Prototheca wickerhamii* from a One Health perspective. *Frontiers in cellular and infection microbiology*, 13: 1152198. doi:10.3389/FCIMB.2023.1152198/FULL.

Guo, J., Jian, J., Wang, L., et al. (2022) Genome Sequences of Two Strains of *Prototheca wickerhamii* Provide Insight Into the Protothecosis Evolution. *Frontiers in Cellular and Infection Microbiology*, 0: 28. doi:10.3389/FCIMB.2022.797017.

Gurevich, A., Saveliev, V., Vyahhi, N., et al. (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29 (8): 1072–1075. doi:10.1093/bioinformatics/btt086.

Hachicha, R., Elleuch, F., Hlima, H. Ben, et al. (2022) Biomolecules from Microalgae and Cyanobacteria: Applications and Market Survey. *Applied Sciences (Switzerland)*, 12 (4). doi:10.3390/app12041924.

Hadariová, L., Vesteg, M., Hampl, V., et al. (2018) Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Current Genetics*, 64: 365–387. doi:10.1007/s00294-017-0761-0.

Hafner, S., Brown, C.C. and Zhang, J. (2013) Green Algal Peritonitis in 2 Cows. *Veterinary Pathology*, 50 (2): 256–259. doi:10.1177/0300985812450722.

Haider, M.J.A., Shave, C.D., Onyishi, C.U., et al. (2023) Species- and strain-specific differences in the phagocytosis of *Prototheca*: insights from live-cell imaging Roy, C.R. (ed.). *Infection and Immunity*, p. e0006623. doi:10.1128/iai.00066-23.

Hamada, M., Schröder, K., Bathia, J., et al. (2018) Metabolic co-dependence drives the evolutionarily ancient *Hydra-Chlorella* symbiosis. *eLife*, 7. doi:10.7554/ELIFE.35122.

Hanschen, E.R. and Starkenburg, S.R. (2020) The state of algal genome quality and diversity. *Algal Research*, 50 (September): 101968. doi:10.1016/j.algal.2020.101968.

Hariprasad, S.M., Prasad, A., Smith, M., et al. (2005) Bilateral choroiditis from Prototheca *wickerhamii algaemia*. *Archives of Ophthalmology*, 123 (8): 1138–1141. doi:10.1001/archopht.123.8.1138.

Hart, J., Mooney, L., Arthur, I., et al. (2014) First case of *Chlorella* wound infection in a

human in Australia. *New Microbes and New Infections*, 2 (4): 132–133. doi:10.1002/nmi2.50.

Hatinguais, R., Willment, J.A. and Brown, G.D. (2020) PAMPs of the fungal cell wall and mammalian PRRs. *Current Topics in Microbiology and Immunology*, 425: 187–223. doi:10.1007/82_2020_201/FIGURES/2.

He, X., Dai, J. and Wu, Q. (2016) Identification of Sporopollenin as the Outer Layer of Cell Wall in Microalga *Chlorella protothecoides*. *Frontiers in Microbiology*, 7 (JUN): 1047. doi:10.3389/FMICB.2016.01047.

Heimann, K. and Huerlimann, R. (2015) Microalgal Classification: Major Classes and Genera of Commercial Microalgal Species. *Handbook of Marine Microalgae: Biotechnology Advances*, pp. 25–41. doi:10.1016/B978-0-12-800776-1.00003-0.

Hennigar, S.R. and McClung, J.P. (2016) Nutritional Immunity: Starving Pathogens of Trace Minerals. *American Journal of Lifestyle Medicine*. 10 (3) pp. 170–173. doi:10.1177/1559827616629117.

Herold, S., Klodt, T., Toelle, D., et al. (2021) Lethal systemic and brain infection caused by *Prototheca zopfii* algae in a patient with acute myeloid leukemia. *Medical Mycology Case Reports*, 32: 17–20. doi:10.1016/J.MMCR.2021.01.004.

Hirayama, D., Iida, T. and Nakase, H. (2018) The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis. *International Journal of Molecular Sciences*, 19 (1). doi:10.3390/IJMS19010092.

Hirose, N., Hua, Z., Kato, Y., et al. (2018) Molecular Characterization of *Prototheca* strains isolated in China revealed the first cases of protothecosis associated with *Prototheca zopfii* genotype 1. *Medical Mycology*, 56 (3): 279–287. doi:10.1093/mmy/myx039.

Hirose, N., Nishimura, K., Inoue-Sakamoto, M., et al. (2013) Ribosomal Internal Transcribed Spacer of *Prototheca wickerhamii* Has Characteristic Structure Useful for Identification and

Genotyping Robinson, D.A. (ed.). *PLoS ONE*, 8 (11): e81223. doi:10.1371/journal.pone.0081223.

Hofbauer, W.K. (2021) Toxic or Otherwise Harmful Algae and the Built Environment. *Toxins*, 13: 465. doi:10.3390/TOXINS13070465.

Horiuchi, Y., Kimura, R., Kato, N., et al. (2003) Evolutional study on acetylcholine expression. *Life Sciences*, 72 (15): 1745–1756. doi:10.1016/S0024-3205(02)02478-5.

Hotez, P.J., Brindley, P.J., Bethony, J.M., et al. (2008) Helminth infections: the great neglected tropical diseases. *The Journal of Clinical Investigation*, 118 (4): 1311. doi:10.1172/JCI34261.

Hovde, B.T., Hanschen, E.R., Steadman Tyler, C.R., et al. (2018) Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Research*, 35: 449–461. doi:10.1016/j.algal.2018.09.012.

Hsieh, J.-C., Hsieh, Y.-F. and Chuang, S.-T. (2020) *Prototheca* from bovine milk and associated minimal algaecide concentration of chlorhexidine and povidone-iodine in Taiwan. *Tierärztliche Praxis Ausgabe G: Großtiere / Nutztiere*, 48 (06): 380–385. doi:10.1055/a-1274-9023.

Huilca-Ibarra, M.P., Vasco-Julio, D., Ledesma, Y., et al. (2022) High Prevalence of Prototheca bovis Infection in Dairy Cattle with Chronic Mastitis in Ecuador. *Veterinary Sciences*, 9: 659. doi:10.3390/VETSCI9120659/S1.

Huss, V.A.R., Holweg, C., Seidel, B., et al. (1994) There is an ecological basis for host/symbiont specificity in *Chlorella*/*Hydra symbioses*. *Endocytobiosis & Cell Research*, 46 (1): 35–46.

Huth, N., Wenkel, R.F., Roschanski, N., et al. (2015) Prototheca zopfii Genotype 2-induced Nasal Dermatitis in a Cat. *Journal of comparative pathology*, 152 (4): 287–290.

doi:10.1016/j.jcpa.2015.02.001.

Iacoviello, V.R., DeGirolami, P.C., Lucarini, J., et al. (1992) Protothecosis complicating prolonged endotracheal intubation: Case report and literature review. *Clinical Infectious Diseases*, 15 (6): 959–967. doi:10.1093/clind/15.6.959.

Irrgang, A., Murugaiyan, J., Weise, C., et al. (2015a) Well-known surface and extracellular antigens of pathogenic microorganisms among the immunodominant proteins of the infectious microalgae Prototheca zopfii. *Frontiers in Cellular and Infection Microbiology*, 5: 67. doi:10.3389/fcimb.2015.00067.

Irrgang, A., Weise, C., Murugaiyan, J., et al. (2015b) Identification of immunodominant proteins of the microalgae Prototheca by proteomic analysis. *New Microbes and New Infections*, 3: 37–40. doi:10.1016/j.nmni.2014.10.004.

Iskra, M., Rudna, J., Bakuła, Z., et al. (2020) *Prototheca vistulensis* sp. nov. – a new microalgal species isolated from the Vistula river, Poland. *Poster presented at: FEMS Online Conference on Microbiology2020*.

Ito, T., Kano, R., Sobukawa, H., et al. (2011) Experimental Infection of Bovine Mammary Gland with *Prototheca zopfii* Genotype 1. *Journal of Veterinary Medical Science*, 73 (1): 117–119. doi:10.1292/jvms.10-0242.

Jagielski, T., Bakuła, Z., Gawor, J., et al. (2019a) The genus *Prototheca* (Trebouxiophyceae, Chlorophyta) revisited: Implications from molecular taxonomic studies. *Algal Research*, 43 (March): 101639. doi:10.1016/j.algal.2019.101639.

Jagielski, T., Bakuła, Z., Pleń, M., et al. (2018a) The activity of silver nanoparticles against microalgae of the *Prototheca* genus. *Nanomedicine*, 13 (9): 1025–1036. doi:10.2217/nnm-2017-0370.

Jagielski, T., Dyląg, M., Roesler, U., et al. (2017a) Isolation of infectious microalga Prototheca

wickerhamii from a carp (*Cyprinus carpio*) - a first confirmed case report of protothecosis in a fish. *Journal of Fish Diseases*, 40 (10): 1417–1421. doi:10.1111/jfd.12614.

Jagielski, T., Gawor, J., Bakuła, Z., et al. (2017b) An optimized method for high quality DNA extraction from microalga *Prototheca wickerhamii* for genome sequencing. *Plant Methods*, 13 (1). doi:10.1186/s13007-017-0228-9.

Jagielski, T., Gawor, J., Bakuła, Z., et al. (2018b) Cytb as a new genetic marker for differentiation of *Prototheca* species. *Journal of Clinical Microbiology*, 56 (10): e00584-18. doi:10.1128/JCM.00584-18.

Jagielski, T., Iskra, M., Bakuła, Z., et al. (2022) Occurrence of *Prototheca* Microalgae in Aquatic Ecosystems with a Description of Three New Species, *Prototheca fontanea*, *Prototheca lentecrescens*, and *Prototheca vistulensis* Villanueva, L. (ed.). *Applied and Environmental Microbiology*, 88 (22): 11. doi:10.1128/AEM.01092-22.

Jagielski, T., Krukowski, H., Bochniarz, M., et al. (2019b) Prevalence of *Prototheca* spp. on dairy farms in Poland – a cross-country study. *Microbial Biotechnology*, 12 (3): 556–566. doi:10.1111/1751-7915.13394.

Jagielski, T. and Lagneau, P.E. (2007) Protothecosis. A pseudofungal infection. *Journal de Mycologie Medicale*, 17: 261–270. doi:https://www.doi.org/10.1016/j.mycmed.2007.08.003.

Jagielski, T., Lassa, H., Ahrholdt, J., et al. (2011) Genotyping of bovine *Prototheca* mastitis isolates from Poland. *Veterinary Microbiology*, 149 (1–2): 283–287. doi:10.1016/j.vetmic.2010.09.034.

Jagielski, T., Niedźwiecka, K., Roeske, K., et al. (2018c) 3-Bromopyruvate as an Alternative Option for the Treatment of Protothecosis. *Frontiers in Pharmacology*, 9 (APR): 375. doi:10.3389/fphar.2018.00375.

Jagielski, T., Roeske, K., Bakuła, Z., et al. (2019c) A survey on the incidence of *Prototheca* mastitis in dairy herds in Lublin province, Poland. *Journal of Dairy Science*, 102 (1): 619–628. doi:10.3168/jds.2018-15495.

Jahangeer, M., Mahmood, Z., Munir, N., et al. (2020) *Naegleria fowleri*: Sources of infection, pathophysiology, diagnosis, and management; a review. *Clinical and Experimental Pharmacology and Physiology*, 47 (2): 199–212. doi:10.1111/1440-1681.13192.

Jánosi, S., Rátz, F., Szigeti, G., et al. (2001) Review of the microbiological, pathological, and clinical aspects of bovine mastitis caused by the alga *Prototheca zopfii*. *Veterinary Quarterly*, 23 (2): 58–61. doi:10.1080/01652176.2001.9695082.

Jinatham, V., Cantoni, D.M., Brown, I.R., et al. (2021) *Prototheca bovis*, a unicellular achlorophyllous trebouxiophyte green alga in the healthy human intestine. *Journal of Medical Microbiology*, 70 (9): 001415. doi:10.1099/jmm.0.001415.

Joerger, T., Sulieman, S., Carson, V.J., et al. (2020) Chronic Meningitis Due to *Prototheca zopfii* in an Adolescent Girl. *Journal of the Pediatric Infectious Diseases Society*. doi:10.1093/JPIDS/PIAA049.

Johnson, N.A., Lahti, D.C. and Blumstein, D.T. (2012) Combating the Assumption of Evolutionary Progress: Lessons from the Decay and Loss of Traits. *Evolution: Education and Outreach*, 5 (1): 128–138. doi:10.1007/S12052-011-0381-Y/FIGURES/3.

Jones, B., Renaut, R.W. and Rosen, M.R. (2003) Silicified Microbes in a Geyser Mound: The Enigma of Low-Temperature Cyanobacteria in a High-Temperature Setting. *PALAIOS*, 18 (2): 87–109.

Jones, J.W., McFadden, H.W., Chandler, F.W., et al. (1983) Green algal infection in a human. *American Journal of Clinical Pathology*, 80 (1): 102–107. doi:10.1093/ajcp/80.1.102.

Joubert, J.J. and Rijkenberg, F.H.J. (1971) Parasitic Green Algae. *Annual Review of*

Phytopathology, 9: 45–64. doi:10.1146/ANNUREV.PY.09.090171.000401.

Józsa, L. and Méhes, G. (1990) Protothecosis. A new, or rarely recognized disease? *Orvosi hetilap*, 131 (33): 1811—1813. Available at: http://europepmc.org/abstract/MED/2399029.

Kalina, T. and Puncocharova, M. (1987) Taxonomy of the subfamily Scotiellocystoideae Fott 1976 (Chlorellaceae, Chlorophyceae). *Archiv für Hydrobiologie. Supplementband. Monographische Beiträge*, 73 (4).

Kano, R., Satoh, K., Yaguchi, T., et al. (2022) Phenotypic Characteristics of *Prototheca* Species Occurring in Humans and Animals. *Medical Mycology Journal*, 63 (1): 17–20. doi:10.3314/mmj.21-00018.

Kato, S., Misumi, O., Maruyama, S., et al. (2023) Genomic analysis of an ultrasmall freshwater green alga, *Medakamo hakoo*. *Communications Biology*, 6: 89. doi:10.1038/s42003-022-04367-9.

Katwilat, P., Chongtrakool, P., Muangsomboon, S., et al. (2019) *Prototheca wickerhamii* prepatellar bursitis in an immunocompetent woman: A case report. *Journal de Mycologie Medicale*, 29 (4): 361–364. doi:10.1016/j.mycmed.2019.100901.

Kawai, S., Araki, S. and Kodama, Y. (2018) No mutual symbiosis following infection of algae-free *Paramecium bursaria* with symbiotic algae from *Mayorella viridis*. *Symbiosis*, 75 (1): 51–59. doi:10.1007/s13199-017-0517-0.

Kawashima, K., Misawa, H., Moriwaki, Y., et al. (2007) Ubiquitous expression of acetylcholine and its biological functions in life forms without nervous systems. *Life Sciences*, 80 (24–25): 2206–2209. doi:10.1016/J.LFS.2007.01.059.

Kayama, M., Chen, J.F., Nakada, T., et al. (2020) A non-photosynthetic green alga illuminates the reductive evolution of plastid electron transport systems. *BMC Biology*, 18: 126. doi:10.1186/S12915-020-00853-W/FIGURES/5.

Keeling, P.J., Mathur, V. and Kwong, W.K. (2021) Corallicolids: The elusive coral-infecting apicomplexans Knoll, L.J. (ed.). *PLOS Pathogens*, 17 (9): e1009845. doi:10.1371/JOURNAL.PPAT.1009845.

Kehl-Fie, T.E., Skaar, E.P., Franz, K., et al. (2010) Nutritional immunity beyond iron: a role for manganese and zinc. *Current Opinion in Chemical Biology*, 14 (2): 218–224. doi:10.1016/J.CBPA.2009.11.008.

Keilin, D. (1921) On the life-history of *Helicosporidium parasiticum*, N.G., N.SP., a new type of protist parasitic in the larva of *Dasyhelea obscura winn.* (diptera, ceratopogonidae) and in some other arthropods. *Parasitology*, 13 (2): 97–113. doi:10.1017/S003118200001235X.

Kessell, A.E., McNair, D., Munday, J.S., et al. (2017) Successful treatment of multifocal pedal *Prototheca wickerhamii* infection in a feline immunodeficiency virus-positive cat with multiple Bowenoid in situ carcinomas containing papillomaviral DNA sequences. *JFMS open reports*, 3 (1): 2055116916688590. doi:10.1177/2055116916688590.

Khan, I.D., Sahni, A.K., Sen, S., et al. (2018) Outbreak of *Prototheca wickerhamii* algaemia and sepsis in a tertiary care chemotherapy oncology unit. *Medical Journal Armed Forces India*, 74 (4): 358–364. doi:10.1016/j.mjafi.2017.07.012.

Kho, K.W., Berselli, G.B. and Keyes, T.E. (2021) A Nanoplasmonic Assay of Oligonucleotide-Cargo Delivery from Cationic Lipoplexes. *Small*, 17 (12): 2005815. doi:10.1002/smll.202005815.

Killam, A. and Myers, J. (1956) A Special Effect of Light on the Growth of *Chlorella vulgaris*. *American Journal of Botany*, 43 (8): 569–572. doi:10.1002/J.1537-2197.1956.TB10535.X.

Kim, H.S., Park, W.K., Lee, B., et al. (2019) Optimization of heterotrophic cultivation of *Chlorella* sp. HS2 using screening, statistical assessment, and validation. *Scientific Reports*, 9: 19383. doi:10.1038/s41598-019-55854-9.

Kim, J.S., Yoon, S.J., Park, Y.J., et al. (2020) Crossing the kingdom border: Human diseases caused by plant pathogens. *Environmental Microbiology*, 22 (7): 2485–2495. doi:10.1111/1462-2920.15028.

Kim, T., Yu, H.-J., Kang, O., et al. (2022) Fatal *Prototheca zopfii* Algaemia in a Patient with Acute Lymphoblastic Leukemia: a Case Report. *Clinical laboratory*, 68 (4). doi:10.7754/CLIN.LAB.2021.210723.

Klintworth, G.K., Fetter, B.F. and Nielsen, H.S. (1968) Protothecosis, an algal infection: report of a case in man. *Journal of medical microbiology*, 1 (2): 211–216. doi:10.1099/00222615-1-2-211.

Knauf, U. and Hachtel, W. (2002) The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Molecular Genetics and Genomics*, 267 (4): 492–497. doi:10.1007/s00438-002-0681-6.

Kodama, Y. and Fujishima, M. (2005) Symbiotic *Chlorella* sp. of the ciliate *Paramecium bursaria* do not prevent acidification and lysosomal fusion of host digestive vacuoles during infection. *Protoplasma*, 225 (3–4): 191–203. doi:10.1007/s00709-005-0087-5.

Kodama, Y. and Fujishima, M. (2016) Differences in infectivity between endosymbiotic *Chlorella variabilis* cultivated outside host *Paramecium bursaria* for 50 years and those immediately isolated from host cells after one year of reendosymbiosis. *Biology Open*, 5 (1): 55–61. doi:10.1242/bio.013946.

Kodama, Y., Nakahara, M. and Fujishima, M. (2007) Symbiotic alga *Chlorella vulgaris* of the ciliate *Paramecium bursaria* shows temporary resistance to host lysosomal enzymes during the early infection process. *Protoplasma*, 230 (1–2): 61–67. doi:10.1007/s00709-006-0193-z.

Köhler, J.R., Casadevall, A. and Perfect, J. (2015) The Spectrum of Fungi That Infects Humans. *Cold Spring Harbor Perspectives in Medicine*, 5 (1). doi:10.1101/CSHPERSPECT.A019273.

Kosmidis, C. and Denning, D.W. (2017) “Opportunistic and Systemic Fungi.” In *Infectious Diseases*. Fourth. Elsevier Ltd. pp. 1681-1709.e3. doi:10.1016/B978-0-7020-6285-8.00189-1.

Krcméry, V. (2000) Systemic chlorellosis, an emerging infection in humans caused by algae. *International Journal of Antimicrobial Agents*, 15 (3): 235–237. doi:10.1016/S0924-8579(00)00150-3.

Krishan, A. (1975) Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *Journal of Cell Biology*, 66 (1): 188–193. doi:10.1083/JCB.66.1.188.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., et al. (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47 (D1): D807–D811. doi:10.1093/NAR/GKY1053.

Kruger, W. (1894) Kurze Charakteristik einiger niedriger Organismen im Saftfluss der Laubbäume. *Hedwigia*, 33: 241–66.

Krukowski, H., Lassa, H., Zastempowska, E., et al. (2020) Etiological agents of bovine mastitis in Poland. *Medycyna Weterynaryjna*, 76 (04): 221–225. doi:10.21521/mw.6339.

Kunthiphun, S., Endoh, R., Takashima, M., et al. (2019) *Prototheca paracutis* sp. nov., a novel oleaginous achlorophyllous microalga isolated from a mangrove forest. *Mycoscience*, 60 (3): 165–169. doi:10.1016/j.myc.2019.02.003.

Kuo, T.T., Hsueh, S., Wu, J.L., et al. (1987) Cutaneous protothecosis. A clinicopathologic study. *Archives of Pathology and Laboratory Medicine*, 111 (8): 737–740.

Kurtzman, C.P., Fell, J., Boekhout, T., et al. (2011) *Yeasts, A Taxonomic Study*. 5th ed. Amsterdam, the Netherlands: Elsevier. doi:10.3314/jjmm.52.107.

Kurumisawa, T., Kano, R., Nakamura, Y., et al. (2018) Is bovine protothecal mastitis related

to persistent infection in intestine? *Journal of Veterinary Medical Science*. 80 (6) pp. 950–952. doi:10.1292/jvms.17-0710.

Kwiecinski, J. (2015) Biofilm formation by pathogenic *Prototheca* algae. *Letters in Applied Microbiology*, 61 (6): 511–517. doi:10.1111/lam.12497.

Langridge, G.C., Phan, M.D., Turner, D.J., et al. (2009) Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. *Genome Research*, 19 (12): 2308. doi:10.1101/GR.097097.109.

Lass-Flörl, C., Fille, M., Gunsilius, E., et al. (2004) Disseminated infection with *Prototheca zopfii* after unrelated stem cell transplantation for leukemia. *Journal of Clinical Microbiology*, 42 (10): 4907–4908. doi:10.1128/JCM.42.10.4907-4908.2004.

Lass-Flörl, C. and Mayr, A. (2007) Human protothecosis. *Clinical Microbiology Reviews*, 20 (2): 230–242. doi:https://www.doi.org/10.1128/CMR.00032-06.

Lassa, H., Jagielski, T. and Malinowski, E. (2011) Effect of Different Heat Treatments and Disinfectants on the Survival of *Prototheca zopfii*. *Mycopathologia*, 171 (3): 177–182. doi:10.1007/s11046-010-9365-7.

Laugé, R. and De Wit, P.J.G.M. (1998) Fungal avirulence genes: structure and possible functions. *Fungal Genetics and Biology*, 24 (3): 285–297. doi:10.1006/FGBI.1998.1076.

Lee, C.M., Jin, S.P., Doh, E.J., et al. (2019) Regional Variation of Human Skin Surface Temperature. *Annals of Dermatology*, 31 (3): 349–352. doi:10.5021/AD.2019.31.3.349.

Lees, A.M., Sejian, V., Lees, J.C., et al. (2019) Evaluating rumen temperature as an estimate of core body temperature in Angus feedlot cattle during summer. *International Journal of Biometeorology*, 63 (7): 939–947. doi:10.1007/S00484-019-01706-0/FIGURES/4.

Leimann, B.C.Q., Monteiro, P.C.F., Lazéra, M., et al. (2004) Protothecosis. *Medical Mycology*, 42 (2): 95–106. doi:10.1080/13693780310001653653.

Lemieux, C., Turmel, M., Otis, C., et al. (2019) A streamlined and predominantly diploid genome in the tiny marine green alga *Chloropicon primus*. *Nature Communications*, 10: 4061. doi:10.1038/s41467-019-12014-x.

Lerche (1952) Eine durch Algen (*Prototheca*) hervorgerufene Mastitis der Kuh. *Berl. Munch. Tierarztl. Wochenschr*, 65: 64–69.

Lesser, M.P., Stat, M. and Gates, R.D. (2013) The endosymbiotic dinoflagellates (*Symbiodinium* sp.) of corals are parasites and mutualists. *Coral Reefs*, 32 (3): 603–611. doi:10.1007/s00338-013-1051-z.

Lewis, L.A. and McCourt, R.M. (2004) Green algae and the origin of land plants. *American Journal of Botany*, 91 (10): 1535–1556. doi:10.3732/AJB.91.10.1535.

Li, S., Tan, H., Liu, B., et al. (2021) Watanabeales ord. nov. and twelve novel species of Trebouxiophyceae (Chlorophyta). *Journal of Phycology*, 57 (4): 1167–1186. doi:10.1111/JPY.13165.

Li, Z., Parris, S. and Saski, C.A. (2020) A simple plant high-molecular-weight DNA extraction method suitable for single-molecule technologies. *Plant Methods*, 16 (1): 38. doi:10.1186/s13007-020-00579-4.

Liao, J. and Liu, Y. (2020) Extraction and detection of DNA from UHT milk during storage. *CyTA - Journal of Food*, 18 (1): 747–752. doi:10.1080/19476337.2020.1839565.

Libisch, B., Picot, C., Ceballos-garzon, A., et al. (2022) Prototheca Infections and Ecology from a One Health Perspective. *Microorganisms*, 10: 938. doi:10.3390/MICROORGANISMS10050938/S1.

Lim, S.L., Chu, W.L. and Phang, S.M. (2010) Use of *Chlorella vulgaris* for bioremediation of textile wastewater. *Bioresource Technology*, 101 (19): 7314–7322. doi:10.1016/J.BIORTECH.2010.04.092.

Lischer, H.E.L. and Shimizu, K.K. (2017) Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, 18: 474. doi:10.1186/s12859-017-1911-6.

Lloyd, C., Tan, K.H., Lim, K.L., et al. (2021) Identification of microalgae cultured in Bold's Basal medium from freshwater samples, from a high-rise city. *Scientific Reports*, 11: 4474. doi:10.1038/s41598-021-84112-0.

Lloyd, D. and Turner, G. (1968) The Cell Wall of *Prototheca zopfii*. *Journal of General Microbiology*, 50: 421–27. doi:10.1271/kagakutoseibutsu1962.4.682.

Lowe, C.D., Minter, E.J., Cameron, D.D., et al. (2016) Shining a Light on Exploitative Host Control in a Photosynthetic Endosymbiosis. *Current Biology*, 26 (2): 207–211. doi:10.1016/j.cub.2015.11.052.

Lu, Y., Kong, R. and Hu, L. (2012) Preparation of protoplasts from *Chlorella protothecoides*. *World Journal of Microbiology and Biotechnology*, 28 (4): 1827–1830. doi:10.1007/s11274-011-0963-4.

Lu, Y., Zhang, X., Ni, F., et al. (2021) Cutaneous Protothecosis with Meningitis Due to *Prototheca wickerhamii* in an Immunocompetent Teenager: Case Report and Literature Review. *Infection and Drug Resistance*, 14: 2787. doi:10.2147/IDR.S320795.

Luo, S., Li, Y., Chen, S., et al. (2011) *Gelechiidae* Moths Are Capable of Chemically Dissolving the Pollen of Their Host Plants: First Documented Sporopollenin Breakdown by an Animal. *PLoS ONE*, 6 (4): e19219. doi:10.1371/JOURNAL.PONE.0019219.

Luo, W., Pröschold, T., Bock, C., et al. (2010) Generic concept in *Chlorella*-related coccoid green algae (Chlorophyta, Trebouxiophyceae). *Plant Biology*, 12 (3): 545–553. doi:10.1111/j.1438-8677.2009.00221.x.

Maboni, G., Elbert, J.A., Stilwell, J.M., et al. (2021) Genomic and Pathologic Findings for

Prototheca cutis Infection in Cat. *Emerging Infectious Diseases*, 27 (3): 979. doi:10.3201/eid2703.202941.

Macartney, L., Rycroft, A.N. and Hammil, J. (1988) Cutaneous protothecosis in the dog: first confirmed case in Britain. *The Veterinary record*, 123 (19): 494–496. doi:10.1136/VR.123.19.494.

Macedo, J.T.S.A., Riet-Correa, F., Dantas, A.F.M., et al. (2008) Cutaneous and nasal protothecosis in a goat. *Veterinary Pathology*, 45 (3): 352–354. doi:10.1354/vp.45-3-352.

Mackenzie, G., Boa, A.N., Diego-Taboada, A., et al. (2015) Sporopollenin, the least known yet toughest natural biopolymer. *Frontiers in Materials*, 2: 66. doi:10.3389/FMATS.2015.00066/BIBTEX.

Magro, S., Visentin, E., Chiarin, E., et al. (2023) Pathogen Detection via Quantitative PCR in Milk of Healthy Cows Collected Using Different Sampling Protocols. *Pathogens*, 12 (7): 935. doi:10.3390/PATHOGENS12070935.

Mancera, N., Douma, L.G., James, S., et al. (2012) Detection of *Helicosporidium* spp. in metagenomic DNA. *Journal of Invertebrate Pathology*, 111 (1): 13–19. doi:10.1016/j.jip.2012.05.002.

Manni, M., Berkeley, M.R., Seppey, M., et al. (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38 (10): 4647–4654. doi:10.48550/arxiv.2106.11799.

Marçais, G. and Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27 (6): 764–770. doi:10.1093/BIOINFORMATICS/BTR011.

Marques, S., Silva, E., Carvalheira, J., et al. (2006) Short communication: In vitro

antimicrobial susceptibility of *Prototheca wickerhamii* and *Prototheca zopfii* isolated from bovine mastitis. *Journal of Dairy Science*, 89 (11): 4202–4204. doi:10.3168/jds.S0022-0302(06)72465-1.

Marques, S., Silva, E., Carvalheira, J., et al. (2010) In vitro susceptibility of *Prototheca* to pH and salt concentration. *Mycopathologia*, 169 (4): 297–302. doi:10.1007/s11046-009-9254-0.

Marsh, E.K. and May, R.C. (2012) *Caenorhabditis elegans*, A model organism for investigating immunity. *Applied and Environmental Microbiology*, 78 (7): 2075–2081. doi:10.1128/AEM.07486-11/ASSET/7F34E3AC-C69C-4646-989B-E952A503D629/ASSETS/GRAPHIC/ZAM9991030940001.JPEG.

Martínez, F., Ascaso, C. and Orús, M.I. (1991) Morphometric and stereologic analysis of *Chlorella vulgaris* under heterotrophic growth conditions. *Annals of Botany*, 67 (3): 239–245. doi:10.1093/oxfordjournals.aob.a088128.

Masuda, M., Hirose, N., Ishikawa, T., et al. (2016) *Prototheca miyajii* sp. nov., isolated from a patient with systemic protothecosis. *International Journal of Systematic and Evolutionary Microbiology*, 66 (3): 1510–1520. doi:10.1099/ijsem.0.000911.

Masuda, M., Jagielski, T., Danesi, P., et al. (2021) Protothecosis in Dogs and Cats - New Research Directions. *Mycopathologia*, 186 (1): 143–152. doi:https://doi.org/10.1007/s11046-020-00508-y.

Matsuda, T. and Matsumoto, T. (1992) Protothecosis: a Report of Two Cases in Japan and a Review of the Literature. *European Jornal of Epidemiology*, 8 (3): 397–406.

May, R.C., Stone, N.R.H., Wiesner, D.L., et al. (2016) *Cryptococcus*: From environmental saprophyte to global pathogen. *Nature Reviews Microbiology*, 14 (2): 106–117. doi:10.1038/nrmicro.2015.6.

Meinke, K.W., Hamedani, F., Wu, S., et al. (2017) *Prototheca zopfii* associated diverticulitis in

an immunosuppressed host, a case presentation and literature review. *Human Pathology: Case Reports*, 10: 43–45. doi:10.1016/j.ehpc.2017.04.005.

Mekvipad, N. and Satjarak, A. (2019) Evolution of organellar genes of chlorophyte algae: Relevance to phylogenetic inference. *PLoS ONE*, 14 (5): e0216608. doi:10.1371/JOURNAL.PONE.0216608.

Melchior, M.B., Vaarkamp, H. and Fink-Gremmels, J. (2006) Biofilms: A role in recurrent mastitis infections? *Veterinary Journal*, 171 (3): 398–407. doi:10.1016/j.tvjl.2005.01.006.

Melo, R.R., Rodrigues, F.T., Baka, J.L.L.C. e S., et al. (2022) An exuberant case of prototecosis: an emergent disease in tropical dermatology. *Dermatology Online Journal*, 28 (1). doi:10.5070/d328157065.

Melville, P.A., Watanabe, E.T., Benites, N.R., et al. (1999) Evaluation of the susceptibility of *Prototheca zopfii* to milk pasteurization. *Mycopathologia*, 146 (2): 79–82. doi:10.1023/A:1007005729711.

Merck&Co (2020) "Bovine Ephemeral Fever - Generalized Conditions." In *MSD Veterinary Manual*. Available at: https://www.msdvetmanual.com/generalized-conditions/bovine-ephemeral-fever/bovine-ephemeral-fever (Downloaded: 6 September 2022).

Méthot, P.O. and Alizon, S. (2014) What is a pathogen? Toward a process view of host-parasite interactions. *Virulence*, 5 (8): 775. doi:10.4161/21505594.2014.960726.

Metting, B. (1981) The Botanical review: The systematics and ecology of the soil algae. *The Botanical Review*, 47 (2): 195–312.

Mettler, F. (1975) Generalisierte Protothecose bei einem Flughund (*Pteropus lylei*). *Veterinary Pathology*, 12 (2): 118–124. doi:10.1177/030098587501200204.

Michel, A.H., Hatakeyama, R., Kimmig, P., et al. (2017) Functional mapping of yeast genomes by saturated transposition. *eLife*, 6. doi:10.7554/ELIFE.23570.

Milanov, D., Petrović, T., Polaček, V., et al. (2016) Mastitis associated with *Prototheca zopfii* - An emerging health and economic problem on dairy farms. *Journal of Veterinary Research (Poland)*, 60 (4): 373–378. doi:10.1515/jvetres-2016-0054.

Milanov, D., Suvajdžić, L., Pušić, I., et al. (2006) Outbreak of endemic form of protothecal mastitis on a dairy farm. *Acta Veterinaria*, 56 (2–3): 259–265. doi:10.2298/AVB0603259M.

Miller, H.C., Morgan, M.J., Walsh, T., et al. (2018) Preferential feeding in *Naegleria fowleri*; intracellular bacteria isolated from amoebae in operational drinking water distribution systems. *Water research*, 141: 126–134. doi:10.1016/J.WATRES.2018.05.004.

Minato, K., Yoshikawa, M., Nakanishi, H., et al. (2020) Long Term Follow-Up of *Prototheca* keratitis: A Case Report. *International Medical Case Reports Journal*, Volume 13: 503–506. doi:10.2147/imcrj.s268696.

Miura, A., Kano, R., Ito, T., et al. (2020) In vitro algaecid effect of itraconazole and ravuconazole on *Prototheca* species. *Medical mycology*, 58 (6): 845–847. doi:10.1093/mmy/myz119.

Moore, R.B., Oborník, M., Janouškovec, J., et al. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, 451 (7181): 959–963. doi:10.1038/nature06635.

Morello, L., Tiroli, T., Aretino, F., et al. (2020) Preliminary Results, Perspectives, and Proposal for a Screening Method of In Vitro Susceptibility of *Prototheca* Species to Antimicrotubular Agents. *Antimicrobial Agents and Chemotherapy*, 63 (3): e01392-19. doi:10.1128/AAC.01392-19.

Munday, B.L. and Peel, B.F. (1983) Severe ulcerative dermatitis in platypus (*Ornithorhynchus anatinus*). *Journal of wildlife diseases*, 19 (4): 363–365. doi:10.7589/0090-3558-19.4.363.

Murphy, K., Travers, P., Walport, M., et al. (2012a) "Chapter 2: Innate Immunity: The First Lines of Defense." In *Janeway's Immunobiology*. 8th ed. New York: Garland Science. pp. 37–

75.

Murphy, K., Travers, P., Walport, M., et al. (2012b) “Chapter 3: The Induced Responses of Innate Immunity.” In *Janeway’s Immunobiology*. 8th ed. New York: Garland Science. pp. 75–126.

Muscatine, L. and McNeil, P.L. (1989) Endosymbiosis in *Hydra* and the evolution of internal defense system. *Integrative and Comparative Biology*, 29 (2): 371–386. doi:10.1093/icb/29.2.371.

Myers, J. and Kratz, W.A. (1955) Relation between pigment content and photosynthetic characteristics in a blue-green algae. *The Journal of general physiology*, 39 (1): 11–22. doi:10.1085/JGP.39.1.11.

Nadakavukaren, M.J. and McCracken, D.A. (1975) Prototheca filamenta: A fungus not an alga. *Mycopathologia*, 57 (2): 89–91. doi:10.1007/BF01365709.

Nadakavukaren, M.J. and McCracken, D.A. (1977) An ultrastructural survey of the genus *Prototheca* with special reference to plastids. *Mycopathologia*, 61 (2): 117–119. doi:10.1007/BF00443840.

Nagatsuka, Y., Kiyuna, T., Kigawa, R., et al. (2017) *Prototheca tumulicola* sp. nov., a novel achlorophyllous, yeast-like microalga isolated from the stone chamber interior of the Takamatsuzuka Tumulus. *Mycoscience*, 58 (1): 53–59. doi:10.1016/j.myc.2016.09.005.

Nair, S.C. and Striepen, B. (2011) What do human parasites do with a chloroplast anyway? *PLoS Biology*, 9 (8): e1001137. doi:10.1371/journal.pbio.1001137.

Nardoni, S., Pisseri, F., Pistelli, L., et al. (2018) In vitro activity of 30 essential oils against bovine clinical isolates of Prototheca zopfii and *Prototheca blaschkeae*. *Veterinary Sciences*, 5 (2). doi:10.3390/vetsci5020045.

Narita, M., Muder, R.R., Cacciarelli, T. V., et al. (2008) Protothecosis after liver

transplantation. *Liver Transplantation*, 14 (8): 1211–1215. doi:10.1002/lt.21565.

Neustupa, J., Němcová, Y., Veselá, J., et al. (2013) Leptochlorella corticola gen. et sp. nov. and *Kalinella apyrenoidosa* sp. nov.: Two novel *Chlorella*-like green microalgae (Trebouxiophyceae, Chlorophyta) from subaerial habitats. *International Journal of Systematic and Evolutionary Microbiology*, 63 (1): 377–387. doi:10.1099/ijs.0.047944-0.

NewEnglandBiolabs (n.d.) *High Molecular Weight (HMW) DNA Extraction - minimizing a bottleneck for emerging long read sequencing applications | NEB*. Available at: https://international.neb.com/nebinspired-blog/hmw-dna---minimizing-the-bottleneck (Accessed: 17 May 2022).

Nickerson, S.C. and Akers, R.M. (2011) Mammary Gland: Anatomy. *Encyclopedia of Dairy Sciences: Second Edition*, pp. 328–337. doi:10.1016/B978-0-12-374407-4.00290-9.

Noda, A., Murata, K., Sakai, H., et al. (2003) Disseminated Protothecosis in a Cape Hyrax. *Japanese Society of Zoo and Wildlife Medicine Japanese*, 8 (2): 121–125.

Normal Rectal Temperatures (n.d.).

Nowicka-Krawczyk, P., Komar, M. and Gutarowska, B. (2022) Towards understanding the link between the deterioration of building materials and the nature of aerophytic green algae. *The Science of the total environment*, 802. doi:10.1016/J.SCITOTENV.2021.149856.

O'Connor, J.P., Nimmo, G.R., Rigby, R.J., et al. (1986) Algal Peritonitis Complicating Continuous Ambulatory Peritoneal Dialysis. *American Journal of Kidney Diseases*, 8 (2): 122–123. doi:10.1016/S0272-6386(86)80123-8.

Odumeru, J., Gao, A., Chen, S., et al. (2001) Use of the bead beater for preparation of *Mycobacterium paratuberculosis* template DNA in milk. *Canadian Journal of Veterinary Research*, 65 (4): 201. Available at: /pmc/articles/PMC1189680/?report=abstract (Accessed: 8 March 2023).

Ogura-Tsujita, Y., Gebauer, G., Hashimoto, T., et al. (2009) Evidence for novel and specialized mycorrhizal parasitism: the orchid *Gastrodia confusa* gains carbon from saprotrophic *Mycena*. *Proceedings. Biological sciences*, 276 (1657): 761–767. doi:10.1098/RSPB.2008.1225.

Okude, H., Ori, D. and Kawai, T. (2021) Signaling Through Nucleic Acid Sensors and Their Roles in Inflammatory Diseases. *Frontiers in Immunology*, 11: 625833. doi:10.3389/FIMMU.2020.625833/BIBTEX.

Oliveira, R.D., Wolking, R.M., Bradway, D.S., et al. (2020) Algal Lymphadenitis in a Dog Caused by Scenedesmus Species. *Veterinary Pathology*, 57 (6): 821–824. doi:10.1177/0300985820948819.

Onozaki, M., Makimura, K., Satoh, K., et al. (2013) Detection and identification of genotypes of *Prototheca zopfii* in clinical samples by quantitative PCR analysis. *Japanese Journal of Infectious Diseases*, 66 (5): 383–390. doi:10.7883/yoken.66.383.

Orsi, C.F., Colombari, B., Callegari, F., et al. (2010) Yessotoxin inhibits phagocytic activity of macrophages. *Toxicon*, 55 (2–3): 265–273. doi:10.1016/J.TOXICON.2009.07.033.

Osumi, T., Kishimoto, Y., Kano, R., et al. (2008) *Prototheca zopfii* genotypes isolated from cow barns and bovine mastitis in Japan. *Veterinary Microbiology*, 131 (3–4): 419–423. doi:10.1016/j.vetmic.2008.04.012.

Paisley, D., Robson, G.D. and Denning, D.W. (2005) Correlation between in vitro growth rate and in vivo virulence in *Aspergillus fumigatus*. *Medical Mycology*, 43 (5): 397–401. doi:10.1080/13693780400005866/2/13693780400005866FIG002.JPEG.

Palaniappan, P.A., Abot, C.A., Mohd Tap, R., et al. (2021) Protothecosis algaemia in a patient presenting with septic arthritis: A rare case of *Prototheca zopfii* isolated from Malaysia. *IDCases*, 24: e01121. doi:10.1016/j.idcr.2021.e01121.

Pawar, N. V., Patil, V.B., Kamble, S.S., et al. (2008) First Report of *Aspergillus niger* as a Plant Pathogen on *Zingiber officinale* from India. *Plant disease*, 92 (9): 1368. doi:10.1094/PDIS-92-9-1368C.

Pegolo, S., Toscano, A., Bisutti, V., et al. (2022) *Streptococcus agalactiae* and *Prototheca* spp. induce different mammary gland leukocyte responses in Holstein cows. *JDS Communications*, 3 (4): 270. doi:10.3168/JDSC.2022-0216.

Perrotti, G., DiLiberto, C. and Pineda, D.M. (2023) The oldest patient infected with cutaneous protothecosis. *Journal of wound care*, 32 (4): 235–237. doi:10.12968/JOWC.2023.32.4.235.

Phair, J.P., Williams, J.E., Bassaris, H.P., et al. (1981) Phagocytosis and Algicidal Activity of Human Polymorphonuclear Neutrophils Against *Prototheca wickerhamii*. *The Journal of Infectious Diseases*, 144 (1): 72–76.

Pirofski, L. anne and Casadevall, A. (2012) Q and A What is a pathogen? A question that begs the point. *BMC Biology*, 10: 6. doi:10.1186/1741-7007-10-6/COMMENTS.

Piyophirapong, S., Linpiyawan, R., Mahaisavariya, P., et al. (2002) Cutaneous protothecosis in an AIDS patient. *British Journal of Dermatology*, 146 (4): 713–715. doi:10.1046/j.1365-2133.2002.47698.x.

Plieger, T. and Wolf, M. (2022) 18S and ITS2 rDNA sequence-structure phylogeny of Prototheca (Chlorophyta, Trebouxiophyceae). *Biologia*, 77: 569–582. doi:10.1007/S11756-021-00971-Y/FIGURES/8.

Pombert, J.-F.F., Blouin, N.A., Lane, C., et al. (2014) A Lack of Parasitic Reduction in the Obligate Parasitic Green Alga *Helicosporidium* Heitman, J. (ed.). *PLoS Genetics*, 10 (5): e1004355. doi:10.1371/journal.pgen.1004355.

Pore, R.S. (1973) Selective Medium for the Isolation of Prototheca. *Applied Microbiology*, 26

(4): 648–649. doi:10.1128/am.26.4.648-649.1973.

Pore, R.S. (1985) *Prototheca* taxonomy. *Mycopathologia*, 90 (3): 129–139. doi:10.1007/BF00436728.

Pore, R.S., Barnett, E.A., Barnes, W.C., et al. (1983) *Prototheca* ecology. *Mycopathologia*, 81 (1): 49–62. doi:10.1007/BF00443909.

Pore, R.S. and Shahan, T.A. (1988) *Prototheca zopfii*: Natural, transient, occurrence in pigs and rats. *Mycopathologia*, 101 (2): 85–88. doi:10.1007/BF00452891.

Di Prinzio, A., Ruf, M., Torre, A.C., et al. (2023) Cutaneous protothecosis in a solid organ transplanted patient. *Anais Brasileiros de Dermatologia*, 98 (5): 723–725. doi:10.1016/J.ABD.2022.10.008.

Pröschold, T., Darienko, T., Silva, P.C., et al. (2011) The systematics of *Zoochlorella* revisited employing an integrative approach. *Environmental Microbiology*, 13 (2): 350–364. doi:10.1111/j.1462-2920.2010.02333.x.

Pröschold, T., Marin, B., Schlösser, U.G., et al. (2001) Molecular Phylogeny and Taxonomic Revision of Chlamydomonas (Chlorophyta). I. Emendation of Chlamydomonas Ehrenberg and Chloromonas Gobi, and Description of Oogamochlamys gen. nov. and Lobochlamys gen. nov. *Protist*, 152 (4): 265–300. doi:10.1078/1434-4610-00068.

Pröschold, T., Pitsch, G. and Darienko, T. (2020) Micractinium tetrahymenae (Trebouxiophyceae, chlorophyta), a new endosymbiont isolated from ciliates. *Diversity*, 12 (5): 200. doi:10.3390/D12050200.

Quigley, L., O'Sullivan, O., Beresford, T.P., et al. (2012) A comparison of methods used to extract bacterial DNA from raw milk and raw milk cheese. *Journal of Applied Microbiology*, 113 (1): 96–105. doi:10.1111/J.1365-2672.2012.05294.X.

Rajan, A., Pai, V. and Shukla, P. (2020) Cutaneous protothecosis in an immunocompetent

host. *Indian Journal of Dermatology, Venereology and Leprology*, 86 (4): 414–417. doi:10.4103/ijdvl.IJDVL_920_18.

Ramanan, P., Virk, A., Boland, J.M., et al. (2018) Answer to August 2018 Photo Quiz. *Journal of Clinical Microbiology*, 56 (8). doi:10.1128/JCM.02418-16.

Ramírez, I., Nieto-Ríos, J.F., Ocampo-Kohn, C., et al. (2016) Protothecal bursitis after simultaneous kidney/liver transplantation: A case report and review. *Transplant Infectious Disease*, 18 (2): 266–274. doi:10.1111/tid.12496.

Ranallo-Benavidez, T.R., Jaron, K.S. and Schatz, M.C. (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11: 1432. doi:10.1038/s41467-020-14998-3.

Rao, P.V., Sethuraman, N., Ramanathan, Y., et al. (2018) Disseminated protothecosis caused by *Prototheca zopfii* in a liver transplant recipient. *Journal of Global Infectious Diseases*, 10 (4): 228–229. doi:10.4103/jgid.jgid_55_17.

Ratzke, C. and Gore, J. (2018) Modifying and reacting to the environmental pH can drive bacterial interactions. *PLOS Biology*, 16 (3): e2004248. doi:10.1371/journal.pbio.2004248.

Raz, R., Rottem, M., Bisharat, N., et al. (1998) Intestinal Protothecosis in a Patient with Chornic Mucocutaneous Candidiasis. *Clinical Infectious Diseases*, 27 (2): 399–400. doi:10.1086/514651.

Ricchi, M., Cammi, G., Garbarino, C.A., et al. (2011) A rapid real-time PCR/DNA resolution melting method to identify *Prototheca* species. *Journal of Applied Microbiology*, 110 (1): 27–34. doi:10.1111/j.1365-2672.2010.04861.x.

Ricchi, M., De Cicco, C., Buzzini, P., et al. (2013) First outbreak of bovine mastitis caused by Prototheca blaschkeae. *Veterinary Microbiology*, 162: 997–999. doi:10.1016/j.vetmic.2012.11.003.

Riet-Correa, F., Carmo, P.M.S. do and Uzal, F.A. (2021) Protothecosis and chlorellosis in sheep and goats: a review. *Journal of Veterinary Diagnostic Investigation*, 33 (2): 283–287. doi:10.1177/1040638720978781.

Riley, L.W. (2014) Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clinical microbiology and infection*, 20: 380–390. doi:10.1111/1469-0691.12646.

Ríos, G., Tadeo, F.R., Leida, C., et al. (2013) Prediction of components of the sporopollenin synthesis pathway in peach by genomic and expression analyses. *BMC Genomics*, 14: 40. doi:10.1186/1471-2164-14-40.

Robert, V.A. and Casadevall, A. (2009) Vertebrate endothermy restricts most fungi as potential pathogens. *The Journal of infectious diseases*, 200 (10): 1623–1626. doi:10.1086/644642.

Rodríguez, F., Feist, S.W., Guillou, L., et al. (2008) Phylogenetic and morphological characterisation of the green algae infesting blue mussel *Mytilus edulis* in the North and South Atlantic oceans. *Diseases of Aquatic Organisms*, 81 (3): 231–240. doi:10.3354/dao01956.

Roesler, U., Möller, A., Hensel, A., et al. (2006) Diversity within the current algal species *Prototheca zopfii*: A proposal for two *Prototheca zopfii* genotypes and description of a novel species, *Prototheca blaschkeae* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 56 (6): 1419–1425. doi:10.1099/ijs.0.63892-0.

Rogers, M.B., Gilson, P.R., Su, V., et al. (2007) The Complete Chloroplast Genome of the Chlorarachniophyte *Bigelowiella natans*: Evidence for Independent Origins of Chlorarachniophyte and Euglenid Secondary Endosymbionts. *Molecular Biology and Evolution*, 24 (1): 54–62. doi:10.1093/MOLBEV/MSL129.

Rosario Medina, I., Román Fuentes, L., Batista Arteaga, M., et al. (2017) Pigeons and their

droppings as reservoirs of *Candida* and other zoonotic yeasts. *Revista Iberoamericana de Micologia*, 34 (4): 211–214. doi:10.1016/j.riam.2017.03.001.

Safi, C., Zebib, B., Merah, O., et al. (2014) Morphology, composition, production, processing and applications of *Chlorella vulgaris*: A review. *Renewable and Sustainable Energy Reviews*, 35: 265–278. doi:10.1016/J.RSER.2014.04.007.

Sambrook, J. (2006) *The condensed protocols from Molecular cloning : a laboratory manual*. 4th ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sands, M., Poppel, D. and Brown, R. (1991) Peritonitis Due to *Prototheca wickerhamii* in a Patient Undergoing Chronic Ambulatory Peritoneal Dialysis. *Reviews of Infectious Diseases*, 13 (3): 376–378. doi:10.1093/CLINIDS/13.3.376.

Sangiovanni, M., Granata, I., Thind, A.S., et al. (2019) From trash to treasure: Detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics*, 20 (Suppl 4): 168. doi:10.1186/s12859-019-2684-x.

Santos, D.W. de C.L., Magalhaes Diniz, Y.C., Silva, R.R. da, et al. (2023) Human protothecosis: A case report in northeastern Brazil. *Medical Mycology Case Reports*, 40: 1–4. doi:10.1016/J.MMCR.2023.02.003.

Sari, S., Dalgic, B., Muehlenbachs, A., et al. (2018) *Prototheca zopfii* Colitis in Inherited CARD9 Deficiency. *Journal of Infectious Diseases*, 218 (3): 485–489. doi:10.1093/infdis/jiy198.

Satoh, K., Ooe, K., Nagayama, H., et al. (2010) *Prototheca cutis* sp. nov., a newly discovered pathogen of protothecosis isolated from inflamed human skin. *International Journal of Systematic and Evolutionary Microbiology*, 60 (5): 1236–1240. doi:10.1099/ijs.0.016402-0.

Schoch, C.L.C., Ciufo, S., Domrachev, M., et al. (2020) *NCBI Taxonomy: a comprehensive update on curation, resources and tools.* doi:10.1093/DATABASE/BAAA062.

Schöniger, S., Roschanski, N., Rösler, U., et al. (2016) *Prototheca* species and *Pithomyces chartarum* as Causative Agents of Rhinitis and/or Sinusitis in Horses. *Journal of Comparative Pathology*, 155 (2–3): 121–125. doi:10.1016/j.jcpa.2016.06.004.

Senoo, H., Sesaki, H. and Iijima, M. (2016) A GPCR Handles Bacterial Sensing in Chemotaxis and Phagocytosis. *Developmental Cell*, 36 (4): 354–356. doi:10.1016/J.DEVCEL.2016.02.006.

Severgnini, M., Lazzari, B., Capra, E., et al. (2018) Genome sequencing of *Prototheca zopfii* genotypes 1 and 2 provides evidence of a severe reduction in organellar genomes. *Scientific Reports*, 8: 14637. doi:10.1038/s41598-018-32992-0.

Shahid, M., Cavalcante, P.A., Knight, C.G., et al. (2020a) Murine and Human Cathelicidins Contribute Differently to Hallmarks of Mastitis Induced by Pathogenic *Prototheca* bovis Algae. *Frontiers in Cellular and Infection Microbiology*, 10: 31. doi:10.3389/fcimb.2020.00031.

Shahid, M., Cobo, E.R., Chen, L., et al. (2020b) *Prototheca zopfii* genotype II induces mitochondrial apoptosis in models of bovine mastitis. *Scientific Reports*, 10: 698. doi:10.1038/s41598-020-57645-z.

Shahid, M., Gao, J., Zhou, Y., et al. (2017a) *Prototheca zopfii* isolated from bovine mastitis induced oxidative stress and apoptosis in bovine mammary epithelial cells. *Oncotarget*, 8 (19): 31938–31947. doi:10.18632/oncotarget.16653.

Shahid, M., Wang, J., Gu, X., et al. (2017b) Prototheca zopfii Induced Ultrastructural Features Associated with Apoptosis in Bovine Mammary Epithelial Cells. *Frontiers in Cellular and Infection Microbiology*, 7: 299. doi:10.3389/fcimb.2017.00299.

Sharma, J.M. (1991) Overview of the avian immune system. *Veterinary Immunology and Immunopathology*, 30 (1): 13–17. doi:10.1016/0165-2427(91)90004-V.

Shave, C.D., Millyard, L. and May, R.C. (2021) Now for something completely different:

Prototheca, pathogenic algae. *PLOS Pathogens*, 17 (4): e1009362. doi:10.1371/journal.ppat.1009362.

Shepherd, J.A., Waigh, R.D. and Gilbert, P. (1988) Antibacterial action of 2-bromo-2-nitropropane-1,3-diol (bronopol). *Antimicrobial Agents and Chemotherapy*, 32 (11): 1693. doi:10.1128/AAC.32.11.1693.

Shihira, I. and Krauss, R.W. (1965) Chlorella: Physiology and Taxonomy of Forty-one Isolates. *Baltimore: Port City Press*, pp. 1–97.

Shu, Y. (2023) *Variation in disinfectant resistance between* Prototheca *species*. University of Birmingham.

Shui, W., Xiong, Y., Xiao, W., et al. (2015) Understanding the Mechanism of Thermotolerance Distinct From Heat Shock Response Through Proteomic Analysis of Industrial Strains of Saccharomyces cerevisiae. *Molecular & Cellular Proteomics*, 14 (7): 1885–1897. doi:10.1074/MCP.M114.045781.

Sievers, F., Wilm, A., Dineen, D., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7 (1): 539. doi:10.1038/MSB.2011.75.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., et al. (2011) *Pseudomonas genomes*: diverse and adaptable. *FEMS Microbiology Reviews*, 35 (4): 652–680. doi:10.1111/J.1574-6976.2011.00269.X.

Sileo, L. and Palmer, N.C. (1973) Probable Cutaneous Protothecosis in a Beaver. *Journal of Wildlife Diseases*, 9 (4): 320–322.

Silveira, C.S., Cesar, D., Keating, M.K., et al. (2018) A Case of *Prototheca zopfii* Genotype 1 Infection in a Dog (*Canis lupus familiaris*). *Mycopathologia*, 183 (5): 853–858. doi:10.1007/s11046-018-0274-5.

Sim, J.Y., Hsueh, P.R., Chuang, P.C., et al. (2019) Fatal disseminated infection caused by Prototheca zopfii in a child with leukemia. *Journal of Microbiology, Immunology and Infection*, 52 (5): 833–835. doi:10.1016/j.jmii.2019.05.001.

Simpson, A.G.B. and Eglit, Y. (2016) Protist Diversification. *Encyclopedia of Evolutionary Biology*, pp. 344–360. doi:10.1016/B978-0-12-800049-6.00247-X.

Siscar-Lewin, S., Hube, B. and Brunke, S. (2019) Antivirulence and avirulence genes in human pathogenic fungi. *Virulence*, 10 (1): 935. doi:10.1080/21505594.2019.1688753.

Sitkiewicz, I. (2018) How to become a killer, or is it all accidental? Virulence strategies in oral streptococci. *Molecular Oral Microbiology*, 33: 1–12. doi:10.1111/omi.12192.

Škaloud, P., Friedl, T., Hallmann, C., et al. (2016) Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus Dictyochloropsis (Trebouxiophyceae, Chlorophyta). *Journal of phycology*, 52 (4): 599–617. doi:10.1111/JPY.12422.

Sobukawa, H., Kano, R., Ito, T., et al. (2011) *In vitro* susceptibility of *Prototheca zopfii* genotypes 1 and 2. *Medical Mycology*, 49 (2): 222–224. doi:10.3109/13693786.2010.511285.

Soil Survey Division Staff (2017) "Examination and Description of Soil Profiles." In *U.S. Department of Agriculture Handbook: Soil Survey Manual*. Available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054253 (Downloaded: 6 September 2022).

Southworth, D. (1974) Solubility of Pollen Exines. *American Journal of Botany*, 61 (1): 36–44. doi:10.1002/j.1537-2197.1974.tb06025.x.

Sprouffske, K. and Wagner, A. (2016) Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17: 172.

doi:10.1186/s12859-016-1016-7.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30 (9): 1312–1313. doi:10.1093/BIOINFORMATICS/BTU033.

Stark, J.R., Cardon, Z.G. and Peredo, E.L. (2020) Extraction of high-quality, high-molecular-weight DNA depends heavily on cell homogenization methods in green microalgae. *Applications in Plant Sciences*, 8 (3): e11333. doi:10.1002/aps3.11333.

Stat, M., Morris, E. and Gates, R.D. (2008) Functional diversity in coral-dinoflagellate symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (27): 9256–9261. doi:10.1073/PNAS.0801328105/SUPPL_FILE/0801328105SI.PDF.

Steenbergen, J.N., Shuman, H.A. and Casadevall, A. (2001) *Cryptococcus neoformans* interactions with amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (26): 15245–15250. doi:10.1073/pnas.261418798.

Stenner, V.J., MacKay, B., King, T., et al. (2007) Protothecosis in 17 Australian dogs and a review of the canine literature. *Medical Mycology*, 45 (3): 249–266. doi:10.1080/13693780601187158.

Stockinger, B.G. and Doster, A.R. (2017) Disseminated Protothecosis in a Ruwenzori Long-Haired Fruit Bat (*Rousettus lanosus*). *Journal of Zoo and Wildlife Medicine*, 48 (4): 1260–1263. doi:10.1638/2017-0070r.1.

Stoecker, D.K. (1999) Mixotrophy among Dinoflagellates. *Journal of Eukaryotic Microbiology*, 46 (4): 397–401. doi:10.1111/J.1550-7408.1999.TB04619.X.

Suarez-Trujillo, A., Hoang, N., Robinson, L., et al. (2022) Effect of circadian system disruption on the concentration and daily oscillations of cortisol, progesterone, melatonin, serotonin,

growth hormone, and core body temperature in periparturient dairy cattle. *Journal of Dairy Science*, 105 (3): 2651–2668. doi:10.3168/JDS.2021-20691.

Sud, I.J. and Feingold, D.S. (1979) Lipid composition and sensitivity of *Prototheca wickerhamii* to membrane-active antimicrobial agents. *Antimicrobial Agents and Chemotherapy*, 16 (4): 486–490. doi:10.1128/AAC.16.4.486.

Sudman, M.S. and Kaplan, W. (1973) Identification of the *Prototheca* species by immunofluorescence. *Applied microbiology*, 25 (6): 981–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4577492 (Accessed: 4 May 2020).

Suvajdžić, B., Dragan, V., Karabasil, N., et al. (2017) Molecular identification of *Prototheca zopfii* genotype 2 mastitis isolates and their influence on the milk somatic cell count. *Veterinarski arhiv*, 87 (3): 249–258. doi:10.24099/vet.arhiv.151219.

Suzuki, S., Endoh, R., Manabe, R.I., et al. (2018) Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae *Prototheca*. *Scientific Reports*, 8: 940. doi:10.1038/s41598-017-18378-8.

Suzuki, T. (2006) Electrophoretic Separation of Chromosomes in an Achlorophyllous Microalga, *Prototheca zopfii*. *Journal of Toyo University*, 50: 13–16.

Syasina, I.G., Kukhlevsky, A.D., Kovaleva, A.L., et al. (2012) Phylogenetic and morphological characterization of the green alga infesting the horse mussel *Modiolus modiolus* from Vityaz Bay (Peter the Great Bay, Sea of Japan). *Journal of Invertebrate Pathology*, 111 (2): 175–181. doi:10.1016/j.jip.2012.08.001.

Takaki, K., Okada, K., Umeno, M., et al. (1996) Chronic prototheca meningitis. *Scandinavian Journal of Infectious Diseases*, 28 (3): 321–323. doi:10.3109/00365549609027183.

Taniyama, H., Okamoto, F., Kurosawa, T., et al. (1994) Disseminated Protothccosis Caused by *Prototheca zopfii* in a Cow. *Veterinary Pathology*, 31 (1): 123–125.

doi:10.1177/030098589403100120.

Tartar, A. (2013) The non-photosynthetic algae *Helicosporidium* spp.: Emergence of a novel group of insect pathogens. *Insects*, 4 (3): 375–391. doi:10.3390/insects4030375.

Tashakkori, N., Rahmani, H.K. and Khoramian, B. (2022) Genotypic and phenotypic diversity of *Prototheca* spp. recovered from bovine mastitis in terms of antimicrobial resistance and biofilm formation ability. *BMC Veterinary Research*, 18: 452. doi:10.1186/S12917-022-03546-4.

I, J.S.K., Wilke, A.E., Overstall, S.M., et al. (2021) *Prototheca wickerhamii* breast implant infection after reconstructive surgery: a new level of complexity. *Medical Mycology Case Reports*, 34: 22–26. doi:10.1016/J.MMCR.2021.08.003.

Telkes, G., Dezsö, K., Doros, A., et al. (2018) Successful Treatment of the Gastrointestinal Manifestation of *Prototheca* in a Kidney Transplant Recipient: A Case Report. *Transplantation Proceedings*, 50 (10): 3928–3931. doi:10.1016/j.transproceed.2018.07.026.

Ten, K.E., Muzahid, N.H., Rahman, S., et al. (2023) Use of the waxworm *Galleria mellonella* larvae as an infection model to study *Acinetobacter baumannii*. *PLOS ONE*, 18 (4): e0283960. doi:10.1371/JOURNAL.PONE.0283960.

Tesson, S.V.M., Skjøth, C.A., Šantl-Temkiv, T., et al. (2016) Airborne microalgae: Insights, opportunities, and challenges. *Applied and Environmental Microbiology*, 82 (7): 1978–1991. doi:10.1128/AEM.03333-15/ASSET/11928430-246E-43C8-9890-941744439536/ASSETS/GRAPHIC/ZAM9991170270001.JPEG.

ThermoScientific (2015) Thermo Scientific PathoProof Complete-16 kit Instructions for Use. *Abcam Protocol Book*, pp. 1–82. Available at: www.abcam.com.

Thiele, D. and Bergmann, A. (2002) Protothecosis in human medicine. *International Journal of Hygiene and Environmental Health*, 204 (5–6): 297–302. doi:10.1078/1438-4639-00105.

Thompson, G., Silva, E., Marques, S., et al. (2009) Algaemia in a dairy cow by *Prototheca blaschkeae*. *Medical Mycology*, 47 (5): 527–531. doi:10.1080/13693780802566341.

Tian, F., Yang, D.C., Meng, Y.Q., et al. (2020) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research*, 48 (D1): D1104–D1113. doi:10.1093/NAR/GKZ1020.

Tillich, M., Lehwark, P., Pellizzer, T., et al. (2017) GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45 (W1): W6–W11. doi:10.1093/NAR/GKX391.

Todd, J.R., King, J.W., Oberle, A., et al. (2012) Protothecosis: Report of a case with 20-year follow-up, and review of previously published cases. *Medical Mycology*, 50 (7): 673–689. doi:10.3109/13693786.2012.677862.

Todd, J.R., Matsumoto, T., Ueno, R., et al. (2018) Medical phycology 2017. *Medical Mycology*, 56 (suppl_1): S188–S204. doi:10.1093/mmy/myx162.

Tomori, M., Nagamine, T., Miyamoto, T., et al. (2019) Evaluation of the Immunomodulatory Effects of Fucoidan Derived from Cladosiphon Okamuranus Tokida in Mice. *Marine Drugs*, 17 (10): 547. doi:10.3390/MD17100547.

Torres, H.A., Bodey, G.P., Tarrand, J.J., et al. (2003) Protothecosis in patients with cancer: Case series and literature review. *Clinical Microbiology and Infection*, 9: 786–792. doi:10.1046/j.1469-0691.2003.00600.x.

Toso, M.A. and Omoto, C.K. (2007) Gregarina niphandrodes may Lack Both a Plastid Genome and Organelle. *Journal of Eukaryotic Microbiology*, 54 (1): 66–72. doi:10.1111/J.1550-7408.2006.00229.X.

Toyotome, T. and Matsui, S. (2022) Analysis of *Prototheca* and yeast species isolated from bulk tank milk collected in Tokachi District, Japan. *Journal of Dairy Science*, 723 (2017): 2014–2018. doi:10.3168/jds.2022-21781.

Trespalacios Sierra, A., Arroyo-Salgado, B., Rodriguez-Blanco, J., et al. (2022) Systemic protothecosis in an immunocompetent patient. *Tropical Diseases, Travel Medicine and Vaccines*, 8: 23. doi:10.1186/S40794-022-00180-8.

Treves, H., Raanan, H., Finkel, O.M., et al. (2013) A newly isolated *Chlorella* sp. from desert sand crusts exhibits a unique resistance to excess light intensity. *FEMS Microbiology Ecology*, 86 (3): 373–380. doi:10.1111/1574-6941.12162.

Triemer, R.E. and Zakryś, B. (2015) "Photosynthetic Euglenoids." In *Freshwater Algae of North America: Ecology and Classification*. 2nd ed. Academic Press. pp. 459–483. doi:10.1016/B978-0-12-385876-4.00010-4.

Trzaska, W.J., Correia, J.N., Villegas, M.T., et al. (2015) pH manipulation as a novel strategy for treating mucormycosis. *Antimicrobial Agents and Chemotherapy*, 59 (11): 6968–6974. doi:10.1128/AAC.01366-15.

Ueno, R. (2009) Visualization of sporopollenin-containing pathogenic green micro-alga *Prototheca wickerhamii* by fluorescent in situ hybridization (FISH). *Canadian Journal of Microbiology*, 55 (4): 465–472. doi:10.1139/W08-155.

Ueno, R., Hanagata, N., Urano, N., et al. (2005) Molecular phylogeny and phenotypic variation in the heterotrophic green algal genus *Prototheca* (Trebouxiophyceae, Chlorophyta). *Journal of Phycology*, 41 (6): 1268–1280. doi:10.1111/j.1529-8817.2005.00142.x.

Ueno, R., Urano, N., Suzuki, M., et al. (2002) Isolation, characterization, and fermentative pattern of a novel thermotolerant *Prototheca zopfii* var. *hydrocarbonea* strain producing ethanol and CO2 from glucose at 40°C. *Archives of Microbiology*, 177 (3): 244–250. doi:10.1007/s00203-001-0384-0.

Uribe-Quero, E. and Rosales, C. (2017) Control of phagocytosis by microbial pathogens.

Frontiers in Immunology, 8: 24. doi:10.3389/fimmu.2017.01368.

UTEX (n.d.) *TAP Medium recipe*. Available at: https://utex.org/products/tap-medium?variant=30991736897626#recipe (Accessed: 31 August 2022).

Vasco-Julio, D., Huilca-Ibarra, M., Ledesma, Y., et al. (2023) The Development of a Multiplex PCR Assay for Fast and Cost-Effective Identification of the Five Most Significant Pathogenic *Prototheca* Species. *Pathogens*, 12 (8): 1018. doi:10.3390/PATHOGENS12081018.

Vazač, J., Füssy, Z., Hladová, I., et al. (2018) Ploidy and Number of Chromosomes in the Alveolate Alga *Chromera velia*. *Protist*, 169 (1): 53–63. doi:10.1016/J.PROTIS.2017.12.001.

Viruel, J., Conejero, M., Hidalgo, O., et al. (2019) A Target Capture-Based Method to Estimate Ploidy From Herbarium Specimens. *Frontiers in Plant Science*, 10: 467104. doi:10.3389/FPLS.2019.00937/BIBTEX.

Wang, F., Kong, L.M., Xie, Y.Y., et al. (2021) Purification, structural characterization, and biological activities of degraded polysaccharides from *Porphyra yezoensis*. *Journal of Food Biochemistry*, 45 (4): e13661. doi:10.1111/JFBC.13661.

Wang, H., Qi, M. and Cutler, A.J. (1993) A simple method of preparing plant samples for PCR. *Nucleic Acids Research*, 21 (17): 4153. doi:10.1093/NAR/21.17.4153.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., et al. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25 (9): 1189–1191. doi:10.1093/BIOINFORMATICS/BTP033.

Wawron, W., Bochniarz, M., Piech, T., et al. (2013) Outbreak of protothecal mastitis in a herd of dairy cows in Poland. *Bulletin of the Veterinary Institute in Pulawy*, 57 (3): 335–339. doi:10.2478/bvip-2013-0058.

Welter, D.K., Ruaud, A., Henseler, Z.M., et al. (2021) Free-Living, Psychrotrophic Bacteria of the Genus *Psychrobacter* Are Descendants of Pathobionts. *mSystems*, 6 (2).

doi:10.1128/msystems.00258-21.

Wesselowski, S., Janacek, B., Landsgaard, K., et al. (2022) Pancarditis as the sole clinical manifestation of protothecosis in a Boxer dog. *Journal of Veterinary Cardiology*, 41: 128–133. doi:10.1016/J.JVC.2022.02.004.

Westblade, L.F., Ranganath, S., Dunne, W.M., et al. (2015) Infection with a Chlorophyllic Eukaryote after a Traumatic Freshwater Injury. *New England Journal of Medicine*, 372 (10): 982–984. doi:10.1056/NEJMc1401816.

Whipple, K.M., Wellehan, J.F., Jeon, A.B., et al. (2020) Cytologic, histologic, microbiologic, and electron microscopic characterization of a canine *Prototheca wickerhamii* infection. *Veterinary Clinical Pathology*, 00: 1–7. doi:10.1111/vcp.12864.

Wingett, S.W. and Andrews, S. (2018) FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, 7: 1338. doi:10.12688/f1000research.15931.2.

Wu, T., Li, L., Jiang, X., et al. (2019) Sequencing and comparative analysis of three *Chlorella* genomes provide insights into strain-specific adaptation to wastewater. *Scientific Reports*, 9: 9514. doi:10.1038/s41598-019-45511-6.

Yagnik, K., Bossé, R., Reppucci, J., et al. (2019) Case report: Olecranon bursitis due to *Prototheca wickerhamii* in an immunocompromised patient. *American Journal of Tropical Medicine and Hygiene*, 100 (3): 703–705. doi:10.4269/ajtmh.18-0818.

Yamashita, M., Ikeda, M., Kato, I., et al. (2022) Protothecosis in the mucosa of the pharynx mimicking pharyngeal cancer in an immunocompetent individual: a case report. *Annals of Clinical Microbiology and Antimicrobials*, 21 (1): 5. doi:10.1186/S12941-022-00495-6.

Yang, H., Genot, B., Duhamel, S., et al. (2022) Organismal and cellular interactions in vertebrate-alga symbioses. *Biochemical Society Transactions*, 50 (1): 609–620. doi:10.1042/BST20210153.

Yang, J., Zhu, J., Kudinha, T., et al. (2020) Differences in in vitro interactions between macrophages with pathogenic and environmental strains of Prototheca. *Future Microbiology*, 15: 427–436. doi:10.2217/fmb-2019-0238.

Yanong, R.P.E., Francis-Floyd, R., Curtis, E., et al. (2002) Algal dermatitis in cichlids. *Journal of the American Veterinary Medical Association*, 220 (9): 1353–1358. doi:10.2460/javma.2002.220.1353.

Yousef, H., Ahangar, E.R. and Varacallo, M. (2022) Physiology, Thermal Regulation. *StatPearls*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK499843/ (Accessed: 6 April 2023).

Yu, J., Li, Z. and Brand, J.J. (2009) Characterization of a green alga isolated from infected human external tissue. *Phycological Research*, 57 (4): 251–258. doi:10.1111/j.1440-1835.2009.00544.x.

Yun, C.H., Jeong, J.H., Ryu, H.R., et al. (2016) Cutaneous protothecosis responds rapidly to voriconazole. *International Journal of Dermatology*, 55 (12): 1373–1377. doi:10.1111/ijd.13160.

Zaitz, C., Miranda Godoy, A., de Sousa, V.M., et al. (2006) Onychoprotothecosis: Report of the first case in Brazil. *International Journal of Dermatology*, 45 (9): 1071–1073. doi:10.1111/j.1365-4632.2006.03038.x.

Zak, I., Jagielski, T., Kwiatkowski, S., et al. (2012) *Prototheca wickerhamii* as a cause of neuroinfection in a child with congenital hydrocephalus. First case of human protothecosis in Poland. *Diagnostic Microbiology and Infectious Disease*. doi:10.1016/j.diagmicrobio.2012.06.015.

Zecconi, A., Dell'Orco, F., Rizzi, N., et al. (2020) Cross-sectional study on the prevalence of contagious pathogens in bulk tank milk and their effects on somatic cell counts and milk

yield. *Italian Journal of Animal Science*, 19 (1): 66–74. doi:10.1080/1828051X.2019.1693282.

Zeitouni, N.E., Chotikatum, S., Köckritz-Blickwede, M. von, et al. (2016) The impact of hypoxia on intestinal epithelial cell functions: consequences for invasion by bacterial pathogens. *Molecular and Cellular Pediatrics*, 3 (1). doi:10.1186/S40348-016-0041-Y.

Zenewicz, L.A. (2017) Oxygen levels and immunological studies. *Frontiers in Immunology*, 8 (MAR): 252040. doi:10.3389/FIMMU.2017.00324/BIBTEX.

Zeng, X., Kudinha, T., Kong, F., et al. (2019) Comparative genome and transcriptome study of the gene expression difference between pathogenic and environmental strains of *Prototheca zopfii*. *Frontiers in Microbiology*, 10 (MAR): 443. doi:10.3389/fmicb.2019.00443.

Zhang, F., Zhuang, J., Li, Z., et al. (2022) Nanoparticle-modified microrobots for in vivo antibiotic delivery to treat acute bacterial pneumonia. *Nature Materials*, 21 (11): 1324–1332. doi:10.1038/s41563-022-01360-9.

Zhang, M., Zhao, M., Qing, Y., et al. (2020) Study on immunostimulatory activity and extraction process optimization of polysaccharides from *Caulerpa lentillifera*. *International Journal of Biological Macromolecules*, 143: 677–684. doi:10.1016/J.IJBIOMAC.2019.10.042.

Zhang, Q., Weng, X., Li, L., et al. (2010) An unusual case of granulomatous lymphadenitis due to *Prototheca zopfii* var. *portoricensis* in an immunocompetent man in China. *International Journal of Infectious Diseases*, 14 (SUPPL. 3): e32–e35. doi:10.1016/j.ijid.2009.10.003.

Zhang, Q.Q., Zhu, L.P., Weng, X.H., et al. (2007) Meningitis due to *Prototheca wickerhamii*: Rare case in China. *Medical Mycology*, 45 (1): 85–88. doi:10.1080/13693780601003835.

Zhang, Z. (2022) *Variation in disinfectant resistance between* Prototheca *species*. University of Birmingham.

Zhao, F., Chen, M. and Fu, Y. (2020) Multiple cutaneous infections caused by *Prototheca wickerhamii*. *Journal of Clinical Laboratory Analysis*, 00: e23492. doi:10.1002/jcla.23492.

Zhao, W., He, F., Barkema, H.W., et al. (2021) *Prototheca* spp. induce an inflammatory response via mtROS-mediated activation of NF-κB and NLRP3 inflammasome pathways in bovine mammary epithelial cell cultures. *Veterinary Research*, 52 (1): 144. doi:10.1186/S13567-021-01014-9.

Zuykov, M., Anderson, J., Archambault, P., et al. (2018) *Mytilus trossulus* and hybrid (*M. edulis-M. trossulus*) – New hosts organisms for pathogenic microalgae *Coccomyxa* sp. from the Estuary and northwestern Gulf of St. Lawrence, Canada. *Journal of Invertebrate Pathology*, 153 (September 2017): 145–146. doi:10.1016/j.jip.2018.02.017.

10) Appendices

Appendix 1. Table of AHP strains possessed by the HAPI lab, including: the internal strain identifier, used to refer to strains in this thesis; the species classification of the strain, as determined by PCR-RFLP for Prototheca *strains and as reported by the supplier for other strains (except for HA3); whether the strain was originally isolated from an infection, which includes symbiosis; which possible host the species was isolated from, which can include which host environment the species was isolated from if the strain was not a clinical isolate; and any external strain names that might have been used to refer to the same strain in other literature.*

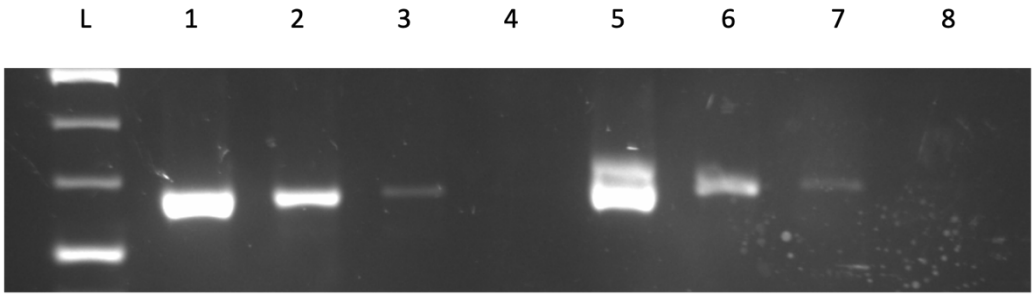
Internal Strain ID	Species	From Infection	Host	Other Names
HP1	*Prototheca wickerhamii*	TRUE	Human	ATCC 30395
HP2	*Prototheca ciferrii*	FALSE	Cattle	SAG2063
HP3	*Prototheca bovis*	TRUE	Cattle	SAG2021
HP4	*Prototheca blaschkeae*	TRUE	Human	SAG2064
HP5	*Prototheca wickerhamii*	TRUE	Human	CBS 157.74, JCM 9643
HP6	*Prototheca wickerhamii*	TRUE	Human	CBS 344.82, JCM 9644
HP7	*Prototheca wickerhamii*	TRUE	Human	
HP8	*Prototheca wickerhamii*	TRUE	Goat	
HP9	*Prototheca wickerhamii*	TRUE	Human	
HP10	*Prototheca wickerhamii*	TRUE	Dog	
HP11	*Prototheca wickerhamii*	TRUE	Cat	VB974275
HP12	*Prototheca wickerhamii*	TRUE	Dog	VB981821
HP13	*Prototheca wickerhamii*	TRUE	Cat	
HP14	*Prototheca bovis*	TRUE	Human	JCM, 2004;42:4907-4908
HP15	*Prototheca bovis*	TRUE	Dog	
HP16	*Prototheca bovis*	TRUE	Dog	
HP17	*Prototheca bovis*	TRUE	Dog	257379
HP18	*Prototheca bovis*	TRUE	Human	
HP19	*Prototheca bovis*	TRUE	Dog	VM894562
HP20	*Prototheca bovis*	TRUE	Dog	
HP21	*Prototheca bovis*	TRUE	Dog	
HP22	*Prototheca bovis*	TRUE	Dog	

HP23	*Prototheca bovis*	TRUE	Dog	
HP24	*Prototheca bovis*	TRUE	Dog	
HP25	*Prototheca bovis*	UNKNOWN	UNKNOWN	UNKNOWN
HP26	*Prototheca bovis*	UNKNOWN	UNKNOWN	UNKNOWN
HP27	*Prototheca miyajii*	TRUE	Human	IFM 53848
HP28	*Prototheca cutis*	TRUE	Human	ATCC PRA-338; JCM 15793
HP29	*Prototheca tumulicola*	FALSE	None	JCM 31123
HP30	*Prototheca tumulicola*	UNKNOWN	UNKNOWN	UNKNOWN
HP31	*Prototheca paracutis*	FALSE	None	JCM 32112; YMTW3-1; TBRC8745
HP32	*Prototheca cookei*	FALSE	None	JCM 8557
HP33	*Prototheca ciferrii*	TRUE	Cattle	P547
HP34	*Prototheca ciferrii*	UNKNOWN	Cattle	125N
HP35	*Prototheca ciferrii*	FALSE	Cattle	48/IV
HP36	*Prototheca ciferrii*	FALSE	None	JCM 8556
HP37	*Prototheca ciferrii*	TRUE	Human	ATCC MYA-4771
HP38	*Prototheca ciferrii*	TRUE	Human	ATCC MYA-4770
HP39	*Prototheca ciferrii*	FALSE	None	JCM 9346
HP40	*Prototheca bovis*	TRUE	Cattle	E9
HP41	*Prototheca bovis*	FALSE	Cattle	SR4
HP42	*Prototheca bovis*	FALSE	Cattle	O1
HP43	*Prototheca bovis*	TRUE	Cattle	P535
HP44	*Prototheca bovis*	TRUE	Cattle	SAG 2290
HP45	*Prototheca blaschkeae*	TRUE	Cattle	PR-24

HP46	*Prototheca blaschkeae*	TRUE	Cattle	BEL-2
HP47	*Prototheca blaschkeae*	FALSE	Cattle	OD
HP48	*Prototheca blaschkeae*	TRUE	Cattle	560
HP49	*Prototheca wickerhamii*	FALSE	None	W4
HP50	*Prototheca wickerhamii*	TRUE	Cattle	UTEX 1437; ATCC 16523; NRRL Y-2464
HP51	*Prototheca moriformis*	UNKNOWN	UNKNOWN	UNKNOWN
HP52	*Prototheca wickerhamii*	FALSE	Human	JCM 9645; CBS 608.66
HP53	*Prototheca xanthoriae*	FALSE	None	CBS 612.66
HP54	*Prototheca xanthoriae*	FALSE	None	JCM 9729; CBS 611.66
HA1	*Auxenochlorella protothecoides*	FALSE	None	CCAP 211/8D; SAG 211-8d
HA2	*Auxenochlorella protothecoides*	FALSE	None	CCAP 211/7A
HA3	*Leptochlorella* sp.	FALSE	None	CCAP 211/54
HA4	*Auxenochlorella protothecoides*	FALSE	None	CCAP 211/7C
HA5	*Auxenochlorella symbiontica*	SYMBIOSIS	*Hydra viridis*	CCAP 211/61
HA6	*Auxenochlorella* sp. 1	UNKNOWN	UNKNOWN	
HA7	*Auxenochlorella* sp. 2	UNKNOWN	UNKNOWN	
HH1	*Helicosporidium* sp. Cs	TRUE	*Cyrtobagus salviniae*	Cs-1; ATCC PRA-374; W strain
HH2	*Helicosporidium* sp. Dm	TRUE	*Dendroctonus micans*	Dm-1; ATCC PRA-375; T strain
HH3	*Helicosporidium* sp. Sj	TRUE	*Simulium jonesi*	Sj-1; ATCC 50920; B strain

Appendix 2. Table of milk samples provided by the NML, including information on when the sample was taken, which pathogens were detected in the sample using the Thermo Scientific PathoProof Complete-16 kit, and whether the sample is of milk pooled between individuals. Location information has not been included in this table to prevent identification of affected farms.

Sample number	Collection date	Prototheca detected by NML	Other pathogens detected by NML	Bulk milk
1	Sep-21	True	True	True
2	Sep-21	True	True	True
3	Sep-21	True	True	True
4	Sep-21	True	True	True
5	Oct-21	True	True	True
6	Oct-21	True	True	True
7	Nov-21	True	True	True
8	Dec-21	False	True	True
9	Dec-21	False	True	True
10	Dec-21	False	True	True
11	Dec-21	False	True	True
12	Dec-21	False	True	True
13	Dec-21	False	True	True
14	Feb-22	False	True	True
15	Feb-22	True	False	False



Appendix 3. PCR products of the cytb gene using DNA purified from milk inoculated with known densities of P. bovis as template. Wells were inoculated to 10^4 (1 and 5), 10^3 (2 and 6), 10^2 (3 and 7) or 10^1 (4 and 8) cells per ml. Wells 1-4 were inoculated with cells one month prior to DNA extraction and were kept at -20°C for the intervening time. Wells 5-8 were inoculated one day prior to DNA extraction and were frozen at -20°C overnight.

Appendix 4. List of taxa searched for mitochondrial DNA to align with cytb primers. Lists of genera are a superset of genera listed on NCBI taxonomy and those found on Wikipedia, excluding members of the AHP lineage as well as Chlorella *and* Micractinium *as sequences for these taxa were already available. The last search for genomes was performed in March 2023.*

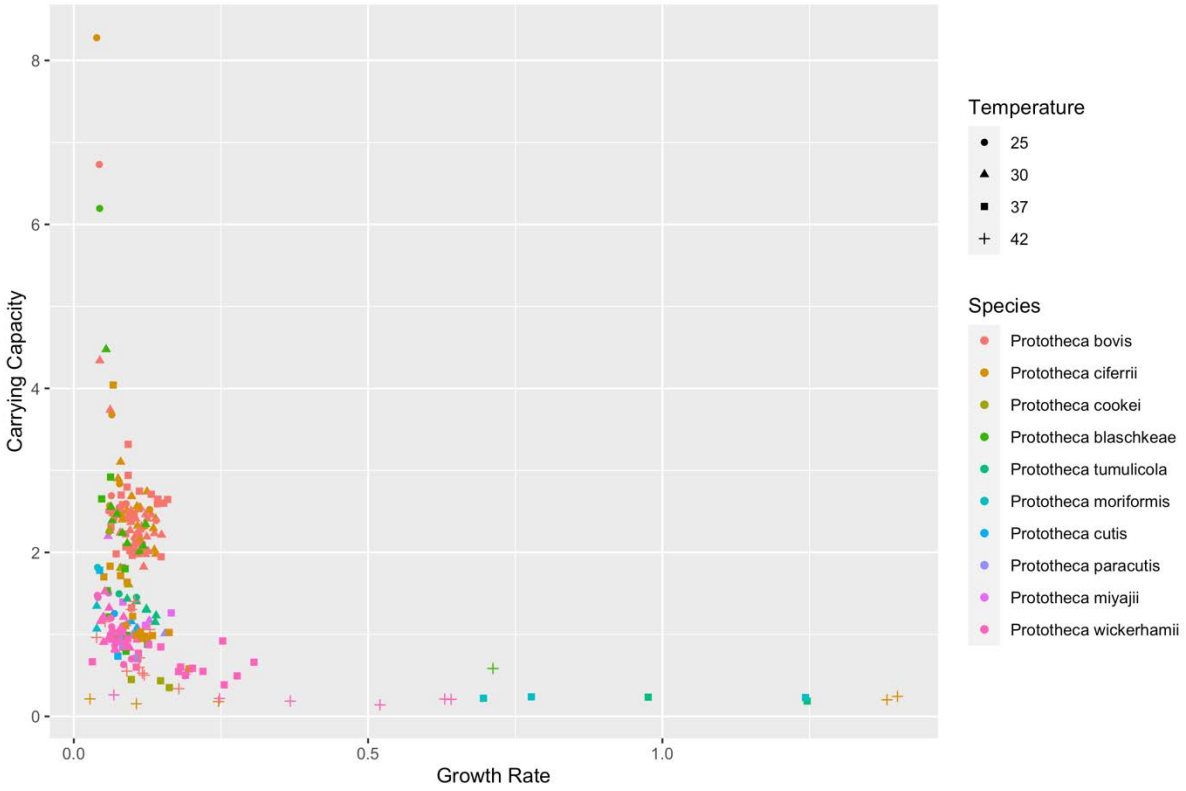
Family	Genera
Chlorellaceae	*Acanthosphaera; Actinastrum; Apatococcus; Apodococcus; Carolibrandtia; Catena; Chloroparva; Closteriopsis; Compactochlorella; Coronacoccus; Coronastrum; Cylindrocelis; Diacanthos; Dicellula; Dicloster; Dictyosphaerium; Didymogenes; Edaphochloris; Endolithella; Fissuricella; Follicularia; Geminella; Gloeotila; Golenkiniopsis; Graesiella; Hegewaldia; Heynigia; Hindakia; Hormospora; Kalenjinia; Keratococcus; Kermatia; Leptochlorella; Lobosphaeropsis; Marasphaerium; Marinichlorella; Marvania; Masaia; Meyerella; Mucidosphaerium; Muriella; Nannochloris; Nanochlorum; Palmellochaete; Parachlorella; Planktochlorella; Podohedra; Pseudochloris; Pseudosiderocelopsis; Pumiliosphaera; Siderocelis; Zoochlorella;* unclassified Chlorellaceae
Leptosiraceae	*Pseudopleurococcus*
Phyllosiphonaceae	None
Oocystaceae	*Amphikrikos; Catenocystis; Cerasterias; Chodatella; Chodatellopsis; Chondrosphaera; Coenolamellus; Conradia; Crucigeniella; Cryocystis; Dactylococcus; Didymocystis; Ecballocystopsis; Ecdysichlamys; Echinocoleum; Elongatocystis; Eremosphaera; Ettliella; Fotterella; Franceia; Glochiococcus; Gloeocystopsis; Gloeotaenium; Gloxidium; Granulocystis; Granulocystopsis; Hemichloris; Hyalochlorella; Jaagichlorella; Juranyiella; Keriochlamys; Kirchneriellosaccus; Makinoella; Makinoelloideae; Micracantha; Mycacanthococcus; Mycotetraedron; Neglectella (done); Nephrochlamys; Nephrocytium; Oocystaenium; Oocystella; Oocystidium; Oocystis; Oocystoideae; Oocystopsis; Oonephris; Ooplanctella; Pachycladella; Palmellococcus; Pilidiocystis; Planctonema; Planctonemopsis; Planktosphaerella; Pseudobohlina; Pseudochlorococcum; Pseudococcomyxa; Quadricoccus; Rayssiella; Reinschiella; Rhombocystis; Saturnella; Schizochlamydella; Scotiella; Selenoderma; Sestosoma; Siderocystopsis; Tetrachlorella; Thelesphaera; Trigonidiella; Trochiscia*; Oocystaceae incertae sedis; unclassified Oocystaceae.
Rhopalosolenaceae	*Rhopalosolen*
incertae sedis	
Unclassified Chlorellales	

Appendix 5. Number of replicates per strain for each condition for all OD growth curves.

Species	Strain	25°C	30°C	33°C	37°C	42°C	pH 2	pH 3	pH 4
Prototheca wickerhamii	HP1	1	2	3	2	1	3	3	3
Prototheca ciferrii	HP2	1	2	3	2	1	2	2	2
Prototheca bovis	HP3	1	2	3	2	1	0	0	0
Prototheca blaschkeae	HP4	1	2	3	2	1	3	3	3
Prototheca wickerhamii	HP5	1	2	0	2	1	2	2	2
Prototheca wickerhamii	HP6	1	2	0	2	1	2	2	2
Prototheca wickerhamii	HP7	1	2	0	2	1	0	0	0
Prototheca wickerhamii	HP8	1	2	0	2	1	0	0	0
Prototheca wickerhamii	HP9	1	2	0	2	1	0	0	0
Prototheca wickerhamii	HP10	1	2	0	2	1	0	0	0
Prototheca wickerhamii	HP12	1	2	0	2	1	0	0	0
Prototheca wickerhamii	HP13	1	2	0	2	1	2	2	2
Prototheca bovis	HP14	1	2	0	2	1	2	2	2
Prototheca bovis	HP15	1	2	0	2	1	0	0	0
Prototheca bovis	HP17	1	2	0	2	1	0	0	0
Prototheca bovis	HP18	1	2	3	2	1	2	2	2
Prototheca bovis	HP19	1	2	0	2	1	0	0	0
Prototheca bovis	HP20	1	2	0	2	1	0	0	0
Prototheca bovis	HP25	1	2	0	2	1	2	2	2
Prototheca bovis	HP26	1	2	0	2	1	2	2	2
Prototheca miyajii	HP27	1	3	3	3	1	3	3	3
Prototheca cutis	HP28	1	3	3	3	1	3	3	3
Prototheca tumulicola	HP29	1	3	3	3	1	3	3	3
Prototheca tumulicola	HP30	1	3	3	3	1	0	0	0
Prototheca paracutis	HP31	1	2	3	3	1	3	3	3
Prototheca cookei	HP32	1	3	3	3	1	3	3	3
Prototheca ciferrii	HP33	1	3	0	3	1	2	2	2
Prototheca ciferrii	HP35	1	2	0	2	1	0	0	0
Prototheca ciferrii	HP36	1	2	0	2	1	2	2	2
Prototheca ciferrii	HP37	2	2	3	2	2	2	2	2
Prototheca ciferrii	HP38	1	2	0	2	1	0	0	0
Prototheca ciferrii	HP39	1	2	0	2	1	2	2	2
Prototheca bovis	HP40	1	3	3	3	1	2	2	2
Prototheca bovis	HP41	1	2	0	2	1	2	2	2
Prototheca bovis	HP42	1	2	0	2	1	2	2	2
Prototheca bovis	HP43	1	2	0	2	1	0	0	0
Prototheca bovis	HP44	1	2	0	2	1	0	0	0
Prototheca blaschkeae	HP45	1	3	3	3	1	3	3	3
Prototheca blaschkeae	HP46	0	2	0	2	0	2	2	2
Prototheca blaschkeae	HP47	1	2	0	2	1	2	2	2
Prototheca wickerhamii	HP49	1	2	0	2	1	2	2	2

Prototheca wickerhamii	HP50	1	2	0	2	1	2	2	2
Prototheca moriformis	HP51	1	3	3	3	1	3	3	3
Prototheca wickerhamii	HP52	0	0	3	0	0	0	0	0
Prototheca xanthoriae	HP53	1	3	3	3	1	3	3	3
Prototheca xanthoriae	HP54	1	3	3	3	1	3	3	3
Auxenochlorella protothecoides	HA1	1	3	3	3	1	3	3	3
Auxenochlorella protothecoides	HA2	0	0	3	0	0	2	2	2
Auxenochlorella symbiontica	HA5	0	0	3	0	0	3	3	3
Auxenochlorella sp. 1	HA6	0	0	3	0	0	0	0	0
Auxenochlorella sp. 2	HA7	0	1	3	1	0	3	3	3

Appendix 6. Scatter plot depicting the relationship between carrying capacity and growth rate for growthcurver models that were not deemed to be outliers. Only models derived from OD_{750} data from cultures exposed to temperature stress are shown.

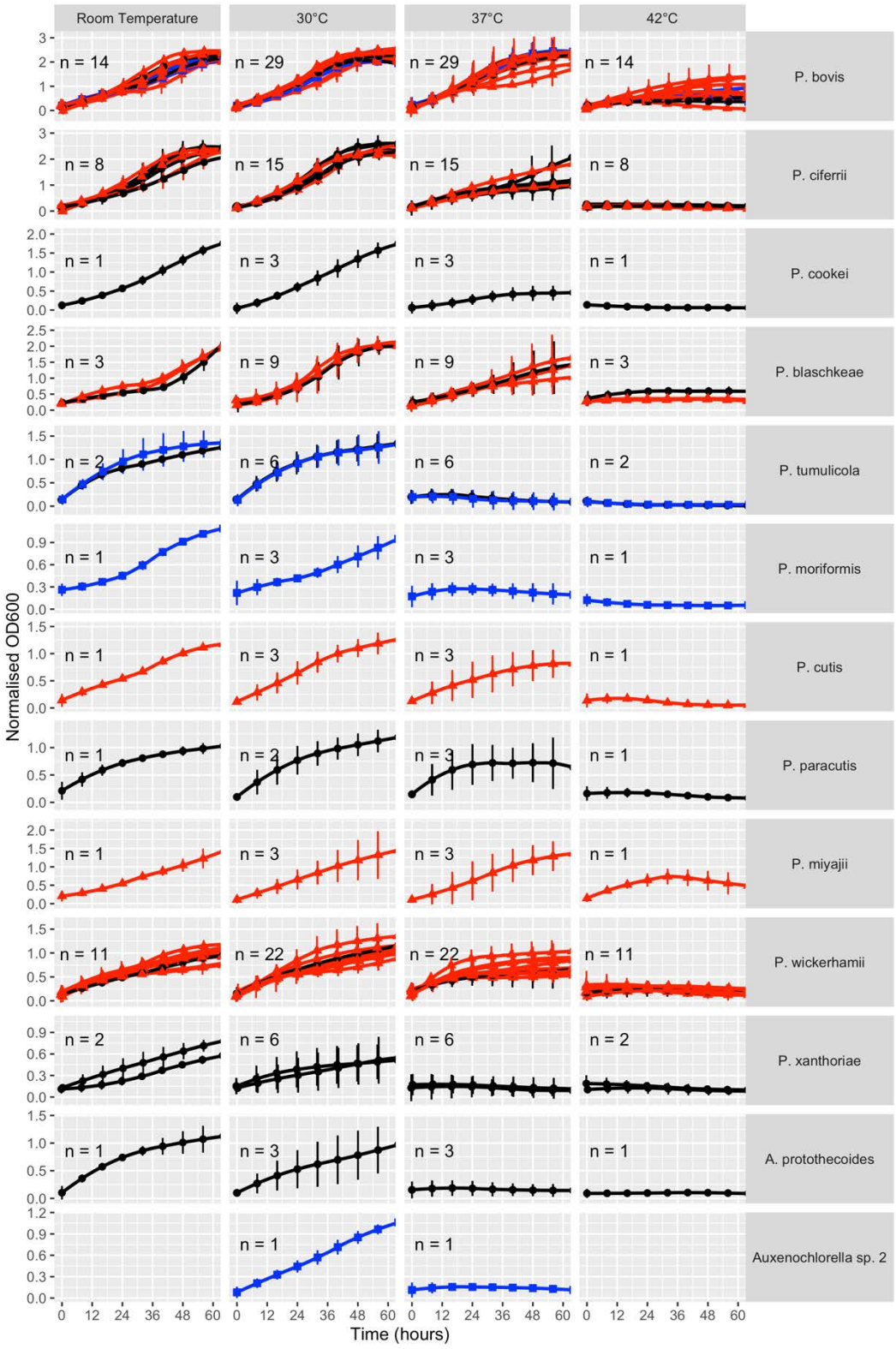


Appendix 7. Detailed breakdown of the number of wells exposed to temperature stress that failed to be fitted to a model, as well as the reason for failure. Failures are grouped according to species and temperature. Only conditions under which at least one model failed are included.

Species	Temperature (°C)	Number of wells in total	Failures due to carrying capacity	Failures due to growth rate	Failures due to sigma	Failure to fit a model	Number of valid wells	Number of invalid wells	Proportion of wells that are invalid
P. blaschkeae	25	6	5	0	0	0	1	5	0.8333
P. bovis	25	28	2	0	0	0	26	2	0.0714
P. xanthoriae	25	4	1	0	0	0	3	1	0.25
A. protothecoides	30	7	1	0	0	0	6	1	0.1429
P. moriformis	30	7	3	0	0	0	4	3	0.4286
P. xanthoriae	30	14	2	0	0	0	12	2	0.1429
A. protothecoides	37	7	0	4	0	0	3	4	0.5714
P. ciferrii	37	31	2	0	0	0	29	2	0.0645
P. moriformis	37	7	0	2	0	0	5	2	0.2857
P. paracutis	37	7	0	0	2	0	5	2	0.2857
P. tumulicola	37	14	0	4	0	0	10	4	0.2857
P. xanthoriae	37	14	2	4	0	5	3	11	0.7857
A. protothecoides	42	2	2	0	0	0	0	2	1
P. blaschkeae	42	6	0	2	0	0	4	2	0.3333
P. bovis	42	28	0	3	0	0	25	3	0.1071
P. ciferrii	42	16	0	4	0	2	10	6	0.375
P. cookei	42	2	0	1	0	1	0	2	1
P. paracutis	42	2	0	0	0	1	1	1	0.5
P. wickerhamii	42	22	1	5	0	3	13	9	0.4091
P. xanthoriae	42	4	0	1	0	0	3	1	0.25

Appendix 8. Table showing the number of biological replicates to analyse Prototheca *and* Auxenochlorella *growth under temperature stress, as measured by OD. For a biological replicate to be included, the model for at least one well was not rejected as an outlier. Species for which fewer than 2 biological replicates existed at 30°C and 37°C are not included.*

Species	N at 25°C	N at 30°C	N at 37°C	N at 42°C
P. bovis	13	29	29	13
P. ciferrii	8	15	15	5
P. cookei	1	3	3	0
P. blaschkeae	1	9	9	1
P. tumulicola	2	6	2	0
P. moriformis	1	2	3	0
P. cutis	1	3	3	0
P. paracutis	1	2	2	0
P. miyajii	1	3	3	1
P. wickerhamii	11	22	22	6



Appendix 9. Growth curves of Prototheca *and* Auxenochlorella *strains under temperature stress, as measured by OD600. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. The variable number of replicates at each temperature is indicated in the top left of each graph and is broken down further in Appendix 5. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. A strain's origin is unknown if, when given the strain, we were not given information regarding the origin of the strain, or if there is evidence that the strain was swapped. Bars show the 95% prediction interval (mean ± 2 standard deviations) for a given strain at 8-hour intervals, slightly offset for improved visibility.*

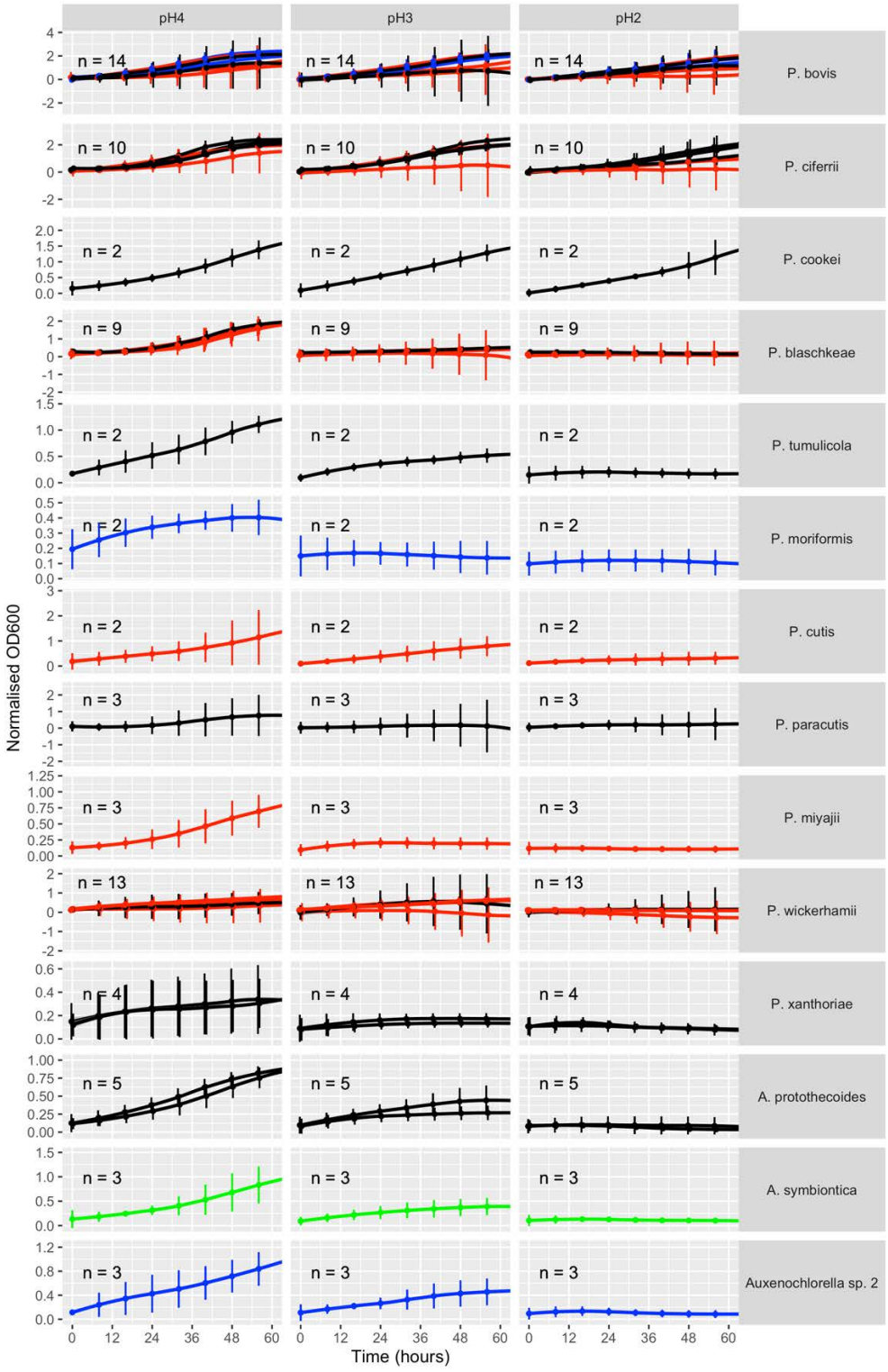
Appendix 10. Detailed breakdown of the number of wells exposed to acid stress that failed to be fitted to a model, as well as the reason for failure. Failures are grouped according to species and pH. Only conditions under which at least one model failed are included.

Species	pH	Number of wells in total	Failures due to carrying capacity	Failures due to growth rate	Failures due to sigma	Failure to fit a model	Number of valid wells	Number of invalid wells	Proportion of wells that are invalid
A. protothecoides	4	10	4	0	0	1	5	5	0.5
Auxenochlorella sp. 2	4	6	3	0	0	0	3	3	0.5
A. symbiontica	4	6	3	0	0	0	3	3	0.5
P. blaschkeae	4	20	3	0	0	0	17	3	0.15
P. bovis	4	28	2	0	0	0	26	2	0.0714
P. ciferrii	4	20	0	0	0	1	19	1	0.05
P. cookei	4	6	1	0	0	0	5	1	0.1667
P. cutis	4	6	1	0	0	0	5	1	0.1667
P. miyajii	4	6	1	0	0	0	5	1	0.1667
P. paracutis	4	6	2	0	0	0	4	2	0.3333
P. wickerhamii	4	26	3	0	0	0	23	3	0.1154
P. xanthoriae	4	12	0	0	0	1	11	1	0.0833
A. protothecoides	3	10	0	0	0	3	7	3	0.3
Auxenochlorella sp. 2	3	6	0	0	0	1	5	1	0.1667
A. symbiontica	3	6	0	0	0	1	5	1	0.1667
P. blaschkeae	3	20	3	0	0	0	17	3	0.15
P. bovis	3	28	3	0	0	3	22	6	0.2143
P. ciferrii	3	20	1	0	0	2	17	3	0.15
P. cookei	3	6	1	0	0	0	5	1	0.1667
P. cutis	3	6	1	0	0	0	5	1	0.1667
P. miyajii	3	6	0	1	0	0	5	1	0.1667
P. moriformis	3	6	0	2	0	1	3	3	0.5
P. paracutis	3	6	2	0	0	0	4	2	0.3333
P. wickerhamii	3	26	4	0	0	0	22	4	0.1538
P. xanthoriae	3	12	1	1	0	0	10	2	0.1667

*A. prototheco*ides	2	10	0	3	0	2	5	5	0.5
Auxenochlorella sp. 2	2	6	0	4	0	0	2	4	0.6667
A. symbiontica	2	6	0	1	0	2	3	3	0.5
P. blaschkeae	2	20	1	4	0	2	13	7	0.35
P. bovis	2	28	0	0	0	6	22	6	0.2143
P. ciferrii	2	20	2	0	0	4	14	6	0.3
P. cookei	2	6	1	0	0	0	5	1	0.1667
P. cutis	2	6	0	1	0	0	5	1	0.1667
P. miyajii	2	6	0	1	0	2	3	3	0.5
P. moriformis	2	6	0	4	0	0	2	4	0.6667
P. paracutis	2	6	3	0	0	0	3	3	0.5
P. wickerhamii	2	26	0	3	0	3	20	6	0.2308
P. xanthoriae	2	12	0	4	0	3	5	7	0.5833

Appendix 11. Table showing the number of biological replicates to analyse Prototheca *and* Auxenochlorella *growth under acid stress, as measured by OD. For a biological replicate to be included, the model for at least one well was not rejected as an outlier. Species for which fewer than 2 biological replicates existed at pH 4 and pH 3 are not included.*

Species	N at pH 4	N at pH 3	N at pH 2
P. bovis	14	12	12
P. ciferrii	10	10	9
P. cookei	3	3	3
P. blaschkeae	10	9	8
P. tumulicola	3	3	3
P. moriformis	3	3	2
P. cutis	3	3	3
P. paracutis	2	3	2
P. miyajii	3	3	2
P. xanthoriae	6	6	3
A. prototheco*ides*	4	4	3
A. symbiontica	2	3	2
Auxenochlorella sp. 2	2	3	1
P. wickerhamii	12	12	11



Appendix 12. Growth curves of Prototheca *and* Auxenochlorella *strains under acid stress, as measured by OD600. Each tile shows the growth of strains of a given species at a particular temperature. Each curve represents data collected from a single strain. Y axis intervals are consistent within a species; X axis intervals are consistent for all tiles. The variable number of replicates at each temperature is indicated in the top left of each graph and is broken down further in Appendix 5. The source of a given strain is indicated by the colour of the points and curves, as well as the shape of the points. A strain's origin is unknown if, when given the strain, we were not given information regarding the origin of the strain, or if there is evidence that the strain was swapped. Bars show the 95% prediction interval (mean ± 2 standard deviations) for a given strain at 8-hour intervals, slightly offset for improved visibility.*

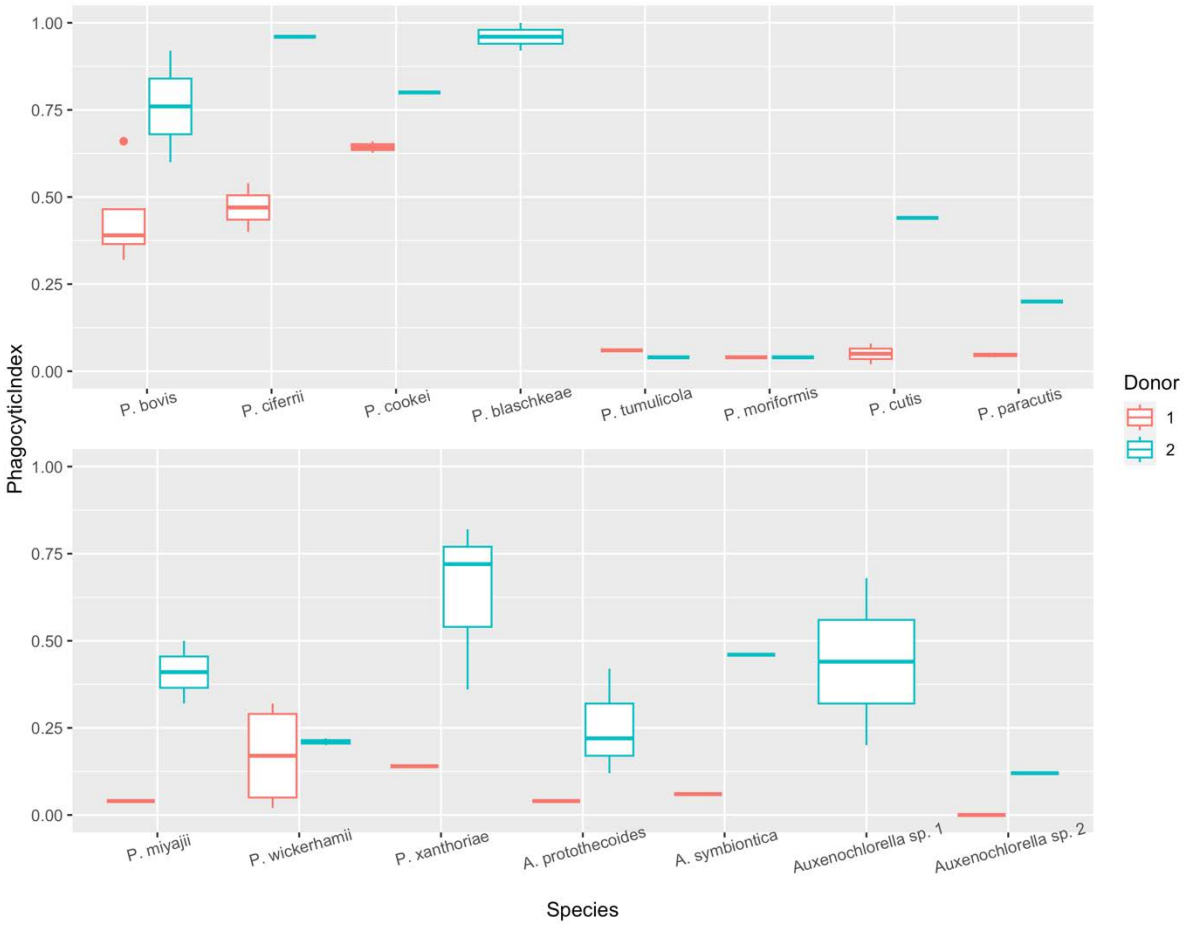
Appendix 13. Number of replicates for phagocytosis assays for each algal species exposed to macrophages and for each type of macrophage used.

Species	Cell Type	Replicates	Strains
Auxenochlorella protothecoides	Human Primary Cells	4	2
Auxenochlorella protothecoides	Murine Cell Line	6	2
Auxenochlorella sp. 1	Human Primary Cells	2	1
Auxenochlorella sp. 1	Murine Cell Line	8	1
Auxenochlorella sp. 2	Human Primary Cells	2	1
Auxenochlorella sp. 2	Murine Cell Line	3	1
Auxenochlorella symbiontica	Human Primary Cells	2	1
Auxenochlorella symbiontica	Murine Cell Line	4	1
Chlamydomonas reinhardtii	Murine Cell Line	2	1
Coccomyxa galuniae	Murine Cell Line	2	1
Crucigeniella apiculata	Murine Cell Line	2	1
Keratococcus bicaudatus	Murine Cell Line	2	1
Prototheca blaschkeae	Human Primary Cells	2	1
Prototheca blaschkeae	Murine Cell Line	2	1
Prototheca bovis	Human Primary Cells	6	2
Prototheca bovis	Murine Cell Line	9	3
Prototheca ciferrii	Human Primary Cells	3	1
Prototheca ciferrii	Murine Cell Line	3	1
Prototheca cookei	Human Primary Cells	3	1
Prototheca cookei	Murine Cell Line	3	1
Prototheca cutis	Human Primary Cells	3	1
Prototheca cutis	Murine Cell Line	3	1
Prototheca miyajii	Human Primary Cells	3	1
Prototheca miyajii	Murine Cell Line	3	1
Prototheca moriformis	Human Primary Cells	2	1
Prototheca moriformis	Murine Cell Line	3	1
Prototheca paracutis	Human Primary Cells	3	1
Prototheca paracutis	Murine Cell Line	3	1
Prototheca tumulicola	Human Primary Cells	2	1
Prototheca tumulicola	Murine Cell Line	3	1
Prototheca wickerhamii	Human Primary Cells	6	2
Prototheca wickerhamii	Murine Cell Line	4	2
Prototheca xanthoriae	Human Primary Cells	4	2
Prototheca xanthoriae	Murine Cell Line	6	2
Trebouxia decolorans	Murine Cell Line	1	1

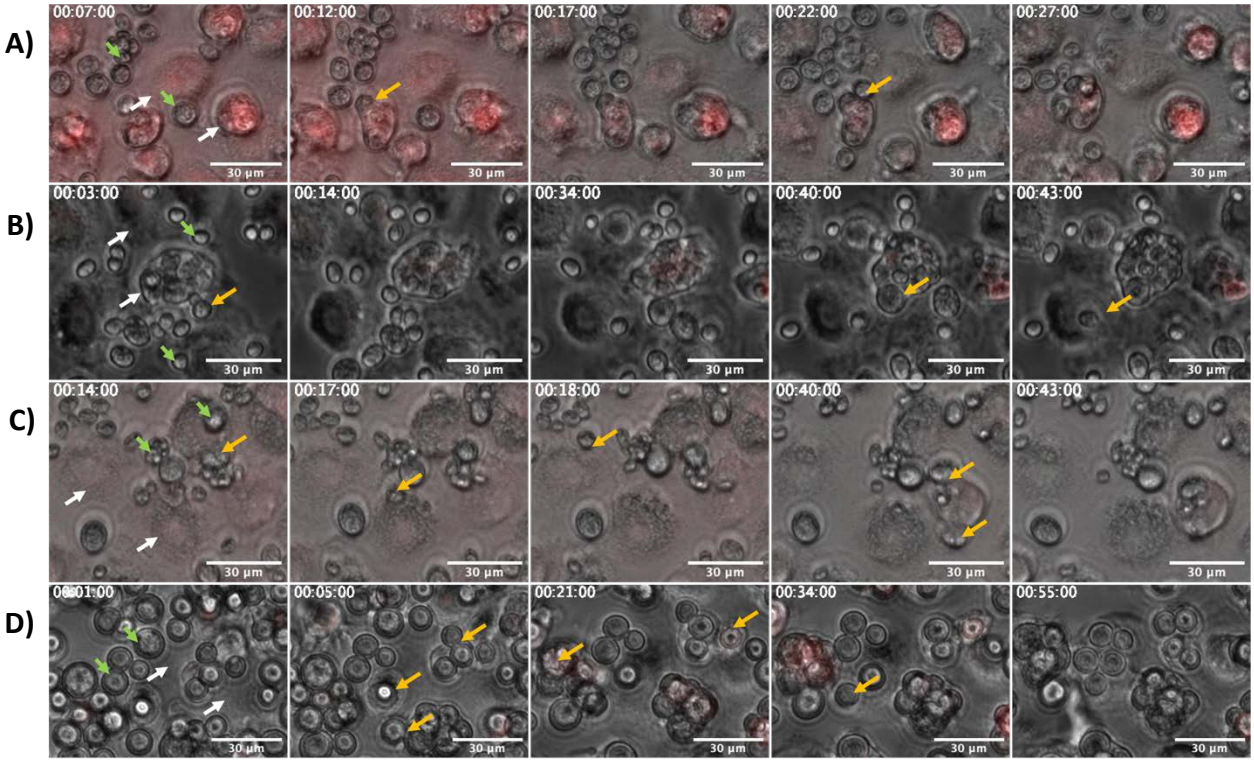
Appendix 14. Pairwise comparisons of phagocytic indexes of pathogenic species of Prototheca *exposed to murine cells.*

Comparison	Cell Type	Significance	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value
P. bovis – P. blaschkeae	Mouse	NS	-0.2	-0.519	0.119	3.82×10^{-1}
P. ciferrii – P. blaschkeae	Mouse	NS	0.0936	-0.278	0.466	9.64×10^{-1}
P. cutis – P. blaschkeae	Mouse	***	-0.847	-1.22	-0.475	1.30×10^{-5}
P. miyajii – P. blaschkeae	Mouse	NS	-0.267	-0.639	0.105	2.53×10^{-1}
P. wickerhamii – P. blaschkeae	Mouse	***	-0.868	-1.22	-0.515	4.44×10^{-6}
P. ciferrii – P. bovis	Mouse	*	0.294	0.0219	0.565	2.99×10^{-2}
P. cutis – P. bovis	Mouse	***	-0.647	-0.918	-0.375	7.03×10^{-6}
P. miyajii – P. bovis	Mouse	NS	-0.0667	-0.338	0.205	9.67×10^{-1}
P. wickerhamii – P. bovis	Mouse	***	-0.668	-0.913	-0.423	1.01×10^{-6}
P. cutis – P. ciferrii	Mouse	***	-0.94	-1.27	-0.608	6.05×10^{-7}
P. miyajii – P. ciferrii	Mouse	*	-0.36	-0.693	-0.0276	2.95×10^{-2}
P. wickerhamii – P. ciferrii	Mouse	***	-0.962	-1.27	-0.651	1.58×10^{-7}
P. miyajii – P. cutis	Mouse	***	0.58	0.247	0.913	3.61×10^{-4}
P. wickerhamii – P. cutis	Mouse	NS	-0.0217	-0.333	0.29	1.00×10^{0}
P. wickerhamii – P. miyajii	Mouse	***	-0.602	-0.913	-0.29	1.06×10^{-4}
P. bovis – P. blaschkeae	Human	NS	-0.413	-0.947	0.121	1.86×10^{-1}
P. ciferrii – P. blaschkeae	Human	NS	-0.327	-0.924	0.27	5.20×10^{-1}
P. cutis – P. blaschkeae	Human	**	-0.78	-1.38	-0.183	6.96×10^{-3}
P. miyajii – P. blaschkeae	Human	*	-0.673	-1.27	-0.0762	2.23×10^{-2}

P. wickerhamii – P. blaschkeae	Human	**	-0.777	-1.31	-0.243	2.63×10^{-3}
P. ciferrii – P. bovis	Human	NS	0.0867	-0.376	0.549	9.90×10^{-1}
P. cutis – P. bovis	Human	NS	-0.367	-0.829	0.0959	1.68×10^{-1}
P. miyajii – P. bovis	Human	NS	-0.26	-0.723	0.203	4.92×10^{-1}
P. wickerhamii – P. bovis	Human	NS	-0.363	-0.741	0.0143	6.31×10^{-2}
P. cutis – P. ciferrii	Human	NS	-0.453	-0.987	0.0808	1.23×10^{-1}
P. miyajii – P. ciferrii	Human	NS	-0.347	-0.881	0.187	3.44×10^{-1}
P. wickerhamii – P. ciferrii	Human	NS	-0.45	-0.913	0.0126	5.91×10^{-2}
P. miyajii – P. cutis	Human	NS	0.107	-0.427	0.641	9.86×10^{-1}
P. wickerhamii – P. cutis	Human	NS	0.00333	-0.459	0.466	1.00×10^{0}
P. wickerhamii – P. miyajii	Human	NS	-0.103	-0.566	0.359	9.77×10^{-1}



Appendix 15. Comparison of phagocytic indexes of AHP species in primary macrophages derived from different human donors. No data is available for the phagocytic index of P. blaschkeae *or* Auxenochlorella *sp. 1 when exposed to human primary cells from donor 1, because these species were not exposed to these cells.*



Appendix 16. Phagocytosis of cattle-associated species of Prototheca *by human primary cells. Algal cells are indicated by green arrows; macrophages are indicated by white arrows; phagocytosis events are indicated by yellow arrows. LTR staining indicates acidified compartments within the macrophages – intended to identify acidified phagolysosomes containing phagocytosed algal cells. Rows show selected frames from videos, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. bovis*, HP3, exposed to cells from donor 2; B)* P. ciferrii*, HP2, exposed to cells from donor 1; C)* P. cookei*, HP32, exposed to cells from donor 1; D)* P. blaschkeae*, HP4, exposed to cells from donor 2.*

Appendix 17. Pairwise comparisons of phagocytic indexes of cattle-associated species of Prototheca *exposed to immune cells.*

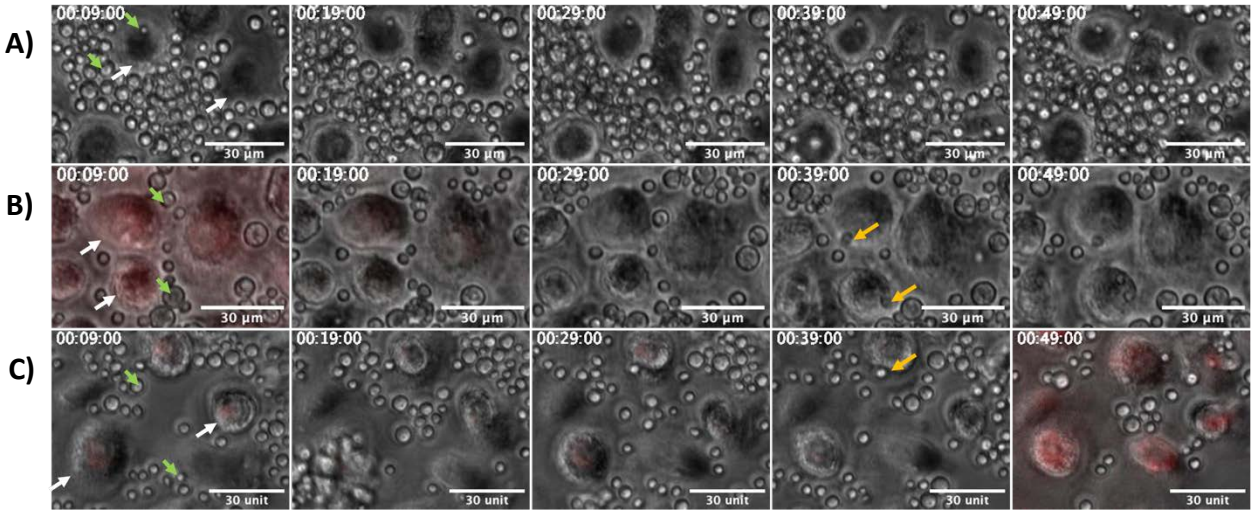
Comparison	Cell Type	Significance	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value
P. bovis - P. blaschkeae	Mouse	NS	-0.2	-0.535	0.135	0.337×10^{-1}
P. ciferrii - P. blaschkeae	Mouse	NS	0.0936	-0.297	0.484	0.894×10^{-1}
P. cookei - P. blaschkeae	Mouse	NS	0.02	-0.371	0.411	0.999×10^{-1}
P. ciferrii - P. bovis	Mouse	*	0.294	0.00813	0.579	0.0431×10^{-2}
P. cookei - P. bovis	Mouse	NS	0.22	-0.0655	0.505	0.158×10^{-1}
P. cookei - P. ciferrii	Mouse	NS	-0.0736	-0.423	0.276	0.925×10^{-1}
P. bovis - P. blaschkeae	Human	NS	-0.413	-0.941	0.114	0.14×10^{-1}
P. ciferrii - P. blaschkeae	Human	NS	-0.327	-0.916	0.263	0.375×10^{-1}
P. cookei - P. blaschkeae	Human	NS	-0.264	-0.854	0.325	0.542×10^{-1}
P. ciferrii - P. bovis	Human	NS	0.0867	-0.37	0.543	0.936×10^{-1}
P. cookei - P. bovis	Human	NS	0.149	-0.308	0.606	0.754×10^{-1}
P. cookei - P. ciferrii	Human	NS	0.0622	-0.465	0.59	0.983×10^{-1}

Appendix 18. Pairwise comparisons of phagocytic indexes of environmental species of Prototheca *exposed to immune cells.*

Comparison	Cell Type	Significance	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value
P. tumulicola - P. moriformis	Mouse	*	-0.0263	-0.0453	-0.00723	1.86×10^{-2}
P. tumulicola - P. moriformis	Human	NS	0.01	-0.033	0.053	4.23×10^{-1}

Appendix 19. Pairwise comparisons of phagocytic indexes of human-associated species of Prototheca *exposed to immune cells.*

Comparison	Cell Type	Significance	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value
P. miyajii - P. cutis	Mouse	***	0.58	0.395	0.765	2.09×10^{-5}
P. paracutis - P. cutis	Mouse	*	0.193	0.00816	0.379	4.07×10^{-2}
P. wickerhamii - P. cutis	Mouse	NS	-0.0217	-0.195	0.152	9.79×10^{-1}
P. paracutis - P. miyajii	Mouse	***	-0.387	-0.572	-0.201	5.14×10^{-4}
P. wickerhamii - P. miyajii	Mouse	***	-0.602	-0.775	-0.428	8.82×10^{-6}
P. wickerhamii - P. paracutis	Mouse	*	-0.215	-0.388	-0.0418	1.63×10^{-2}
P. miyajii - P. cutis	Human	NS	0.107	-0.298	0.511	8.56×10^{-1}
P. paracutis - P. cutis	Human	NS	-0.0822	-0.487	0.322	9.26×10^{-1}
P. wickerhamii - P. cutis	Human	NS	0.00333	-0.347	0.354	1.00×10^{0}
P. paracutis - P. miyajii	Human	NS	-0.189	-0.593	0.216	5.21×10^{-1}
P. wickerhamii - P. miyajii	Human	NS	-0.103	-0.454	0.247	8.11×10^{-1}
P. wickerhamii - P. paracutis	Human	NS	0.0856	-0.265	0.436	8.81×10^{-1}



Appendix 20. Lack of chemotaxis of human primary cells towards human-associated species of Prototheca*. Algal cells are indicated by green arrows; macrophages are indicated by white arrows. Chemotaxis towards algal cells, of the sort seen in murine cells, is not readily apparent, despite phagocytosis occurring at a greater rate for* P. wickerhamii *and* P. cutis*. Yellow arrows indicate phagocytosis events. Rows show selected frames from a single video, at times indicated by the time stamp in the top left. The species and strain of the algal cells in each row are: A)* P. cutis*, HP28, exposed to cells from donor 1; B)* P. paracutis*, HP31, exposed to cells from donor 1; C)* P. wickerhamii*, HP50, exposed to cells from donor 1.*

Appendix 21. Pairwise comparisons of phagocytic indexes of Auxenochlorella *species and closely related* Prototheca *species exposed to murine cells.*

Comparison	Cell Type	Significance	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value
Auxenochlorella sp. 1 - *A. protothecoides*	Mouse	NS	0.0895	-0.136	0.315	7.76×10^{-1}
Auxenochlorella sp. 2 - *A. protothecoides*	Mouse	NS	-0.164	-0.491	0.164	5.98×10^{-1}
A. symbiontica - A. protothecoides	Mouse	NS	-0.137	-0.415	0.141	6.08×10^{-1}
P. xanthoriae - A. protothecoides	Mouse	*	0.298	0.0337	0.562	2.12×10^{-2}
Auxenochlorella sp. 2 - *Auxenochlorella* sp. 1	Mouse	NS	-0.253	-0.56	0.0531	1.43×10^{-1}
A. symbiontica - Auxenochlorella sp. 1	Mouse	NS	-0.227	-0.48	0.0258	9.44×10^{-2}
P. xanthoriae - Auxenochlorella sp. 1	Mouse	NS	0.208	-0.0291	0.446	1.07×10^{-1}
A. symbiontica - Auxenochlorella sp. 2	Mouse	NS	0.0264	-0.32	0.373	9.99×10^{-1}
P. xanthoriae - Auxenochlorella sp. 2	Mouse	**	0.462	0.126	0.797	3.50×10^{-3}
P. xanthoriae - A. symbiontica	Mouse	**	0.435	0.148	0.723	1.21×10^{-3}
Auxenochlorella sp. 1 - *A. protothecoides*	Human	NS	0.24	-0.501	0.981	8.08×10^{-1}
Auxenochlorella sp. 2 - *A. protothecoides*	Human	NS	-0.14	-0.881	0.601	9.65×10^{-1}
A. symbiontica - A. protothecoides	Human	NS	0.06	-0.681	0.801	9.99×10^{-1}
P. xanthoriae - A. protothecoides	Human	NS	0.31	-0.295	0.915	4.67×10^{-1}
Auxenochlorella sp. 2 - *Auxenochlorella* sp. 1	Human	NS	-0.38	-1.24	0.476	5.90×10^{-1}
A. symbiontica - Auxenochlorella sp. 1	Human	NS	-0.18	-1.04	0.676	9.50×10^{-1}
P. xanthoriae - Auxenochlorella sp. 1	Human	NS	0.07	-0.671	0.811	9.97×10^{-1}
A. symbiontica - Auxenochlorella sp. 2	Human	NS	0.2	-0.656	1.06	9.28×10^{-1}

P. xanthoriae - *Auxenochlorella* sp. 2	Human	NS	0.45	-0.291	1.19	3.20×10^{-1}
P. xanthoriae - *A. symbiontica*	Human	NS	0.25	-0.491	0.991	7.86×10^{-1}

Appendix 22. Pairwise comparisons of number of internalised algae between sub-lineages of the AHP lineage exposed to macrophages.

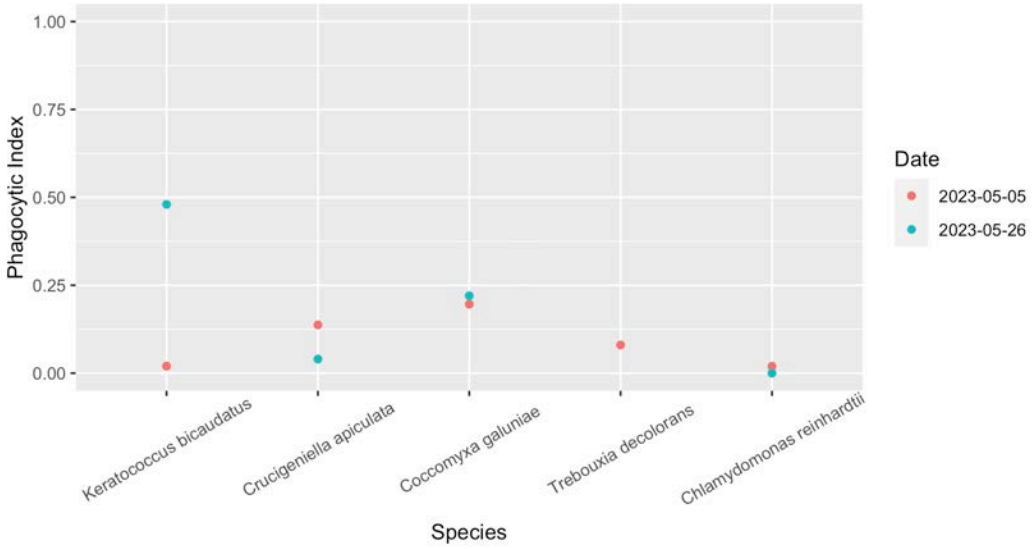
Group Comparison	Cell Type	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value	Significance
Cattle-*Auxenochlorella*	Human Primary	1.851	0.6251	3.078	1.24×10^{-3}	**
Human-*Auxenochlorella*	Human Primary	-0.6837	-1.89	0.5227	4.37×10^{-1}	NS
Environmental-*Auxenochlorella*	Human Primary	-0.7659	-2.586	1.054	6.76×10^{-1}	NS
Human-Cattle	Human Primary	-2.535	-3.718	-1.352	5.64×10^{-6}	***
Environmental-Cattle	Human Primary	-2.617	-4.422	-0.8123	1.99×10^{-3}	**
Environmental-Human	Human Primary	-0.08221	-1.874	1.709	9.99×10^{-1}	NS
Cattle-*Auxenochlorella*	Murine	1.507	0.6152	2.398	2.24×10^{-4}	***
Human-*Auxenochlorella*	Murine	-0.2373	-1.267	0.7926	9.28×10^{-1}	NS
Environmental-*Auxenochlorella*	Murine	-0.2151	-1.967	1.537	9.88×10^{-1}	NS
Human-Cattle	Murine	-1.744	-2.858	-0.6298	6.70×10^{-4}	***
Environmental-Cattle	Murine	-1.722	-3.525	0.08125	6.63×10^{-2}	NS
Environmental-Human	Murine	0.02217	-1.853	1.897	1.00×10^{0}	NS

Appendix 23. Pairwise comparisons of phagocytic rates of Auxenochlorella *species and closely related* Prototheca *species exposed to murine cells.*

Species Comparison	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value	Significance
Auxenochlorella sp. 1- *Auxenochlorella protothecoides*	-1233	-2077	-389.7	2.83×10^{-3}	**
Auxenochlorella sp. 2- *Auxenochlorella protothecoides*	683.7	-796.3	2164	6.33×10^{-1}	NS
Auxenochlorella symbiontica- *Auxenochlorella protothecoides*	738.7	-431.3	1909	3.44×10^{-1}	NS
Prototheca xanthoriae- *Auxenochlorella protothecoides*	-668.5	-1575	237.8	2.11×10^{-1}	NS
Auxenochlorella sp. 2- *Auxenochlorella* sp. 1	1917	572.8	3261	3.55×10^{-3}	**
Auxenochlorella symbiontica- *Auxenochlorella* sp. 1	1972	979.3	2965	1.13×10^{-4}	***
Prototheca xanthoriae- *Auxenochlorella* sp. 1	564.9	-96.99	1227	1.15×10^{-1}	NS
Auxenochlorella symbiontica- *Auxenochlorella* sp. 2	55	-1515	1625	1.00×10^{0}	NS
Prototheca xanthoriae- *Auxenochlorella* sp. 2	-1352	-2737	32.19	5.73×10^{-2}	NS
Prototheca xanthoriae- *Auxenochlorella symbiontica*	-1407	-2454	-360.7	5.94×10^{-3}	**

Appendix 24. Pairwise comparisons of phagocytic rates between sub-lineages of the AHP lineage exposed to macrophages.

Group Comparison	Cell Type	Difference	95% CI lower bound	95% CI upper bound	Adjusted p value	Significance
Cattle-*Auxenochlorella*	Human Primary	-511.3	-978	-44.58	2.69×10^{-2}	*
Human-*Auxenochlorella*	Human Primary	-5.585	-464.8	453.6	1.00×10^{0}	NS
Environmental-*Auxenochlorella*	Human Primary	-78.15	-771	614.7	9.90×10^{-1}	NS
Human-Cattle	Human Primary	505.7	55.42	956	2.23×10^{-2}	*
Environmental-Cattle	Human Primary	433.2	-253.8	1120	3.43×10^{-1}	NS
Environmental-Human	Human Primary	-72.56	-754.5	609.3	9.92×10^{-1}	NS
Cattle-*Auxenochlorella*	Murine	-756.4	-1318	-194.9	4.22×10^{-3}	**
Human-*Auxenochlorella*	Murine	423.4	-218.8	1066	3.08×10^{-1}	NS
Environmental-*Auxenochlorella*	Murine	1343	272.3	2413	8.56×10^{-3}	**
Human-Cattle	Murine	1180	506.9	1853	1.40×10^{-4}	***
Environmental-Cattle	Murine	2099	1010	3188	2.92×10^{-5}	***
Environmental-Human	Murine	919.2	-213.5	2052	1.50×10^{-1}	NS

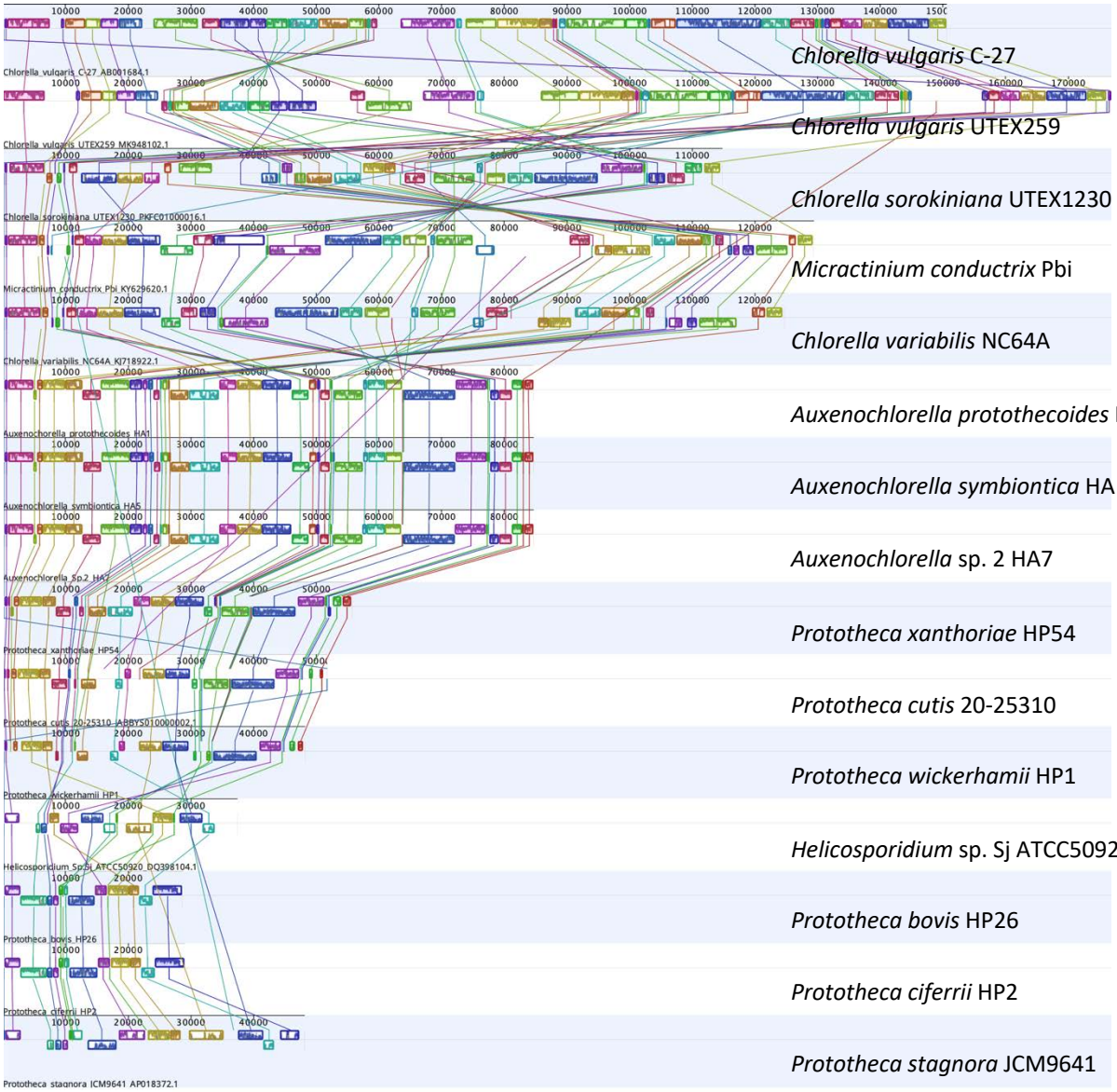


Appendix 25. Phagocytic activity of murine cells in response to non-AHP green algae. Each point represents the phagocytic index of a given strain from a single replicate. Points are coloured according to the date on which the replicate took place.

Strain	Number of reads	Median insert size	Coverage
HP1	2,441,159	211	41
HP2	2,832,620	55	40
HP3	1,859,647	116	27
HP5	9,144,591	686	145
HP6	7,904,718	664	184
HP14	4,970,926	684	59
HP18	3,826,816	654	41
HP25	2,630,486	669	50
HP26	1,888,940	663	25
HP53	208,263	707	8
HP54	1,567,128	695	32
HA1	5,602,077	673	96
HA5	4,356,700	773	65
HA7	2,551,136	813	45

Appendix 26. Sequencing statistics for Illumina sequencing of Prototheca *and* Auxenochlorella *strains.*

Appendix 27. Arrangement of syntenic blocks of mitochondrial sequences from the CHAMP lineage, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.



Appendix 28. Arrangement of syntenic blocks of chloroplast sequences from the CHAMP lineage, as visualised in MAUVE. Similar regions are indicated by colours of the blocks, and by connecting lines. Blocks above the central line represent genes encoded on the forward strand, while blocks below the central line represent genes encoded on the reverse strand. Numbers indicate the length of the sequence in base pairs.