

GENOMIC COMPOSITION OF GREEN ALGAE GROWN IN HIGH ALKALINE CONDITIONS

by

Calvin Lee Cicha Goemann

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology and Cell Biology

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2023

©COPYRIGHT

by

Calvin Lee Cicha Goemann

2023

All Rights Reserved

DEDICATION

I dedicate this dissertation to my family for providing me with the opportunities and support to pursue my passions in science. Thank you, Hannah, Mom, Dad, Jarrod, Olivia, Michael, Josie, Little Grace, Heidi, Carry, Aaron, and Molly. Words cannot express my gratitude and appreciation.

ACKNOWLEDGEMENTS

I am so thankful to be able to say that I have enjoyed my time performing research, teaching, and learning at MSU. Montana State University is full of incredibly talented people who have contributed to my career development during graduate school. I would first like to thank Dr. Blake Wiedenheft for his advice, patience, and confidence in my ability. Your passion for science is contagious and motivational. Thank you to the Wiedenheft group and CBE faculty, students, and staff for the thoughtful conversations and friendship. Having a positive working environment is such a blessing. Thank you to my committee members, Dr. Robin Gerlach, Dr. Ross Carlson, Dr. Matthew Fields, and Dr. Mensur Dlakic for your insight, mentorship, and support. I would also like to thank Kari Cargill for your mentorship, friendship, and trust in my teaching ability. I strive to one day be as passionate about teaching as you are.

Most of all, thank you to Hannah for taking on this challenge with me. Your help, motivation, positivity, and adventurous spirit never fail to brighten my day and remind me of what is important in life. I couldn't and wouldn't want to do it without you.

TABLE OF CONTENTS

1. PREFACE AND OUTLINE	1
2. INTRODUCTION AND LITERATURE REVIEW	3
Classification and evolution of algae	3
Repurposing algal metabolism to make useful bioproducts	6
Impacts of stress on algal growth and biomass production	8
Current limits of algal cultivation and lipid production	12
High alkaline cultivation strategies	15
Gaps in our knowledge	17
Hypothesis	18
References	19
3. GENOME SEQUENCE, PHYLOGENETIC ANALYSIS, AND STRUCTURE-BASED ANNOTATION REVEALS METABOLIC POTENTIAL OF <i>CHLORELLA SP.</i> SLA-04	27
Contribution of Authors and Co-Authors	27
Manuscript information	29
Abstract	31
Introduction	32
Methods	34
Strain information	34
Strain cultivation conditions	34
Axenic isolation	34
Genome sequencing and assembly	35
High molecular weight DNA extraction	35
RNA extraction	35
DNA and RNA sequencing	36
Genome assembly and annotation	37
Comparative analysis	39
Phylogenetic analysis	39
Results	40
General genome characteristics	40
<i>Chlorella sp.</i> SLA-04 genome assembly statistics	40
Genome completeness analysis	41
Annotation of <i>Chlorella sp.</i> SLA-04	41
Annotation using protein 3D structural homology	42
Phylogenetic analysis of <i>Chlorella sp.</i> SLA-04	43
Genomic analysis of metabolic pathways in <i>Chlorella sp.</i> SLA-04	44
Discussion	46
Conclusion	49

TABLE OF CONTENTS CONTINUED

Keywords	49
Abbreviations	49
Data access	50
Competing interest statement.....	50
Acknowledgments.....	50
Author contributions	50
Supplemental Figures.....	52
References	58
 4. METABOLIC REPROGRAMMING OF CARBON METABOLISM IN CHLORELLA SP. SLA-04 UNDER MULTIPLE ENVIRONMENTAL STRESSORS	 67
Contribution of Authors and Co-Authors	67
Manuscript information	68
Abstract	70
Introduction	71
Results	73
Physiological data	73
Cell growth rates	73
Lipid and carbohydrate composition	74
Transcriptome analysis of carbon metabolism	77
Photosynthesis.....	77
Starch and sugar metabolism	78
Fatty acid synthesis	80
Lipid synthesis	80
Mitochondrial carbon metabolism	81
Genes involved in adaptation to nitrogen depletion	82
Genes involved in adaptation to alkaline conditions	85
Discussion	87
Conclusion	91
Methods.....	92
Strain cultivation conditions	92
Quantifying media composition.....	92
Dissolved inorganic carbon.....	92
Media nitrate concentration	93
Cell growth.....	93
Cell count.....	93
Total suspended solids	93
Biomass composition	94
Chlorophyll concentration	94
Nile Red fluorescence	94

TABLE OF CONTENTS CONTINUED

Lyophilization of algal biomass	95
FAME analysis.....	95
Carbohydrate analysis	95
Transcriptome analysis	96
RNA extraction and quantification	96
RNA sequencing	97
Differential gene expression	97
Protein structural prediction.....	98
Supplemental Figures.....	99
References.....	117
 5. CONCLUSIONS AND FUTURE STUDIES	 123
Future Studies	125
Nile red optimization	125
Tracking carbon partitioning in SLA-04.....	127
Identifying antiviral genes in green algae	128
Characterizing transcription factors in SLA-04.....	129
Concluding remarks	129
References.....	131
 APPENDICES	 133
APPENDIX A: Pedagogical Research.....	134
APPENDIX B: Nanopore Sequencing Publication	153
APPENDIX C: Co-Author Publications	159
Temporal detection and phylogenetic assessment of SARS-COV-2 in municipal wastewater	160
SARS-COV-2 genomic surveillance identifies naturally occurring truncation of ORF7A that limits immune suppression	192
Intrinsic signal amplification by type III CRISPR-CAS system provides a sequence-specific SARS-COV-2 diagnostic	237
 REFERENCES CITED.....	 281

LIST OF TABLES

Table	Page
S3.1. Genome statistics of <i>Chlorella</i> sp. SLA-04 genome.....	53
S3.2. Summary of repetitive elements identified by RepeatMasker	55
S3.3. Active transcription of lipid synthesis genes within <i>Chlorella</i> sp. SLA-04 genome.....	57
S4.1. Fatty acid composition and statistics	99
S4.2. Differential genes between pre- versus post-nitrogen depletion.....	103
S4.3. Differential genes between neutral versus alkaline	111

LIST OF FIGURES

Figure	Page
2.1. The light-dependent reaction of photosynthesis in algae	4
2.2. The light-independent reaction of photosynthesis and carbon fixation in algae	5
2.3. Transesterification of triacylglycerols (TAGs) to produce fatty acid methyl esters (FAMES)	7
2.4. Henry's Law	15
2.5. Dissolution of inorganic carbon across pH	16
3.1. Complete genome of Chlorella sp. SLA-04	40
3.2. Annotation using protein structural similarity improves functional understanding of SLA-04 proteome	42
3.3. Phylogenomic analysis of green algae	44
3.4. Flow of carbon from fixation to lipid synthesis in Chlorella sp. SLA-04	46
S3.1. DNA isolation and sequencing parameters for Chlorella sp. SLA-04	52
S3.2. Phylogenetic tree of LTR retrotransposon capsid domains	54
S3.3. Phylogenomic analysis of green algae performed Using 18S rDNA	56
S3.4. Presence of ATP citrate lyase and sphingolipid enzymes across algal genomes	57
4.1. Experimental conditions and biomass production of SLA-04	74
4.2. Biomass composition of SLA-04 in neutral versus alkaline conditions	76
4.3. Differential expression of carbon flow before and after nitrogen depletion	79

LIST OF FIGURES CONTINUED

Figure	Page
4.4. Genes with the greatest differential expression between conditions.....	83
S4.1. Sampling time points and sequencing coverage	100
S4.2. PCA analysis of sequencing	100
S4.3. Schematic of differential carbon flow genes	101
S4.4. Volcano plots showing differential expression of pre vs post-nitrogen depletion	102
S4.5. Volcano plots showing differential expression of neutral vs alkaline conditions	102

ABSTRACT

Algae are responsible for 50% of global oxygen production and sequestration of CO₂ from the atmosphere. Algal photosynthesis plays a critical role in all aquatic ecosystems converting sunlight and CO₂ into usable biomass. Algal growth and biomass production can be coopted to produce industrially relevant bioproducts like triacylglycerol (TAGs) that can be converted into biodiesel and provide a sustainable carbon-neutral alternative to fossil fuels. In high-stress environments, algae produce high levels of TAGs. Multiple stresses including nitrogen limitation and high pH impact algae physiology, but little is known about how algae shift their metabolism to produce TAGs in response to these stresses. This topic remains relatively unexplored due to the limited availability of complete algae genomes. Here we sequence and annotate the complete telomere-to-telomere genome of an alkali-tolerant green algae *Chlorella* sp. SLA-04. Genomic analysis supports a reclassification of Chlorophyta green algae and illuminates how SLA-04 adapts to diverse environmental conditions. Additionally, transcriptomic analysis revealed how *Chlorella* sp. SLA-04 rewires carbon metabolism in high alkaline and nutrient-deplete conditions to produce TAGs while minimizing photosynthetic oxidative stress. Together, we double the amount of publicly available telomere-to-telomere green algal genomes and use this resource to explore how algae respond to diverse environmental conditions in their native and industrial settings.

CHAPTER ONE

PREFACE AND OUTLINE

Chapter two introduces algal biofuels and cultivation methods used to optimize algal biomass composition. This chapter gives background information on photosynthesis, the diversity of algal metabolism, and how environmental stresses can be utilized to trigger lipid production in algae. Here we focus on high-alkalinity cultivation systems that improve algal growth by increasing CO₂ availability and reducing culture contamination in outdoor cultivation systems. This chapter summarizes our poor understanding of algal metabolism in high-alkalinity systems and addresses the need for high-quality genomic and transcriptomic data to identify the metabolic foundation of how algae adapt and produce lipids in extreme environments. This gap in our knowledge is explored in chapters two and three.

Chapter three introduces the first complete genome sequence and annotation of an alkali-tolerant green alga and explores genotypes that may contribute to adaptation during environmental variability. *Chlorella* sp. SLA-04 is a green alga that grows to high cell density in a wide variety of environments – high and neutral pH, high and low alkalinity, and high and low salinity. SLA-04's ability to grow in high pH and high alkalinity media without external CO₂ supply is favorable for large-scale algal biomass production. Genomic analysis of SLA-04 identifies the genes capable of utilizing TCA cycle intermediates to replenish cytosolic acetyl-CoA pools for lipid production. We also identify a complete metabolic pathway for sphingolipid anabolism that may allow SLA-04 to readily adapt to changing environmental conditions. Together this analysis provides the genomic resources necessary to explore algal metabolism in alkaline systems and identifies genotypes that facilitate robust cultivation in mass-production systems.

Chapter four expands this analysis to identify genotypes that contribute to growth in alkaline conditions and explore how carbon is reallocated after nitrogen depletion in *Chlorella sp.* SLA-04. A combined approach tracking changes in physiological composition and transcriptomic expression identifies that SLA-04 rewires carbon metabolism to rapidly produce carbohydrates and triacylglycerols in nutrient deplete environments. Additionally, SLA-04 appears to cycle carbon between TAG storage and the TCA cycle to consume excess photosynthetic reducing power through a proposed futile cycle. In extreme alkaline conditions, SLA-04 maintains high cellular concentrations of carbohydrates, unsaturated lipids, and upregulated ion transporters that likely function to maintain intracellular homeostasis when experiencing alkaline-induced osmotic stress. These combined analyses provide an in-depth understanding of the adaptations during growth in alkaline conditions and how SLA-04 reallocates carbon in stressful conditions.

Chapter five summarizes the contributions made throughout this dissertation and their impact on algal biofuel production. This chapter also proposes future work to expand our understanding of carbon flow in extreme alkaline conditions.

Appendix A contains pedagogical research performed at MSU investigating how different teaching styles impact the motivation and engagement of non-major students.

Appendix B contains a first-author publication summarizing the sequencing and annotation of phage using nanopore sequencing.

Appendix C contains co-author publications that address the detection and genetic variation of SARS-CoV2 in Gallatin Valley and repurposing CRISPR systems for rapid SARS-CoV2 diagnostics

CHAPTER TWO

INTRODUCTION AND LITERATURE REVIEW

Classification and Evolution of Algae

Algae are a large group of photosynthetic eukaryotes that contains an estimated 800,000 species (Guiry, 2012). Algae are incredibly diverse ranging from unicellular microorganisms to macroalgae spanning 60 meters in length. Algal diversity stems from primary and secondary endosymbiotic events approximately 1.6 billion years ago when an ancient cyanobacterium was retained within the cytoplasm of ancestral algal species leading to the chloroplast (Yoon et al., 2004). These endosymbiotic events occurred across the eukaryotic tree of life leading to a paraphyletic clade of algae with immense nuclear genetic diversity. The wide range of algal genotypes and phenotypes has resulted in algae being a ubiquitous and necessary component of all aquatic and many terrestrial ecosystems today.

Across these ecosystems, algae play fundamental roles in nutrient processing and carbon fixation. Algae take up nutrients such as nitrogen and phosphorus from the water, incorporating these nutrients into their biomass. Without these photoautotrophs, nutrients would not be recycled and heterotrophic growth across the planet would be unsustainable. Algae are the primary source of nitrogen for marine ecosystems assimilating six times more inorganic nitrogen into organic biomolecules compared to terrestrial plants (Kuypers et al., 2018). When algae die or are consumed by other organisms, these nutrients are released back into the water, contributing to the nutrient cycling process. While algae play critical roles in nutrient cycling, their photosynthetic capabilities and resulting carbon sequestration are also central to the survival of aquatic ecosystems.

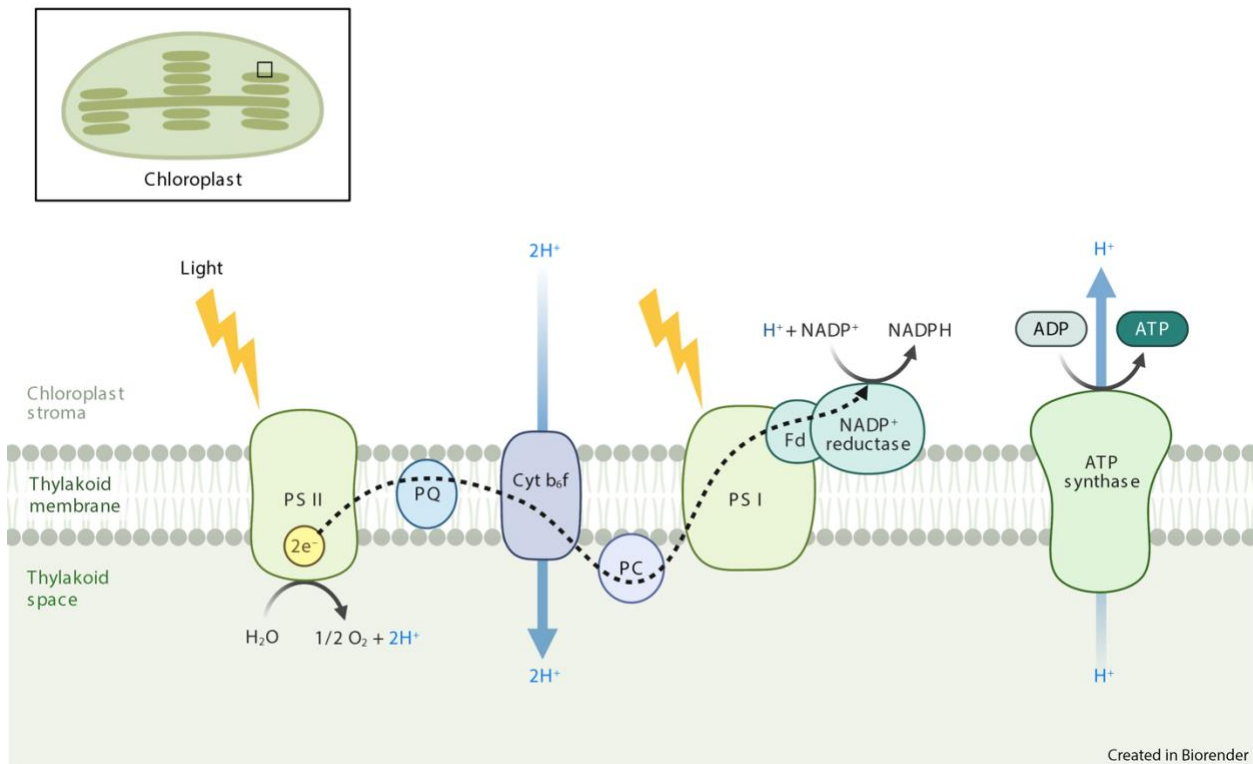


Figure 1: The light-dependent reaction of photosynthesis in algae

Algae are primary producers and produce organic matter through photosynthesis. This photosynthetic capability provides the biomolecules that most organisms use to generate energy and survive. During the light-dependent reactions of photosynthesis algae strip electrons away from H_2O and generate O_2 (**Fig. 1**). These electrons are transferred to the light-harvesting center of photosystem II within the chloroplast (Hall & Rao, 1999). Light from the photosynthetically active radiation spectrum (400-700 nm) is absorbed by photosynthetic pigments and used to excite these electrons to a higher energy state (McCree, 1981). High-energy electrons subsequently pass through the photosynthetic electron transport chain generating energy-rich molecules like ATP (adenosine triphosphate) and NADPH (nicotinamide adenine dinucleotide phosphate) that are used as reducing power to generate biomass in the cell. Oxygen is a byproduct of photosynthesis and is

released into the atmosphere where it can be utilized by heterotrophic organisms to generate energy through oxidative phosphorylation. Algal photosynthesis contributes to about 50% of global oxygen generation and is a necessary step in maintaining an oxygen-rich atmosphere (Chapman, 2013; Walker, 1980).

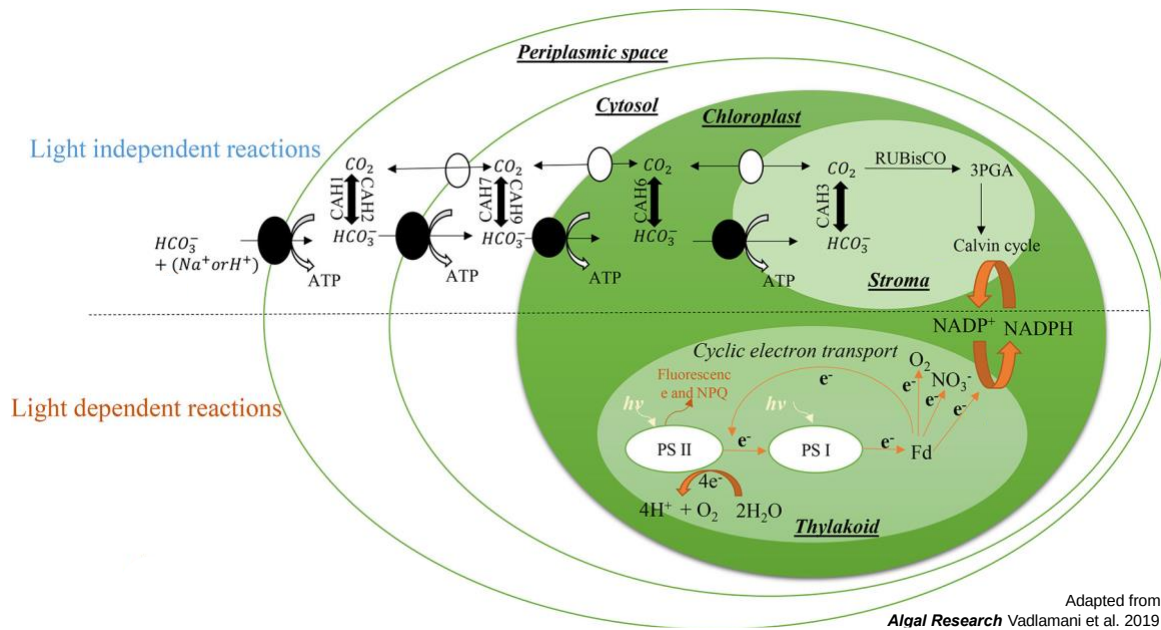


Figure 2: The light-independent reaction of photosynthesis and carbon fixation in algae.

The reducing power generated through the light-dependent reaction of photosynthesis is then used to sequester carbon and produce biomass in the light-independent reaction of photosynthesis (**Fig. 2**). Algae perform 50% of global biological carbon sequestration, generating most of the biologically available carbon in aquatic ecosystems (Behrenfeld, 2001; Cai et al., 2019). CO₂ is an inorganic carbon source that is highly abundant, but biologically unavailable for many cells. During carbon fixation, CO₂ naturally diffuses or is converted to bicarbonate (HCO₃⁻) by carbonic anhydrases (CAH) and actively transported into the chloroplast of the cell (**Fig. 2**) (Supuran, 2008). Here ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBisCO)

converts concentrated CO₂ into biologically available 3-phosphoglycerinaldehyde (3PGA) using the reducing power of the NADPH and ATP produced from photosynthesis (Hall & Rao, 1999). 3PGA is then converted to glucose through the Calvin cycle and enters central carbon metabolism to synthesize all the necessary organic compounds such as carbohydrates, nucleic acids, lipids, and proteins needed for cell growth and division. Sequestered carbon can then be used by heterotrophic organisms when consumed or stored in sediment when algae sink to the bottom of the ocean or lake, helping to regulate the global climate (Tsai et al., 2015).

Repurposing Algal Metabolism to Make Useful Bioproducts

The genomic diversity of algae contains a wide variety of metabolic pathways that produce many different carbon compounds from photosynthesis. Algae can produce antioxidants, proteins, and fats like triacylglycerols (TAGs) from CO₂ and sunlight. While this biomass serves as the foundation of all life across the planet, algae have also recently been pursued as a potential biomass feedstock for industrially relevant products. Algal metabolism can be used to produce a range of useful bioproducts, including biofuels, food additives, pharmaceuticals, bioplastics, fertilizers, and animal feed. Producing TAGs from algae has been a major focus due to their chemical and physical similarities to diesel. TAGs can be extracted from algae, transesterified, and the resulting products, fatty acid methyl esters (FAMES), can be utilized as biodiesel (**Fig. 3**) (Yodsuwan et al., 2017).

Sourcing biodiesel from TAGs instead of fossil fuels has the potential to reduce global carbon emissions significantly. Fossil fuel utilization releases 7.8 billion tons of carbon per year into the atmosphere which is not offset by biological carbon fixation processes (Ciais & Jones, 2013). Human-produced carbon emissions have had measurable and visible effects on the earth's ecosystems, increasing CO₂ concentration from 278 ppm in 1960 to 415 ppm in 2020 (Mehta et

al., 2023). Additionally, current rates of fossil fuel consumption and increasing world populations are expected to exhaust the fossil fuel supply by 2085 (Stephens et al., 2010). TAG-derived biodiesel can directly supplement petroleum-derived diesel, be used in existing engines, contains fewer contaminants than existing fuel sources, and provides a carbon-neutral renewable fuel source that could provide continuous energy for future generations (Song et al., 2022).

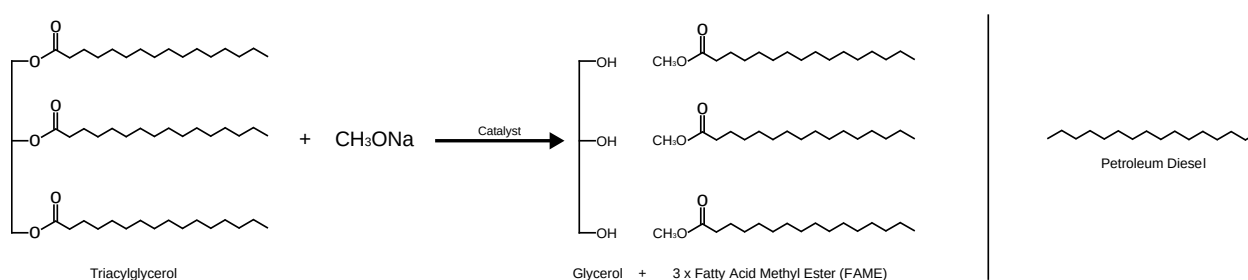


Figure 3: Transesterification of triacylglycerols (TAGs) to produce fatty acid methyl esters (FAMES). FAMES have similar structures to petroleum diesel and can be burned in existing diesel engines.

Current “biofuel crops” such as corn and oil palms can only produce a fraction of the biodiesel necessary to replace fossil fuels (Chisti, 2007). Further, the growth of these crops directly competes with cropland availability for food (Chisti, 2007). Algae can grow much faster with a doubling time of about 10-24 hours (Hoek et al., 1995). Additionally, algae produce up to 70% of their biomass as useable oil which results in 10-23 times more TAGs produced per hectare than alternative biofuel crops (Sharma et al., 2018). Algae can also grow in arid and nonarable landscapes to further limit competition with food crops. As such, algae are a highly productive biofuel crop and would only require 5.0% of existing US cropland to provide 100% of all transportation fuel needed for the United States (Chisti, 2007).

Impacts of Stress on Algal Growth and Biomass Production

While algae can grow in a wide range of conditions (e.g., temperatures 0-56 °C, pH 0-11.5, and salt concentrations 0.02-3 M, (Evans, 1958; Varshney et al., 2015) most algal species grow best at circum-neutral pH (pH 7), and between 20 and 30 °C (Pawlita-Posmyk et al., 2018). In ideal conditions where nutrients are in excess, algae can grow exponentially, replicating about 1-3 times every 24 hours (Hoek et al., 1995). While log phase growth produces high levels of biomass, algal cells typically have low levels of highly reduced storage compounds like TAGs. In these conditions, photosynthesis produces high levels of energy and needs to be tightly regulated to prevent cellular damage. Low levels of photosynthesis cause cells to be energy limited, while too much photosynthesis results in oxidative stress. In ideal conditions, algae can maintain this homeostasis by rapidly consuming excess photosynthetic energy through cell growth and division. Constant cell growth utilizes the NADPH created during photosynthesis and replenishes NADP⁺ pools that serve as an electron sink in the light-dependent reaction of photosynthesis (Sharma et al., 2012). If these pools are not regenerated, electrons from the light-dependent cycle build up and create oxidative radicals that would destroy photosynthetic machinery, proteins, and cell membranes (Pérez-Pérez et al., 2012). Therefore, algal photosynthesis, carbon fixation, and cell growth are tightly regulated and coupled to minimize the stresses associated with converting light to chemical energy.

However, when algae run out of nutrients and environmental conditions become unfavorable/stressful, algae stop growing and enter stationary phase (Luo & Liu, 2011; Song et al., 2022). Stationary phase cells cannot maintain photosynthetic homeostasis through cell growth, and excess reducing power can create oxidative radicals. In stressful conditions, algae must rewire

carbon metabolism to consume NADPH and replenish NADP⁺ pools through mechanisms other than cellular division (Shi et al., 2020). Multiple studies reviewed below have explored how algal physiology changes in response to different stresses including high pH, nutrient depletion, extreme temperatures, and hypersalinity to identify how algae minimize photooxidation in stressful conditions.

Salinity Stress

Salinity stress has been shown to affect the growth and biomass composition of algae. When exposed to hypersaline conditions, green algae *Ulva prolifera* shows a significant decrease in growth rate and high levels of oxidative damage (Luo & Liu, 2011). In response to high salinity, *U. prolifera* accumulated high levels of antioxidants like glutathione and increased the activities of reactive oxygen-scavenging enzymes. Exposure to saline conditions can also alter the carbon allocation and synthesis of storage molecules in algae. Salinity stress increased the accumulation of triacylglycerols (TAGs) and antioxidant pigments including astaxanthin in green algae *Chromochloris zofingiensis* (Kou et al., 2020). Additionally, *Chlorella sorokiniana* SDEC-18, a freshwater green alga, produced 2.16 times more lipids when cultivated in saline conditions compared to a mesosaline control (Zhang et al., 2018). These findings provide evidence that salinity stress triggers the accumulation of carbon storage compounds in algae.

Temperature Stress

Temperature is a crucial factor influencing algal productivity, particularly during mass cultivation in open ponds. Temperature fluctuations can also stress metabolic processes, resulting in altered growth, productivity, and biomass composition (Calhoun et al., 2021; Nalley et al., 2018; Raven & Geider, 1988; Singh & Singh, 2015). Cold stress has been found to reduce growth and

productivity in various algal taxa, but enhances lipid production within algal cells (Sushchik et al., 2003; Wang et al., 2016). Cold stress can also lead to an increase in unsaturated fatty acids that play a crucial role in maintaining membrane fluidity and permeability (Ma et al., 2017). In contrast, high temperatures enhance lipid production but increase the saturation of fatty acids (Renaud et al., 2002; Sharma et al., 2012).

pH Stress

Several studies have been conducted to understand the influence of different pH levels on algal growth and lipid production. Mixed algal cultures grown with varying pH levels had significantly higher lipid productivities (26.93%) at pH 8 compared to 14% lipid in lower pH systems (Alkhamis et al., 2022). Similarly, the eukaryotic alga *Rhodomonas salina* had lower productivity in high pH conditions but produced higher protein and total fatty acid content compared to neutral conditions (Latsos et al., 2022).

Nutrient Depletion

The most common environmental stress that impacts algal growth is the limitation of nutrients. When nutrients become depleted in the environment, algal growth can become severely limited. Algae are commonly nitrogen or phosphorous-limited within their native ecosystems which can lead to a decrease in biomass production and a transition to stationary phase (Sharma et al., 2018). Due to the prevalence of nutrient limitation and the resulting impacts on algal growth, this is one of the most well-studied algal stressors. Lipid accumulation is observed in select diatoms and green algae during phosphorous starvation. *Scenedesmus sp.* grown in phosphorous depleted conditions increases TAG accumulation by 53% (Xin et al., 2010). Phosphorous limitation in *Chlorella kessleri* also triggers TAG production of up to 39.3% of biomass (El-Sheekh & Rady,

1995). Interestingly, phosphorus limitation shows the opposite effect for other green algae imposing lower TAG accumulation in *Nannochloris atomus* and *Tetraselimis* sp. (Reitan et al., 1994). In contrast, all reported studies of nitrogen limitation show increased lipid production in algae (Arisz et al., 2000; Widjaja et al., 2009). *Chlorella vulgaris* increases lipid content by 40% in low nitrogen conditions (Illman et al., 2000). Lipid accumulation is also observed when coupling nitrogen depletion and the addition of inorganic carbon in *Scenedesmus* sp., *Chlamydomonas reinhardtii*, and *Phaeodactylum tricornutum* (Gardner et al., 2012; Gardner et al., 2013). This trend of increased TAG accumulation triggered by nitrogen deficiency is seen across a wide diversity of algae including diatoms, red algae, and green algae (Hsieh & Wu, 2009; Praveenkumar et al., 2012; Yeh & Chang, 2011).

Stress-Induced Lipid Accumulation Strategies

Stresses can have significant impacts on algal growth and biomass production. While these studies focused on a wide variety of algae and stress conditions one common similarity is that stressful conditions appear to shift carbon metabolism to producing highly reduced storage compounds like TAGs, starch, and pigments. This shift to synthesizing high-energy storage compounds helps to recycle NADP⁺ electron acceptor pools for photosynthesis and minimize oxidative stress even when cells are not growing (Sharma et al., 2012).

Intentionally inducing stressful conditions to remodel algal metabolism is an effective method to maximize the production of desired bioproducts like TAGs in algae (Shi et al., 2020). Researchers have implemented a two-stage growth system to first maximize biomass production, then impose a stressor to increase lipid composition (Sharma et al., 2018). By optimizing the amount of algal biomass through initial growth in optimal conditions followed by optimizing

biomass composition through stress induction researchers have been able to significantly improve the productivity and economic viability of algal biofuels.

Nitrogen starvation is the most used and extensively researched nutrient starvation method to maximize biofuel production in algae. Nitrogen is the primary limiting factor for the growth of eukaryotic microalgae and is one of the first nutrients to be exhausted during algae cultivation. Therefore, nitrogen stress is naturally applied to cultivation systems through algal growth. Additionally, TAG accumulation appears to increase in all studies testing nitrogen depletion to date, making it the most effective technique for inducing lipid production (Hsieh & Wu, 2009; Praveenkumar et al., 2012; Yeh & Chang, 2011). However, achieving high lipid production after nitrogen stress can take between 2-5 days and is often accompanied by minimal growth rates and low cell counts, which ultimately impact the overall biomass and lipid productivity (Widjaja et al., 2009). Although algae are currently the best potential source of renewable biofuel, their desired growth and lipid phenotypes seen at smaller scales do not always translate when scaled up (Teresa Lopes da Silva, 2015). As a result, large-scale biofuel production with existing algae strains is not economically feasible due to low biomass and oil productivity leading to high costs.

Current Limits of Algal Cultivation and Lipid Production

Algal cultivation for lipid production has many potential advantages, such as high biomass productivity, low land requirements, and the ability to utilize non-arable land and waste streams. However, there are also several limitations and hurdles that need to be overcome to make algal biofuels and other lipid-derived products economically and environmentally sustainable. While culturing experiments have provided proof of concept that algae can serve as a sustainable platform for producing biodiesel, cultivation costs are still too expensive (J. Chen, Li J., Dong, W., Zhang,

X., et al., 2018; Delrue et al., 2012). Large-scale algae cultivation can occur in either open raceway ponds or enclosed photobioreactors (Chisti, 2007). Photobioreactors have significant overhead and operating costs making them economically viable only when growing algae for high-value products like astaxanthin pigment (Kothari et al., 2017). Outdoor ponds provided much cheaper scale-up and are currently the only potentially economic choice for large-scale algae growth for biofuel production. The Department of Energy (DOE) has invested millions of dollars to reduce the cultivation costs of outdoor raceway systems to \$3 per gallon gasoline equivalent (Xin et al., 2016; Yousuf, 2020). However, when using raceway ponds the current algal biodiesel production costs are currently estimated to range from \$6.27 - \$14.43 per gallon (Energy, 2012; Hoffman et al., 2017; Prieto et al., 2017; Quinn & Davis, 2015; Rafa et al., 2021; Yousuf, 2020).

Many factors that contribute to high costs are non-biological and include purchasing materials and media nutrients, operating expenses, supplying water, and harvesting biomass. High costs are also due to the underperformance of algal growth in raceway ponds (Quinn & Davis, 2015). The underperformance of algal growth and lipid production in outdoor raceway ponds is mostly due to two major factors, low CO₂ transfer from the atmosphere into solution, and biological contamination (Wilson et al., 2009). CO₂ is the primary source of carbon for photosynthetic organisms and is required for algal growth. Algae cultures are traditionally grown at near neutral pH where total dissolved inorganic carbon concentration is low (0.07 mM) and dissolution rates struggle to replenish used CO₂ (Vadlamani et al., 2019; Yuvraj & Padmanabhan, 2017). As such CO₂ is often the limiting resource for algae in a raceway pond, and biomass production is lower than expected.

Open raceway ponds at neutral pH are also easily subject to contamination of unwanted algae, bacteria, protozoa, and viruses via airborne transmission which can reduce lipid productivity and quality (McBride et al., 2014). Contamination of competing microorganisms and parasites results in annual yield losses of ~30% in outdoor algae cultures (Newby et al., 2016; Rowe et al., 2013). Contamination of predators such as *Daphnia* (water fleas) can lead to a crash of an entire pond or facility leading to devastating economic losses (Vadlamani et al., 2017). Improving carbon transfer and minimizing contamination is necessary to make algal biofuels economically feasible.

To overcome CO₂ limitation, algal cultures are often sparged with external concentrated CO₂. This process improves inorganic carbon availability and restores carbon fixation and biomass production in raceway systems. However, purchasing and delivering external CO₂ sources significantly increases costs, contributing to 25-65% of cultivation-related operating costs (Energy et al., 2017; Eustance et al., 2020; Ryan Davis, 2016; Vadlamani et al., 2017). Cheaper sources of CO₂ such as flue gas from factories and refineries can be used, but flue gas contains toxins such as nitrogen and sulfur oxides that build up over time and slow algal biomass production (Doucha et al., 2005; Lara-Gil et al., 2014; Quinn & Davis, 2015; Ryan Davis, 2016). Further, this method does not address culture contamination. New algal strains are being identified with higher tolerance to toxin build-up, but even if biomass loss to toxin build-up can be overcome the required co-location of algae farms next to viable flue gas sources limits production area and this process could only displace 5% of total US petroleum imports (M. H. Langholtz, 2016; Michael Huesemann, 2017). An alternative growth method is required to improve carbon transfer and reduce contamination that does not rely on factory flue gas.

High Alkaline Cultivation Strategies

One potential solution to improve cultivation costs is growing algae in high alkaline, high pH conditions. Extreme cultivation conditions (> pH 10) have been shown to impose stress to trigger lipid production, but more importantly, they significantly improve atmospheric CO₂ dissolution into an algal culture via increased hydroxyl ion [OH⁻] concentrations (Vadlamani et al., 2017). CO₂ is an ideal gas and the maximum concentration of dissolved CO₂ can be calculated using Henry's Law (**Fig. 4**). Henry's law states that the concentration of dissolved CO₂ is dependent on the atmospheric partial pressure of CO₂ and temperature. Therefore, the concentration of CO₂ is constant across all pH ranges if the temperature and partial pressure remain unchanged. The maximum concentration of dissolved CO₂ at standard temperature and pressure is incredibly low at 0.013 mM. Once CO₂ reaches this very low concentration, dissolution from the atmosphere stops causing CO₂ to be a limiting growth factor for algae in neutral pH conditions.

$$C = K_H * P$$

Figure 4: Henry's Law, C = concentration of dissolved gas, K_H = Henry's Law constant, P = partial pressure of gas

When CO₂ dissolves into solution it can become deprotonated forming bicarbonate (HCO₃⁻, pK_a=6.3) or carbonate (CO₃⁻², pK_a=10.3 **Fig. 4a**). The ratio of inorganic carbon speciation is dependent on pH. Below pH 6.3 the major inorganic carbon species is CO₂, between pH 6.3 and 10.3 the major species is bicarbonate, and above 10.3 the major species is carbonate (**Fig. 5b,c**). Bicarbonate and carbonate are non-gaseous and can reach much higher maximum concentrations within solution compared to CO₂. At lower pH where CO₂ is the major inorganic carbon species, CO₂ dissolves and quickly reaches its maximum but very low concentration of 0.013 mM. This

causes CO_2 dissolution from the atmosphere to stop and limits algal growth. However, at higher pH where HCO_3^- and CO_3^{2-} are the major inorganic carbon species, when CO_2 dissolves it is converted to these other forms of inorganic carbon whose maximum concentrations are much greater. At a pH greater than nine, CO_2 is instantly converted into bicarbonate and carbonate, and high levels of CO_2 can dissolve before reaching the maximum concentration of 0.013 mM. As a result, atmospheric dissolution continues at high levels resulting in up to 10,000 times greater dissolved inorganic carbon (**Fig. 5d**). Inorganic carbon concentrations in high pH conditions can naturally reach 6 mM to 100 mM and eliminate the need for an external CO_2 supply.

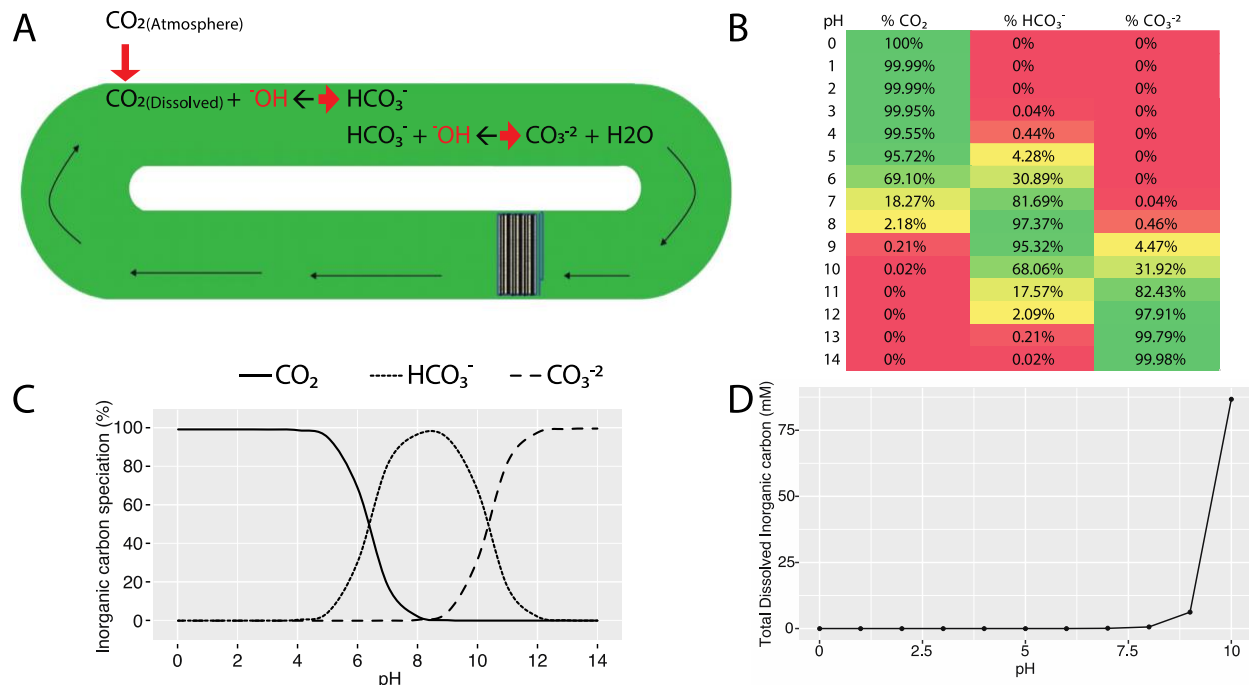


Figure 5: Dissolution of inorganic carbon across pH. (A) Schematic of inorganic carbon dissolution in algal cultivation systems. (B,C) Speciation of CO_2 , bicarbonate (HCO_3^-), and carbonate (CO_3^{2-}) across pH. Increasing hydroxyl (OH^-) concentrations increase carbon conversion to bicarbonate and carbonate. (D) Quantification of total dissolved inorganic carbon across different pH.

High pH systems have also been shown to effectively prevent culture contamination and culture crashes caused by competing microorganisms and predators like *Daphnia* (Gimmler, 2001; Vadlamani et al., 2017). Most microorganisms that contaminate algal cultivation systems grow rapidly at neutral pH but die in high pH environments. While most contaminating microorganisms cannot grow in extreme alkaline conditions, industrial algal strains have been isolated for natural alkaliphilic environments like Soap Lake, WA (100 mM inorganic carbon, pH 10) and Mono Lake, CA (pH 10) (Canon-Rubio et al., 2016; Gimmler, 2001). *Chlorella* sp. SLA-04 is an alkali-tolerant green alga isolated from Soap Lake, WA that can grow up to pH 11 while producing TAGs (Goemann et al., 2023; Vadlamani et al., 2017). SLA-04 is one of a select few alkali-tolerant model green algae that are being used to optimize alkaline cultivation systems. These alkali-tolerant organisms have comparable growth rates when grown in high alkaline conditions with natural CO₂ transfer compared to neutral conditions with supplemented CO₂ sparging (Canon-Rubio et al., 2016; Vadlamani et al., 2017). Currently, high alkaline systems are one of the best solution to naturally overcome CO₂ limitation without sparging and improve culture stability by limiting culture crashes.

Gaps in our Knowledge

Despite the potential benefits of high alkalinity cultivation conditions for algal growth and lipid production, there are still many unknowns and challenges that need to be addressed. While it is well established that some alkali-tolerant algae can thrive in extreme alkaline conditions, most research has focused on algal physiological composition and isolating novel algal strains with desired phenotypes of high alkaline growth and lipid production. Although high alkalinity conditions can induce a stress response in algae, it is not yet clear how this response is regulated

or how high alkalinity affects algal metabolism and lipid biosynthesis (Gimmler, 2001). Further research is needed to identify the key genes and metabolic pathways involved in the alkalinity-induced stress response. Understanding the mechanisms of growth, and how algal lipids are produced and stored within this extreme environment will help lower cultivation costs, reduce culture crashes, and provide targets for genome engineering and biomass optimization.

Hypothesis

I hypothesize that *Chlorella sp.* SLA-04 remodels its carbon metabolism to respond and adapt to changing environmental conditions and subsequent stresses. To test this hypothesis, I first sequenced and annotated the genome of *Chlorella sp.* SLA-04 to identify the genes and metabolic potential of this organism. This analysis is described in Chapter Three of this thesis. Using this genome, I next performed differential gene expression to identify how carbon is reallocated when SLA-04 is grown in high alkalinity conditions and encounters nutrient depletion. This transcriptomic analysis is described in Chapter Four of this thesis. Together this analysis provides evidence of genotypes that may contribute to environmental adaptation and global remodeling of lipid synthesis to combat photosynthetic stress and maintain viable cells throughout nutrient depletion and growth in extreme alkaline conditions.

References Cited

- Alkhamis, Y., Mathew, R., Ganesan, N., Rahman, S., & Rahman, M. M. (2022, 10/14). pH induced stress enhances lipid accumulation in microalgae grown under mixotrophic and autotrophic condition. *Frontiers in Energy Research*, 10, 1033068.
<https://doi.org/10.3389/fenrg.2022.1033068>
- Arisz, S. A., van Himbergen, J. A., Musgrave, A., van den Ende, H., & Munnik, T. (2000, Jan). Polar glycerolipids of *Chlamydomonas moewusii*. *Phytochemistry*, 53(2), 265-270.
[https://doi.org/10.1016/s0031-9422\(99\)00505-1](https://doi.org/10.1016/s0031-9422(99)00505-1)
- Behrenfeld, M. J. (2001). Biospheric Primary Production During an ENSO Transition. *Science*, 291(5513), 2594-2597. <https://doi.org/10.1126/science.1055071>
- Cai, H., Li, Q., Fang, X., Li, J., Curtis, N. E., Altenburger, A., Shibata, T., Feng, M., Maeda, T., Schwartz, J. A., Shigenobu, S., Lundholm, N., Nishiyama, T., Yang, H., Hasebe, M., Li, S., Pierce, S. K., & Wang, J. (2019). A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Scientific Data*, 6(1), 190022.
<https://doi.org/10.1038/sdata.2019.22>
- Calhoun, S., Bell, T. A. S., Dahlin, L. R., Kunde, Y., Labutti, K., Louie, K. B., Kuftin, A., Treen, D., Dilworth, D., Mihaltcheva, S., Daum, C., Bowen, B. P., Northen, T. R., Guarnieri, M. T., Starkenburg, S. R., & Grigoriev, I. V. (2021). A multi-omic characterization of temperature stress in a halotolerant *Scenedesmus* strain for algal biotechnology. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-021-01859-y>
- Canon-Rubio, K. A., Sharp, C. E., Bergerson, J., Strous, M., & De la Hoz Siegler, H. (2016, Feb). Use of highly alkaline conditions to improve cost-effectiveness of algal biotechnology. *Applied Microbiology and Biotechnology*, 100(4), 1611-1622.
<https://doi.org/10.1007/s00253-015-7208-7>
- Chapman, R. L. (2013). Algae: the world's most important "plants"—an introduction. *Mitigation and Adaptation Strategies for Global Change*, 18(1), 5-12.
<https://doi.org/10.1007/s11027-010-9255-9>
- Chen, J., Li J., Dong, W., Zhang, X., Tyagi, R. D., Drogui, P., & Surampalli, R. Y. (2018, 2018/07/01/). The potential of microalgae in biodiesel production. *Renewable and sustainable energy reviews*, 90, 336-346.
<https://doi.org/https://doi.org/10.1016/j.rser.2018.03.073>
- Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, 25(3), 294-306.
<https://doi.org/10.1016/j.biotechadv.2007.02.001>

- Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C., & Jones, C. L. Q., R.B. Myneni, S. Piao and P. Thornton,. (2013). *Carbon and Other Biogeochemical Cycles*. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.
- Delrue, F., Setier, P.-A., Sahut, C., Cournac, L., Roubaud, A., Peltier, G., & Froment, A.-K. (2012). An economic, sustainability, and energetic model of biodiesel production from microalgae. *Bioresource Technology*, 111, 191-200.
- Doucha, J., Straka, F., & Lívanský, K. (2005). Utilization of flue gas for cultivation of microalgae *Chlorella* sp.) in an outdoor open thin-layer photobioreactor. *Journal of Applied Phycology*, 17(5), 403-412. <https://doi.org/10.1007/s10811-005-8701-7>
- El-Sheekh, M., & Rady, A. (1995, 01/01). Effect of phosphorus starvation on growth, photosynthesis and some metabolic processes in the unicellular green alga *Chlorella kessleri*. *Phyton*, 35, 139-151.
- Energy, D. o. (2012). *algae biofuels, E.E.R.E U.S. Department of Energy in: Growing America's Energy Future*. .
- Energy, U. S. D. o., Energy, O. o. E. E. a. R., & Office, B. T. (2017). Algae Cultivation for Carbon Capture and Utilization Workshop Summary Report.
- Eustance, E., Lai, Y.-J. S., Shesh, T., & Rittmann, B. E. (2020, 2020/10/01/). Improved CO₂ utilization efficiency using membrane carbonation in outdoor raceways. *Algal Research*, 51, 102070. <https://doi.org/https://doi.org/10.1016/j.algal.2020.102070>
- Evans, J. H. (1958). The Survival of Freshwater Algae During Dry Periods: Part I. An Investigation of the Algae of Five Small Ponds. *The Journal of Ecology*, 46(1), 149. <https://doi.org/10.2307/2256910>
- Gardner, R. D., Cooksey, K. E., Mus, F., Macur, R., Moll, K., Eustance, E., Carlson, R. P., Gerlach, R., Fields, M. W., & Peyton, B. M. (2012). Use of sodium bicarbonate to stimulate triacylglycerol accumulation in the chlorophyte *Scenedesmus* sp. and the diatom *Phaeodactylum tricornutum*. *Journal of Applied Phycology*, 24(5), 1311-1320. <https://doi.org/10.1007/s10811-011-9782-0>
- Gardner, R. D., Lohman, E., Gerlach, R., Cooksey, K. E., & Peyton, B. M. (2013). Comparison of CO₂ and bicarbonate as inorganic carbon sources for triacylglycerol and starch accumulation in *Chlamydomonas reinhardtii*. *Biotechnology and Bioengineering*, 110(1), 87-96. <https://doi.org/10.1002/bit.24592>

- Gimmler, H., Degenhardt, B. (2001). Alkaliphilic and Alkali-Tolerant Algae. (Algal Adaptations to Environmental Stresses). https://doi.org/10.1007/978-3-642-59491-5_10
- Goemann, C. L. C., Wilkinson, R., Henriques, W., Bui, H., Goemann, H. M., Carlson, R. P., Viamajala, S., Gerlach, R., & Wiedenheft, B. (2023, 2023/01/01/). Genome sequence, phylogenetic analysis, and structure-based annotation reveal metabolic potential of *Chlorella* sp. SLA-04. *Algal Research*, 69, 102943. <https://doi.org/https://doi.org/10.1016/j.algal.2022.102943>
- Guiry, M. D. (2012). HOW MANY SPECIES OF ALGAE ARE THERE? *Journal of Phycology*, 48(5), 1057-1063. <https://doi.org/10.1111/j.1529-8817.2012.01222.x>
- Hall, D. O., & Rao, K. (1999). *Photosynthesis*. Cambridge University Press.
- Hoek, C., Mann, D., Jahns, H. M., & Jahns, M. (1995). *Algae: an introduction to phycology*. Cambridge university press.
- Hoffman, J., Pate, R. C., Drennen, T., & Quinn, J. C. (2017). Techno-economic assessment of open microalgae production systems. *Algal Research*, 23, 51-57. <https://doi.org/10.1016/j.algal.2017.01.005>
- Hsieh, C.-H., & Wu, W.-T. (2009, 2009/09/01/). Cultivation of microalgae for oil production with a cultivation strategy of urea limitation. *Bioresource Technology*, 100(17), 3921-3926. <https://doi.org/https://doi.org/10.1016/j.biortech.2009.03.019>
- Illman, A. M., Scragg, A. H., & Shales, S. W. (2000, Nov 1). Increase in *Chlorella* strains calorific values when grown in low nitrogen medium. *Enzyme Microb Technol*, 27(8), 631-635. [https://doi.org/10.1016/s0141-0229\(00\)00266-0](https://doi.org/10.1016/s0141-0229(00)00266-0)
- Kothari, R., Pandey, A., Ahmad, S., Kumar, A., Pathak, V. V., & Tyagi, V. V. (2017). Microalgal cultivation for value-added products: a critical enviro-economical assessment. *3 Biotech*, 7(4). <https://doi.org/10.1007/s13205-017-0812-8>
- Kou, Y., Liu, M., Sun, P., Dong, Z., & Liu, J. (2020, 2020/09/01/). High light boosts salinity stress-induced biosynthesis of astaxanthin and lipids in the green alga *Chromochloris zofingiensis*. *Algal Research*, 50, 101976. <https://doi.org/https://doi.org/10.1016/j.algal.2020.101976>
- Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, 16(5), 263-276. <https://doi.org/10.1038/nrmicro.2018.9>

- Lara-Gil, J. A., Alvarez, M. M., & Pacheco, A. (2014, Feb). Toxicity of flue gas components from cement plants in microalgae CO₂ mitigation systems. *Journal of Applied Phycology*, 26(1), 357-368. <https://doi.org/10.1007/s10811-013-0136-y>
- Latsos, C., Wassenaar, E., Moerdijk, T., Coleman, B., Robbens, J., Van Roy, S., Bastiaens, L., Van Houcke, J., & Timmermans, K. R. (2022). Effect of pH on *Rhodomonas salina* growth, biochemical composition, and taste, produced in semi-large scale under sunlight conditions. *Journal of Applied Phycology*, 34(3), 1215-1226. <https://doi.org/10.1007/s10811-022-02730-1>
- Luo, M. B., & Liu, F. (2011, 2011/12/01/). Salinity-induced oxidative stress and regulation of antioxidant defense system in the marine macroalga *Ulva prolifera*. *Journal of Experimental Marine Biology and Ecology*, 409(1), 223-228. <https://doi.org/https://doi.org/10.1016/j.jembe.2011.08.023>
- M. H. Langholtz, B. J. S., and L. M. Eaton (2016). 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy. *U.S. Department of Energy*.
- Ma, Z., Tian, M., Tan, Y., Cui, G., Feng, Y., Cui, Q., & Song, X. (2017). Response mechanism of the docosahexaenoic acid producer *Aurantiochytrium* under cold stress. *Algal Research*, 25, 191-199.
- McBride, R. C., Lopez, S., Meenach, C., Burnett, M., Lee, P. A., Nohilly, F., & Behnke, C. (2014). Contamination Management in Low Cost Open Algae Ponds for Biofuels Production. 10(3), 221-227. <https://doi.org/10.1089/ind.2013.0036>
- McCree, K. J. (1981). Photosynthetically Active Radiation. In O. L. Lange, P. S. Nobel, C. B. Osmond, & H. Ziegler (Eds.), *Physiological Plant Ecology I: Responses to the Physical Environment* (pp. 41-55). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68090-8_3
- Mehta, P., Sahil, K., Sarao, L. K., Jangra, M. S., & Bhardwaj, S. K. (2023). Algal Biofuels: Clean Energy to Combat the Climate Change. In (pp. 187-210). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6810-5_7
- Michael Huesemann, J. B. (2017). Microalgae Biofuels Production on CO₂ from Air.
- Nalley, J. O., O'Donnell, D. R., & Litchman, E. (2018). Temperature effects on growth rates and fatty acid content in freshwater algae and cyanobacteria. *Algal Research*, 35, 500-507.
- Newby, D. T., Mathews, T. J., Pate, R. C., Huesemann, M. H., Lane, T. W., Wahlen, B. D., Mandal, S., Engler, R. K., Feris, K. P., & Shurin, J. B. (2016). Assessing the potential of polyculture to accelerate algal biofuel production. *Algal Research*, 19, 264-277. <https://doi.org/10.1016/j.algal.2016.09.004>

- Pawlita-Posmyk, M., Wzorek, M., & Płaczek, M. (2018). The influence of temperature on algal biomass growth for biogas production. MATEC Web of Conferences,
- Pérez-Pérez, M. E., Lemaire, S. D., & Crespo, J. L. (2012). Reactive oxygen species and autophagy in plants and algae. *Plant Physiology*, 160(1), 156-164.
- Praveenkumar, R., Shameera, K., Mahalakshmi, G., Akbarsha, M. A., & Thajuddin, N. (2012, 2012/02/01/). Influence of nutrient deprivations on lipid accumulation in a dominant indigenous microalga *Chlorella* sp., BUM11008: Evaluation for biodiesel production. *Biomass and Bioenergy*, 37, 60-66.
<https://doi.org/https://doi.org/10.1016/j.biombioe.2011.12.035>
- Prieto, C. V. G., Ramos, F. D., Estrada, V., Villar, M. A., & Diaz, M. S. (2017). Optimization of an integrated algae-based biorefinery for the production of biodiesel, astaxanthin and PHB. *Energy*, 139, 1159-1172.
- Quinn, J. C., & Davis, R. (2015). The potentials and challenges of algae based biofuels: A review of the techno-economic, life cycle, and resource assessment modeling. *Bioresource Technology*, 184, 444-452. <https://doi.org/10.1016/j.biortech.2014.10.075>
- Rafa, N., Ahmed, S., Badruddin, I., Rahman, M. M., & Kamangar, S. (2021, 08/23). Strategies to Produce Cost-Effective Third-Generation Biofuel From Microalgae. *Frontiers in Energy Research*. <https://doi.org/10.3389/fenrg.2021.749968>
- Raven, J. A., & Geider, R. J. (1988). Temperature and algal growth. *New Phytologist*, 110(4), 441-461.
- Reitan, K. I., Rainuzzo, J. R., & Olsen, Y. (1994). EFFECT OF NUTRIENT LIMITATION ON FATTY ACID AND LIPID CONTENT OF MARINE MICROALGAE1. *Journal of Phycology*, 30(6), 972-979. <https://doi.org/10.1111/j.0022-3646.1994.00972.x>
- Renaud, S. M., Thinh, L.-V., Lambrinidis, G., & Parry, D. L. (2002). Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures. *Aquaculture*, 211(1-4), 195-214.
- Rowe, J. M., Dunigan, D. D., Blanc, G., Gurnon, J. R., Xia, Y., & Van Etten, J. L. (2013). Evaluation of higher plant virus resistance genes in the green alga, *Chlorella variabilis* NC64A, during the early phase of infection with *Paramecium bursaria chlorella virus-1*. *Virology*, 442(2), 101-113. <https://doi.org/10.1016/j.virol.2013.04.018>
- Ryan Davis, J. M., Christopher Kinchin, Nicholas Grundl, Eric Tan, David Hymbird. (2016). Process Design and Economics for the Production of Algal Biomass: Algal Biomass

Production in Open Pond Systems and Processing Through Dewatering for Downstream Conversion.

- Sharma, J., Kumar, S. S., Bishnoi, N. R., & Pugazhendhi, A. (2018). Enhancement of lipid production from algal biomass through various growth parameters. *Journal of Molecular Liquids*, 269, 712-720. <https://doi.org/10.1016/j.molliq.2018.08.103>
- Sharma, K. K., Schuhmann, H., & Schenk, P. M. (2012). High Lipid Induction in Microalgae for Biodiesel Production. *Energies*, 5(5), 1532-1553. <https://doi.org/10.3390/en5051532>
- Shi, T. Q., Wang, L. R., Zhang, Z. X., Sun, X. M., & Huang, H. (2020). Stresses as First-Line Tools for Enhancing Lipid and Carotenoid Production in Microalgae. *Front Bioeng Biotechnol*, 8, 610. <https://doi.org/10.3389/fbioe.2020.00610>
- Singh, S., & Singh, P. (2015). Effect of temperature and light on the growth of algae species: A review. *Renewable and sustainable energy reviews*, 50, 431-444.
- Song, X., Liu, B.-F., Kong, F., Ren, N.-Q., & Ren, H.-Y. (2022, 2022/08/01/). Overview on stress-induced strategies for enhanced microalgae lipid production: Application, mechanisms and challenges. *Resources, Conservation and Recycling*, 183, 106355. <https://doi.org/https://doi.org/10.1016/j.resconrec.2022.106355>
- Stephens, E., Ross, I. L., Mussgnug, J. H., Wagner, L. D., Borowitzka, M. A., Posten, C., Kruse, O., & Hankamer, B. (2010). Future prospects of microalgal biofuel production systems. *Trends in Plant Science*, 15(10), 554-564. <https://doi.org/10.1016/j.tplants.2010.06.003>
- Supuran, C. T. (2008). Carbonic anhydrases-an overview. *Current pharmaceutical design*, 14(7), 603-614.
- Sushchik, N., Kalacheva, G., Zhila, N., Gladyshev, M., & Volova, T. (2003). A Temperature Dependence of the Intra-and Extracellular Fatty-Acid Composition of Green Algae and Cyanobacterium. *Russian journal of plant physiology*, 50(3).
- Teresa Lopes da Silva, A. R. (2015). scale-up problems for the large scale production of algae. https://doi.org/10.1007/978-3-319-22813-6_6
- Tsai, D. D.-W., Chen, P. H., Chou, C. M.-j., Hsu, C.-F., & Ramaraj, R. (2015, 2015/11/01/). Carbon sequestration by alga ecosystems. *Ecological Engineering*, 84, 386-389. <https://doi.org/https://doi.org/10.1016/j.ecoleng.2015.09.024>
- Vadlamani, A., Pendyala, B., Viamajala, S., & Varanasi, S. (2019). High Productivity Cultivation of Microalgae without Concentrated CO₂ Input. *ACS Sustainable Chemistry & Engineering*, 7(2), 1933-1943. <https://doi.org/10.1021/acssuschemeng.8b04094>

- Vadlamani, A., Viamajala, S., Pendyala, B., & Varanasi, S. (2017). Cultivation of Microalgae at Extreme Alkaline pH Conditions: A Novel Approach for Biofuel Production. *ACS Sustainable Chemistry & Engineering*, 5(8), 7284-7294.
<https://doi.org/10.1021/acssuschemeng.7b01534>
- Varshney, P., Mikulic, P., Vonshak, A., Beardall, J., & Wangikar, P. P. (2015, May). Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour Technol*, 184, 363-372. <https://doi.org/10.1016/j.biortech.2014.11.040>
- Walker, J. C. G. (1980). The Oxygen Cycle. In (pp. 87-104). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-662-24940-6_5
- Wang, Y., He, B., Sun, Z., & Chen, Y.-F. (2016). Chemically enhanced lipid production from microalgae under low sub-optimal temperature. *Algal Research*, 16, 20-27.
- Widjaja, A., Chien, C.-C., & Ju, Y.-H. (2009, 2009/01/01/). Study of increasing lipid production from fresh water microalgae *Chlorella vulgaris*. *Journal of the Taiwan Institute of Chemical Engineers*, 40(1), 13-20.
<https://doi.org/https://doi.org/10.1016/j.jtice.2008.07.007>
- Wilson, W. H., Van Etten, J. L., & Allen, M. J. (2009). The Phycodnaviridae: The Story of How Tiny Giants Rule the World. In (pp. 1-42). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-540-68618-7_1
- Xin, C., Addy, M. M., Zhao, J., Cheng, Y., Cheng, S., Mu, D., Liu, Y., Ding, R., Chen, P., & Ruan, R. (2016). Comprehensive techno-economic analysis of wastewater-based algal biofuel production: A case study. *Bioresour Technol*, 211, 584-593.
<https://doi.org/10.1016/j.biortech.2016.03.102>
- Xin, L., Hong-ying, H., Ke, G., & Ying-xue, S. (2010, 2010/07/01/). Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga *Scenedesmus* sp. *Bioresour Technol*, 101(14), 5494-5500.
<https://doi.org/https://doi.org/10.1016/j.biortech.2010.02.016>
- Yeh, K. L., & Chang, J. S. (2011, Nov). Nitrogen starvation strategies and photobioreactor design for enhancing lipid content and lipid production of a newly isolated microalga *Chlorella vulgaris* ESP-31: implications for biofuels. *Biotechnol J*, 6(11), 1358-1366.
<https://doi.org/10.1002/biot.201000433>
- Yodsuwan, N., Sawayama, S., & Sirisansaneeyakul, S. (2017). Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom *Phaeodactylum tricornutum*. *Agriculture and Natural Resources*, 51(3), 190-197.
<https://doi.org/10.1016/j.anres.2017.02.004>

- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004, May). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol*, 21(5), 809-818. <https://doi.org/10.1093/molbev/msh075>
- Yousuf, A. (2020). Microalgae Cultivation for Biofuels Production Preface. *Microalgae Cultivation for Biofuels Production*, Xv-Xvi. https://doi.org/Book_Doi 10.1016/C2018-0-01358-4
- Yuvraj, & Padmanabhan, P. (2017). Technical insight on the requirements for CO₂-saturated growth of microalgae in photobioreactors. *3 Biotech*, 7(2). <https://doi.org/10.1007/s13205-017-0778-6>
- Zhang, L., Pei, H., Chen, S., Jiang, L., Hou, Q., Yang, Z., & Yu, Z. (2018, 2018/02/01/). Salinity-induced cellular cross-talk in carbon partitioning reveals starch-to-lipid biosynthesis switching in low-starch freshwater algae. *Bioresource Technology*, 250, 449-456. <https://doi.org/https://doi.org/10.1016/j.biortech.2017.11.067>

CHAPTER THREE

GENOME SEQUENCE, PHYLOGENETIC ANALYSIS, AND STRUCTURE-BASED
ANNOTATION REVEALS METABOLIC POTENTIAL OF CHLORELLA SP. SLA-04

Contribution of Authors and Co-Authors

Author: Calvin L.C. Goemann

Contributions: Conceptualization, Methodology, Investigation & Data Collection, Genomics & Bioinformatic Analysis, Writing – Original Draft, Writing – Review & Editing

Co-Author: Royce Wilkinson

Contributions: Methodology, Investigation & Data Collection, Writing – Review & Editing

Co-Author: William Henriques

Contributions: Genomics & Bioinformatic Analysis, Writing – Review & Editing

Co-Author: Huyen Bui

Contributions: Genomics & Bioinformatic Analysis, Writing – Review & Editing Co-Author:

Hannah M. Goemann

Contributions: Investigation & Data Collection, Genomics & Bioinformatic Analysis, Writing – Review & Editing

Co-Author: Ross P. Carlson

Contributions: Conceptualization, Writing – Review & Editing

Co-Author: Sridhar Viamajala

Contributions: Conceptualization, Writing – Review & Editing

Co-Author: Robin Gerlach

Contributions: Conceptualization, Writing – Review & Editing

Co-Author: Blake Wiedenheft

Contributions: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing

Manuscript Information

Calvin L. C. Goemann, Royce Wilkinson, William Henriques, Huyen Bui, Hannah M. Goemann,
Ross P. Carlson, Sridhar Viamajala, Robin Gerlach, Blake Wiedenheft

Algal Research

Status of Manuscript: [Put an x in one of the options below, delete this]

☐ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☒ Published in a peer-reviewed journal

Algal Research

Submitted 17 February 2022

Published January 2023

Volume 69

DOI: <https://doi.org/10.1016/j.algal.2022.102943>

Genome sequence, phylogenetic analysis, and structure-based annotation reveals metabolic potential of *Chlorella* sp. SLA-04

Calvin L. C. Goemann^{a,c}, Royce Wilkinson^a, William Henriques^a, Huyen Bui^{b,c}, Hannah M. Goemann^{a,c}, Ross P. Carlson^{b,c}, Sridhar Viamajala^d, Robin Gerlach^{b,c}, Blake Wiedenheft^{a*}

^aDepartment of Microbiology and Cell Biology, Montana State University, Bozeman, Montana, USA

^bDepartment of Chemical and Biological Engineering, Montana State University, Bozeman, Montana, USA

^cCenter for Biofilm Engineering, Montana State University, Bozeman, Montana, USA

^dDepartment of Chemical Engineering, University of Toledo, Toledo Ohio, USA

*Correspondence: bwiedenheft@gmail.com

Running Head: *Chlorella* sp. SLA-04 genomic analysis

Abstract

Algae are a broad class of photosynthetic eukaryotes that are phylogenetically and physiologically diverse. Most of the phylogenetic diversity has been inferred from 18S rDNA sequencing since there are only a few complete genomes available in public databases. Here we use ultra-long-read Nanopore sequencing to determine a gapless, telomere-to-telomere complete genome sequence of *Chlorella sp.* SLA-04, previously described as *Chlorella sorokiniana* SLA-04. *Chlorella sp.* SLA-04 is a green alga that grows to high cell density in a wide variety of environments – high and neutral pH, high and low alkalinity, and high and low salinity. SLA-04's ability to grow in high pH and high alkalinity media without external CO₂ supply is favorable for large-scale algal biomass production. Phylogenetic analysis performed using ribosomal DNA and conserved protein sequences consistently reveal that *Chlorella sp.* SLA-04 forms a distinct lineage from other strains of *Chlorella sorokiniana*. We complement traditional genome annotation methods with high throughput structural predictions and demonstrate that this approach expands functional prediction of the SLA-04 proteome. Genomic analysis of the SLA-04 genome identifies the genes capable of utilizing TCA cycle intermediates to replenish cytosolic acetyl-CoA pools for lipid production. We also identify a complete metabolic pathway for sphingolipid anabolism that may allow SLA-04 to readily adapt to changing environmental conditions and facilitate robust cultivation in mass production systems. Collectively, this work clarifies the phylogeny of *Chlorella sp.* SLA-04 within Trebouxiophyceae and demonstrates how structural predictions can be used to improve annotation beyond sequence-based methods.

Introduction

Algae are diverse photosynthetic eukaryotes that assimilate six times more nitrogen than terrestrial plants, perform 45% of global oxygenic photosynthesis, and 50% of total CO₂ fixation (Behrenfeld, 2001; Field, 1998; Kuypers et al., 2018; Walker, 1980). The phylogenetic and physiological diversity of algae is mirrored by the diversity of aquatic ecosystems where algae are found (e.g., temperatures 0-56°C, pH 0-11.5, and salt concentrations 0.02-3M) (Evans, 1958; Varshney et al., 2015). However, it is currently unclear how selective pressures from disparate aquatic environments shape algal genotypes and corresponding phenotypes.

Genomic analysis of algae across different environments is essential to understand how algae survive and adapt to changing conditions. As of September 2022, there are 170 green algae genomes in the National Center for Biotechnology Information (NCBI) database. Most of these genomes are highly fragmented and only 16 contain defined chromosome sequences. This resource limitation restricts efforts to confidently identify genotypes that drive functional diversity. Understanding the phylogenetic and functional diversity of algae will benefit from a broader collection of complete algal genomes that can be used for comparative analyses.

To date, most algal genomes have been sequenced using short-read sequencing platforms (Blanc-Mathieu et al., 2014; Merchant et al., 2007). While these platforms are accurate, PCR bias and short read lengths frequently result in incomplete assemblies and omission of repetitive regions, such as retroelements that can represent up to 90% of a eukaryotic genome (Amarasinghe et al., 2020; Kingsford et al., 2010; Koren et al., 2012; Mehrotra & Goyal, 2014). In contrast, ultra-long-read sequencing platforms generate megabase-length reads that span repetitive regions, and these methods do not rely on PCR (Deamer et al., 2016; Payne et al., 2018). While the error rates

for long-read technologies are higher, consensus correction software significantly increases accuracy as a function of read depth (Dohm et al., 2020; Wang et al., 2021). Long-read sequence data with >100x coverage often results in highly accurate chromosome level assemblies (Sutton et al., 2021).

Here we determine a complete genome sequence of *Chlorella. sp.* SLA-04, previously described as *Chlorella sorokiniana* SLA-04, using ultra-long-read Nanopore sequencing. *Chlorella sp.* SLA-04 is an oleaginous green alga that was isolated from a brackish, alkaline lake (100mM HCO_3^- , pH 10) in eastern Washington (USA) (Vadlamani et al., 2017). SLA-04 has growth and composition characteristics that are attractive for biofuel production. Phylogenetic analysis performed using either ribosomal DNA or conserved protein sequences reveal that SLA-04 forms a distinct lineage from other strains of the same species. While sequenced-based annotation methods confidently assign putative function to a majority of the SLA-04 genes, 44.2% of the genes were classified as “hypothetical”. Using a high-throughput structure prediction software (i.e., AlphaFold) we predict the 3D protein structures encoded by the “hypothetical” proteome. Structural predictions are queried against the protein data bank (PDB) to identify functional homologs and expand functional predictions for 677 SLA-04 genes that are widely conserved in algae and a subset of genes that are unique to *Chlorella sp.* SLA-04. Collectively, this work expands our phylogenetic and functional understanding of algae.

Methods

2.1 Strain Information

2.1.1 Strain Cultivation Conditions. *Chlorella* sp. SLA-04 cultures were grown in Modified Bold's Basal Medium pH 8.7 (Bolds) containing NaNO₃ (2.94 mM), K₂HPO₄ (1.43 mM), KH₂PO₄ (1.43 mM), MgSO₄·7H₂O (0.30 mM), CaCl₂·2H₂O (0.17 mM), NaCl (0.42 mM), 1 mL/L of EDTA solution, 1 mL/L iron solution, and 2 mL/L trace metal solution. The EDTA solution contained 50 g/L Na₂EDTA and 31 g/L KOH. The iron solution contained 4.98 g/L FeSO₄·7H₂O and 1 mL/L concentrated H₂SO₄. The trace metal solution contained H₃BO₃ (9.7 mM), MnCl₂·4H₂O (1.26 mM), ZnCl₂ (0.15 mM), CuCl₂·2H₂O (0.11 mM), Na₂MoO₄·2H₂O (0.07 mM), CoCl₂·6H₂O (0.06 mM), NiCl₂·6H₂O (0.04 mM), V₂O₅ (0.01 mM) and KBr (0.08 mM) (Bold, 1949). Media was titrated to pH 8.7 with NaOH. Cultures were grown at 23°C with 400 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ light intensity cycling with a 14:10 hour light/dark cycle using T5 fluorescent tubes.

2.1.2 Axenic Isolation. An axenic culture of *Chlorella* sp. SLA-04 was generated through antibiotic treatment modified from Mustapa et al. (Mustapa et al., 2016). Briefly, clonal isolates of SLA-04 were streaked onto plates containing Bolds agar with 700 $\mu\text{g}/\text{mL}$ ampicillin and 200 $\mu\text{g}/\text{mL}$ cefotaxime. Individual colonies were picked and grown in liquid Bolds. Plate antibiotic treatment and isolation was repeated a second time and algal colonies were picked and grown in liquid Bolds with 10 $\mu\text{g}/\text{mL}$ tetracycline. Axenic cultures were confirmed using amplicon sequencing of the V4 region of the 16S rRNA on an Illumina MiSeq platform (**Sup. Fig. 1**) (Goemann et al., 2021). A small percentage of 16S rRNA reads within the axenic culture mapped to *Ralstonia* species, but *Ralstonia* sequences are common contaminants of DNA purification

columns (Weyrich et al., 2019). Axenic stock cultures of SLA-04 were maintained on agar slants and in liquid culture of Bolds with 10 µg/mL tetracycline.

2.2 Genome Sequencing and Assembly

2.2.1 High Molecular Weight DNA Extraction. *Chlorella sp.* SLA-04 was grown in Bolds to log phase. Cells from 100 mL of culture were spun down at 3,000 x g for 10 minutes and transferred to a liquid N₂ chilled mortar. The pellet was ground for 30 minutes with liquid N₂ added as needed. Ground algal biomass was transferred to 15 mL of lysis buffer (20 mM EDTA pH 8.25, 10 mM Tris, 500 mM guanidine HCl, 200 mM NaCl, 1% w/w Triton X-100, and 9 mg hemicellulase) and incubated at 37°C for 1 hour. To remove RNA, RNase A was added to a final concentration of 20 µg/mL and the mixture was incubated for an additional 30 minutes at 37°C. To remove proteins, Proteinase K (12 µg) was added and the mixture was incubated at 50°C for 2 hours. Samples were pelleted at 15,000 x g at 4°C for 20 minutes and the supernatant was added to an equilibrated Qiagen Genomic-tip 100/G gravity column (Cat. No.10243). DNA was washed and eluted per the manufacturer's instructions. DNA was precipitated with 3.5 mL molecular grade isopropyl alcohol (IPA). The DNA pellet was collected by centrifugation (5850 x g) at 4°C for 30 minutes and resuspended in molecular grade water.

2.2.2 RNA Extraction. *Chlorella sp.* SLA-04 RNA was extracted using the Qiagen RNeasy powerplant RNA isolation kit (Cat. No.13500-50) with modifications. SLA-04 culture was grown to log phase (2×10^7 cells/mL) in Bolds. Cells from 200 mL of culture were pelleted at 3,000 x g for 5 minutes at 4°C and the pellet was ground for 10 minutes in a liquid N₂ chilled mortar using a pestle before being transferred to 1 mL Trizol in a PowerBead tube provided in the Qiagen RNeasy powerplant kit. Potassium acetate (1 M) was added to a final concentration of 0.2 M to

improve polysaccharide precipitation. Phenolic separation solution (50 μ L) from the Qiagen powerplant kit was added. The sample was shaken at 6.5 meters per second for 60 seconds using a MP Biomedicals FastPrep-24 Classic bead beater, incubated at room temp for 20 minutes, and centrifuged at 13,000 x g for 10 minutes at 4°C. Protein and cell debris was removed by adding 200 μ L chloroform, shaking for 30 seconds, centrifuging at 13,000 x g for 10 minutes at 4°C and transferring the aqueous phase to a new tube. RNA was extracted and purified from the aqueous phase using the RNeasy powerplant RNA isolation kit per manufacturer's instructions starting at the inhibitor removal step (step 6). All glassware was washed with 2% diethylpyrocarbonate (DEPC) in water (v/v) and autoclaved before use.

2.2.3 DNA and RNA Sequencing. The *Chlorella sp.* SLA-04 genome was sequenced using Nanopore sequencing. High molecular weight DNA was sequenced using a MinION sequencer and the ligation sequencing kit (SQK-LSK109 by Oxford Nanopore) per manufacturer's instructions. Briefly, 1 μ g of genomic DNA was used for library preparation that included end preparation and ligation to Nanopore adapters. 350 ng of the DNA library was used and sequenced on the MinION sequencer. Sequencing generated 1.97 M reads totaling 5.64 gigabases (Gb) of data.

Chlorella sp. SLA-04 mRNA sequencing was performed using a MinION sequencer and the direct RNA sequencing kit (SQK-RNA002 by Oxford Nanopore) per manufacturer's instructions. mRNA was isolated and purified from total RNA using magnetic Dynabeads coated with Oligo(dT)₂₅ adapters. Purification using magnetic beads yielded 338 ng of mRNA from 22 μ g of total RNA. 338 ng of mRNA was used for Nanopore library preparation that included reverse transcription of cDNA scaffold and ligation of Nanopore adapters. 168 ng of mRNA library was

used and sequenced on the MinION Nanopore sequencer. Sequencing generated 1.12 M reads totaling 1.35 Gb of data.

2.2.4 Genome Assembly and Annotation. Nanopore raw DNA reads (5.64 GB) were basecalled and analyzed for quality using MinKNOW 4.1.22 software in fast basecalling mode. Quality reads were submitted to CANU v1.9, FLYE v2.5-g0c3de5b, and Miniasm v0.3-r179 programs for long-read genome assembly (Kolmogorov et al., 2019; Koren et al., 2017; Li, 2016). CANU, FLYE, and Miniasm assembly pipelines each generated contig level assemblies with minor gaps. Manual curation of assemblies was performed by aligning contigs from the three assembly pipelines using minimap2 v2.17-r954-dirty (Li, 2018a). Each gap within the individual assemblies was completely bridged by contigs from at least one other genome assembly. Assemblies were merged to form a chromosome level genome with no gaps. Chromosomes were considered complete if they had flanking telomeric repeats (CCCTAAA)_n on both termini with no gaps. Coverage depth of DNA reads was calculated using samtools v1.13 (Danecek et al., 2021).

RepeatMasker v4.0.9 was used to identify repetitive elements in the *Chlorella* sp. SLA-04 genome using Repbase and Dfam as reference databases (Bao et al., 2015; Smit et al., 1996; Storer et al., 2021). LTRharvest v1.5.10 was used to identify and fetch the sequences of long-terminal repeats in the SLA-04 genome by performing a six open reading frame (ORF) translation of the genome and scanning each ORF longer than 75 amino acids for homology to LTR retrotransposon protein domains from Pfam (El-Gebali et al., 2019; Ellinghaus et al., 2008; Rice et al., 2000). To identify active retrotransposons, we extracted 13 sequences that contained *gag* and *pol* domains flanked by long-terminal repeats and aligned them in MAFFT v7.429 (Rozewicki et al., 2019). To verify LTR retrotransposon families, we aligned the capsid domains of diverse families of LTR

retrotransposons from Repbase and those identified in SLA-04. SLA-04 retrotransposons clustered with retrotransposons from Copia retrotransposons found in plants.

Raw mRNA Nanopore reads were mapped to the SLA-04 genome using minimap2 and a consensus sequence and transcript counts were generated using Stringtie v2.0 and Gffread v0.11.6 (Pertea & Pertea, 2020; Pertea et al., 2015). The mRNA consensus data was used to aid the *de novo* annotation prediction software MAKER v3.01.03 to generate a transcriptome of SLA-04 (Campbell et al., 2014; Cantarel et al., 2007). Gene functions were assigned by submitting protein sequence to the Kyoto Encyclopedia for Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) and Egnog mapper webserver with default settings (Cantalapiedra et al., 2021; Moriya et al., 2007). Genome completeness was assessed by BUSCO v5.4.2 using 1519 single copy orthologs found in the Chlorophyta clade (Manni et al., 2021; Simão et al., 2015). A schematic of the SLA-04 genome (**Fig. 1**) was generated using the ggplot2 v3.3.3 package within RStudio v1.3.959 (Hadley, 2016). GC content was determined using the RStudio seqinr package v4.2-5 (Charif & Lobry, 2007).

Structural-based protein annotation was performed following the ColabFold-FoldSeek-EggNOG (CoFFE) pipeline (Ruperti et al., 2022). Briefly, amino acid sequences were submitted to the ColabFold AlphaFold2-batch notebook through google colab with default settings (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022). The highest-ranking predicted protein data base (pdb) structure with predicted local distance difference test on the C α atoms (pLDDT) > 70 was selected for each analyzed protein. PDB structures were submitted to Foldseek v3.915ef7d for structural comparison against the AlphaFold database of protein structures. For each SLA-04 protein we kept the AlphaFold database hit with the greatest bit score. We enforced a e-value cutoff

of 1xe-05. Sequences of AlphaFold database best hits of each SLA-04 protein were annotated using EggNOG mapper webserver with default settings.

2.3 Comparative Analyses

2.3.1 Phylogenetic Analysis. Phylogenetic analysis was performed on all chromosome level green algal genomes (n=10) (**Sup Fig. 3**) and all Trebouxiophyceae genomes (n=25) (**Fig. 2**) within the NCBI database. 18S rDNA sequences were identified and extracted from chromosome level algal genomes (n=10) using SSU-align v0.1.1 (Cannone et al., 2002; Eddy & Nawrocki, 2009). Ribosomal sequences were trimmed with trimal v1.4.rev22 with the –nogap flag, aligned with MAFFT with the –auto flag, and a maximum likelihood phylogenetic tree of the 18S multiple sequence alignment was generated using IQtree v1.6.1 using default settings (Katoh & Standley, 2013; Nguyen et al., 2015). Phylogeny based on orthologous proteins was performed on all Trebouxiophyceae genomes (n=25) using OrthologR v0.4.0 in RStudio (Drost et al., 2015). Briefly, OrthologR performed a reciprocal best hit blast for orthology inference between algal species. Orthologous sequences were aligned based on amino acid sequence using the Needleman-Wunsch method and codons were aligned using pal2nal (Suyama et al., 2006). The Comeron's method was used to predict dN/dS estimations. Orthologous protein sequences were aligned with MAFFT and a maximum likelihood phylogenetic tree was generated using IQtree (Katoh & Standley, 2013; Nguyen et al., 2015). Trees were rendered using the RStudio ggtree v2.0.4 package (Yu et al., 2017).

Results

3.1 General Genome Characteristics

3.1.1 *Chlorella* sp. SLA-04 Genome Assembly Statistics. To determine a telomere-to-telomere genome of *Chlorella* sp. SLA-04, high molecular weight DNA was extracted from SLA-04 cultures for long-read sequencing using a MinION sequencer from Oxford Nanopore (**Sup. Fig 1**). Nanopore sequencing resulted in 5.64 Gb of sequence data in a total of 1.97 M reads (**Sup. Table 1**). Sequencing reads were assembled using a combination of FLYE, Miniasm, and CANU (Kolmogorov et al., 2019; Koren et al., 2017; Li, 2016), resulting in a 38.8 Mb complete genome with an average of 117x coverage. The SLA-04 genome consists of 13 linear chromosomes, ranging from 1.20 to 5.25 Mbp in length with an average GC content of 65.87%, and circular mitochondrion and chloroplast sequences (**Fig. 1, Sup. Table 1**). This chromosomal architecture is consistent with previously published *Chlorella* genomes, which contain 12-16 nuclear chromosomes (Arriola et al., 2018; Higashiyama et al., 1995; Hovde et al., 2018).

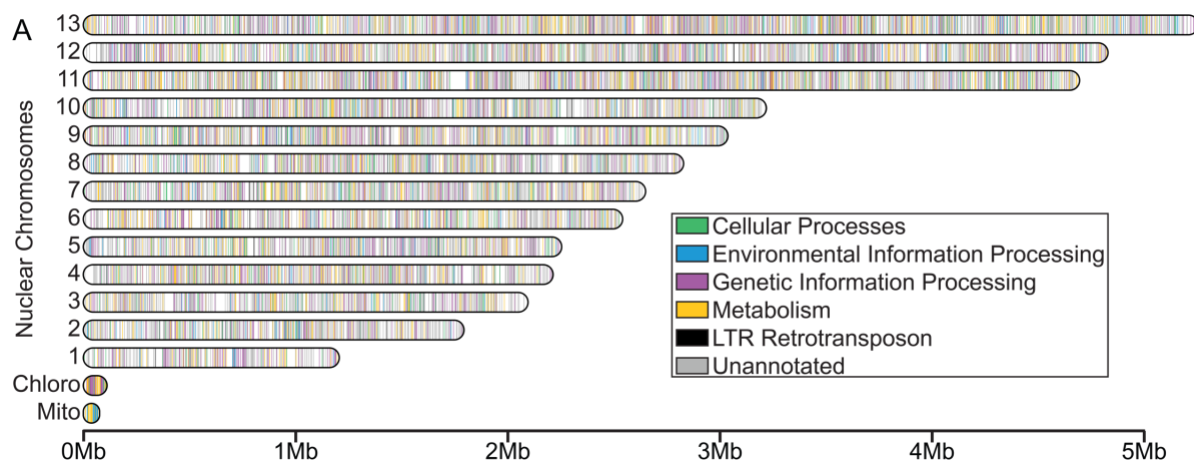


Figure 1. Complete genome of *Chlorella* sp. SLA-04. (A) Chromosome level assembly of *Chlorella* sp. SLA-04 reveals thirteen nuclear chromosomes, and circular chloroplast and mitochondrial genomes. Genes were annotated with KEGG automatic annotation server (KAAS) and color coded according to six different functional categories.

3.1.2 Genome Completeness Analysis. Genome completeness was evaluated using several distinct but complementary metrics. First, BUSCO was used to calculate the percentage of orthologous Chlorophyta genes within the *Chlorella* sp. SLA-04 genome (Manni et al., 2021). The SLA-04 genome contains 93.4% of Chlorophyta orthologs (n=1418) which surpasses the 90% BUSCO cutoff considered to represent “good” genome assemblies (**Sup. Table 1**) (Hanschen et al., 2020). Second, repetitive elements were identified and mapped to unique chromosomal locations, indicating that this assembly provides sufficient coverage to unambiguously assign chromosomal location to repetitive elements that have historically frustrated complete genome assemblies (**Fig. 1, Sup. Fig 2, Sup. Table 2**) (Smit et al., 1996; Tarailo-Graovac & Chen, 2009). Finally, each nuclear chromosome is composed of a single contiguous well-supported sequence flanked by telomeric repeats (CCCTAAA)_n (Fulnečková et al., 2012) (**Fig. 1**). Collectively, these metrics suggest that the SLA-04 genome constitutes a complete telomere-to-telomere algal genome.

3.1.3 Annotation of *Chlorella* sp. SLA-04. *Chlorella* sp. SLA-04 mRNA was extracted, and directly sequenced using a MinION sequencer. Sequences were submitted to MAKER v3.01.02 and aligned to the SLA-04 genome to predict protein coding sequences (CDSs) and splice sites (Campbell et al., 2014; Holt & Yandell, 2011). The SLA-04 genome contains 10,310 CDSs that span 45.73% of the genome with an average gene length of 3,475 bp (Sup. Table 1). EggNOG mapper was used to assign gene function resulting in 5,757 (55.8%) CDSs with predicted function (Fig. 2A) (Cantalapiedra et al., 2021). Annotated gene functions appear evenly distributed across the nuclear and plastid chromosomes with little functional clustering (Fig. 1). Genes coding for

proteins that participate in protein synthesis, amino acid metabolism and chaperones are the most abundant gene categories in the SLA-04 genome (Fig. 2).

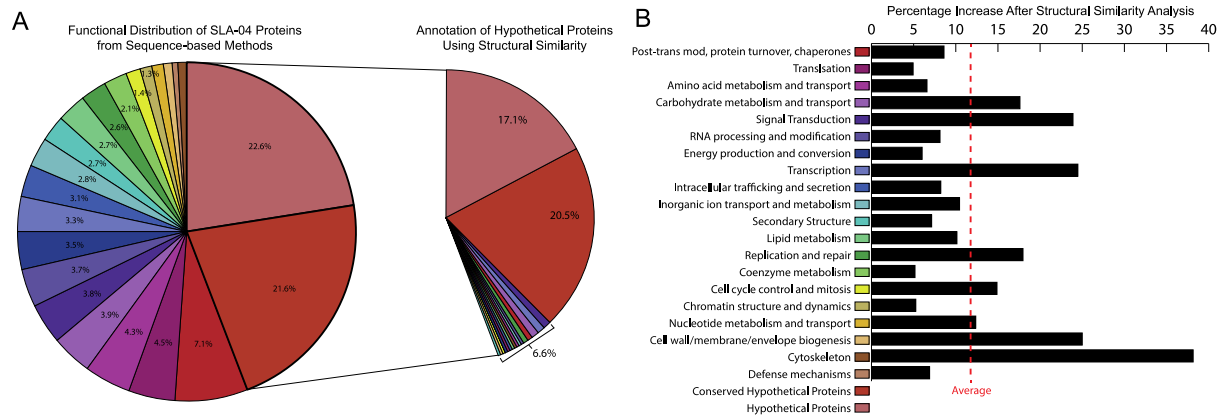


Figure 2. Annotation using protein structural similarity improves functional understanding of SLA-04 proteome. A) 10,310 protein sequences from SLA-04 were submitted to EggNog mapper for annotation using sequence-based methods resulting in 5,757 sequences (55.8%) with predicted function. Remaining sequences (2,226 conserved hypothetical and 2,327 hypothetical proteins) were submitted to Colabfold structural prediction and resulting structures were compared to the AlphaFold and Uniprot structural databases using Foldseek to identify structural homologs of hypothetical SLA-04 proteins. Sequences of structural homologs were submitted to EggNog Mapper resulting in the annotation of an additional 6.6% of the SLA-04 proteome. B) The percent increase in annotation after structural similarity analysis was calculated by dividing the number of proteins annotated using structural prediction methods by the number of original annotated proteins from sequence-based methods. Proteins were grouped based on their Clusters of Orthologous Genes (COG) functionality.

3.1.4 Annotation using Protein 3D Structural Homology. The 4,553 (44.2%) CDSs that could not be assigned putative functions using sequence-based annotation, consisted of “hypothetical” and “conserved hypothetical” proteins with unknown functions. Hypothetical proteins are CDSs with no predicted function and no homology to sequences in the EggNOG database, whereas conserved hypothetical proteins are annotated CDSs that have homology to sequences within the EggNOG database, but the function of these homologous sequences remains unknown. To further improve annotation of the SLA-04 genome, we used the ColabFold-

FoldSeek-EggNOG (CoFFE) pipeline to predict the 3D structures of hypothetical and conserved hypothetical proteins to query the protein databank for structural homologs (Ruperti et al., 2022). Consistent with functional importance, 3D structures of proteins are more conserved during evolution than primary amino-acid sequences (Illergård et al., 2009). Thus, we hypothesized that the integration of structure prediction into genome annotation would improve our functional understanding of the SLA-04 genome. In fact, structure-based annotation resulted in high-confidence functional annotations for 677 of the CDSs that were previously assigned to the hypothetical and conserved hypothetical categories using EggNOG mapper (Cantalapiedra et al., 2021) (**Fig. 2A**). Structure-based annotation identifies proteins primarily associated with formation of the cytoskeleton and cell wall, transcription, and signal transduction (**Fig. 2B**).

3.1.5 Phylogenetic Analysis of *Chlorella* sp. SLA-04.Initial taxonomic classification of *Chlorella* sp. SLA-04 was based on 5.8S and internal transcribed spacer (ITS) regions of the ribosomal small subunit (Vadlamani et al., 2017). ITS regions contain highly variable sequences and can be used to distinguish genera; however, phylogenetic analysis performed using a single gene can limit species level classification and lack statistical support for certain phylogenetic nodes (Álvarez & Wendel, 2003; Delsuc et al., 2005; Evans et al., 2007). Phylogenomic analysis performed using orthologous proteins or 18S rDNA sequences, reveal that SLA-04 shares a closest common ancestor with *Chlorella vulgaris* UTEX 395 (**Fig. 3, Sup. Fig. 3**). Further, the *Chlorella* genus (n=8) forms a paraphyletic clade interspersed with *Micractinium conductrix* SAG 241.80 (**Fig. 3**). These data suggest that the current taxonomy of Trebouxiophyceae may not capture the true phylogenetic history of these algae and we anticipate that the addition of more complete algal genome sequences will provide new taxonomic perspectives.

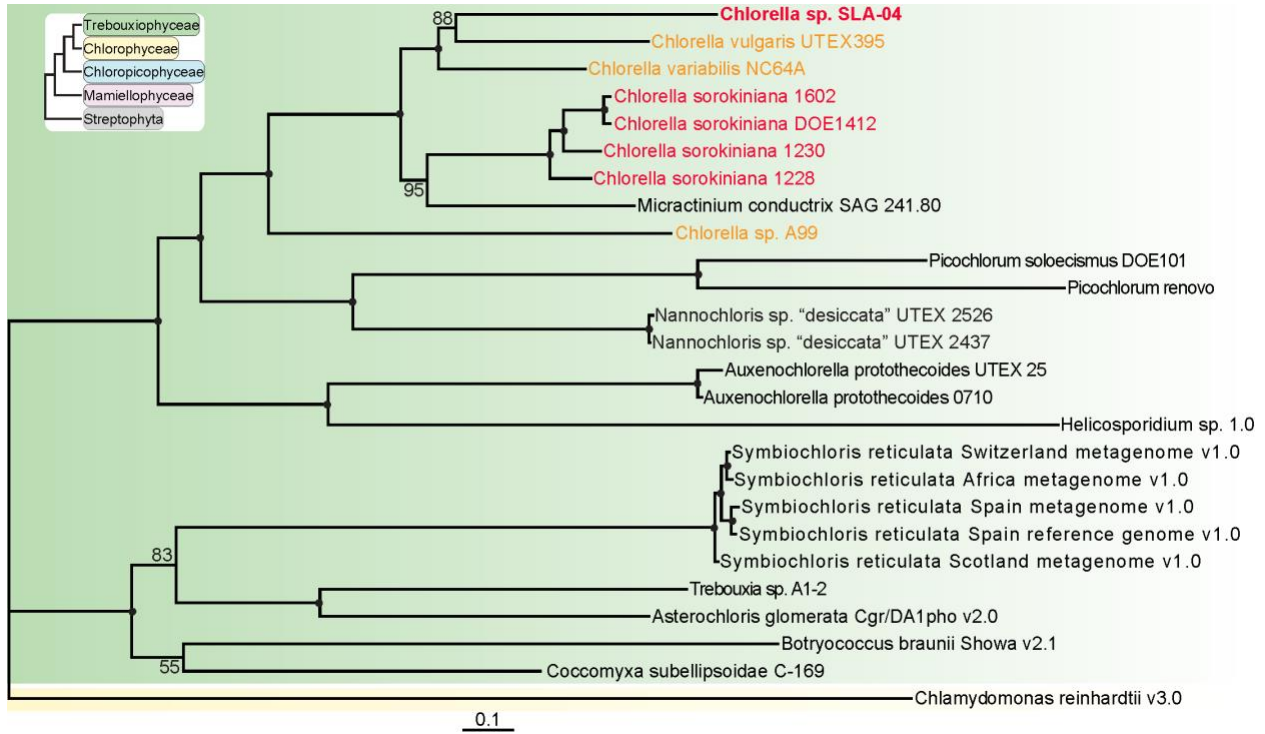


Figure 3. Phylogenomic analysis of green algae performed using orthologous genes indicates that SLA-04 is distinct from other *Chlorella sorokiniana* species. Maximum likelihood phylogenetic analysis based on a concatenation of 215 orthologous proteins sequences of available Trebouxiophyceae genomes from NCBI and PhycoCosm (JGI) databases. *Chlamydomonas reinhardtii* was used as an outgroup. Tree scale indicates the number of amino acid substitutions per site between samples. Bootstrap values of 100 are represented by black dots on nodes. Bootstrap values less than 100 are provided. Classes of algae are denoted via color on tree branches. Red text denotes *Chlorella sorokiniana* species. Orange text represents remaining algae classified within the *Chlorella* genus.

3.2 Genomic Analysis of Metabolic Pathways in *Chlorella* sp. SLA-04

To evaluate the metabolic capability encoded within the *Chlorella* sp. SLA-04 genome, we submitted the complete chromosome level genome to the Kegg Automatic Annotation Server (KAAS) and KofamKOALA for metabolic mapping (Aramaki et al., 2020; Moriya et al., 2007). These programs annotate and map proteins to established metabolic pathways. According to this analysis, the SLA-04 genome contains all genes necessary for core metabolic pathways including photosynthesis, fatty acid synthesis, carbohydrate synthesis, and the citric acid cycle.

Genes involved in lipid synthesis are critical for biofuel production and stress response to environmental changes (De Bigault Du Granrut & Cacas, 2016; Li et al., 2017; Resemann et al., 2021). The SLA-04 genome encodes ATP citrate lyase EC 2.3.3.8 (ACLY), involved in replenishing cytosolic acetyl-CoA pools from citrate (**Fig. 4**). This enzyme is present in some, but not all green algal genomes (**Sup. Fig. 4**). Cytosolic acetyl-CoA pools are necessary for fatty acid synthesis within the endoplasmic reticulum. The presence and transcription of ATP citrate lyase ACLY within the SLA-04 genome suggests that this organism can utilize TCA cycle intermediates for fatty acid synthesis in addition to using photosynthate directly (**Sup. Table 3**).

Metabolic network analysis of SLA-04 genes involved in lipid synthesis also reveals the presence and transcription of genes coding for the proteins necessary for sphingolipid biosynthesis (**Fig. 4, Sup. Table 3**). Similar to ACLY, enzymes involved in sphingolipid synthesis are only present in a subset of green algal genomes (**Sup. Fig. 4**). Phytoceramide and phytosphingosine are two sphingolipids synthesized within the endoplasmic reticulum (ER) of land plants and algae that remodel the cell membranes to maintain fluidity in response to temperature and osmotic stress (De Bigault Du Granrut & Cacas, 2016; Kong et al., 2015; Li et al., 2017). The ability to synthesize phytoceramide and phytosphingosine suggest that SLA-04 maintains the ability to synthesize and remodel cell membrane lipids in response to changes in temperature and osmotic stress.

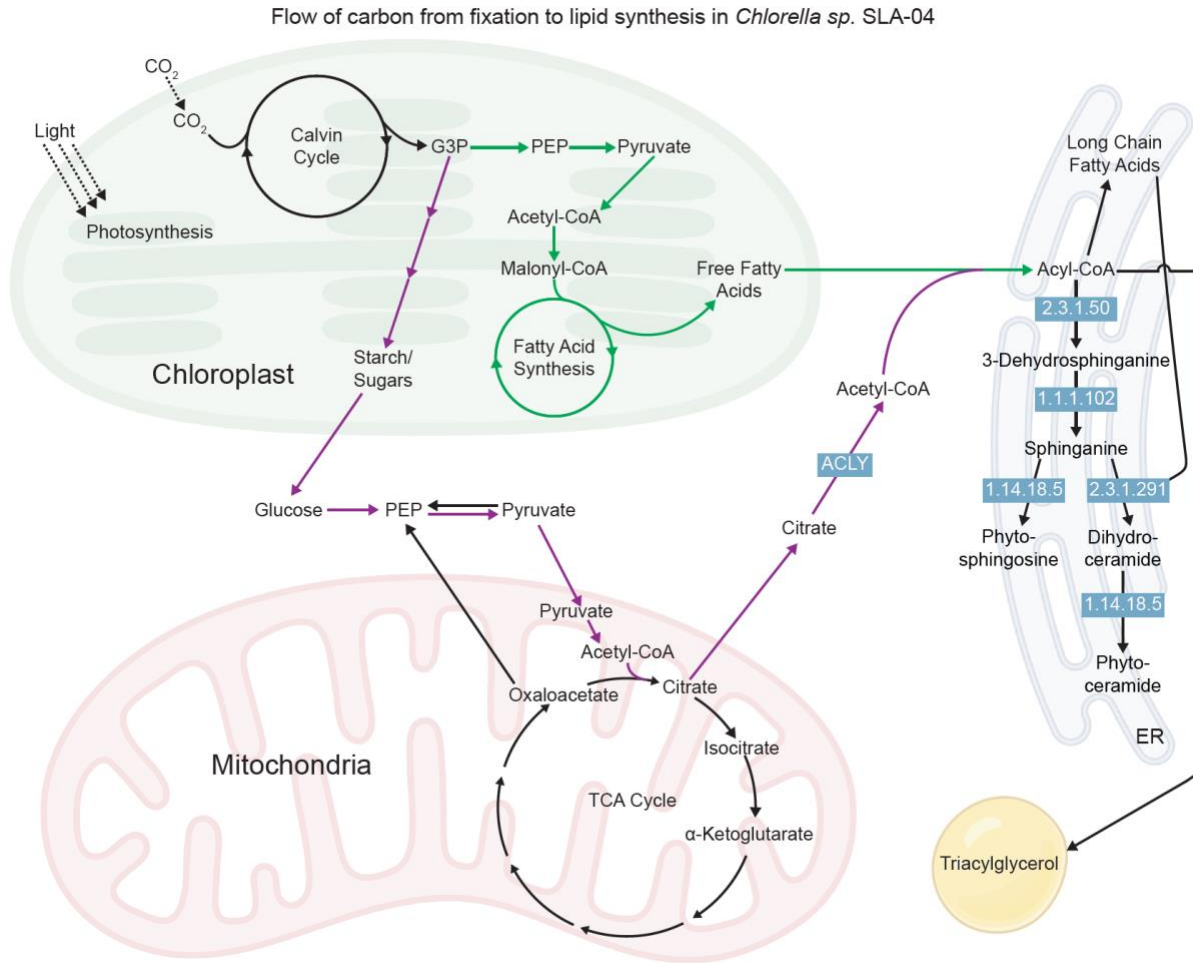


Figure 4. Flow of carbon from fixation to lipid synthesis in *Chlorella* sp. SLA-04. Schematic of carbon flow from fatty acid biosynthesis and TCA cycle to triacylglycerol and sphingolipid biosynthesis in algae. Green arrows represent the flow of carbon from photosynthesis directly to fatty acids within the chloroplast. Purple arrows represent the flow of carbon from photosynthesis to fatty acids through starch synthesis and the TCA cycle. Black arrows represent additional pathways involved in central carbon metabolism. Blue boxes show enzymes of interest. Enzymes are labeled using KEGG and International Union of Biochemistry and Molecular Biology (IUBMB) nomenclature.

Discussion

Algae thrive in diverse aquatic ecosystems and identifying the genotypes responsible for growth across diverse environments requires a comprehensive understanding of genomic and metabolic variation across habitats. Using ultra-long-read Nanopore sequencing, we determined a

complete chromosome level assembly of the *Chlorella* sp. SLA-04 genome (**Fig. 1**). Annotation of the SLA-04 genome using sequence-based methods resulted in an incomplete representation of SLA-04 metabolism with just over half of the proteins having assigned functions. While traditional sequence-based annotation methods identified core metabolic pathways within SLA-04, a large portion of the proteome remained hypothetical with unknown function (**Fig. 2A**).

The structure of a protein dictates its function (Kendrew et al., 1958). By predicting the structure of hypothetical proteins and comparing them to a structural database we were able to predict the function of proteins that were unrecognizable at the sequence level. While structural annotation identified additional proteins within all functional categories, some categories such as cytoskeleton and cell wall proteins were overrepresented in this analysis (**Fig. 2**). The overrepresentation of cytoskeleton and cell wall proteins could be a consequence of functional diversification or an underrepresentation of cytoskeleton and cell wall protein sequences within databases used for sequence-based annotation (NCBI or EggNOG). Regardless, we found that using structure-based annotation methods aid in generating a more complete representation of algal metabolism. Interestingly, 37.6% of SLA-04 proteins continue to escape functional predictions following structural annotation, and 20.5% of these sequences are conservation within other algae and plants. Evolutionary conservation and retention of proteins across algae and plants suggests these proteins are functionally important, and this analysis helps focus future efforts on conserved aspect of the proteome that remain functionally uncharacterized.

While annotating the SLA-04 proteome, we identified lipid synthesis pathways that may allow SLA-04 to respond to changing environmental conditions. ATP citrate lyase (ACLY) generates cytosolic acetyl-CoA from citrate and is a critical enzyme for fatty acid (FA)

biosynthesis and modification (Bauer et al., 2005; Hynes & Murray, 2010; Liang & Jiang, 2016). Algal fatty acids are initially synthesized in the chloroplast and exported into the cytosol where they are modified into triacylglycerols (TAG) or sphingolipids within the ER (Avidan & Pick, 2015; Bellou & Aggelis, 2013; De Bigault Du Granrut & Cacas, 2016; Fan et al., 2015; Goncalves et al., 2016; Tan & Lee, 2016) (**Fig. 4**). Modification and synthesis of additional fatty acids within the cytosol and ER requires cytosolic acetyl-CoA. The presence and transcription of genes encoding *ACLY* suggests that SLA-04 can utilize starch-derived TCA intermediates as a precursor for lipid synthesis. The ability to interconvert between storage molecules enables the metabolic flexibility to synthesize desired biomolecules regardless of the molecular resources available within a cell. Since starches and lipids are important precursors for bioproducts like biodiesel and ethanol, this work demonstrates how genome sequences may improve our ability to link bioproduct production to specific organisms and environmental conditions.

Acetyl-CoA has many cytosolic functions, but we hypothesize that SLA-04 may generate additional cytosolic acetyl-CoA in part to synthesize long term storage lipids and sphingolipids. Sphingolipids are membrane lipids synthesized within the ER that contain long chain FAs. Sphingolipids modify the algal cell membrane to maintain membrane fluidity in response to temperature and osmotic stress (De Bigault Du Granrut & Cacas, 2016; Li et al., 2017; Resemann et al., 2021). The synthesis of sphingolipids such as phytoceramide and phytosphingosine requires cytosolic acetyl-CoA to modify and elongate free FAs that originated from the chloroplast (Aid, 2020) (**Fig. 4**). Similar to *ACLY*, genes encoding enzymes responsible for phytoceramide and phytosphingosine biosynthesis are present and being transcribed in the SLA-04 genome (**Fig. 4, Sup. Table 3**). Together, the capacity for sphingolipid biosynthesis and cytosolic acetyl-CoA

generation by ACLY suggests that SLA-04 may have the metabolic capacity to modulate cell membrane fluidity to rapidly adapt to diurnal and seasonal temperature and osmotic changes (e.g., alkalinity and/or salinity) within the environment (Cael et al., 2017; Eakins, 2010).

Conclusion

Chlorella sp. SLA-04 grows well in a wide range of media conditions such as (i) high (>100 mM) (Vadlamani et al., 2019) and low (<5 mM) (Vadlamani et al., 2017) alkalinities, (ii) circumneutral pH (Hanifzadeh et al., 2018) and high (>10.2) pH (Vadlamani et al., 2019), and (iii) in media containing 0 to 35,000 ppm salt (Hanifzadeh, 2018). The genome analysis presented here, together with prior experimental results indicate that SLA-04 has robust and adaptable growth characteristics, with potential for use in diverse engineered growth systems for biofuel and bioproduct applications. More broadly, this comprehensive view of metabolism within SLA-04 has demonstrated how complete algal genomes can be leveraged to discover new genotypes and identify algal strains most suitable for production of desired bioproducts.

Keywords:

Chlorella sp., Ultra-long-read sequencing, Complete genome, Protein 3D structural annotation, ATP citrate lyase, Sphingolipids

Abbreviations:

NCBI, National Center for Biotechnology Information; KEGG, Kyoto Encyclopedia for Genes and Genomes; KAAS, KEGG Automatic Annotation Server; ITS, Internal transcribed spacer;

TCA, Citric Acid Cycle; ACLY, ATP citrate lyase; ER, Endoplasmic reticulum; TAG, Triacylglycerol; FA, Fatty acids

Data Access:

The raw and processed genomic and transcriptomic data generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under BioProject accession number PRJNA766673 and BioSample accession number SAMN21857598.

Competing Interest Statement:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

B.W. is the founder of SurGene and VIRIS Detection Systems and is an inventor on patent applications related to CRISPR-Cas systems and applications thereof

Acknowledgments:

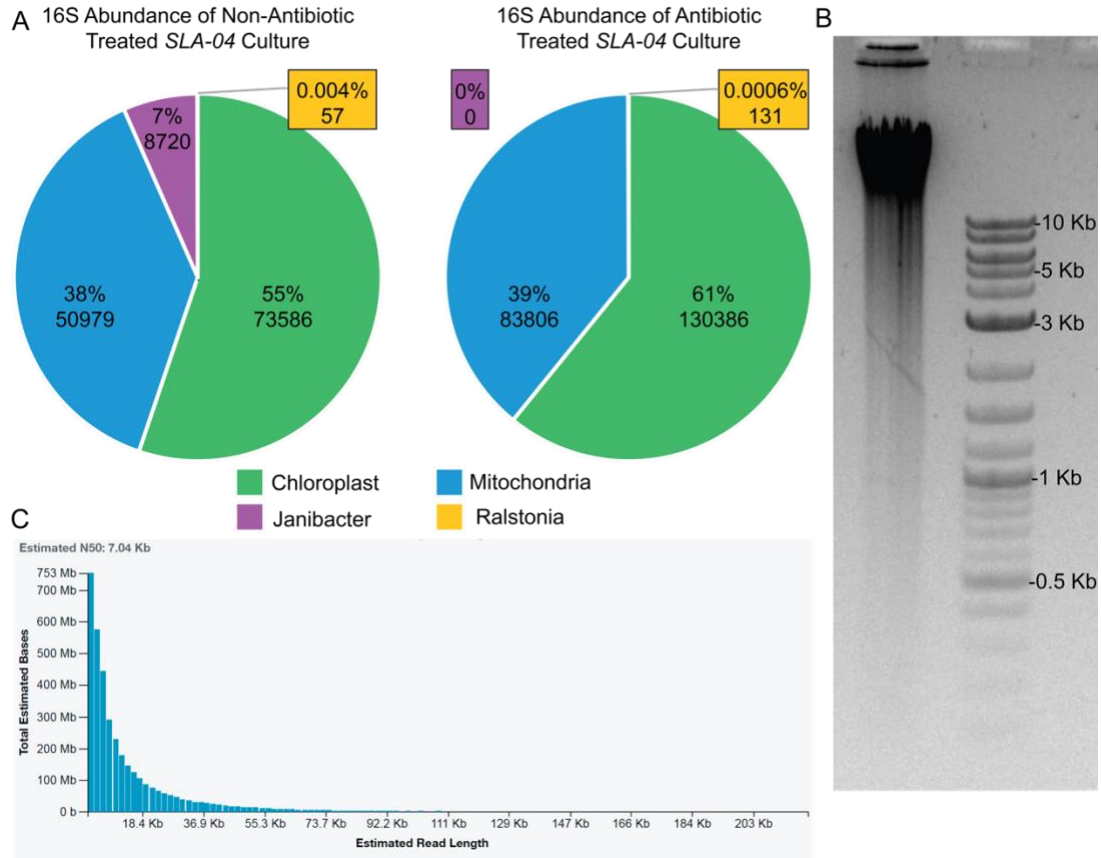
This work was supported by the Bioenergy Technologies Office (BETO) of the U.S. Department of Energy (DOE) for the Productivity Enhanced Algae and ToolKits (PEAK) program [grant number #DE-EE0008247].

Author Contributions:

Conceptualization, B.W., C.G., R.C., R.G., S.V.; Methodology, B.W., C.G., R.W.; Investigation & Data Collection, C.G., H.G., R.W.; Genomics & Bioinformatic Analysis, C.G., H.B., H.G.,

W.H.; Writing – Original Draft, B.W., C.G; Writing – Review & Editing, B.W., C.G., H.B.,
H.G., R.C., R.G, R.W., S.V., W.H.

Supplementary Data

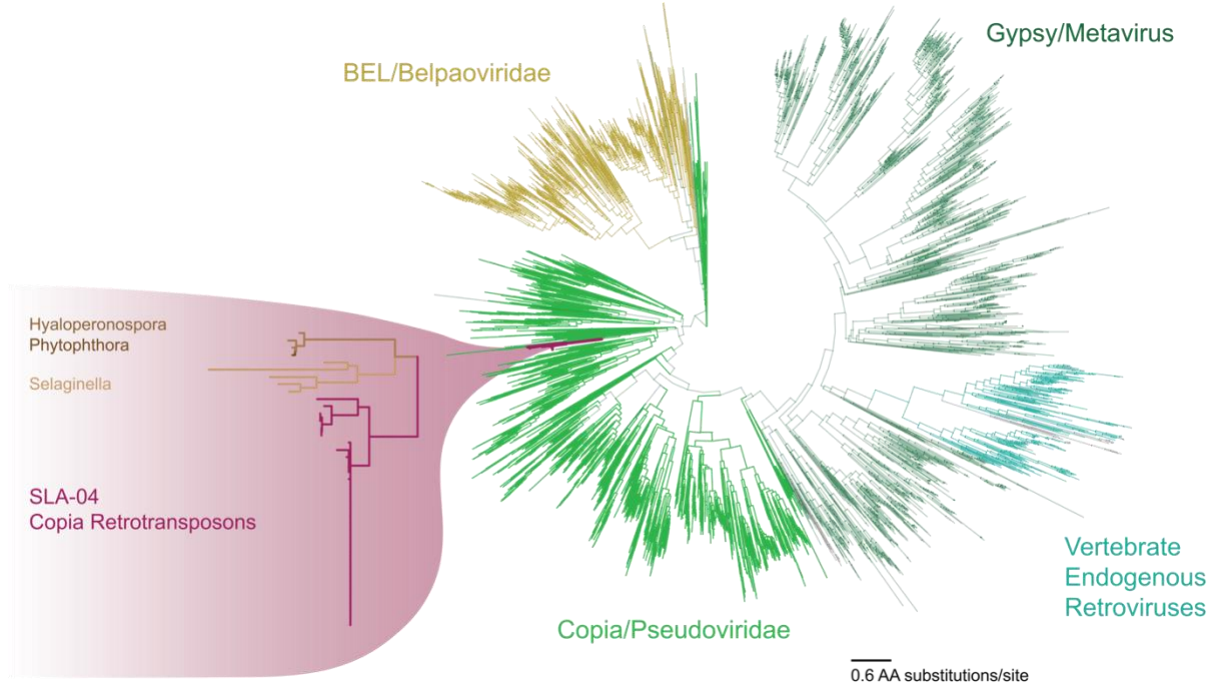


Supplementary Figure 1: DNA isolation and sequencing parameters for *Chlorella* sp. SLA-04. (A) An axenic culture of SLA-04 was confirmed by sequencing the V4 region of the 16S rRNA gene on an Illumina MiSeq platform [21]. Pie charts indicate raw read numbers from an untreated (left) and antibiotic treated culture (right). The antibiotic-treated culture was used for DNA extraction and genomic analysis. *Ralstonia* sequences are common contaminants of DNA purification columns [22]. (B) Gel of high molecular weight DNA from *Chlorella* sp. SLA-04 using liquid N₂ grinding and gravity column purification. (C) Nanopore output of sequence lengths from SLA-04 sequencing run.

Assembly and annotation statistics of *Chlorella* sp. SLA-04

SLA-04 Genome Assembly:	
Assembled genome size	38.8 Mb
Chromosomes	13
Average contig size	1,436,022 bp
N50	2.20 Mb
N90	762,653 bp
GC content	65.87%
Coverage	117 X
SLA-04 Genome Annotation:	
Annotated protein coding genes	10,310
Annotated tRNA genes	100
Average gene length	3,475 bp
Average CDS length	1,745 bp
% of genome assembly as genes	91.07%
% of genome assembly as CDS	45.73%
Busco: Complete genes	93.4%
Busco: Fragmented genes	2.4%
Busco: Missing genes	4.2%
Genes with KEGG function	6,434 (62.4%)
Cellular processing genes	770
Environmental info. processing genes	307
Genetic info. processing genes	1976
Metabolism genes	2427
Unannotated genes	3,876 (37.6%)

Supplementary Table 1: Genome statistics of *Chlorella* sp. SLA-04 genome.

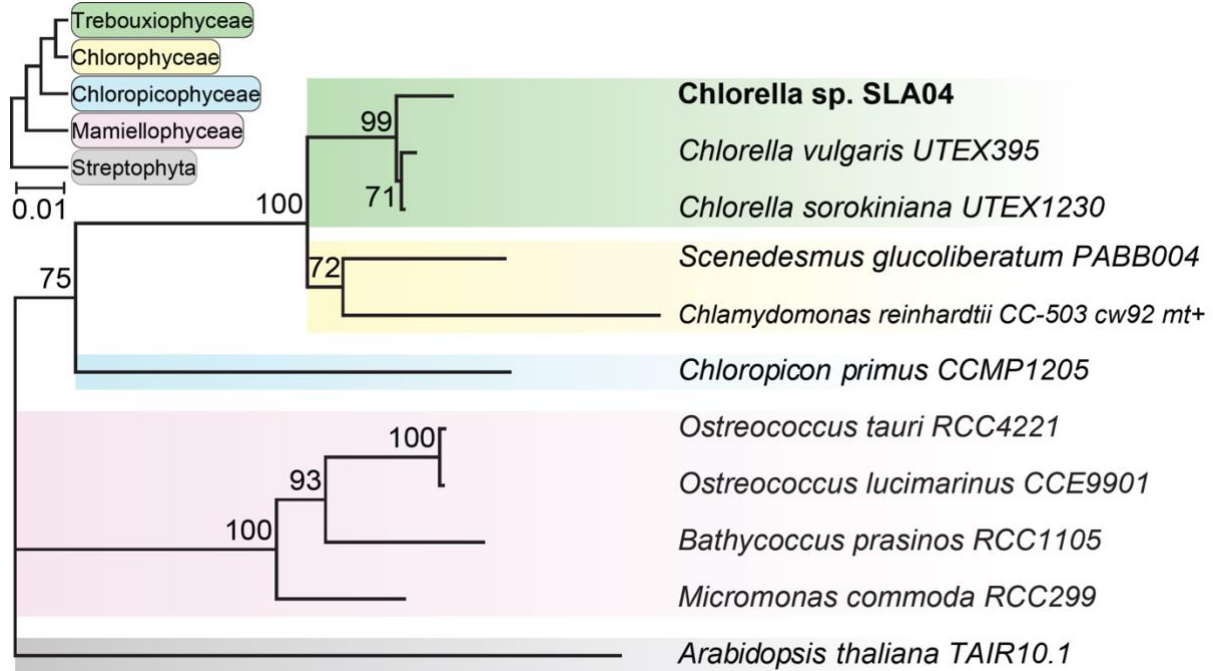


Supplementary Figure 2: Phylogenetic tree of LTR retrotransposon capsid domains places SLA-04 retrotransposons among Copia/Pseudoviridae LTR retrotransposons. Hidden Markov Models for LTR retroelement capsid domains were downloaded from Pfam. 752,604 ORFs from NCBI Virus and Repbase were queried to identify LTR retroelements capsid genes in the SLA-04 genome. Redundant sequences were removed with CD-Hit. Sequences were aligned in MAFFT and trimmed using Trimal and manual curation. Final alignment of 75 AA core capsid domain residues from 5520 sequences were used to build a maximum likelihood phylogenetic tree in Fasttree2. Tree was annotated in ggtree.

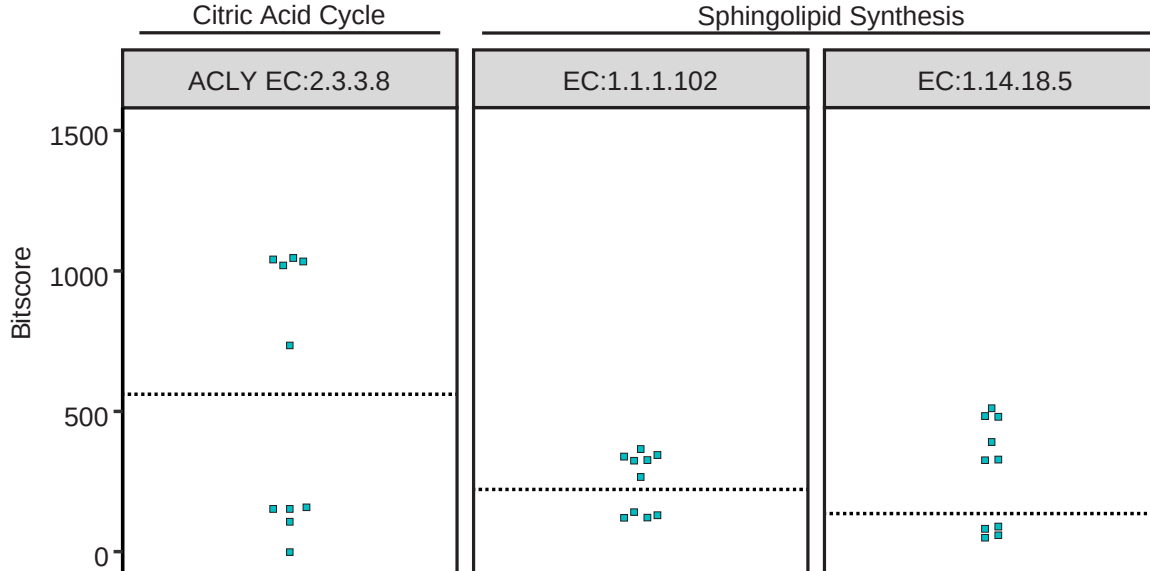
total length: 38807484 bp (38807484 bp excl N/X-runs)
 GC level: 65.87 %
 bases masked: 403707 bp (1.04 %)

	number of elements*	length occupied	percentage of sequence
=====			
Retroelements	2312	302824 bp	0.78 %
SINES:	21	1111 bp	0.00 %
Penelope	14	747 bp	0.00 %
LINES:	1347	133696 bp	0.34 %
CRE/SLACS	1	77 bp	0.00 %
L2/CR1/Rex	30	1469 bp	0.00 %
R1/LOA/Jockey	336	25654 bp	0.07 %
R2/R4/NeSL	7	361 bp	0.00 %
RTE/Bov-B	68	5553 bp	0.01 %
L1/CIN4	71	22474 bp	0.06 %
LTR elements:	944	168017 bp	0.43 %
BEL/Pao	31	3818 bp	0.01 %
Ty1/Copia	378	128222 bp	0.33 %
Gypsy/DIRS1	438	29928 bp	0.08 %
Retroviral	82	4306 bp	0.01 %
DNA transposons	772	52810 bp	0.14 %
hobo-Activator	124	8667 bp	0.02 %
Tc1-IS630-Pogo	99	6362 bp	0.02 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	2	62 bp	0.00 %
Tourist/Harbinger	41	3105 bp	0.01 %
Other (Mirage, P-element, Transib)	3	439 bp	0.00 %
Rolling-circles	116	7169 bp	0.02 %
Unclassified:	12	4049 bp	0.01 %
Total interspersed repeats:		359683 bp	0.93 %
Small RNA:	108	39406 bp	0.10 %
Satellites:	67	7508 bp	0.02 %
Simple repeats:	0	0 bp	0.00 %
Low complexity:	0	0 bp	0.00 %
=====			

Supplementary Table 2: Summary of repetitive elements identified by RepeatMasker using Dfam and Repbase as a combined reference database.



Supplementary Figure 3: Phylogenomic analysis of green algae performed using 18S rDNA. 18S rDNA gene sequences of *Chlorella* sp. SLA-04 and complete algae genomes from NCBI were used to make a phylogenetic tree using maximum likelihood. Branch lengths represent the number of nucleotide substitutions per site. *Arabidopsis thaliana* was used as an outgroup. Bootstrap values are labeled next to each node. Classes of algae are denoted via color on tree branches.



Supplementary Figure 4: Presence of ATP citrate lyase and sphingolipid enzymes across algal genomes. Algal protein coding sequences from all chromosome level algal genomes (n=10) within the NCBI database were submitted to KofamKOALA to predict function. KofamKOALA assigned KEGG Orthologs (KOs) to protein sequences by homology search against a database of profile hidden Markov models (KOfam) with pre-computed adaptive score thresholds. Each datapoint represents the likelihood of a functional protein being present within a single algal genome. Genes are considered functional if the bitscore is above the adaptive threshold (dotted line). Genes of interest involved in fatty acid and sphingolipid synthesis are reported.

Gene ID	Gene Name	KEGG Ortholog	Coverage	FPKM	TPM
SLA-04_00282-RA	ATP citrate (pro-S)-lyase [EC:2.3.3.8]	K01648	97.7	136.7	171.5
SLA-04_07917-RA	ATP citrate (pro-S)-lyase [EC:2.3.3.8]	K01648	82.4	115.3	144.6
SLA-04_04864-RA	serine palmitoyltransferase [EC:2.3.1.50]	K00654	11.2	15.6	19.6
SLA-04_04303-RA	serine palmitoyltransferase [EC:2.3.1.50]	K00654	22.9	32.0	40.2
SLA-04_04491-RA	3-dehydrosphinganine reductase [EC:1.1.1.102]	K04708	16.4	22.9	28.7
SLA-04_04653-RA	sphingoid base N-palmitoyltransferase [EC:2.3.1.291]	K23727	8.3	11.6	14.5
SLA-04_00395-RA	sphingoid base N-palmitoyltransferase [EC:2.3.1.291]	K23727	38.2	53.5	67.1
SLA-04_04461-RA	sphingolipid C4-monoxygenase [EC:1.14.18.5]	K04713	32.9	46.0	57.7

Supplementary Table 3: Active Transcription of lipid synthesis genes within the *Chlorella* sp. SLA-04 genome. mRNA transcript counts were determined by mapping the mRNA reads to the SLA-04 genome using Integrated Genome Viewer (IGV). Coverage represents the total number of mRNA reads mapped back to each gene. Fragments Per Kilobase Million (FPKM) and Transcripts Per Million (TPM) adjusted coverage values are also reported for each gene.

References Cited

- Aid, F. (2020). Plant Lipid Metabolism. In. IntechOpen.
<https://doi.org/10.5772/intechopen.81355>
- Álvarez, I., & Wendel, J. F. (2003, 2003/12/01/). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, 29(3), 417-434.
[https://doi.org/https://doi.org/10.1016/S1055-7903\(03\)00208-2](https://doi.org/https://doi.org/10.1016/S1055-7903(03)00208-2)
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1). <https://doi.org/10.1186/s13059-020-1935-5>
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7), 2251-2252.
<https://doi.org/10.1093/bioinformatics/btz859>
- Arriola, M. B., Velmurugan, N., Zhang, Y., Plunkett, M. H., Hondzo, H., & Barney, B. M. (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. <https://doi.org/10.1111/tpj.13789>
- Avidan, O., & Pick, U. (2015). Acetyl-CoA synthetase is activated as part of the PDH-bypass in the oleaginous green alga *Chlorella desiccata*. *Journal of Experimental Botany*, 66(22), 7287-7298. <https://doi.org/10.1093/jxb/erv424>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1). <https://doi.org/10.1186/s13100-015-0041-9>
- Bauer, D. E., Hatzivassiliou, G., Zhao, F., Andreadis, C., & Thompson, C. B. (2005). ATP citrate lyase is an important component of cell growth and transformation. *Oncogene*, 24(41), 6314-6322. <https://doi.org/10.1038/sj.onc.1208773>
- Behrenfeld, M. J. (2001). Biospheric Primary Production During an ENSO Transition. *Science*, 291(5513), 2594-2597. <https://doi.org/10.1126/science.1055071>
- Bellou, S., & Aggelis, G. (2013). Biochemical activities in *Chlorella* sp. and *Nannochloropsis salina* during lipid and sugar synthesis in a lab-scale open pond simulating reactor. *Journal of Biotechnology*, 164(2), 318-329. <https://doi.org/10.1016/j.jbiotec.2013.01.010>
- Blanc-Mathieu, R., Verhelst, B., Derelle, E., Rombauts, S., Bouget, F.-Y., Carré, I., Château, A., Eyre-Walker, A., Grimsley, N., Moreau, H., Piégu, B., Rivals, E., Schackwitz, W., Van

- De Peer, Y., & Piganeau, G. (2014). An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics*, 15(1), 1103. <https://doi.org/10.1186/1471-2164-15-1103>
- Bold, H. C. (1949). The Morphology of *Chlamydomonas chlamydogama*, Sp. Nov. *Bulletin of the Torrey Botanical Club*, 76(2), 101. <https://doi.org/10.2307/2482218>
- Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth's lakes. *Geophysical Research Letters*, 44(1), 209-218. <https://doi.org/10.1002/2016gl071378>
- Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48(1). <https://doi.org/10.1002/0471250953.bi0411s48>
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., & Gutell, R. R. (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3(1), 2. <https://doi.org/10.1186/1471-2105-3-2>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825-5829. <https://doi.org/10.1093/molbev/msab293>
- Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., & Yandell, M. (2007). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18(1), 188-196. <https://doi.org/10.1101/gr.6743907>
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In (pp. 207-232). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35306-5_10
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- De Bigault Du Granrut, A., & Cacas, J.-L. (2016). How Very-Long-Chain Fatty Acids Could Signal Stressful Conditions in Plants? *Frontiers in Plant Science*, 7, 1490-1490. <https://doi.org/10.3389/fpls.2016.01490>

- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518-524. <https://doi.org/10.1038/nbt.3423>
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361-375. <https://doi.org/10.1038/nrg1603>
- Dohm, J. C., Peters, P., Stralis-Pavese, N., & Himmelbauer, H. (2020). Benchmarking of long-read correction methods. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/nargab/lqaa037>
- Drost, H.-G., Gabel, A., Grosse, I., & Quint, M. (2015). Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Molecular Biology and Evolution*, 32(5), 1221-1231. <https://doi.org/10.1093/molbev/msv012>
- Eakins, B. W., G.F. Sharman. (2010). *Volumes of the World's Oceans from ETOPO1* NOAA National Geophysical Data Center, Boulder, CO.
- Eddy, S. R., & Nawrocki, E. P. (2009). Structural rna homology search and alignment using covariance models.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Erik, L., Hirsh, L., Paladin, L., Piovesan, D., Silvio, C., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427-D432. <https://doi.org/10.1093/nar/gky995>
- Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9(1), 18. <https://doi.org/10.1186/1471-2105-9-18>
- Evans, J. H. (1958). The Survival of Freshwater Algae During Dry Periods: Part I. An Investigation of the Algae of Five Small Ponds. *The Journal of Ecology*, 46(1), 149. <https://doi.org/10.2307/2256910>
- Evans, K. M., Wortley, A. H., & Mann, D. G. (2007, 2007/07/18/). An Assessment of Potential Diatom “Barcode” Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). *Protist*, 158(3), 349-364. <https://doi.org/https://doi.org/10.1016/j.protis.2007.04.001>
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., & Hassabis, D. (2021). *Protein complex prediction with AlphaFold-Multimer*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2021.10.04.463034>

- Fan, J., Ning, K., Zeng, X., Luo, Y., Wang, D., Hu, J., Li, J., Xu, H., Huang, J., Wan, M., Wang, W., Zhang, D., Shen, G., Run, C., Liao, J., Fang, L., Huang, S., Jing, X., Su, X., Wang, A., Bai, L., Hu, Z. M., Xu, J., & Li, Y. (2015). Genomic Foundation of Starch to Lipid Switch in Oleaginous *Chlorella*. *Plant Physiology*, pp.01174.02015.
<https://doi.org/10.1104/pp.15.01174>
- Field, C. B. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374), 237-240. <https://doi.org/10.1126/science.281.5374.237>
- Fulnečková, J., Hasíková, T., Fajkus, J., Lukešová, A., Eliáš, M., & Sýkorová, E. (2012). Dynamic Evolution of Telomeric Sequences in the Green Algal Order Chlamydomonadales. *Genome Biology and Evolution*, 4(3), 248-264.
<https://doi.org/10.1093/gbe/evs007>
- Goemann, H. M., Gay, J. D., Mueller, R. C., Brookshire, E. N. J., Miller, P., Poulter, B., & Peyton, B. M. (2021, Dec 2021 2021-11-08). Aboveground and belowground responses to cyanobacterial biofertilizer supplement in a semi-arid, perennial bioenergy cropping system. *Global Change Biology. Bioenergy*, 13(12), 1908-1923. <https://doi.org/https://doi.org/10.1111/gcbb.12892>
- Goncalves, E. C., Wilkie, A. C., Kirst, M., & Rathinasabapathi, B. (2016). Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnology Journal*, 14(8), 1649-1660.
<https://doi.org/10.1111/pbi.12523>
- Hadley, W. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.
- Hanifzadeh, M., Garcia, E. C., & Viamajala, S. (2018). Production of lipid and carbohydrate from microalgae without compromising biomass productivities: Role of Ca and Mg. *Renewable Energy*, 127, 989-997. <https://doi.org/10.1016/j.renene.2018.05.012>
- Hanifzadeh, M.-M. (2018). *Approaches for Sustainable Production of Microalgae with High Productivity and Flexible Composition* The University of Toledo].
- Hanschen, E. R., Hovde, B. T., & Starkenburg, S. R. (2020, 2020/10/01/). An evaluation of methodology to determine algal genome completeness. *Algal Research*, 51, 102019.
<https://doi.org/https://doi.org/10.1016/j.algal.2020.102019>
- Higashiyama, T., Maki, S., & Yamada, T. (1995). Molecular organization of *Chlorella vulgaris* chromosome I: presence of telomeric repeats that are conserved in higher plants. *Molecular and General Genetics MGG*, 246(1), 29-36.
<https://doi.org/10.1007/bf00290130>

- Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12(1), 491. <https://doi.org/10.1186/1471-2105-12-491>
- Hovde, B. T., Hanschen, E. R., Steadman Tyler, C. R., Lo, C.-C., Kunde, Y., Davenport, K., Daligault, H., Msanne, J., Canny, S., Eyun, S.-I., Riethoven, J.-J. M., Polle, J., & Starkenburg, S. R. (2018). Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Research*, 35, 449-461. <https://doi.org/10.1016/j.algal.2018.09.012>
- Hynes, M. J., & Murray, S. L. (2010). ATP-Citrate Lyase Is Required for Production of Cytosolic Acetyl Coenzyme A and Development in *Aspergillus nidulans*. *Eukaryotic Cell*, 9(7), 1039-1048. <https://doi.org/10.1128/ec.00080-10>
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009, Nov 15). Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins*, 77(3), 499-508. <https://doi.org/10.1002/prot.22458>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, 181(4610), 662-666. <https://doi.org/10.1038/181662a0>
- Kingsford, C., Schatz, M. C., & Pop, M. (2010). Assembly complexity of prokaryotic genomes using short reads. *BMC Bioinformatics*, 11(1), 21. <https://doi.org/10.1186/1471-2105-11-21>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540-546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kong, J. N., Hardin, K., Dinkins, M., Wang, G., He, Q., Mujadzic, T., Zhu, G., Bielawski, J., Spassieva, S., & Bieberich, E. (2015). Regulation of *Chlamydomonas* flagella and

- ependymal cell motile cilia by ceramide-mediated translocation of GSK3. *Molecular Biology of the Cell*, 26(24), 4451-4465. <https://doi.org/10.1091/mbc.e15-06-0371>
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., & Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*, 30(7), 693-700. <https://doi.org/10.1038/nbt.2280>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research*, 27(5), 722-736. <https://doi.org/10.1101/gr.215087.116>
- Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, 16(5), 263-276. <https://doi.org/10.1038/nrmicro.2018.9>
- Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14), 2103-2110. <https://doi.org/10.1093/bioinformatics/btw152>
- Li, H. (2018a). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, Y., Lou, Y., Mu, T., Ke, A., Ran, Z., Xu, J., Chen, J., Zhou, C., Yan, X., Xu, Q., & Tan, Y. (2017). Sphingolipids in marine microalgae: Development and application of a mass spectrometric method for global structural characterization of ceramides and glycosphingolipids in three major phyla. *Analytica Chimica Acta*, 986, 82-94. <https://doi.org/10.1016/j.aca.2017.07.039>
- Liang, M.-H., & Jiang, J.-G. (2016, 2016/12/01/). Characterization and nitrogen deficiency response of ATP-citrate lyase from unicellular alga *Dunaliella tertiolecta*. *Algal Research*, 20, 77-86. <https://doi.org/https://doi.org/10.1016/j.algal.2016.09.025>
- Manni, M. e., Matthew, Seppey, M., Felipe, & Evgeny. (2021, 2021-06-22). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *arXiv pre-print server*. <https://doi.org/None>
arxiv:2106.11799
- Mehrotra, S., & Goyal, V. (2014). Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics, proteomics & bioinformatics*, 12(4), 164-171. <https://doi.org/10.1016/j.gpb.2014.07.003>

- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Marechal-Drouard, L., Marshall, W. F., Qu, L. H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C. L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I. T., Pazour, G., Purton, S., Ral, J. P., Riano-Pachon, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C. J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W. C. A., Otilar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S., & Grossman, A. R. (2007). The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science*, 318(5848), 245-250. <https://doi.org/10.1126/science.1143609>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007, Jul). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35(Web Server issue), W182-185. <https://doi.org/10.1093/nar/gkm321>
- Mustapa, M., Sallehudin, N., Mohamed, M., Mohammad Noor, N., & Raus, R. (2016, 01/01). Decontamination of *Chlorella* sp. Culture using antibiotics and antifungal cocktail treatment. *ARPJ Journal of Engineering and Applied Sciences*, 11.
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274. <https://doi.org/10.1093/molbev/msu300>
- Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2018). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, 35(13), 2193-2198. <https://doi.org/10.1093/bioinformatics/bty841>

- Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, 9, 304. <https://doi.org/10.12688/f1000research.23297.2>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290-295. <https://doi.org/10.1038/nbt.3122>
- Resemann, H. C., Herrfurth, C., Feussner, K., Hornung, E., Ostendorf, A. K., Gömann, J., Mittag, J., Van Gessel, N., Vries, J. D., Ludwig-Müller, J., Markham, J., Reski, R., & Feussner, I. (2021). Convergence of sphingolipid desaturation across over 500 million years of plant evolution. *Nature Plants*, 7(2), 219-232. <https://doi.org/10.1038/s41477-020-00844-3>
- Rice, P., Longden, I., & Bleasby, A. (2000, Jun). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16(6), 276-277. [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2)
- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019, Jul 2). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*, 47(W1), W5-w10. <https://doi.org/10.1093/nar/gkz342>
- Ruperti, F., Papadopoulos, N., Musser, J., & Arendt, D. (2022). *Beyond sequence similarity: cross-phyla protein annotation by structural prediction and alignment*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2022.07.05.498892>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- [Record #564 is using a reference type undefined in this output style.]
- Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, 12(1). <https://doi.org/10.1186/s13100-020-00230-y>
- Sutton, J. M., Millwood, J. D., Case McCormack, A., & Fierst, J. L. (2021). Optimizing experimental design for genome sequencing and assembly with Oxford Nanopore Technologies. *Gigabyte*, 2021, 1-26. <https://doi.org/10.46471/gigabyte.27>
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server), W609-W612. <https://doi.org/10.1093/nar/gkl315>

- Tan, K. W. M., & Lee, Y. K. (2016). The dilemma for lipid productivity in green microalgae: importance of substrate provision in improving oil yield without sacrificing growth. *Biotechnology for Biofuels*, 9(1). <https://doi.org/10.1186/s13068-016-0671-2>
- Tarailo-Graovac, M., & Chen, N. (2009, Mar). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*, Chapter 4, Unit 4.10. <https://doi.org/10.1002/0471250953.bi0410s25>
- Vadlamani, A., Pendyala, B., Viamajala, S., & Varanasi, S. (2019). High Productivity Cultivation of Microalgae without Concentrated CO₂ Input. *ACS Sustainable Chemistry & Engineering*, 7(2), 1933-1943. <https://doi.org/10.1021/acssuschemeng.8b04094>
- Vadlamani, A., Viamajala, S., Pendyala, B., & Varanasi, S. (2017). Cultivation of Microalgae at Extreme Alkaline pH Conditions: A Novel Approach for Biofuel Production. *ACS Sustainable Chemistry & Engineering*, 5(8), 7284-7294. <https://doi.org/10.1021/acssuschemeng.7b01534>
- Varshney, P., Mikulic, P., Vonshak, A., Beardall, J., & Wangikar, P. P. (2015, May). Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour Technol*, 184, 363-372. <https://doi.org/10.1016/j.biortech.2014.11.040>
- Walker, J. C. G. (1980). The Oxygen Cycle. In (pp. 87-104). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-24940-6_5
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Weyrich, L. S., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., Handsley-Davis, M., Adler, C. J., Breen, J., & Cooper, A. (2019). Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources*, 19(4), 982-996. <https://doi.org/10.1111/1755-0998.13011>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28-36. <https://doi.org/10.1111/2041-210x.12628>

CHAPTER FOUR

METABOLIC REPROGRAMMING OF CARBON METABOLISM IN CHLORELLA SP.
SLA-04 UNDER MULTIPLE ENVIRONMENTAL STRESSORS

Contribution of Authors and Co-Authors

Author: Calvin L.C. Goemann

Contributions: Conceptualization, Methodology, Investigation & Data Collection, Transcriptome
& Bioinformatic Analysis, Writing – Original Draft

Co-Author: Huyen Bui

Contributions: Methodology, Investigation & Data Collection

Co-Author: Sandra Rincon Miranda

Contributions: Methodology, Investigation & Data Collection

Co-Author: Adrienne Arnold

Contributions: Methodology, Investigation & Data Collection

Co-Author: Ross P. Carlson

Contributions: Conceptualization, Methodology

Co-Author: Sridhar Viamajala

Contributions: Conceptualization

Co-Author: Robin Gerlach

Contributions: Conceptualization, Methodology

Co-Author: Blake Wiedenheft

Contributions: Conceptualization, Methodology, Writing – Original Draft

Manuscript Information

Calvin L. C. Goemann, Huyen Bui, Sandra Rincon Miranda, Adrienne Arnold, Ross P. Carlson,
Sridhar Viamajala, Robin Gerlach, Blake Wiedenheft

Algal Research

Status of Manuscript: [Put an x in one of the options below, delete this]

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

Metabolic reprogramming of carbon metabolism in *Chlorella sp.* SLA-04 under multiple environmental stressors

Calvin L. C. Goemann, Huyen Bui, Sandra Rincon Miranda, Adrienne Arnold, Ross P. Carlson, Sridhar Viamajala, Robin Gerlach, Blake Wiedenheft

Highlights

- In response to nitrogen depletion SLA-04 rapidly produces high levels of carbohydrates within 24 hours followed by a slow accumulation of triacylglycerols for the remainder of batch cultivation.
- SLA-04 appears to cycle carbon between TAG storage and the TCA cycle to consume excess photosynthetic reducing power through a proposed futile cycle.
- SLA-04 appears to maintain high cellular concentrations of carbohydrates, unsaturated lipids, and upregulated ion transporters to maintain intracellular homeostasis when experiencing alkaline-induced osmotic stress.
- Nitrogen limitation triggers global repression of cellular division while alkaline growth triggers upregulation of retroelements and antiviral immune systems.

Abstract

During stress induced by nutrient limitation, algae rewire metabolism and begin to accumulate carbon storage compounds like triacylglycerols (TAGs). TAGs can be converted into biodiesel and are a sustainable alternative to petroleum-based fuels. Intentionally inducing stress in algal cultivation systems is a way to maximize TAG accumulation and biofuel yields. High-alkaline cultivation is emerging as an effective strategy for growing algae because high alkaline stress triggers lipid production while also minimizing culture contamination and eliminating the need for external CO₂ supply. *Chlorella* sp. SLA-04 is one of a handful of algae that can grow in extreme alkaline conditions but it is unclear what genotypes promote survival in extreme alkalinity and how SLA-04 remodels metabolism after stress to produce lipids. Here we grow SLA-04 in neutral and high alkalinity conditions sampling both physiological composition and mRNA expression. This analysis reveals how SLA-04 reallocates carbon to accumulate both carbohydrates and lipids after nitrogen depletion. We also identify a proposed futile cycle through which SLA-04 recycles TAGs to mitigate photooxidative stress. Further, SLA-04 appears to respond to alkaline conditions by synthesizing high levels of intracellular carbohydrates and unsaturated lipids while rewiring the expression of ion transporters and immune system genes. Together, this analysis identifies genotypes that algae may use to survive extreme alkalinity and provides a foundation to optimize biomass production through high alkalinity cultivation and genome editing.

Introduction

Photosynthesis is an essential process in algae, generating most of the reducing power (NADPH) and ATP needed to synthesize biomass (Hohmann-Marriott & Blankenship, 2011; Losada, 1976). This process is tightly regulated due to the high potential of photobleaching and the production of oxidative radicals that can damage the cell (Paul & Foyer, 2001). Algae growing in exponential phase sink photosynthetic electrons through cell division and the consumption of reducing power through central carbon metabolism (Aro & Andersson, 2006). Utilizing photosynthetic energy to power cellular growth and division minimizes the buildup of excess electrons that can result in reactive oxygen species. However, when algae encounter stressful conditions where cell growth is limited, photosynthetic energy production is no longer balanced with cellular division (Sharma et al., 2012). Algal cells in stressful conditions must maintain photosynthetic activity while minimizing the oxidative damage associated with imbalanced photosynthesis.

During stress induced by nutrient limitation, cells must respond through means other than cellular division to mitigate excess photosynthetic reducing power. One mechanism that algae utilize to consume photosynthetic energy during stationary phase is the synthesis of reduced carbon storage compounds like starches, pigments, and triacylglycerols (TAGs). In stressful conditions including nutrient depletion, or extreme temperature, light, salinity, or pH, algae can accumulate up to 80% of their biomass as TAGs to deal with the excess photosynthetic energy burden (Chisti, 2007).

TAGs are industrially relevant molecules that can be converted into biodiesel and supplement petroleum-based fuels. Algal TAG production provides a carbon-neutral fuel source

that can help power a growing human population without negatively affecting global climate (Dragone et al., 2010; Perera & Nadeau, 2022). Stressful conditions have been coopted to maximize the TAG accumulation in industrial algae strains (Alkhamis et al., 2022; Calhoun et al., 2021; Luo & Liu, 2011). Researchers implement a two-stage growth system to first optimize biomass production by growing algae in stress-free conditions and then imposing stress such as nutrient depletion to convert cellular biomass to biomolecules of interest like TAGs (Sharma et al., 2018). Cultivation strategies often use nitrogen depletion as a stress, causing algal cells to enter stationary phase and begin to synthesize carbon storage molecules like TAGs to mitigate excess photosynthetic energy.

While stress-induced growth has improved TAG production, algal biofuel is still too expensive to feasibly compete with fossil fuels (J. Chen, Li J., Dong, W., Zhang, X., et al., 2018; Delrue et al., 2012). Current cultivation strategies are costly in part due to the external delivery of CO₂, and culture contamination (Wilson et al., 2009). Algae are typically grown at circum-neutral pH where natural atmospheric CO₂ dissolution is poor. These conditions rely on external CO₂ delivery, contributing up to 65% of operating expenses (Vadlamani et al., 2019). Further, in outdoor raceway cultivation systems at circum-neutral pH, 30% of cultures are lost due to the growth of unwanted bacteria, predatory protozoa, or viruses (Newby et al., 2016; Rowe et al., 2013). To minimize costs associated with CO₂ and contamination researchers have proposed growing algae in high pH and high alkalinity growth conditions (Vadlamani et al., 2017). High-alkalinity systems not only provide the stress triggers to improve lipid production, but these conditions also eliminate the need for external CO₂ and the growth of unwanted microbial contamination (Gimmler, 2001; Vadlamani et al., 2017).

Chlorella sp. SLA-04 is an alkali-tolerant green alga that grows up to pH 11 while producing high levels of biomass and lipid content (Goemann et al., 2023). *Chlorella* sp. SLA-04 is one of a handful of algae that can grow in extreme alkaline conditions and serves as a model organism to understand adaptations to this environment. While SLA-04 shows desirable phenotypes in alkaline conditions it is unclear what genotypes promote survival in extreme alkalinity and how SLA-04 remodels metabolism after stress to produce lipids and mitigate photosynthetic oxidation. To test this, we grew SLA-04 in neutral and high alkaline conditions sampling both physiological composition and mRNA expression. Transcriptomic data were combined with biomass composition to identify genotypes that promote lipid accumulation and growth in high-alkaline environments. This analysis reveals how SLA-04 reallocates carbon to carbohydrates and lipids after nitrogen depletion. We also identify a proposed futile cycle where SLA-04 recycles TAGs to consume excess photosynthetic reducing power. Further, SLA-04 appears to respond to alkaline conditions by synthesizing high levels of intracellular carbohydrates, and unsaturated lipids while rewiring the expression of ion transporters, mobile genetic elements, and antiviral immunity.

Results

Physiological Data

Cell Growth. *Chlorella* sp. SLA-04 was grown in Bold's Basal Media at pH 7 without added inorganic carbon (neutral condition) and pH 10 with 100 mM inorganic carbon (alkaline condition) (**Fig 1 A, B**). SLA-04 grew faster in neutral conditions producing more cells and biomass. Log phase doubling time in neutral conditions was 10.52 hours and alkaline cultures doubled every 16.45 hours (**Fig 1 E, F**). Neutral cultures consumed all nitrate triggering stationary

phase within 3 days while alkaline cultures took 6 days to reach nitrate depletion (**Fig. 1 C**). While neutral conditions produced more cells and biomass, individual cells in alkaline conditions were larger and weighed on average 1.89 times more than cells grown in neutral conditions. Total chlorophyll concentrations increased significantly faster and to higher levels in neutral conditions reaching a maximum total chlorophyll concentration on day four before reducing to similar concentrations as the alkaline condition on day six (**Fig 1 D**). While total chlorophyll concentrations were higher in the neutral condition, normalized chlorophyll concentration per cell was on average 2.67 times greater in the alkaline.

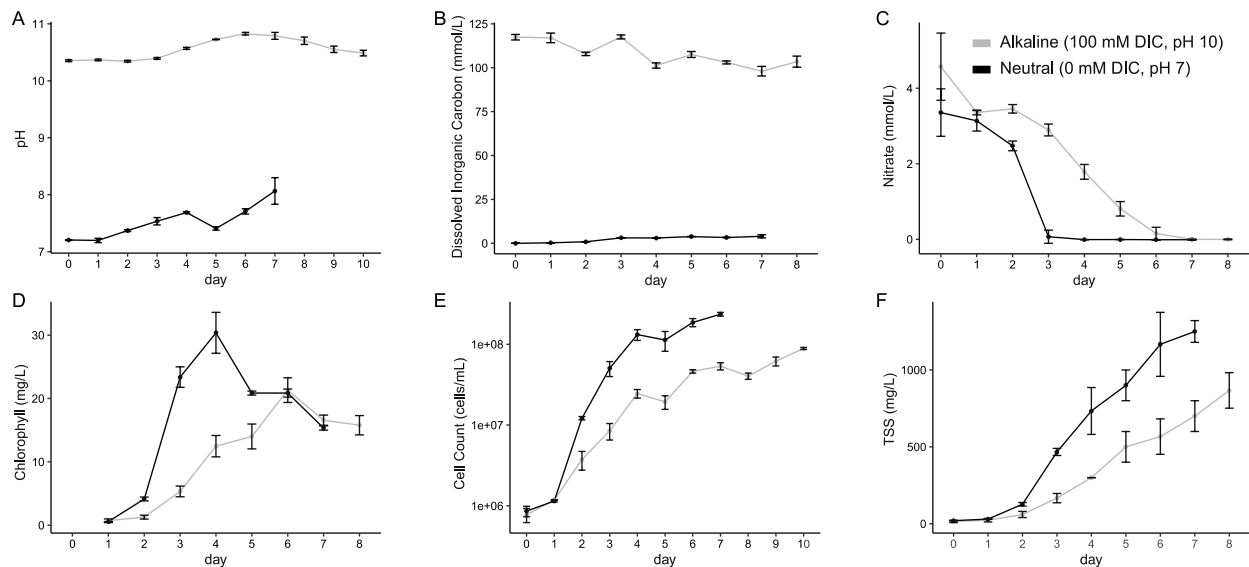


Figure 1: Experimental conditions and biomass production of SLA-04. SLA-04 was grown in neutral (pH 7, 0 mM added inorganic carbon) and alkaline (pH 10, 100 mM added inorganic carbon) conditions. (A) pH, (B) inorganic carbon concentration, (C) nitrate concentration, (D) chlorophyll content, (E) cell count, (F) and total suspended solids were monitored and quantified for both conditions.

Lipid and Carbohydrate Composition. Higher Nile red fluorescence is observed in *Chlorella* sp. SLA-04 cultures grown in alkaline conditions (**Fig 2 A**). Nile red fluorescence is only observed in non-polar environments like lipid bodies and has shown a strong correlation to

lipid concentration in algal cultures (Cooksey et al., 1987). While Nile red cannot be used to quantify lipid concentrations, it provides a quick low-cost way to compare lipid accumulation between algal strains and cultivation conditions. In contradiction to increased Nile red fluorescence, increased lipid accumulation is not observed when quantifying lipid content with GCMS. After centering lipid accumulation around nitrogen depletion (Neutral = Day 3, Alkaline = Day 6), cells from both cultures had near identical lipid accumulation after nitrogen depletion only showing significantly lower lipid content for the alkaline conditions four days after nitrogen depletion (**Fig 2 B**). In these conditions, Nile red fluorescence appears not correlated to lipid content as seen previously. We hypothesize that differential Nile red fluorescence may be due to the increase of other non-polar storage molecules in alkaline conditions.

To test this hypothesis, we measured the total carbohydrate content of SLA-04 across neutral and alkaline conditions. Prior to nitrogen depletion SLA-04 biomass grown in alkaline conditions contains approximately three times more carbohydrates (21.4%) than SLA-04 grown in neutral conditions (7.0%). After nitrogen depletion, both conditions show a significant increase in total carbohydrate concentration reaching similar maximum concentrations of 30.5% for neutral conditions and 29.1% for alkaline conditions (**Fig 2C**). Carbohydrate content does not appear to be what is impacting Nile red fluorescence in alkaline conditions.

SLA-04 has five major fatty acid species in both neutral and alkaline conditions. Fatty acids range from C16-C18 in length and contain 0-3 degrees of unsaturation (**Fig 2D**). The overall fatty acid length within SLA-04 appears to be impacted by nitrogen availability. Prior to nitrogen depletion, fatty acids within the neutral and alkaline conditions are composed of 54.66% and 57.73% C18 fatty acids respectively with the remaining fatty acids being C16. After nitrogen

depletion, both cultures increase in C18 fatty acid composition to 66.28% and 66.93% of total fatty acids. In addition to increasing length after nitrogen depletion, fatty acids in both conditions become more saturated. After nitrogen depletion, both cultures show a decrease in C18:3 and C16:3 fatty acids and an increase in C18:1 compared to log phase cultures.

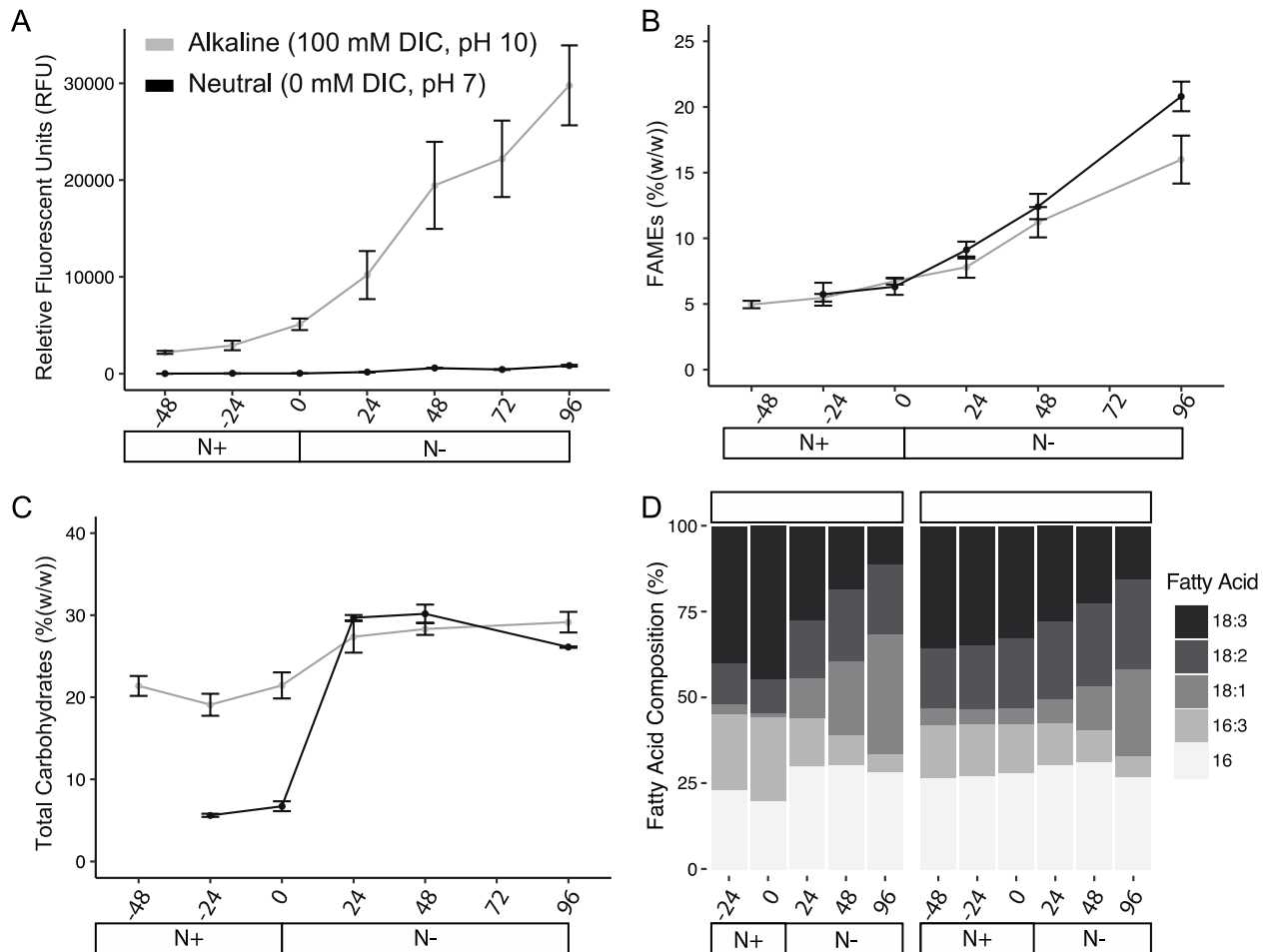


Figure 2: Biomass composition of SLA-04 in neutral versus alkaline conditions. Biomass composition was centered on nitrogen depletion (nitrogen depletion = day 3 for neutral and day 6 for alkaline conditions). N+ represent nutrient replete conditions. N- represent nutrient deplete conditions. X axis denote hours before or after nitrogen depletion. (A) Nile Red fluorescence was used as proxy to compare the amount of lipids within SLA-04 grown in neutral and alkaline conditions. Lyophilized SLA-04 biomass was used to calculate (B) Fatty acid methyl ester (FAME) content and (C) total carbohydrate content of SLA-04. Fatty acids were extracted through transesterification and quantified using a GCMS. (D) fatty acid composition was quantified and mapped throughout the experiment for SLA-04 in neutral and alkaline conditions.

While SLA-04 grown in neutral and alkaline conditions show similar trends of increased saturation and length after nitrogen depletion, fatty acid composition between conditions is different. During log phase growth, SLA-04 grown in neutral conditions contains significantly more unsaturated fatty acids, with statistically more C18:3 and C16:3 (**Supp Table 1**). After nitrogen depletion, this trend flips and fatty acids from SLA-04 grown in alkaline conditions are significantly more unsaturated. Post-nitrogen depletion, both conditions have identical C16/C16:3 fatty acid profiles but SLA-04 grown in alkaline conditions has significantly more C18:3 and C18:2 while neutral conditions select C18:1 as the major C18 fatty acid.

Transcriptome Analysis of Carbon Metabolism

Photosynthesis.mRNA was extracted from *Chlorella sp.* SLA-04 before and after nitrogen depletion from both neutral and alkaline conditions to identify differential expression (**Supp. Figure 1**). Principal component analysis of samples showed that biological replicates of each condition formed distinct clusters (**Supp. Figure 2**). Sequenced reads were mapped to corresponding genes within the SLA-04 genome and expression changes were calculated using the log2 fold change of normalized read counts between conditions (Goemann et al., 2023). Differential expression was compared between neutral and alkaline cultures while also comparing nitrogen-replete versus nitrogen-deplete. To explain the mechanism of fatty acid and carbohydrate content increases after nitrogen depletion we explored how central carbon metabolism is differentially expressed between conditions (**Fig. 3**). The highest levels of differential expression within carbon metabolism were observed between pre- and post-nitrogen depletion for both neutral and alkaline cultures (**Supp. Fig. 3**).

Most photosynthetic genes decrease in expression once both neutral and alkaline conditions enter nitrogen limitation (**Fig. 3**). Downregulation was observed in genes coding for PSII, PSI, and ATP synthase proteins (**Supp. Fig. 3**). Select photosynthetic genes involved in photooxidation repair showed opposite expression changes having the highest upregulation of all genes involved in carbon assimilation after nitrogen depletion. Carbon fixation genes show similar trends with a down-regulation of most genes after nitrogen depletion. Terpenoid biosynthesis is downregulated in nitrogen limited conditions suggesting that photosynthetic accessory pigment production is also reduced. In contrast, tetrapyrrole genes involved in synthesizing chlorophyll appear to be upregulated after nitrogen depletion.

Starch and Sugar Metabolism. The expression of genes involved in starch production in SLA-04 is also altered between conditions. Starch synthesis shows a slight upregulation of genes involved in starch elongation in nutrient deplete conditions. However, ADP-glucose pyrophosphatase which produces ADP-glucose, the first committed step of starch synthesis, is downregulated (**Supp. Fig. 3**). A clearer trend of upregulation is observed in starch degradation genes (**Fig. 3**). Genes involved in starch degradation have their lowest levels of expression in neutral nitrogen replete conditions with upregulation in both alkaline conditions and nitrogen deplete environments. Three genes involved in starch degradation, two coding for amylase and one for alpha-galactosidase, show strong upregulation in only alkaline conditions and are expressed at low levels even in neutral post-nitrogen deplete conditions.

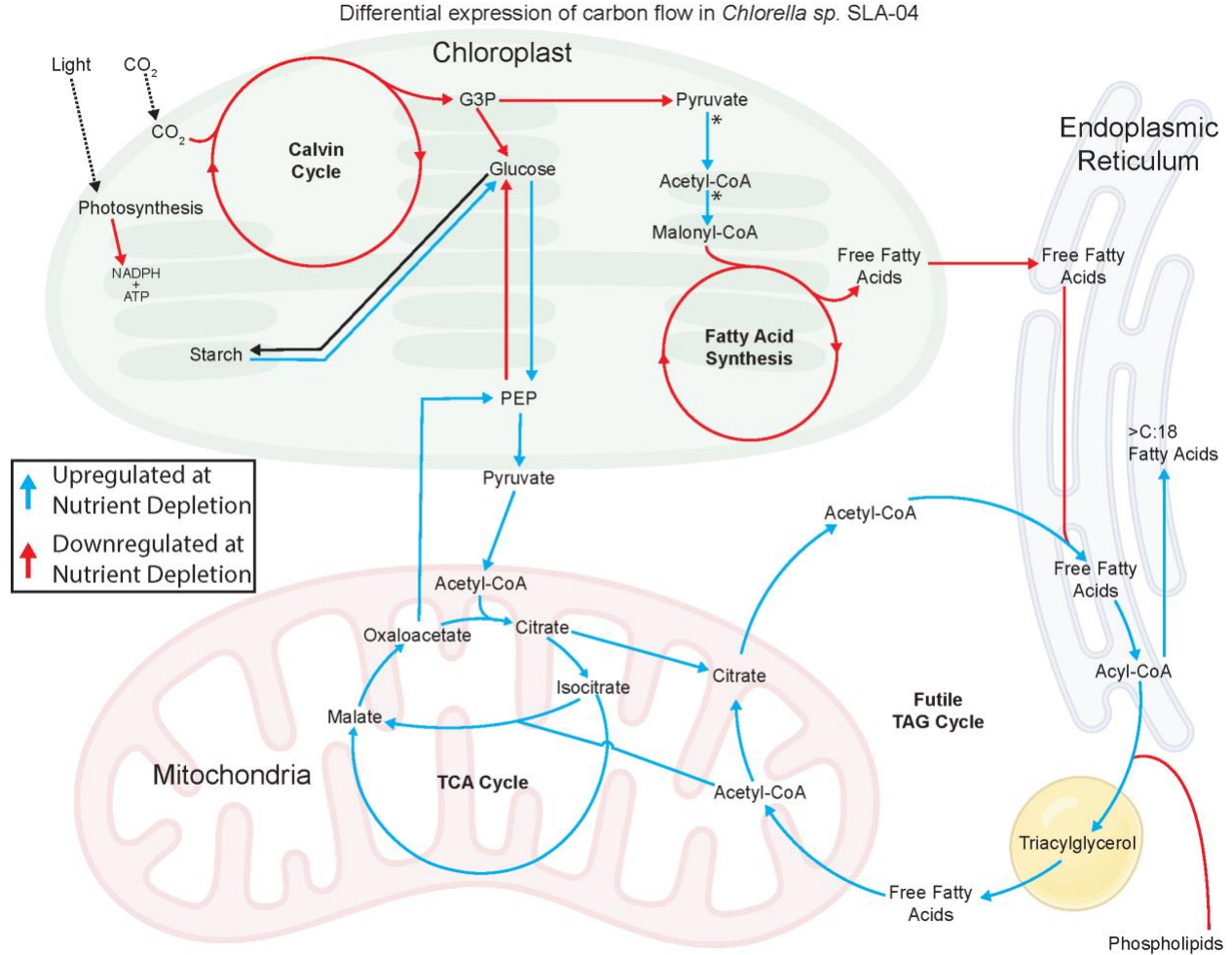


Figure 3: Differential expression of carbon flow before and after nitrogen depletion. Blue arrows indicate upregulated pathways in nutrient-deplete conditions, Red arrows indicate downregulated pathways. Black arrows indicate pathways with no differential expression.

Genes related to sugar degradation are also upregulated in nitrogen-depleted conditions. All irreversible steps of glycolysis are significantly upregulated in neutral and alkaline conditions after nitrogen depletion (**Fig. 3, Supp. Fig. 3**). Upregulation is also observed in the oxidation of pyruvate to acetyl-CoA. While catabolic reactions are upregulated, gluconeogenesis is downregulated after nitrogen depletion. Together, the upregulation of starch degradation, glycolysis, and pyruvate oxidation in addition to lower gluconeogenesis expression suggests that

SLA-04 may be breaking down starches to produce other macromolecules in the cell after nitrogen depletion.

Fatty Acid Synthesis. Chloroplast fatty acid synthesis appears to be down-regulated in stress-inducing environments (**Fig. 3, Supp. Fig. 3**). Almost all fatty acid synthesis genes have the highest levels of expression in neutral pre-nitrogen depletion conditions. Levels of expression are slightly lower in alkaline pre-nitrogen depletion conditions, but the most significant downregulation is seen once both neutral and alkaline cultures transition to nitrogen deplete environments. While fatty acid synthesis is downregulated after nitrogen depletion it appears that the production of fatty acid precursors has different expression trends. Fatty acid synthesis requires acetyl-CoA and malonyl-CoA to elongate fatty acids. Genes involved in the production of acetyl-CoA from pyruvate and the conversion of acetyl-CoA to malonyl-CoA within the chloroplast are slightly upregulated in both alkaline conditions and downregulation in both neutral conditions.

Lipid Synthesis. While it appears that carbon flow in the chloroplast is downregulated after nutrient depletion, carbon flow in other cellular compartments is significantly upregulated. The synthesis, elongation, and degradation of triacylglycerol (TAG) storage lipids are upregulated in nutrient deplete conditions (**Fig. 3, Supp. Fig. 3**). TAG biosynthesis and degradation are also upregulated in nitrogen-replete alkaline conditions suggesting that both nitrogen availability and alkalinity impact fatty acid expression (**Supp. Fig. 3**). In contrast to fatty acid synthesis, elongation, and degradation, genes related to the conversion of phospholipids to TAGs is downregulated in nitrogen-depleted conditions.

Mitochondrial Carbon Metabolism. TAGs can be stored in lipid bodies or catabolized into fatty acids. Fatty acids produced by TAG degradation are further degraded to acetyl-CoA via beta-oxidation within the mitochondria. In SLA-04 we see an upregulation of beta-oxidation genes after nitrogen depletion (**Fig. 3, Supp. Fig. 3**). Acetyl-CoA generated from beta-oxidation can either enter the tricarboxylic acid (TCA) cycle to generate energy or enter the glyoxylate shunt to replenish TCA intermediate biomolecules. In addition to upregulated beta-oxidation, there is a significant increase in the number of transcripts related to the glyoxylate shunt in SLA-04 in nitrogen-deplete conditions.

Genes related to the TCA cycle are upregulated after nitrogen depletion in both neutral and alkaline SLA-04 cultures (**Fig. 3**). Expression of TCA cycle genes is lowest in neutral pre-nitrogen deplete conditions (**Supp. Fig. 3**). Intermediate levels of expression are observed for TCA genes in nitrogen-replete alkaline conditions. The highest levels of expression are present in neutral nitrogen deplete conditions followed closely by SLA-04 harvested from alkaline nitrogen deplete conditions. In addition to the upregulation of central TCA genes, we also see an upregulation of genes capable of converting TCA cycle intermediates to cytosolic acetyl-CoA. ATP citrate lyase (ACLY) is upregulated in alkaline nutrient-replete and both nutrient-deplete conditions compared to SLA-04 from nitrogen-replete neutral conditions. ACLY is an essential gene for replenishing cytosolic acetyl-CoA for TAG synthesis and elongation.

Overall, the strongest trends of differential expression in SLA-04 carbon flow are observed in pre- versus post-nitrogen conditions. Only a select few genes show differential expression trends that correlate with neutral versus alkaline conditions (**Supp. Fig. 3**). When comparing pre- versus post-nitrogen depletion photosynthesis, carbon fixation, and chloroplast fatty acid synthesis are

strongly downregulated after nitrogen depletion (**Fig. 3**). In the endoplasmic reticulum TAG synthesis and elongation are upregulated after nitrogen depletion. Up-regulation is also observed in the mitochondria with increased TCA cycle and glyoxylate shunt gene expression. Lastly, increased TAG degradation, beta-oxidation, and ACLY expression in nutrient deplete conditions suggest that SLA-04 increases the cycling of carbon between lipids and the TCA cycle in stressful conditions.

Genes Involved in Adaptation to Nitrogen Depletion

Genes with the greatest differential expression were identified between neutral versus alkaline conditions and pre- versus post-nitrogen depletion. To select for gene expression that correlates with nutrient availability we analyzed the neutral and alkaline conditions separately and identified the most significant differentially expressed genes after nutrient depletion. After nutrient depletion in the neutral condition, 644 genes are significantly downregulated while 969 are upregulated compared to the nutrient-rich condition (**Fig. 4, Supp. Fig. 4a**). In the alkaline condition after nutrient depletion, 214 genes are downregulated while 270 are upregulated (**Fig. 4, Supp. Fig. 4b**). 169 downregulated and 176 upregulated genes were shared between neutral and alkaline conditions after nitrogen depletion. Of the shared genes, 126 downregulated genes and 97 upregulated genes have predicted functions. We hypothesize that shared genes play critical roles in adaptation to nutrient deplete conditions.

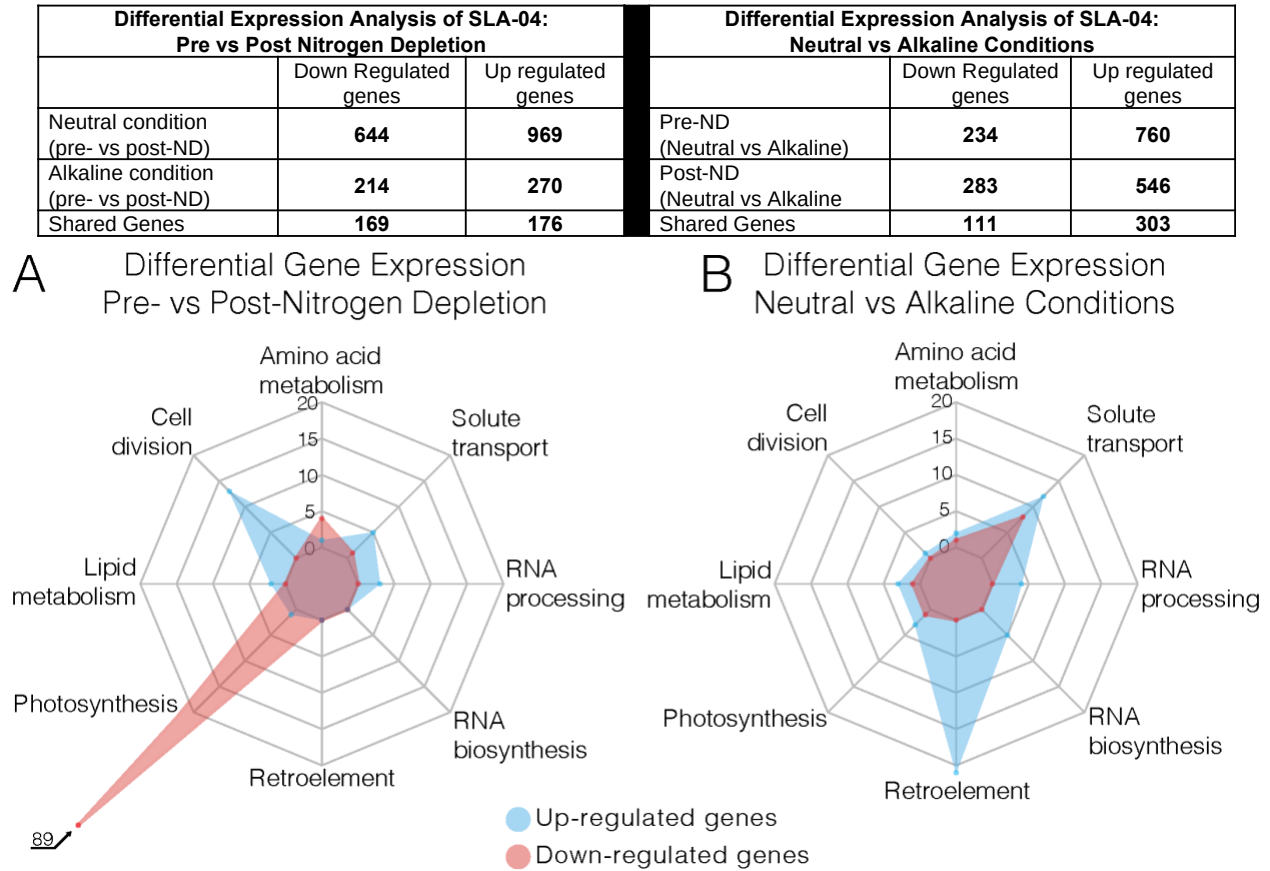


Figure 4: Genes with the greatest differential expression between conditions. (A) genes with greatest up or downregulation in post-nitrogen depletion conditions compared to pre-nitrogen depletion. Red is down regulated and blue is upregulated. Total differential genes are noted in left table. (B) Genes with greatest up or downregulation in alkaline conditions compared to neutral conditions. Total differential genes are noted in right table.

Shared genes that were differentially expressed after nutrient depletion in neutral and alkaline conditions clustered into multiple metabolic pathways. Most strikingly, of the 126 functionally annotated down-regulated genes 89 were involved in photosynthesis (**Fig. 4a, Supp. Table 3**). The downregulation of photosynthetic genes is complemented by significant upregulation of oxidative stress and photosynthetic repair genes. The synthesis of antioxidants including carotenoids and serotonin is upregulated (**Supp. Table 3**). Additionally, superoxide

dismutase is highly upregulated in nutrient deplete conditions. This suggests a strong global downregulation of photosynthesis and increased photo-repair in nutrient deplete conditions.

The second highest downregulation was observed in genes involved in amino acid synthesis and translation. In nutrient-deplete conditions algae likely do not have the necessary nitrogen for high levels of protein synthesis. SLA-04 appears to upregulate genes related to nitrogen scavenging in nutrient-deplete conditions. Significant upregulation was also seen in genes involved in the transport of nitrogen-rich compounds including amino acids, uric acid, and nucleotides (**Fig 4a, Supp. Table 3**). Additionally, catabolic ammonium generation through deaminase activity is significantly upregulated. Transporting nitrogen-rich molecules and recycling amines may be used to supplement the low nitrogen availability in stationary phase (Dündar & Bush, 2009).

Nitrogen depletion also impacts global regulatory systems in SLA-04. The most significant upregulation after nitrogen depletion was seen in cell cycle regulatory genes. Genes critical for inhibiting G2/M transition during mitosis (WEE1 kinase) and condensing chromosomal DNA (chromosome structural maintenance protein SMC2-1) are significantly upregulated (**Supp. Table 3**). Global transcription is also impacted by nutrient availability. Multiple transcription factors are upregulated in nutrient deplete conditions (**Supp. Table 3**). These genes may serve as the driving factors for expression changes. In addition to regulation at the transcriptional level, nutrient depletion also causes upregulation in histone modifications that may alter gene expression and chromosomal condensation through structural genomic rearrangements.

This analysis also identified genes involved in fatty acid biosynthesis and TAG accumulation. While we see a downregulation in plastid fatty acid synthesis, genes involved in

TAG accumulation are significantly upregulated. Diacylglycerol acetyl transferase (DGAT) is highly upregulated in nutrient deplete conditions (**Supp. Table 3**). This gene has previously been described to be involved in nutrient-deplete lipid accumulation and serves as a benchmark that our analysis is identifying genes critical for lipid acclimation to low-nutrient conditions. In addition to DGAT, we also see up-regulation of a caleosin-type peroxygenase. Caleosin-type peroxygenase enzymes have appreciated roles in abiotic stress response and lipid droplet storage in plants (Rahman et al., 2018). The increased expression may suggest that peroxygenase activity plays a critical role in TAG accumulation or processing in SLA-04 after nitrogen depletion.

Genes Involved in Adaptation to Alkaline Conditions

To identify genes that are associated with adaptation to alkaline growth, we analyzed differential gene expression between neutral and alkaline conditions. Cultures growing in log phase Pre-ND conditions and stationary Post-ND conditions were compared independently to minimize the effects of nitrogen depletion on gene expression (neutral Pre-ND vs alkaline Pre-ND and neutral Post-ND vs alkaline Post-ND). Shared up- or down-regulated genes between these two comparisons should play a role in adaptation to alkaline growth. When comparing the neutral versus alkaline conditions prior to nitrogen depletion, 234 genes are significantly downregulated while 760 are upregulated (**Fig. 4, Supp. Fig. 5a**). Comparing the alkaline versus neutral condition after nutrient depletion showed 283 genes are significantly downregulated and 546 are upregulated (**Fig. 4, Supp. Fig. 5b**). 111 downregulated and 303 upregulated genes were shared and may be involved in adaptation to growth in alkaline conditions. Of the shared genes, 47 downregulated genes and 149 upregulated genes have predicted functions.

Interestingly, retroelements are significantly upregulated in alkaline conditions. The SLA-04 genome contains 26 Copia-like retrotransposons integrated throughout the 13 nuclear chromosomes (Goemann et al., 2023). Of the 26 retrotransposons, 21 are highly upregulated in alkaline conditions (**Fig. 4b, Supp. Table 4**). Retroelements are commonly activated in stressful conditions. Alkaline-induced stress is supported by the upregulation of superoxide dismutase and desiccation response proteins (**Supp. Table 4**). In addition to retroelements, we also see an upregulation in RNA processing and biosynthesis. Specifically, we see an upregulation in RNase activity including dicer and two T2 Rases. Dicer generates small interfering siRNAs involved in antiviral RNA interference (RNAi) which function to suppress target viral or transposon genes within algae (Rowe et al., 2013). The upregulation of pathogenic and antiviral genes may suggest that alkaline conditions impose stresses that trigger immune responses in algae.

Solute transport appears to be rewired in SLA-04 during alkaline growth. Multiple potassium, sodium, sulfate, calcium, and phosphate transporters are highly upregulated in alkaline conditions (**Fig. 4b, Supp. Table 4**). In contrast, iron accumulation and oxidation genes are downregulated. SLA-04 may be adapting to the higher osmotic stresses of alkaline conditions by utilizing alternative transport mechanisms to acquire necessary nutrients and maintain intracellular homeostasis. We also see that a strong up-regulation of a member two ammonium transporter and a down-regulation of a member one ammonium transporter (Yuan et al., 2007). Transcriptional profiles of ammonium transporters are highly diverse; however, some have shown strong differential expression in response to nutrient and light stress in plants (Gazzarrini et al., 1999; Yan et al., 2013). In high pH conditions above 9.25, the major nitrogen species shift from

ammonium to ammonia. Differential expression of ammonium transport may be triggered by alkaline stress and the resulting shift in nitrogen speciation.

Alkaline conditions also seem to trigger the degradation and modification of macromolecules. Genes involved in sugar degradation and the expression of proteases and thioesterases involved in protein degradation are upregulated in alkaline conditions (**Supp. Table 4**). We also see the select degradation of nitrogen rich metabolites to produce ammonium in alkaline conditions. This may suggest that in alkaline conditions algae are rewiring carbon metabolism and recycling limiting nutrients. While sugars and proteins show increased degradation, alkaline conditions appear to alter the saturation of lipids. In alkaline conditions, there is a significant upregulation of desaturase genes. These genes correlate with the increased levels of unsaturation seen in alkaline conditions after nitrogen depletion (**Fig. 2**).

Discussion

Our analysis suggests that SLA-04 responds to nitrogen depletion by synthesizing both carbohydrates and lipids. In nutrient deplete conditions SLA-04 has an excess of photosynthetic energy due to the inability to consume NADPH through growth and cell division (Sharma et al., 2018). Synthesis of carbohydrates and lipids likely serves as an electron sink to mitigate photooxidative stress associated with excess photosynthesis (McKie-Krisberg et al., 2018). SLA-04 responds quickly to nutrient depletion by synthesizing high levels of carbohydrates (**Fig. 2**). However, once carbohydrate concentration reaches about 30%, cells transition to producing lipids. Carbohydrates are moderately reduced biomolecules while lipids are highly reduced biomolecules that require 50% more reducing power to synthesize. In excess energy states, lipids serve as an ideal molecule to maximize the consumption of NADPH per carbon molecule in the cell. Here it

appears that in initial nutrient depletion, SLA-04 opts to synthesize high levels of moderately reduced carbohydrates. As nitrogen depletion continues for extended periods of time SLA-04 then transitions to maximize carbon reduction and starts accumulating highly reduced lipids.

We hypothesize that during early nutrient depletion algal cells still have most of the enzymes needed for carbon metabolism. In these conditions, algae have enough enzymatic turnover to synthesize high volumes of carbohydrates. After extended periods of nitrogen depletion, the cell is unable to replenish all enzymes and needs to maximize the utilization of photosynthetic reducing power with limited enzymatic activity. By prioritizing TAG synthesis cells can maintain lower levels of biosynthesis while maintaining higher levels of NADPH consumption to minimize oxidative stress in the cell. This transition is supported by our transcriptomic analysis showing the upregulation of TAG biosynthesis with minor expression changes in starch synthesis 24 hours after nitrogen depletion in both alkaline and neutral conditions. Starch synthesis was likely significantly upregulated directly following nitrogen depletion; however, our sampling regiment did not capture these immediate metabolic shifts. Our transcriptome also suggests that carbon for TAG synthesis is coming in part from starch degradation and less from *de novo* fatty acid synthesis (**Fig. 3**).

In addition to the upregulation of TAG synthesis, we also see an upregulation in TAG degradation, beta-oxidation, the glyoxylate shunt, the TCA cycle, and cytosolic acetyl-CoA generation. Together these metabolisms form a cyclical pathway between storage lipids and the TCA cycle (**Fig. 3**). In nutrient-deplete conditions cell growth slows and *de novo* synthesis of carbon is reduced. This limits carbon availability for synthesizing lipids and consuming excess photosynthetic NADPH. We hypothesize that upregulated carbon cycling between the TCA cycle

and TAG storage may allude to a recycling pathway that SLA-04 uses to replenish precursors for TAG synthesis from carbon already stored in the cell. This proposed mechanism of breaking apart TAGs just to passage carbon through the TCA cycle back to TAG synthesis is a major waste of energy and reducing power in a cell. Cells growing in nutrient-rich conditions would have no reason to use this strategy and waste photosynthetic energy. However, in nutrient-limited conditions, SLA-04 has excess photosynthetic energy that is harmful to the cell. The use of futile cycles to mitigate excess energy in a cell is common in other over-encumbered microbes (Boer et al., 2010; Ren et al., 2009). This pathway may be an effective futile cycle within algae to consume excess reducing power in the cell while maintaining a store of carbon in TAGs and starch.

Genes involved in photosynthesis, translation, amino acid synthesis, and cell cycle regulation were the most abundant gene categories with differential expression in nutrient deplete conditions. In nitrogen-limited environments SLA-04 cannot produce the necessary enzyme to grow and divide rapidly. SLA-04 responds to limited nutrients through the global downregulation of protein synthesis and termination of cellular division. In addition, SLA-04 appears to recycle and scavenge for nitrogen via upregulated transport of nitrogen-rich molecules including amino acids, uric acid, and nucleotides. The downregulation of photosynthesis is likely due to limited nitrogen availability, but also to minimize oxidative stress in the cell. We hypothesize that in nutrient deplete conditions SLA-04 is maintaining a minimal amount of photosynthesis to provide enough energy to survive without over-oxidizing the stationary algal cell.

Interestingly, while most differentially expressed genes in nitrogen deplete conditions are directly related to nitrogen availability, scavenging, and the cell cycle, alkaline conditions induce other expression trends in SLA-04. Alkaline conditions appear to impose stresses that activate

mobile genetic elements and associated antiviral genes. In alkaline conditions, we see almost all retroelements in the SLA-04 genome transition from no expression to high levels of expression. Active retroelements can insert throughout the genome causing genome rearrangement and modification. SLA-04 was isolated from alkaline conditions and likely experienced constant activation of retroelements. While retroelements provide a mechanism for SLA-04 to adapt to extreme environments through genome rearrangement, highly active mobile genetic elements can be detrimental to the cell. Steps must be taken to minimize the parasitic burden of retroelements. In alkaline conditions, we see high levels of RNAi antiviral genes including dicer. Other than RNAi very few other algal immune systems have been identified (Rowe et al., 2013). If high alkalinity activates mobile genetic elements, then this condition may be used to screen for unknown antiviral systems in algae.

Other upregulated genes may play unappreciated roles in the algal immune defense. T2 Rnases are predicted to respond to pathogens in plants but have not been appreciated to provide protection against pathogens in green algae (Deshpande & Shankar, 2002; Singh et al., 2020). In addition to RNase activity, we also see an upregulation of aldehyde oxidase, an enzyme that oxidizes aldehydes to produce H_2O_2 (**Supp. Table 4**). This gene combats viral infection in plants through the production of H_2O_2 and could be used by algae to combat pathogens (Yergaliyev et al., 2016). T2 Rnase, aldehyde oxidase, and other upregulated genes with unknown functions may be candidates for immune defense in algae and expand our understanding of how algae evade the constant barrage of viruses and mobile genetic elements (Deshpande & Shankar, 2002; Singh et al., 2020).

Alkaline growth also impacted the physiological characteristics of SLA-04 causing high levels of carbohydrate accumulation prior to nitrogen depletion. We hypothesize that increased carbohydrates play an important role in osmoregulation in alkaline conditions. Extreme alkaline conditions have over 100 times more inorganic carbon ions than neutral conditions. This dramatically increases the osmotic pressure experienced by cells in alkaline conditions. While carbohydrates serve as an electron sink after nitrogen depletion, non-structural sugars can also act in osmoregulation to maintain intracellular homeostasis (Chen et al., 2011; Hartmann & Trumbore, 2016). Increased carbohydrates in alkaline conditions prior to nitrogen depletion may help combat the osmotic pressure of growing in alkaline conditions. Additionally, SLA-04 in alkaline conditions post-nitrogen depletion has a more unsaturated fatty acid profile compared to neutral conditions. Unsaturated acids are more susceptible to oxidative damage and are reduced in stressful conditions. Maintaining higher levels of unsaturation suggests that unsaturation and membrane fluidity may be necessary to maintain homeostasis in alkaline conditions. To further support that osmotic stress is a problem for algae growing in alkaline conditions we see an upregulation of ionic transport in alkaline conditions. Due to the increased ionic load of alkaline conditions, SLA-04 appears to increase ion transport to maintain homeostasis. While alkaline conditions impose stress on algae, transcriptional changes are distinct from those of nutrient depletion and tend to focus on the activation of mobile genetic elements, antiviral systems, and maintenance of intracellular homeostasis.

Conclusion

Chlorella sp. SLA-04 is a productive green alga that grows well in neutral and alkaline conditions. Here we show how SLA-04 rearranges central carbon metabolism to produce

carbohydrates and TAGs in nutrient deplete conditions. Additionally, SLA-04 appears to recycle fatty acids through a futile cycle to consume excess photosynthetic energy while maintaining stores of highly reduced TAGs. This analysis further highlights how SLA-04 adapts to the different stresses imposed by nitrogen limitation and extreme alkalinity. SLA-04 survives nutrient depletion by increasing nutrient scavenging and lowering photosynthetic activity while extreme alkaline stresses are overcome through antiviral activity and osmoregulation. Together, this analysis identifies genotypes that algae may use to survive in extreme alkaline and provides a foundation to optimize alkaline biomass production through condition optimization and genome editing.

Methods

Strain Cultivation Conditions

Chlorella sp. SLA-04 was grown in 1.25 L tube photobioreactors with 400 $\mu\text{mol}/\text{m}^2\text{s}$ light using T5 fluorescent bulbs with a 14:10 light dark cycle. Cultures received 0.4 L/min of air and temperature was maintained at 23°C. Alkaline cultures were grown in Bold's Basal media with 100 mM added inorganic carbon in the form of bicarbonate. Sodium hydroxide was added to titrate cultures to a pH of 10.3. neutral cultures were grown in bolds basal media titrated to pH 7. 5% CO_2 was mixed inline with the air supply as necessary to buffer the neutral condition if pH rose above 7.2 and would stop once pH returned to 7.0.

Quantifying Media Composition

Dissolved Inorganic Carbon. 10 mL of culture was harvested and filter sterilized. Supernatant of alkaline conditions was diluted 10 times to bring the inorganic carbon concentration below 100 mg carbon/L. Neutral sample did not need to be diluted. Samples were run on a

SKALAR formacs HT high-temperature instrument to quantify media inorganic carbon concentration. Instrument oxygen flow was set to 200 and fresh 4% phosphoric acid (H_3PO_4) was supplied every 60 samples. Samples were compared to a standard dilution series containing 0 - 1000 mg carbon/L composed of molar equivalents of sodium carbonate and sodium bicarbonate.

Media Nitrate Concentration. Szechrome NAS reagent from Polysciences, Inc. (Cat# 08762) was prepared by dissolving 0.5 g of NAS in 100 mL of equal parts concentrated phosphoric H_3PO_4 and sulfuric H_2SO_4 acid. Acid mixture was prepared 1 week before adding NAS to minimize background NO_3 concentrations in the acid. One mL of culture was harvested and filter sterilized with a 0.22 μm filter. Samples were diluted as necessary to be lower than 3 mmol or 42 mg/L NO_3^- . 20 μL of sample was transferred to a 96 well plate along with nitrate standards containing 0 – 50 mg/L NO_3^- . 200 μL of NAS acidified NAS reagent was added to each sample starting with the most concentrated standard. Samples were mixed with a pipet >20 times and let stand for 30 min. After 30 minutes sample absorbance was measured at 507nm, and the concentration was determined using a calibration curve of known standards.

Cell Growth

Cell Count. One mL of culture was taken and diluted to approximately $1\text{E}6$ cells/mL. 12 μL of diluted sample was added to a hemocytometer with a 10^{-4} mL counting volume and samples were counted in triplicate. Data were converted to cells/mL by multiplying the calculated average cell count by 10^4 .

Total Suspended Solids. A 0.2 μm filter disk with 25 mm diameter was pre-weighed. Ten mL of culture sample was collected and filtered through pre-weighed filter using a vacuum

filtration setup followed by washing the filter 3 times with 5 mL diH₂O pH 6. Filters containing algal biomass were dried at 80°C for 24 hours and weighed. Total suspended solid values (mg/L) was calculated by subtracting the weight of clean filter from the weight of dried filter with algal biomass and adjusted for the volume of filtrate used.

Biomass Composition

Chlorophyll Concentration. One mL of algae culture was pelleted in a 15 mL falcon tube @ 500 x G for 10 minutes. The supernatant was discarded and the pellet was resuspended in 5 mL of 100% methanol. Sample was heated 5 minutes at 70 °C followed by centrifugation for 10 minutes at 500 x G. Supernatant was removed and saved. Five mL of methanol was again added to the pellet and the extraction was repeated. One mL of combined methanol (10 mL total) was added to a cuvette with a one cm path length and absorbance values were collected at 470, 653, and 666 nm using a spectrophotometer. One hundred percent methanol was used as a blank. Chlorophyll A, B, and carotenoid concentrations were calculated using the following equations.

- $\text{Chl a (mg/L)} = 17.12 \cdot A_{666} - 8.86 \cdot A_{653}$
- $\text{Chl b (mg/L)} = 32.23 \cdot A_{653} - 14.55 \cdot A_{666}$
- $\text{Chl a+b (mg/L)} = 2.57 \cdot A_{666} + 23.6 \cdot A_{653}$
- $\text{Car (mg/L)} = (1000 \cdot A_{470} - 2.86 \cdot (15.65 \cdot A_{666} - 7.34 \cdot A_{653}) - 129.2 \cdot (27.05 \cdot A_{653} - 11.21 \cdot A_{666})) / 221$

Nile Red Fluorescence. 40 uL of Nile Red (25 ng/uL) dissolved in acetone was added to 1 mL of algae culture. Nile red samples were mixed by vortex and 200 uL was transferred to a 96 well dark walled plate and incubated in the dark for exactly 10 minutes. Microplate was immediately transferred to a Citation5 microplate reader with Gen5 software to record

fluorescence using an excitation of 480 nm, emission of 580 nm, gain of 75, and read height of 7.0 mm. Relative absorbance units were recorded and used to compare lipid content between samples.

Lyophilization of Algal Biomass for FAME and Carbohydrate Analysis. 50 mL samples were collected from cultures for biomass composition analysis. Samples were pelleted at 3000 x G for 10 minutes and the supernatant was removed. Pellets were frozen to -80 °C and stored for lyophilization. Caps were removed and tubes with frozen pellets were transferred to a pre-chilled Millrock Technologies freeze dryer set at -40 °C. Samples were lyophilized for 72 hours at -40 °C and 100 mTorr. Freeze-dried samples were stored at -20 °C and used for FAME and carbohydrate analysis.

FAME Analysis. 5-15 mg of dried algal biomass was transferred to pre-weighed glass vials with Teflon caps. 0.5 mL of toluene and 1 mL of sodium methoxide was added to each sample. Closed vials were incubated for 30 min at 90°C vortexing every 10 minutes. 1 mL of boron trifluoride (14% BF₃ in methanol) was added and incubation step was repeated. Tubes were cooled to room temperature before adding 0.4 mL of hexanes and 0.4 mL of concentrated salt solution (15% NaCl w/w). tubes were incubated for 10 minutes at 90°C and organic layer was separated via centrifugation of 500 g for 2 minutes. Organic layer was transferred to capped GC vials and used for GCMS analysis of fatty acid methyl ester (FAME) content.

Carbohydrate Analysis. Total carbohydrate content was calculated using a colorimetric anthrone assay (Beck et al., 2018). Anthrone reagent was prepared by mixing 10 mg of anthrone and 250 uL of fresh absolute ethanol per reaction. 75% sulfuric acid was added to a final volume of 5 mL and stirred until anthrone was dissolved. 0.5-1.0 mg of dried algal biomass was transferred

to pre-weighed glass vials with Teflon caps. Samples were re-suspended in 1 mL of water and placed on ice. Glucose standards were prepared from a 1 mg/mL stock. Standards ranged from 10-600 ug/mL with a total volume of 1 mL. 5 mL of ice-cold anthrone reagent was added to each sample and standard. Samples and standards were chilled on ice for 5 minutes, vortexed, and incubated in boiling water for 10 minutes. Tubes were placed back on ice for 10 minutes to cool and vortexed. 200 uL of each sample and standard were transferred to a 96-well plate and absorbance was measured at 625 nm. Standards were used to make a calibration curve to calculate the total carbohydrate content of samples.

Transcriptome Analysis

RNA Extraction and Quantification. Biomass samples were collected for RNA sequencing. 50 mL samples were spun down (4k rfc @ 4°C for 5 min) and flash frozen in liquid nitrogen to limit RNA degradation. Samples flanking the time of nitrogen depletion were taken on days 2 and 4 for the neutral sample and on days 5 and 7 for the highly alkaline condition to identify gene expression changes associated with nitrogen depletion and alkalinity changes under these two conditions. Total RNA extraction was performed by grinding algal biomass with a mortar and pestle in liquid nitrogen. Frozen, ground biomass was transferred to Trizol extraction buffer and total RNA was purified using the Zymo research direct-zol RNA miniprep per manufacturer's instructions. Purification yielded 7.9-45.5 mg of total RNA per extraction for all 18 samples with a 260nm/280nm ratio ranging from 2-2.3 and a 260nm/230nm ratio ranging from 2.1-2.5. RNA integrity was also confirmed using a 2100 Bioanalyzer (Agilent). Samples had distinct 25s and 18s rRNA bands with RNA integrity numbers (RIN) of 7.8-9.5.

RNA Sequencing and Processing. Stranded RNASeq library(s) were created and quantified by qPCR. Sequencing was performed using an Illumina instrument at the Joint Genome Institute (JGI). Raw fastq file reads were filtered and trimmed using the JGI QC pipeline (Vasanth Singan). Using BBduk, raw reads were evaluated for artifact sequence by kmer matching (kmer=25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads (*BBduk*). RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever was longer).

Filtered reads from each library were aligned to the reference genome using HISAT2 version 2.2.0 (Kim et al., 2015). 98.8% of reads mapped back to the SLA-04 genome. Strand-specific coverage bigWig files (fwd and rev) were generated using deepTools v3.1 (Ramírez et al., 2014). featureCounts was used to generate the raw gene counts file using gff3 annotations (Liao et al., 2014). Only primary hits assigned to the reverse strand were included in the raw gene counts (-s 2 -p --primary options). Raw gene counts were used to evaluate the level of correlation between biological replicates using Pearson's correlation and determine which replicates would be used in the DGE analysis. FPKM and TPM normalized gene counts are also provided.

Differential Gene Expression. DESeq2 (version 1.30.0) was subsequently used to determine which genes were differentially expressed between pairs of conditions (Love et al., 2014). The parameters used to call a gene DE between conditions were p-value < 0.05 and log2 fold change < 2. SLA-04 genes were annotated using Mercator 4.5 and mapped to metabolic pathways using

KofamKOALA and KEGG mapper (Aramaki et al., 2020; Lohse et al., 2014; Moriya et al., 2007).

Volcano plots were generated in RStudio using ggplot2 (Hadley, 2016).

Protein Structural Prediction. Structural-based protein annotation was performed following the ColabFold-FoldSeek-EggNOG (CoFFE) pipeline (Ruperti et al., 2022). Briefly, amino acid sequences were submitted to the ColabFold Alphafold2-mmseqs2 notebook through google colab with default settings (Evans et al., 2021; Jumper et al., 2021; Mirdita et al., 2022). The highest-ranking predicted protein data base (pdb) structure with predicted local distance difference test on the C α atoms (pLDDT) > 70 was selected for each analyzed protein. PDB structures were submitted to Foldseek webserver for structural comparison against the AlphaFold database of protein structures (van Kempen et al., 2022).

Supplementary data

Percent Biomass as Fatty acids (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	4.95%	0.29%	NA
ND-1	5.74%	0.88%	5.47%	0.29%	0.6613
ND	6.32%	0.63%	6.73%	0.27%	0.3765
ND+ 1	8.75%	0.41%	7.81%	0.81%	0.1726
ND+ 2	12.42%	0.97%	11.23%	1.15%	0.2429
ND+ 4	20.81%	1.13%	16.00%	1.82%	0.0358

Percent Fatty Acid as 18:3 (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	35.31%	1.40%	NA
ND-1	39.66%	0.21%	34.52%	0.73%	0.0039
ND	44.59%	0.78%	32.49%	0.36%	0.0004
ND+1	27.22%	0.42%	27.72%	1.64%	0.6568
ND+2	18.27%	0.64%	22.24%	2.07%	0.0687
ND+4	10.92%	0.36%	15.31%	1.28%	0.0190

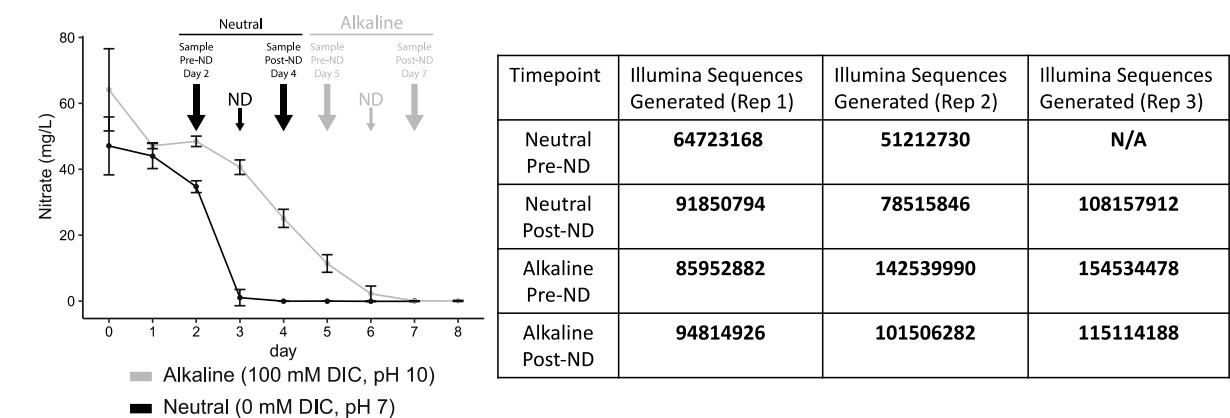
Percent Fatty Acid as 16:3 (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	15.32%	0.70%	NA
ND-1	22.08%	0.45%	14.97%	0.17%	0.0003
ND	24.58%	0.07%	14.31%	0.47%	0.0004
ND+ 1	14.13%	0.32%	12.32%	1.06%	0.0017
ND+ 2	8.91%	0.77%	9.40%	0.76%	0.4720
ND+ 4	5.26%	0.48%	5.99%	0.43%	0.2225

Percent Fatty Acid as 18:2 (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	17.61%	0.31%	NA
ND-1	12.00%	0.36%	18.68%	0.39%	2.89E-05
ND	9.67%	0.12%	20.50%	0.38%	0.0004
ND+1	16.94%	0.34%	22.59%	0.44%	0.0001
ND+2	21.08%	0.19%	24.38%	0.50%	0.0034
ND+4	20.54%	0.20%	26.44%	0.21%	0.0003

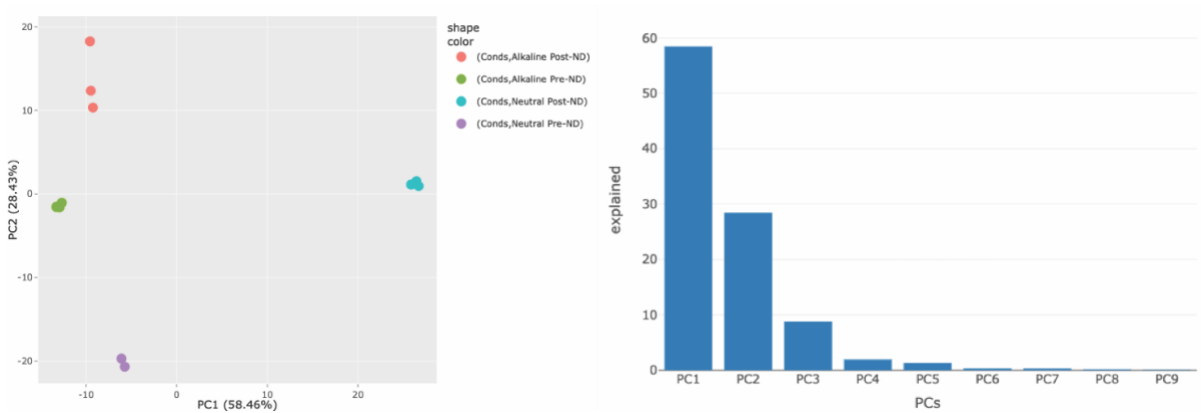
Percent Fatty Acid as 16:0 (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	26.78%	1.58%	NA
ND-1	23.25%	0.87%	27.30%	0.26%	0.0100
ND	19.93%	0.63%	28.14%	0.30%	0.0004
ND+ 1	30.05%	0.62%	30.36%	0.69%	0.6051
ND+ 2	30.33%	0.34%	31.30%	0.96%	0.2164
ND+ 4	28.46%	0.26%	27.08%	1.21%	0.1806

Percent Fatty Acid as 18:1 (w/w)					
Time	Neutral	Neutral stdev	Alkaline	Alk stdev	Pvalue
ND-2	NA	NA	4.98%	0.68%	NA
ND-1	3.00%	0.33%	4.53%	0.43%	0.00962
ND	1.23%	0.13%	4.56%	0.18%	2.96E-05
ND+1	11.65%	0.39%	7.02%	1.60%	0.0317
ND+2	21.41%	1.26%	12.67%	3.31%	0.0320
ND+4	34.82%	0.38%	25.18%	3.02%	0.0288

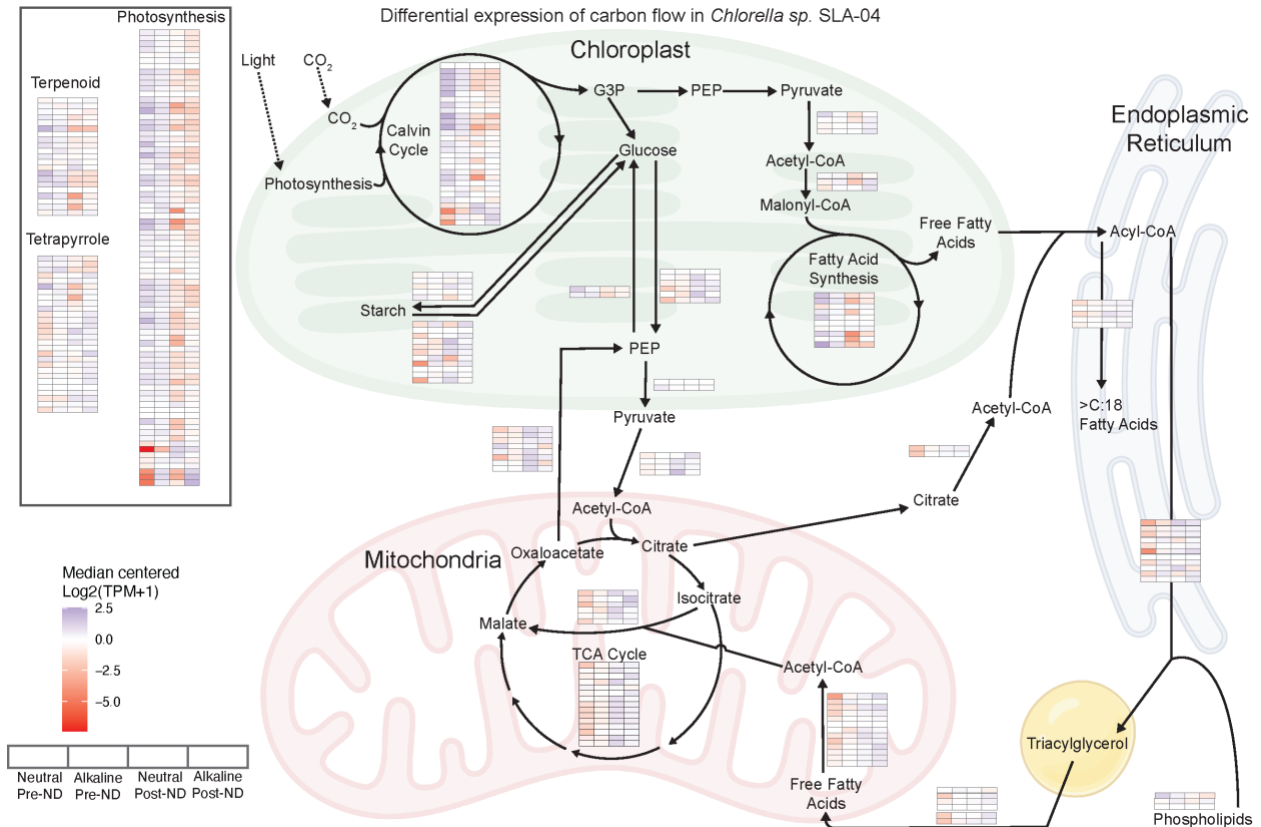
Supplementary Table 1: Fatty acid composition and statistics



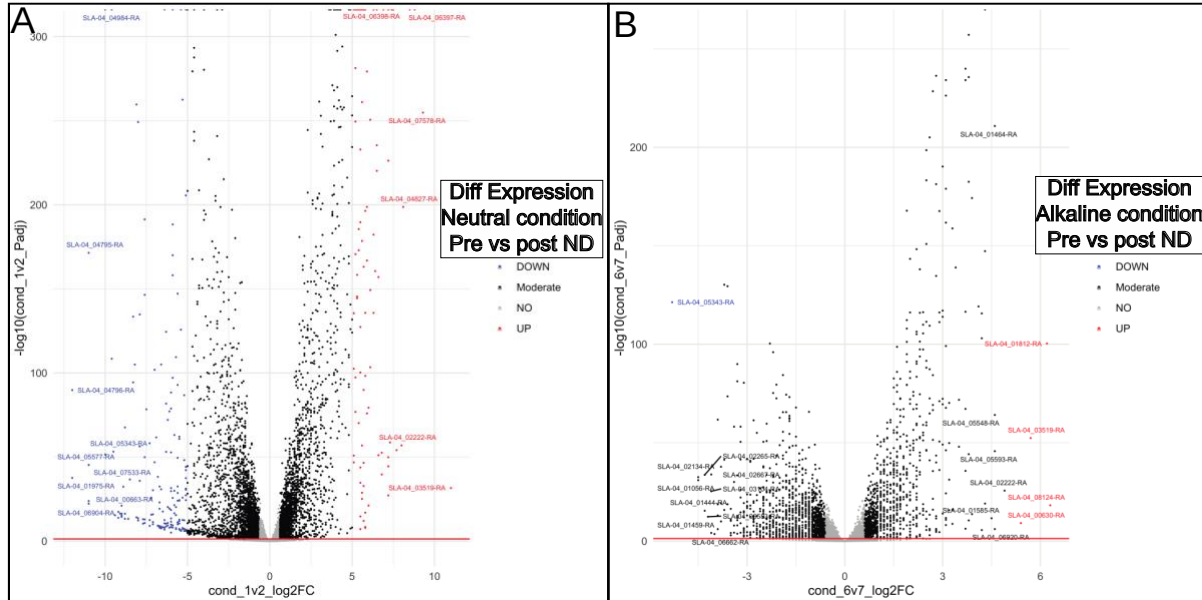
Supplementary Figure 1: Sampling timepoints and sequencing coverage



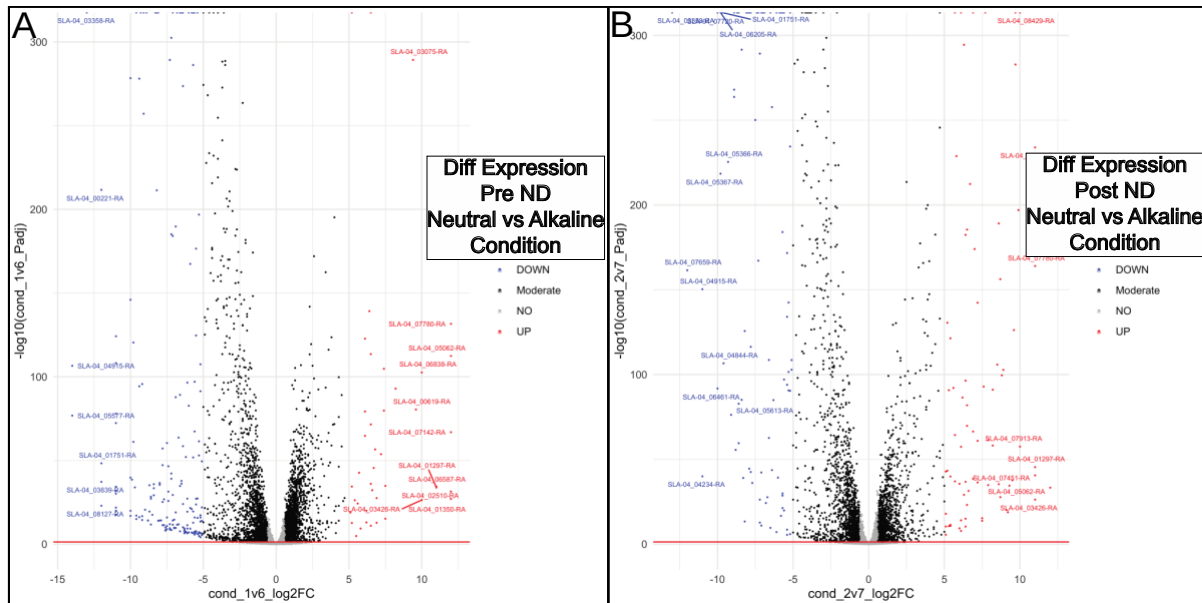
Supplementary Figure 2: PCA analysis of sequencing



Supplementary Figure 3: Schematic of differential carbon flow genes involved in growth in alkaline conditions. Between neutral and alkaline condition data.



Supplementary Figure 4: Volcano plots showing differential expression of pre vs post nitrogen depletion



Supplementary Figure 5: Volcano plots showing differential expression of neutral vs alkaline

Supplementary Table 2: Table of all differential genes between pre- versus post-nitrogen depletion

Gene ID	Average DGE	Annotation	Functional Category
SLA-04_03519-RA	-8.35	NDPK2 Nucleoside diphosphate kinase 2, chloroplastic (Nicotiana tabacum OX=4097)	Nucleotide metabolism
SLA-04_02222-RA	-6.45	OEP24A Outer envelope pore protein 24A, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_05548-RA	-6.15	CAH1 Carbonic anhydrase 1 (Chlamydomonas reinhardtii OX=3055)	Photosynthesis
SLA-04_04827-RA	-6	efeO Iron uptake system component EfeO (Shigella boydii serotype 4 (strain Sb227) OX=300268)	not assigned
SLA-04_05897-RA	-5.65	CYP38 Peptidyl-prolyl cis-trans isomerase CYP38, chloroplastic (Arabidopsis thaliana OX=3702)	Protein modification
SLA-04_00630-RA	-5.55	MAN1 Mannan endo-1,4-beta-mannosidase 1 (Oryza sativa subsp. japonica OX=39947)	Cell wall organisation
SLA-04_06397-RA	-5.55	metE 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (Synechococcus sp. (strain JA-2-3B'a(2-13)) OX=321332)	not assigned
SLA-04_05896-RA	-5.55	TLP40 Peptidyl-prolyl cis-trans isomerase, chloroplastic (Spinacia oleracea OX=3562)	not assigned
SLA-04_06398-RA	-5.4	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_04008-RA	-4.95	unknown function	Photosynthesis
SLA-04_06418-RA	-4.95	FKBP16-3 Peptidyl-prolyl cis-trans isomerase FKBP16-3, chloroplastic (Arabidopsis thaliana OX=3702)	Protein modification
SLA-04_05086-RA	-4.85	SAMC2 Probable S-adenosylmethionine carrier 2, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_02428-RA	-4.75	ASP1 Aspartate aminotransferase, mitochondrial (Arabidopsis thaliana OX=3702)	Nutrient uptake
SLA-04_02292-RA	-4.7	RPI3 Probable ribose-5-phosphate isomerase 3, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02300-RA	-4.7	ISPD 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, chloroplastic (Arabidopsis thaliana OX=3702)	Secondary metabolism
SLA-04_03993-RA	-4.65	Ppif Peptidyl-prolyl cis-trans isomerase F, mitochondrial (Mus musculus OX=10090)	Enzyme classification
SLA-04_01094-RA	-4.65	Sc5d Lathosterol oxidase (Rattus norvegicus OX=10116)	not assigned
SLA-04_07250-RA	-4.6	CIN2 Beta-fructofuranosidase, insoluble isoenzyme 2 (Oryza sativa subsp. indica OX=39946)	Enzyme classification
SLA-04_03682-RA	-4.6	GC1 Epimerase family protein SDR39U1 homolog, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01455-RB	-4.4	NEK8 Serine/threonine-protein kinase Nek8 (Homo sapiens OX=9606)	not assigned
SLA-04_08304-RA	-4.35	unknown function	Coenzyme metabolism
SLA-04_01464-RA	-4.25	CSR1 Phosphatidylinositol transfer protein CSR1 (Ashbya gossypii (strain ATCC 10895 / CBS 109.51 / FGSC 9923 / NRRL Y-1056) OX=284811)	not assigned
SLA-04_01585-RA	-4.25	Cyp4x1 Cytochrome P450 4X1 (Rattus norvegicus OX=10116)	Enzyme classification
SLA-04_02716-RA	-4.25	fabD Malonyl CoA-acyl carrier protein transacylase (Bacillus subtilis (strain 168) OX=224308)	Lipid metabolism
SLA-04_00555-RA	-4.25	fabZ 3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ (Bacillus velezensis (strain DSM 23117 / BGSC 10A6 / FZB42) OX=326423)	Lipid metabolism
SLA-04_05179-RA	-4.25	PSB27-1 Photosystem II repair protein PSB27-H1, chloroplastic (Arabidopsis thaliana OX=3702)	Photosynthesis
SLA-04_07251-RA	-4.2	CIN6 Beta-fructofuranosidase, insoluble isoenzyme 6 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_05771-RA	-4.15	At3g63540 Thylakoid lumenal 19 kDa protein, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned

SLA-04_05353-RA	-4.15	PETF Ferredoxin, chloroplastic (<i>Chlamydomonas reinhardtii</i> OX=3055)	Photosynthesis
SLA-04_01549-RA	-4.15	PPD5 PsbP domain-containing protein 5, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_01061-RA	-4.1	fdhC Probable formate transporter (<i>Methanothermobacter thermautotrophicus</i> OX=145262)	not assigned
SLA-04_04639-RA	-4.1	Osl_00941 GDT1-like protein 1, chloroplastic (<i>Oryza sativa</i> subsp. <i>indica</i> OX=39946)	Solute transport
SLA-04_06791-RA	-4.1	Triosephosphate isomerase, chloroplastic (<i>Secale cereale</i> OX=4550)	Enzyme classification
SLA-04_07203-RA	-4.05	MGD A Probable monogalactosyldiacylglycerol synthase, chloroplastic (<i>Glycine max</i> OX=3847)	Lipid metabolism
SLA-04_05593-RA	-4.05	TIC62 Protein TIC 62, chloroplastic (<i>Pisum sativum</i> OX=3888)	Photosynthesis
SLA-04_01928-RA	-4	SECY Preprotein translocase subunit SECY, chloroplastic (<i>Pisum sativum</i> OX=3888)	Protein translocation
SLA-04_07391-RA	-3.95	HEME2 Uroporphyrinogen decarboxylase 2, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Coenzyme metabolism
SLA-04_06507-RA	-3.95	rplF 50S ribosomal protein L6 (<i>Gloeotheca citriformis</i> (strain PCC 7424) OX=65393)	Protein biosynthesis
SLA-04_02363-RA	-3.9	Setd4 SET domain-containing protein 4 (<i>Mus musculus</i> OX=10090)	not assigned
SLA-04_03544-RA	-3.85	PGK Phosphoglycerate kinase, chloroplastic (<i>Volvox carteri</i> OX=3067)	not assigned
SLA-04_01355-RA	-3.8	slI0418 2-methyl-6-phytyl-1,4-hydroquinone methyltransferase (<i>Synechocystis</i> sp. (strain PCC 6803 / <i>Kazusa</i>) OX=1111708)	not assigned
SLA-04_02273-RA	-3.75	PPD1 PsbP domain-containing protein 1, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_04404-RA	-3.75	PPL1 PsbP-like protein 1, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Photosynthesis
SLA-04_06422-RA	-3.6	ppa1 Soluble inorganic pyrophosphatase 1, chloroplastic (<i>Chlamydomonas reinhardtii</i> OX=3055)	Multi-process regulation
SLA-04_00455-RA	-3.6	CPX Oxygen-dependent coproporphyrinogen-III oxidase, chloroplastic (<i>Glycine max</i> OX=3847)	Coenzyme metabolism
SLA-04_01971-RA	-3.6	DVR Divinyl chlorophyllide a 8-vinyl-reductase, chloroplastic (<i>Cucumis sativus</i> OX=3659)	Coenzyme metabolism
SLA-04_08354-RA	-3.55	HEMC Porphobilinogen deaminase, chloroplastic (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Coenzyme metabolism
SLA-04_00646-RA	-3.55	Os09g0505700 Ribulose-phosphate 3-epimerase, cytoplasmic isoform (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Carbohydrate metabolism
SLA-04_05938-RA	-3.5	FRAAL3718 Putative S-adenosyl-L-methionine-dependent methyltransferase FRAAL3718 (<i>Frankia alni</i> (strain ACN14a) OX=326424)	not assigned
SLA-04_01993-RA	-3.5	IMPL1 Phosphatase IMPL1, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_06896-RA	-3.5	TATA Sec-independent protein translocase protein TATA, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Protein translocation
SLA-04_01799-RA	-3.45	FFC Signal recognition particle 54 kDa protein, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Protein translocation
SLA-04_06400-RA	-3.45	GDCST Aminomethyltransferase, mitochondrial (<i>Mesembryanthemum crystallinum</i> OX=3544)	Photosynthesis
SLA-04_06077-RA	-3.4	CHLM Magnesium protoporphyrin IX methyltransferase, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Coenzyme metabolism
SLA-04_06097-RC	-3.4	PORA Protochlorophyllide reductase, chloroplastic (<i>Cucumis sativus</i> OX=3659)	Coenzyme metabolism
SLA-04_03264-RA	-3.4	CPX Oxygen-dependent coproporphyrinogen-III oxidase, chloroplastic (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Enzyme classification
SLA-04_07504-RA	-3.35	CRTISO1 Prolycopene isomerase 1, chloroplastic (<i>Oncidium hybrid</i> cultivar OX=141207)	not assigned
SLA-04_00946-RA	-3.35	AKR2A Ankyrin repeat domain-containing protein 2A (<i>Arabidopsis thaliana</i> OX=3702)	Protein translocation
SLA-04_06404-RA	-3.3	ctpA C-terminal processing peptidase, chloroplastic (<i>Tetrademus obliquus</i> OX=3088)	not assigned

SLA-04_01860-RA	-3.25	CLEB3J9 Thylakoid lumenal 29 kDa protein, chloroplastic (Solanum lycopersicum OX=4081)	not assigned
SLA-04_01463-RA	-3.25	PAT1 Anthranilate phosphoribosyltransferase, chloroplastic (Arabidopsis thaliana OX=3702)	Amino acid metabolism
SLA-04_02504-RA	-3.25	slI1737 Tic20 family protein (Synechocystis sp. (strain PCC 6803 / Kazusa) OX=1111708)	Protein translocation
SLA-04_08090-RA	-3.2	Fructose-bisphosphate aldolase, cytoplasmic isozyme (Spinacia oleracea OX=3562)	not assigned
SLA-04_07769-RA	-3.2	GAPA Glyceraldehyde-3-phosphate dehydrogenase A, chloroplastic (Spinacia oleracea OX=3562)	Photosynthesis
SLA-04_02169-RA	-3.2	TRPA1 Tryptophan synthase alpha chain (Arabidopsis thaliana OX=3702)	Amino acid metabolism
SLA-04_00800-RA	-3.15	LHCB5 Chlorophyll a-b binding protein CP26, chloroplastic (Arabidopsis thaliana OX=3702)	Photosynthesis
SLA-04_07646-RA	-3.15	PETE Plastocyanin, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Photosynthesis
SLA-04_06521-RA	-3.15	asd Aspartate-semialdehyde dehydrogenase (Aquifex aeolicus (strain VF5) OX=224324)	Amino acid metabolism
SLA-04_06522-RA	-3.15	asd Aspartate-semialdehyde dehydrogenase (Synechocystis sp. (strain PCC 6803 / Kazusa) OX=1111708)	Amino acid metabolism
SLA-04_01182-RA	-3.15	DIT2-1 Dicarboxylate transporter 2.1, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_01067-RA	-3.15	HST Homogentisate solanesyltransferase, chloroplastic (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_01679-RA	-3.1	VTE5 Phytol kinase 1, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01990-RA	-3.05	At2g44920 Thylakoid lumenal 15 kDa protein 1, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01680-RA	-3.05	VPS4 Vacuolar protein sorting-associated protein 4 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)	not assigned
SLA-04_01227-RA	-3	ABA4 Protein ABA DEFICIENT 4, chloroplastic (Arabidopsis thaliana OX=3702)	Secondary metabolism
SLA-04_06151-RA	-3	ALDCHL Fructose-bisphosphate aldolase 1, chloroplastic (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_01589-RA	-3	Caur_2266 Succinyl-CoA--D-citramalate CoA-transferase (Chloroflexus aurantiacus (strain ATCC 29366 / DSM 635 / J-10-fl) OX=324602)	not assigned
SLA-04_05621-RA	-3	CEST Protein CHLOROPLAST ENHANCING STRESS TOLERANCE, chloroplastic (Oryza sativa subsp. indica OX=39946)	Photosynthesis
SLA-04_01502-RA	-3	efp Elongation factor P (Synechococcus sp. (strain JA-2-3B'a(2-13)) OX=321332)	Protein biosynthesis
SLA-04_03443-RA	-2.95	alkJ Alcohol dehydrogenase [acceptor] (Pseudomonas oleovorans OX=301)	not assigned
SLA-04_03827-RA	-2.95	PRXIIIE Peroxiredoxin-2E, chloroplastic (Arabidopsis thaliana OX=3702)	Redox homeostasis
SLA-04_01497-RA	-2.95	PSBP Oxygen-evolving enhancer protein 2, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Photosynthesis
SLA-04_06003-RA	-2.9	CYP26-2 Peptidyl-prolyl cis-trans isomerase CYP26-2, chloroplastic (Arabidopsis thaliana OX=3702)	Secondary metabolism
SLA-04_00224-RA	-2.85	CHLD Magnesium-chelatase subunit ChlD, chloroplastic (Pisum sativum OX=3888)	Coenzyme metabolism
SLA-04_07137-RA	-2.8	CRTZ Beta-carotene 3-hydroxylase, chloroplastic (Fragment) (Haematococcus lacustris OX=44745)	Secondary metabolism
SLA-04_01528-RA	-2.8	PSAO Photosystem I subunit O (Arabidopsis thaliana OX=3702)	Photosynthesis
SLA-04_01551-RA	-2.75	At2g31400 Pentatricopeptide repeat-containing protein At2g31400, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04973-RA	-2.75	PSAN Photosystem I reaction center subunit N, chloroplastic (Volvox carteri OX=3067)	Photosynthesis
SLA-04_08148-RA	-2.75	UROS Uroporphyrinogen-III synthase, chloroplastic (Arabidopsis thaliana OX=3702)	Coenzyme metabolism
SLA-04_06154-RA	-2.65	LIL3.1 Light-harvesting complex-like protein 3 isotype 1, chloroplastic (Arabidopsis thaliana OX=3702)	Photosynthesis

SLA-04_00647-RA	-2.65	PNSL5 Photosynthetic NDH subunit of lumenal location 5, chloroplastic (Arabidopsis thaliana OX=3702)	Enzyme classification
SLA-04_08213-RA	-2.65	CPX Oxygen-dependent coproporphyrinogen-III oxidase, chloroplastic (Glycine max OX=3847)	Coenzyme metabolism
SLA-04_03847-RA	-2.6	DDB_G0271742 UPF0548 protein (Dictyostelium discoideum OX=44689)	Solute transport
SLA-04_02007-RA	-2.55	bioC Malonyl-[acyl-carrier protein] O-methyltransferase (Paenibacillus sp. (strain JDR-2) OX=324057)	not assigned
SLA-04_06175-RA	-2.55	GSA Glutamate-1-semialdehyde 2,1-aminomutase, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Coenzyme metabolism
SLA-04_00452-RA	-2.55	YLMG2 YlmG homolog protein 2, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04611-RA	-2.5	CAS Calcium sensing receptor, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02904-RA	-2.5	Chlorophyll a-b binding protein of LHCII type I, chloroplastic (Lemna gibba OX=4470)	Photosynthesis
SLA-04_04193-RA	-2.45	CTH1 Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase 2, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Coenzyme metabolism
SLA-04_00828-RA	-2.45	LHCA1 Chlorophyll a-b binding protein 6, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_06070-RA	-2.45	UPP Uracil phosphoribosyltransferase (Nicotiana tabacum OX=4097)	Nucleotide metabolism
SLA-04_06540-RA	-2.4	ALB3.1 Inner membrane ALBINO3-like protein 1, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Protein translocation
SLA-04_04590-RA	-2.4	Dclre1a DNA cross-link repair 1A protein (Mus musculus OX=10090)	not assigned
SLA-04_00454-RB	-2.4	LTD Protein LHCP TRANSLOCATION DEFECT (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01567-RA	-2.4	(S)-coclaurine N-methyltransferase (Thalictrum flavum subsp. glaucum OX=150095)	not assigned
SLA-04_06267-RA	-2.4	chlI Magnesium-chelatase subunit ChlI (Cyanophora paradoxa OX=2762)	Coenzyme metabolism
SLA-04_03645-RA	-2.4	CHLP Geranylgeranyl diphosphate reductase, chloroplastic (Arabidopsis thaliana OX=3702)	Coenzyme metabolism
SLA-04_07705-RA	-2.4	Lhcb4 Chlorophyll a-b binding protein CP29 (Chlamydomonas reinhardtii OX=3055)	Photosynthesis
SLA-04_07643-RA	-2.35	CYN Cyanate hydratase (Phaeodactylum tricornutum (strain CCAP 1055/1) OX=556484)	Amino acid metabolism
SLA-04_01335-RA	-2.35	RCABP89 Chlorophyll a-b binding protein, chloroplastic (Oryza sativa subsp. japonica OX=39947)	Photosynthesis
SLA-04_02486-RA	-2.35	Chlorophyll a-b binding protein, chloroplastic (Petunia hybrida OX=4102)	not assigned
SLA-04_02410-RA	-2.35	PSAH Photosystem I reaction center subunit VI, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Photosynthesis
SLA-04_03243-RA	-2.3	HEMB Delta-aminolevulinic acid dehydratase, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Enzyme classification
SLA-04_01343-RA	-2.3	LHCA4 Chlorophyll a-b binding protein 4, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02485-RA	-2.3	LHCA4 Chlorophyll a-b binding protein 4, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02422-RA	-2.3	LUT2 Lycopene epsilon cyclase, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_07400-RA	-2.25	CAB8 Chlorophyll a-b binding protein 8, chloroplastic (Solanum lycopersicum OX=4081)	Photosynthesis
SLA-04_01827-RA	-2.25	PSBY Photosystem II core complex proteins psbY, chloroplastic (Spinacia oleracea OX=3562)	Photosynthesis
SLA-04_02375-RA	-2.2	ageIM Modification methylase AgeI (Thalassobius gelatinovorus OX=53501)	not assigned
SLA-04_06216-RA	-2.2	CHLH Magnesium-chelatase subunit ChlH, chloroplastic (Oryza sativa subsp. indica OX=39946)	Coenzyme metabolism
SLA-04_01075-RA	-2.15	CSP4 Cold shock domain-containing protein 4 (Arabidopsis thaliana OX=3702)	External stimuli response

SLA-04_00400-RA	-2.15	LHCA5 Photosystem I chlorophyll a/b-binding protein 5, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01786-RA	-2.15	PPOX1 Protoporphyrinogen oxidase, chloroplastic (Oryza sativa subsp. japonica OX=39947)	Coenzyme metabolism
SLA-04_02757-RA	-2.15	hemE Uroporphyrinogen decarboxylase (Synechococcus sp. (strain ATCC 27264 / PCC 7002 / PR-6) OX=32049)	Coenzyme metabolism
SLA-04_06403-RA	-2.15	STR11 Rhodanese-like domain-containing protein 11, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02993-RA	-2.1	PSAG Photosystem I reaction center subunit V, chloroplastic (Arabidopsis thaliana OX=3702)	Photosynthesis
SLA-04_01097-RA	2.1	ASZ1 Ankyrin repeat, SAM and basic leucine zipper domain-containing protein 1 (Neofelis nebulosa OX=61452)	not assigned
SLA-04_03102-RA	2.35	DGAT2A Diacylglycerol O-acyltransferase 2A (Umbelopsis ramanniana OX=41833)	not assigned
SLA-04_05737-RA	2.4	exo1 Exodeoxyribonuclease 1 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)	not assigned
SLA-04_00425-RA	2.4	ZEU1 Thymidylate kinase (Arabidopsis thaliana OX=3702)	Nucleotide metabolism
SLA-04_02881-RA	2.4	erkB Extracellular signal-regulated kinase 2 (Dictyostelium discoideum OX=44689)	Multi-process regulation
SLA-04_07489-RA	2.45	At1g16220 Probable protein phosphatase 2C 6 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02555-RA	2.5	HLJ1 Protein HLJ1 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)	not assigned
SLA-04_03372-RA	2.5	ODA7 Leucine-rich repeat-containing protein ODA7 (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_05523-RA	2.5	RCL1 RNA 3'-terminal phosphate cyclase-like protein (Bos taurus OX=9913)	not assigned
SLA-04_00908-RA	2.55	AURKA Aurora kinase A (Bos taurus OX=9913)	Enzyme classification
SLA-04_06780-RA	2.55	RAD54 DNA repair and recombination protein RAD54 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_02323-RA	2.55	SPBC1711.16 Uncharacterized WD repeat-containing protein C17D11.16 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)	not assigned
SLA-04_02512-RA	2.65	SPAC3F10.06c Uncharacterized protein C3F10.06c (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)	not assigned
SLA-04_00412-RA	2.65	RH20 DEAD-box ATP-dependent RNA helicase 20 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_05674-RA	2.65	SAT17 Taurine hydroxylase-like protein SAT17 (Stachybotrys chartarum (strain CBS 109288 / IBT 7711) OX=1280523)	not assigned
SLA-04_07089-RA	2.75	obr1 P25 protein (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)	not assigned
SLA-04_02691-RA	2.75	RFS2 Probable galactinol--sucrose galactosyltransferase 2 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04752-RA	2.8	Os03g0728800 DEAD-box ATP-dependent RNA helicase 16 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_06935-RA	2.8	pus10 Putative tRNA pseudouridine synthase Pus10 (Xenopus laevis OX=8355)	RNA processing
SLA-04_05716-RA	2.85	rumi O-glucosyltransferase rumi (Drosophila pseudoobscura pseudoobscura OX=46245)	not assigned
SLA-04_06254-RA	2.9	atg26 Sterol 3-beta-glucosyltransferase (Aspergillus clavatus (strain ATCC 1007 / CBS 513.65 / DSM 816 / NCTC 3887 / NRRL 1) OX=344612)	not assigned
SLA-04_03497-RA	2.9	EMB2001 GTP-binding protein At2g22870 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02094-RA	2.9	Tbc1d2 TBC1 domain family member 2A (Mus musculus OX=10090)	not assigned
SLA-04_01706-RA	2.95	SGT1 Peptidyl serine alpha-galactosyltransferase (Chlamydomonas reinhardtii OX=3055)	Cell wall organisation
SLA-04_03447-RA	3	LDL1 Lysine-specific histone demethylase 1 homolog 1 (Arabidopsis thaliana OX=3702)	not assigned

SLA-04_05807-RA	3	RAD51D DNA repair protein RAD51 homolog 4 (Arabidopsis thaliana OX=3702)	DNA damage response
SLA-04_02624-RA	3	TRANK1 TPR and ankyrin repeat-containing protein 1 (Homo sapiens OX=9606)	not assigned
SLA-04_04262-RA	3.1	Wdr78 WD repeat-containing protein 78 (Rattus norvegicus OX=10116)	not assigned
SLA-04_05965-RA	3.15	ackA Acetate kinase (Desulfovibrio alaskensis (strain G20) OX=207559)	not assigned
SLA-04_03340-RA	3.15	DDB_G0277071 Probable serine/threonine-protein kinase DDB_G0277071 (Dictyostelium discoideum OX=44689)	Enzyme classification
SLA-04_05290-RA	3.15	Fam221b Protein FAM221B (Rattus norvegicus OX=10116)	not assigned
SLA-04_05719-RA	3.2	AVT6E Amino acid transporter AVT6E (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02226-RA	3.2	FSM Chromatin assembly factor 1 subunit FSM (Oryza sativa subsp. japonica OX=39947)	Chromatin organisation
SLA-04_06790-RA	3.2	ORC4 Origin of replication complex subunit 4 (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_05195-RA	3.2	POLE2 DNA polymerase epsilon subunit 2 (Gallus gallus OX=9031)	not assigned
SLA-04_00948-RA	3.25	At1g60750 Probable aldo-keto reductase 6 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_08285-RA	3.25	FAP3 Fatty-acid-binding protein 3, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04267-RA	3.25	pfl Formate acetyltransferase (Clostridium pasteurianum OX=1501)	not assigned
SLA-04_05194-RA	3.3	DPB2 DNA polymerase epsilon subunit B (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_01577-RA	3.3	HAC1 Protein HIGH ARSENIC CONTENT 1, mitochondrial (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04954-RA	3.3	TOUSLED Serine/threonine-protein kinase TOUSLED (Arabidopsis thaliana OX=3702)	Protein modification
SLA-04_07685-RA	3.35	ANP2 Mitogen-activated protein kinase kinase kinase 2 (Arabidopsis thaliana OX=3702)	Enzyme classification
SLA-04_08008-RA	3.35	ROA2 DNA replication licensing factor MCM3 homolog 2 (Zea mays OX=4577)	Cell division
SLA-04_01411-RA	3.35	mad2l1-1 Mitotic spindle assembly checkpoint protein MAD2A (Dictyostelium discoideum OX=44689)	not assigned
SLA-04_00799-RA	3.45	hndD NADP-reducing hydrogenase subunit HndD (Desulfovibrio fructosivorans OX=878)	not assigned
SLA-04_01333-RA	3.5	BTR1 Protein BTR1 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_00747-RA	3.5	HSP17.6-L 17.6 kDa class I heat shock protein (Glycine max OX=3847)	not assigned
SLA-04_06343-RA	3.5	ODA9 Dynein, 78 kDa intermediate chain, flagellar outer arm (Chlamydomonas reinhardtii OX=3055)	Cytoskeleton organisation
SLA-04_07765-RA	3.55	GLX1 Lactoylglutathione lyase GLX1 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01515-RA	3.55	NDC80 Kinetochore protein NDC80 homolog (Gallus gallus OX=9031)	Cell division
SLA-04_02206-RA	3.55	GDHA Glutamate dehydrogenase A (Nicotiana glauca OX=4092)	Nutrient uptake
SLA-04_04090-RA	3.6	DFR Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase (Pyrus communis OX=23211)	not assigned
SLA-04_06383-RA	3.6	KIN12F Kinesin-like protein KIN-12F (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_01460-RA	3.6	Osl_36121 DNA replication licensing factor MCM2 (Oryza sativa subsp. indica OX=39946)	Cell division
SLA-04_00840-RA	3.65	MCM7 DNA replication licensing factor MCM7 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_05798-RA	3.7	NHX6 Sodium/hydrogen exchanger 6 (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_01940-RA	3.75	DDB_G0291133 Probable protein kinase DDB_G0291133 (Dictyostelium discoideum OX=44689)	Enzyme classification

SLA-04_02556-RA	3.8	GOLS2 Galactinol synthase 2 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_05191-RA	3.85	Os05g0235800 DNA replication licensing factor MCM6 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_05975-RA	3.85	rnaseh2b Ribonuclease H2 subunit B (Xenopus laevis OX=8355)	Cell division
SLA-04_01393-RA	3.9	Os01g0868300 DNA polymerase alpha catalytic subunit (Oryza sativa subsp. japonica OX=39947)	Cell division
SLA-04_02148-RA	3.9	SUVH4 Histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH4 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_03446-RA	3.95	NIC2 Nicotinamidase 2 (Arabidopsis thaliana OX=3702)	Coenzyme metabolism
SLA-04_05190-RA	3.95	MCM6 DNA replication licensing factor MCM6 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_08326-RA	3.95	SMC2-1 Structural maintenance of chromosomes protein 2-1 (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_07143-RA	4	At4g14490 FHA domain-containing protein At4g14490 (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_07796-RA	4.05	MYB3R5 Transcription factor MYB3R-5 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_03318-RA	4.05	SQD2 Sulfoquinovosyl transferase SQD2 (Arabidopsis thaliana OX=3702)	Lipid metabolism
SLA-04_05478-RA	4.05	CBSX3 CBS domain-containing protein CBSX3, mitochondrial (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02134-RA	4.15	CPK1 Calcium-dependent protein kinase 1 (Arabidopsis thaliana OX=3702)	Enzyme classification
SLA-04_00798-RA	4.15	NAD11 NADH-ubiquinone oxidoreductase 75 kDa subunit (Reclinomonas americana OX=48483)	not assigned
SLA-04_02009-RA	4.25	AZG1 Adenine/guanine permease AZG1 (Arabidopsis thaliana OX=3702)	Nucleotide metabolism
SLA-04_00766-RA	4.25	CHTF8 Chromosome transmission fidelity protein 8 homolog (Homo sapiens OX=9606)	not assigned
SLA-04_07342-RA	4.25	PXG1 Peroxygenase 1 (Arabidopsis thaliana OX=3702)	Lipid metabolism
SLA-04_02503-RA	4.3	MCM5 DNA replication licensing factor MCM5 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_05159-RA	4.35	ARC5 Dynamin-like protein ARC5 (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_01293-RA	4.4	CDCA7L Cell division cycle-associated 7-like protein (Homo sapiens OX=9606)	not assigned
SLA-04_03371-RA	4.4	ODA7 Leucine-rich repeat-containing protein ODA7 (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_04145-RA	4.4	unknown function	RNA processing
SLA-04_05443-RA	4.45	POL2A DNA polymerase epsilon catalytic subunit A (Arabidopsis thaliana OX=3702)	Cell division
SLA-04_03860-RA	4.5	incenp-b Inner centromere protein B (Xenopus laevis OX=8355)	not assigned
SLA-04_01459-RA	4.5	WEE1 Wee1-like protein kinase (Oryza sativa subsp. japonica OX=39947)	Cell division
SLA-04_07760-RA	4.55	NAT2 Nucleobase-ascorbate transporter 2 (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_01727-RA	4.6	NEK3 Serine/threonine-protein kinase Nek3 (Homo sapiens OX=9606)	Enzyme classification
SLA-04_00142-RA	4.8	alxA Alternative oxidase, mitochondrial (Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) OX=227321)	not assigned
SLA-04_03910-RA	4.9	MFSD6 Major facilitator superfamily domain-containing protein 6 (Sus scrofa OX=9823)	not assigned
SLA-04_02667-RA	4.95	CDC45 Cell division control protein 45 homolog (Homo sapiens OX=9606)	Cell division
SLA-04_07268-RA	5.15	AAO2 Indole-3-acetaldehyde oxidase (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02138-RA	5.3	MSD1 Superoxide dismutase [Mn] 1, mitochondrial (Arabidopsis thaliana OX=3702)	Redox homeostasis

SLA-04_08060-RA	5.35	CURT1C Protein CURVATURE THYLAKOID 1C, chloroplastic (Arabidopsis thaliana OX=3702)	Photosynthesis
SLA-04_02943-RA	5.55	TPH1 Tryptophan 5-hydroxylase 1 (Oryctolagus cuniculus OX=9986)	Amino acid metabolism
SLA-04_06583-RA	5.65	BT_0258 Putative glucosamine-6-phosphate deaminase-like protein BT_0258 (Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582	not assigned
SLA-04_01056-RA	5.7	uapA Uric acid-xanthine permease (Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) OX=227321)	not assigned
SLA-04_00479-RA	5.8	CCD8 Carotenoid cleavage dioxygenase 8, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_03252-RA	5.85	SKOR Potassium channel SKOR (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_08414-RA	5.95	ODA2 Dynein gamma chain, flagellar outer arm (Chlamydomonas reinhardtii OX=3055)	Cytoskeleton organisation
SLA-04_01975-RA	7.65	BAT1 Amino-acid permease BAT1 (Arabidopsis thaliana OX=3702)	Solute transport

Supplementary Table 3: Table of all differential genes between neutral versus alkaline

Gene ID	Average DGE	Annotation	Functional Category
SLA-04_05062-RA	-12	GINT1 Glucosamine inositolphosphorylceramide transferase 1 (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Lipid metabolism
SLA-04_02510-RA	-10.55	FRAS1 Extracellular matrix protein FRAS1 (<i>Homo sapiens</i> OX=9606)	not assigned
SLA-04_03426-RA	-10.5	SULTR2 Proton/sulfate cotransporter 2 (<i>Chlamydomonas reinhardtii</i> OX=3055)	Solute transport
SLA-04_07142-RA	-10.35	NRAMP2 Metal transporter Nramp2 (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	not assigned
SLA-04_03075-RA	-9.55	CID11 Polyadenylate-binding protein-interacting protein 11 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_01350-RA	-9	ftt1 High affinity iron permease 1 (<i>Rhizopus delemar</i> (strain RA 99-880 / ATCC MYA-4621 / FGSC 9543 / NRRL 43880) OX=246409)	Nutrient uptake
SLA-04_06838-RA	-8.6	Hephaestin-like protein (<i>Acropora millepora</i> OX=45264)	Nutrient uptake
SLA-04_04453-RA	-7.7	PTPRJ Receptor-type tyrosine-protein phosphatase eta (<i>Gallus gallus</i> OX=9031)	not assigned
SLA-04_00948-RA	-7.4	At1g60750 Probable aldo-keto reductase 6 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_07701-RA	-7	rtn4ip1 Reticulon-4-interacting protein 1 homolog, mitochondrial (<i>Danio rerio</i> OX=7955)	not assigned
SLA-04_08285-RA	-6.95	FAP3 Fatty-acid-binding protein 3, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_07612-RA	-6.65	mtcA1 Beta-carbonic anhydrase 1 (<i>Mycobacterium tuberculosis</i> (strain ATCC 25618 / H37Rv) OX=83332)	Photosynthesis
SLA-04_00007-RA	-6.5	APK3 Adenylyl-sulfate kinase 3 (<i>Arabidopsis thaliana</i> OX=3702)	Coenzyme metabolism
SLA-04_02085-RA	-6.45	BASS4 Probable sodium/metabolite cotransporter BASS4, chloroplastic (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Solute transport
SLA-04_03940-RA	-6.25	Wnk4 Serine/threonine-protein kinase WNK4 (<i>Mus musculus</i> OX=10090)	Enzyme classification
SLA-04_07913-RA	-6.15	lsd18 Putative epoxidase LasC (<i>Streptomyces lasaliensis</i> OX=324833)	not assigned
SLA-04_03422-RA	-5.9	CLIPB4 CLIP domain-containing serine protease 14D (<i>Anopheles gambiae</i> OX=7165)	not assigned
SLA-04_02308-RA	-5.2	GIT1 Glycerophosphoinositol permease 1 (<i>Candida albicans</i> (strain SC5314 / ATCC MYA-2876) OX=237561)	Solute transport
SLA-04_01551-RA	-5.1	At2g31400 Pentatricopeptide repeat-containing protein At2g31400, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_01782-RA	-5.05	DMBT1 Deleted in malignant brain tumors 1 protein (<i>Homo sapiens</i> OX=9606)	Enzyme classification
SLA-04_03452-RA	-5.05	pho-4 Phosphate-repressible phosphate permease pho-4 (<i>Neurospora crassa</i> (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=36)	Solute transport
SLA-04_06237-RA	-4.55	BC10 Glycosyltransferase BC10 (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	not assigned
SLA-04_08428-RA	-4.35	ZIP3 Zinc transporter 3 (<i>Oryza sativa</i> subsp. <i>japonica</i> OX=39947)	Solute transport
SLA-04_02956-RA	-4.25	AMT1-1 Ammonium transporter 1 member 1 (<i>Arabidopsis thaliana</i> OX=3702)	Solute transport
SLA-04_00872-RA	-4.25	BZ1 Anthocyanidin 3-O-glucosyltransferase (<i>Zea mays</i> OX=4577)	Enzyme classification
SLA-04_01056-RA	-3.95	uapA Uric acid-xanthine permease (<i>Emericella nidulans</i> (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) OX=227321)	not assigned
SLA-04_02185-RA	-3.9	FMO1 Probable flavin-containing monooxygenase 1 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_07121-RA	-3.9	NUP58 Nuclear pore complex protein NUP58 (<i>Arabidopsis thaliana</i> OX=3702)	Protein homeostasis
SLA-04_03834-RA	-3.65	hcp Hydroxylamine reductase (<i>Burkholderia vietnamiensis</i> (strain G4 / LMG 22486) OX=269482)	not assigned
SLA-04_01728-RA	-3.65	Nek1 Serine/threonine-protein kinase Nek1 (<i>Mus musculus</i> OX=10090)	Enzyme classification
SLA-04_03444-RB	-3.6	p80 Protein P80 (<i>Dictyostelium discoideum</i> OX=44689)	not assigned

SLA-04_03833-RA	-3.2	hcp Hydroxylamine reductase (Shewanella amazonensis (strain ATCC BAA-1098 / SB2B) OX=326297)	not assigned
SLA-04_02218-RA	-3.2	P4H7 Probable prolyl 4-hydroxylase 7 (Arabidopsis thaliana OX=3702)	Protein modification
SLA-04_00199-RA	-3.05	AIL5 AP2-like ethylene-responsive transcription factor AIL5 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_06909-RA	-3.05	drkA Probable serine/threonine-protein kinase drkA (Dictyostelium discoideum OX=44689)	Multi-process regulation
SLA-04_01080-RA	-2.95	slc38a7 Putative sodium-coupled neutral amino acid transporter 7 (Xenopus laevis OX=8355)	not assigned
SLA-04_00254-RA	-2.8	ANK1 Ankyrin-1 (Homo sapiens OX=9606)	not assigned
SLA-04_05996-RA	-2.8	PGDH1 D-3-phosphoglycerate dehydrogenase 1, chloroplastic (Arabidopsis thaliana OX=3702)	Amino acid metabolism
SLA-04_07568-RA	-2.75	MAP3K11 Mitogen-activated protein kinase kinase kinase 11 (Homo sapiens OX=9606)	Enzyme classification
SLA-04_05716-RA	-2.7	rumi O-glucosyltransferase rumi (Drosophila pseudoobscura pseudoobscura OX=46245)	not assigned
SLA-04_02849-RA	-2.65	BOU Mitochondrial carnitine/acylcarnitine carrier-like protein (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_03703-RA	-2.55	SYD Chromatin structure-remodeling complex protein SYD (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_07566-RA	-2.5	aq_1088 Uncharacterized protein aq_1088 (Aquifex aeolicus (strain VF5) OX=224324)	not assigned
SLA-04_07963-RA	-2.4	Slc8b1 Mitochondrial sodium/calcium exchanger protein (Mus musculus OX=10090)	Solute transport
SLA-04_04137-RA	-2.35	PTAC14 Protein PLASTID TRANSCRIPTIONALLY ACTIVE 14 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_05677-RA	-2.25	Rad54l DNA repair and recombination protein RAD54-like (Mus musculus OX=10090)	not assigned
SLA-04_04397-RB	-2.05	AP17 AP-2 complex subunit sigma (Arabidopsis thaliana OX=3702)	Protein modification
SLA-04_07786-RA	2.1	APUM3 Pumilio homolog 3 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_07002-RA	2.15	BUB1 Mitotic checkpoint serine/threonine-protein kinase BUB1 (Arabidopsis thaliana OX=3702)	Enzyme classification
SLA-04_06950-RA	2.15	SETH3 Probable arabinose 5-phosphate isomerase (Arabidopsis thaliana OX=3702)	Cell wall organisation
SLA-04_04766-RA	2.2	DDB_G0281745 Probable serine/threonine-protein kinase DDB_G0281745 (Dictyostelium discoideum OX=44689)	Enzyme classification
SLA-04_01383-RA	2.2	PROC1 Pyrroline-5-carboxylate reductase (Arabidopsis thaliana OX=3702)	Amino acid metabolism
SLA-04_03964-RA	2.25	PRY1 Protein PRY1 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)	Phytohormone action
SLA-04_03922-RA	2.25	unknown function	Solute transport
SLA-04_03802-RA	2.3	At3g26670 Probable magnesium transporter NIPA8 (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_06811-RA	2.3	ddiA Ribonuclease Ddi (Dictyostelium discoideum OX=44689)	RNA processing
SLA-04_04636-RB	2.3	SPC25 Kinetochore protein Spc25 (Bos taurus OX=9913)	Cell division
SLA-04_01043-RA	2.35	gap3 Glyceraldehyde-3-phosphate dehydrogenase 3 (Nostoc sp. (strain PCC 7120 / SAG 25.82 / UTEX 2576) OX=103690)	Enzyme classification
SLA-04_07031-RA	2.35	PGPS2 CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase 2 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01097-RA	2.4	ASZ1 Ankyrin repeat, SAM and basic leucine zipper domain-containing protein 1 (Neofelis nebulosa OX=61452)	not assigned
SLA-04_04347-RA	2.4	Epn2 Epsin-2 (Rattus norvegicus OX=10116)	not assigned
SLA-04_05060-RA	2.45	CAO Chlorophyllide a oxygenase, chloroplastic (Chlamydomonas reinhardtii OX=3055)	Coenzyme metabolism
SLA-04_04596-RA	2.45	KIN7D Kinesin-like protein KIN-7D, mitochondrial (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_06554-RA	2.5	hh Protein hedgehog (Drosophila virilis OX=7244)	not assigned

SLA-04_06641-RA	2.5	SBP2 Squamosa promoter-binding protein 2 (<i>Antirrhinum majus</i> OX=4151)	RNA biosynthesis
SLA-04_01558-RA	2.5	speA Biosynthetic arginine decarboxylase (<i>Nostoc</i> sp. (strain PCC 7120 / SAG 25.82 / UTEX 2576) OX=103690)	Polyamine metabolism
SLA-04_07382-RA	2.55	PDPK1 3-phosphoinositide-dependent protein kinase 1 (<i>Arabidopsis thaliana</i> OX=3702)	Protein modification
SLA-04_07755-RA	2.65	ODA1 Outer dynein arm protein 1 (<i>Chlamydomonas reinhardtii</i> OX=3055)	not assigned
SLA-04_06178-RA	2.7	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_06929-RA	2.8	At5g53320 Probable inactive receptor kinase At5g53320 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_02074-RA	2.85	BC10 Glycosyltransferase BC10 (<i>Oryza sativa</i> subsp. japonica OX=39947)	not assigned
SLA-04_04963-RA	2.85	bp26 26 kDa periplasmic immunogenic protein (<i>Brucella suis</i> biovar 1 (strain 1330) OX=204722)	not assigned
SLA-04_07120-RA	2.85	SERPINA3-1 Serpin A3-1 (<i>Bos taurus</i> OX=9913)	Protein homeostasis
SLA-04_01685-RA	2.85	TPPP2 Tubulin polymerization-promoting protein family member 2 (<i>Homo sapiens</i> OX=9606)	not assigned
SLA-04_00896-RA	2.9	GALT5 Hydroxyproline O-galactosyltransferase GALT5 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_05404-RA	2.9	At5g01750 Protein LURP-one-related 15 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_01240-RA	3	SPL4 Squamosa promoter-binding-like protein 4 (<i>Arabidopsis thaliana</i> OX=3702)	RNA biosynthesis
SLA-04_02442-RA	3	ugt3a1 UDP-glucuronosyltransferase 3A1 (<i>Xenopus laevis</i> OX=8355)	Enzyme classification
SLA-04_07489-RA	3.05	At1g16220 Probable protein phosphatase 2C 6 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_01243-RA	3.05	tagA Serine protease/ABC transporter B family protein tagA (<i>Dictyostelium discoideum</i> OX=44689)	not assigned
SLA-04_02721-RA	3.1	ecaA Carbonic anhydrase (<i>Nostoc</i> sp. (strain PCC 7120 / SAG 25.82 / UTEX 2576) OX=103690)	Photosynthesis
SLA-04_01931-RA	3.1	RECQL5 ATP-dependent DNA helicase Q-like 5 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_06499-RA	3.15	KIN7I Kinesin-like protein KIN-7I (<i>Oryza sativa</i> subsp. japonica OX=39947)	Cytoskeleton organisation
SLA-04_02581-RA	3.15	unknown function	Vesicle trafficking
SLA-04_08002-RA	3.2	Histone H4 (<i>Chelidonium majus</i> OX=71251)	Chromatin organisation
SLA-04_04070-RA	3.25	fabG 3-oxoacyl-[acyl-carrier-protein] reductase FabG (<i>Thermotoga maritima</i> (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099) OX=243274)	not assigned
SLA-04_07524-RA	3.25	GGP1 Gamma-glutamyl peptidase 1 (<i>Arabidopsis thaliana</i> OX=3702)	Protein homeostasis
SLA-04_04410-RA	3.25	SPATA20 Spermatogenesis-associated protein 20 (<i>Homo sapiens</i> OX=9606)	not assigned
SLA-04_02471-RA	3.3	CNR11 Cell number regulator 11 (<i>Zea mays</i> OX=4577)	not assigned
SLA-04_07102-RA	3.3	MRS2-7 Magnesium transporter MRS2-7 (<i>Arabidopsis thaliana</i> OX=3702)	Solute transport
SLA-04_04687-RA	3.3	ODC Ornithine decarboxylase (<i>Capsicum annuum</i> OX=4072)	Polyamine metabolism
SLA-04_05571-RA	3.35	SIS8 Probable serine/threonine-protein kinase SIS8 (<i>Arabidopsis thaliana</i> OX=3702)	Enzyme classification
SLA-04_07041-RA	3.35	ANK1 Ankyrin-1 (<i>Homo sapiens</i> OX=9606)	not assigned
SLA-04_07228-RA	3.35	omh1 O-glycoside alpha-1,2-mannosyltransferase omh1 (<i>Schizosaccharomyces pombe</i> (strain 972 / ATCC 24843) OX=284812)	not assigned
SLA-04_05947-RA	3.35	PHT4;4 Ascorbate transporter, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	Solute transport
SLA-04_02239-RA	3.35	PRSS12 Neurotrypsin (<i>Pongo pygmaeus</i> OX=9600)	not assigned
SLA-04_07012-RA	3.35	SKN1 Beta-glucan synthesis-associated protein SKN1 (<i>Candida albicans</i> OX=5476)	not assigned

SLA-04_08345-RA	3.4	TUB Tubby protein homolog (Homo sapiens OX=9606)	RNA biosynthesis
SLA-04_03261-RA	3.45	GATA12 GATA transcription factor 12 (Arabidopsis thaliana OX=3702)	RNA biosynthesis
SLA-04_03294-RA	3.5	PIN1A Auxin efflux carrier component 1a (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_03506-RA	3.55	FAD4L1 Fatty acid desaturase 4-like 1, chloroplastic (Arabidopsis thaliana OX=3702)	Lipid metabolism
SLA-04_02334-RA	3.55	MYB44 Transcription factor MYB44 (Arabidopsis thaliana OX=3702)	RNA biosynthesis
SLA-04_02077-RA	3.55	proB Glutamate 5-kinase (Acaryochloris marina (strain MBIC 11017) OX=329726)	not assigned
SLA-04_00858-RB	3.55	SEC Probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SEC (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_07931-RA	3.65	Irrc40 Leucine-rich repeat-containing protein 40 (Xenopus tropicalis OX=8364)	not assigned
SLA-04_03770-RA	3.65	SYNPCC7002_A0175 Uncharacterized protein SYNPCC7002_A0175 (Synechococcus sp. (strain ATCC 27264 / PCC 7002 / PR-6) OX=32049)	not assigned
SLA-04_06256-RA	3.85	FAM210B Protein FAM210B, mitochondrial (Homo sapiens OX=9606)	not assigned
SLA-04_08149-RA	3.9	Bckdk [3-methyl-2-oxobutanoate dehydrogenase [lipoamide]] kinase, mitochondrial (Mus musculus OX=10090)	Cellular respiration
SLA-04_06486-RA	3.9	At3g61320 UPF0187 protein At3g61320, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_00152-RA	3.9	CAS Calcium sensing receptor, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01995-RA	3.95	D-galactoside-specific lectin (Heliocidaris crassispina OX=1043166)	not assigned
SLA-04_02671-RA	3.95	PSBS2 Photosystem II protein PSBS2 (Chlamydomonas reinhardtii OX=3055)	not assigned
SLA-04_01706-RA	3.95	SGT1 Peptidyl serine alpha-galactosyltransferase (Chlamydomonas reinhardtii OX=3055)	Protein modification
SLA-04_07847-RA	4.05	At3g61320 UPF0187 protein At3g61320, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_01065-RA	4.05	MLO11 MLO-like protein 11 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_05419-RA	4.05	PF13_0198 Reticulocyte-binding protein 2 homolog a (Plasmodium falciparum (isolate 3D7) OX=36329)	not assigned
SLA-04_07424-RA	4.05	PSR2 Probable phosphatase PSR2 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)	Lipid metabolism
SLA-04_00319-RA	4.1	CG15309 Protein yippee-like CG15309 (Drosophila melanogaster OX=7227)	not assigned
SLA-04_01154-RB	4.15	ELI3 Probable mannitol dehydrogenase (Mesembryanthemum crystallinum OX=3544)	Enzyme classification
SLA-04_07047-RA	4.2	RPS20 40S ribosomal protein S20 (Macaca fascicularis OX=9541)	Protein biosynthesis
SLA-04_01692-RA	4.2	CBR Carotene biosynthesis-related protein CBR, chloroplastic (Dunaliella salina OX=3046)	Photosynthesis
SLA-04_01238-RA	4.2	Extracellular ribonuclease LE (Solanum lycopersicum OX=4081)	RNA processing
SLA-04_03923-RA	4.25	Os01g0917400 Zinc finger CCCH domain-containing protein 12 (Oryza sativa subsp. japonica OX=39947)	RNA processing
SLA-04_07960-RB	4.25	PDC2 Pyruvate decarboxylase 2 (Nicotiana tabacum OX=4097)	Enzyme classification
SLA-04_00807-RA	4.3	At3g06270 Probable protein phosphatase 2C 35 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_03192-RA	4.4	SERINC1 Serine incorporator 1 (Bos taurus OX=9913)	sphingolipid
SLA-04_07960-RA	4.5	PDC2 Pyruvate decarboxylase 2 (Fragment) (Pisum sativum OX=3888)	not assigned
SLA-04_05494-RA	4.55	COL5 Zinc finger protein CONSTANS-LIKE 5 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_04861-RA	4.55	ORLIKE Protein ORANGE-LIKE, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_02291-RA	4.6	Desiccation-related protein PCC13-62 (Craterostigma plantagineum OX=4153)	not assigned
SLA-04_00955-RA	4.65	BOR2 Probable boron transporter 2 (Arabidopsis thaliana OX=3702)	Solute transport

SLA-04_05164-RA	4.7	glcA Glucan endo-1,3-beta-glucosidase A1 (Bacillus circulans OX=1397)	not assigned
SLA-04_01626-RA	4.75	At2g45870 UPF0187 protein At2g45870, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_00878-RB	4.8	PD_1893 UPF0394 membrane protein PD_1893 (Xylella fastidiosa (strain Temecula1 / ATCC 700964) OX=183190)	not assigned
SLA-04_00220-RA	4.85	DCL1 Endoribonuclease Dicer homolog 1 (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_06355-RA	4.9	Barwin (Hordeum vulgare OX=4513)	not assigned
SLA-04_01309-RA	4.95	bgaB Beta-galactosidase bgaB (Geobacillus kaustophilus OX=1462)	not assigned
SLA-04_08116-RA	5	Ankrd6 Ankyrin repeat domain-containing protein 6 (Mus musculus OX=10090)	not assigned
SLA-04_05329-RA	5	Os09g0459600 Probable protein phosphatase 2C 69 (Oryza sativa subsp. japonica OX=39947)	not assigned
SLA-04_01632-RA	5.15	At3g61320 UPF0187 protein At3g61320, chloroplastic (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_08082-RA	5.25	At4g11730 Putative ATPase, plasma membrane-like (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_00008-RA	5.35	APK1 Adenylyl-sulfate kinase 1, chloroplastic (Arabidopsis thaliana OX=3702)	Coenzyme metabolism
SLA-04_01455-RB	5.35	NEK8 Serine/threonine-protein kinase Nek8 (Homo sapiens OX=9606)	not assigned
SLA-04_01139-RA	5.35	Thermitase (Thermoactinomyces vulgaris OX=2026)	protein degradation
SLA-04_06004-RA	5.4	CBR Carotene biosynthesis-related protein CBR, chloroplastic (Dunaliella salina OX=3046)	Photosynthesis
SLA-04_06618-RA	5.45	unknown function	Phytohormone action
SLA-04_07383-RA	5.55	FPV245 Putative ankyrin repeat protein FPV245 (Fowlpox virus (strain NVSL) OX=928301)	not assigned
SLA-04_03359-RA	5.55	LYPLA2 Acyl-protein thioesterase 2 (Homo sapiens OX=9606)	not assigned
SLA-04_02472-RA	5.55	PGL3 6-phosphogluconolactonase 3, chloroplastic (Arabidopsis thaliana OX=3702)	not assigned
SLA-04_01126-RA	5.8	ACOT9 Acyl-coenzyme A thioesterase 9, mitochondrial (Homo sapiens OX=9606)	not assigned
SLA-04_07341-RA	5.85	AQP5 Aquaporin-5 (Milnesium tardigradum OX=46460)	not assigned
SLA-04_01081-RA	5.85	ARB_07867 WSC domain-containing protein ARB_07867 (Arthroderma benhamiae (strain ATCC MYA-4681 / CBS 112371) OX=663331)	not assigned
SLA-04_03630-RA	5.85	Desiccation-related protein PCC13-62 (Craterostigma plantagineum OX=4153)	not assigned
SLA-04_02883-RA	5.85	ISA1 Isoamylase 1, chloroplastic (Arabidopsis thaliana OX=3702)	starch degradation
SLA-04_02661-RA	5.9	ABCG22 ABC transporter G family member 22 (Arabidopsis thaliana OX=3702)	Solute transport
SLA-04_04052-RA	5.9	p80 Protein P80 (Dictyostelium discoideum OX=44689)	not assigned
SLA-04_05575-RA	6	Histamine oxidase (Arthrobacter globiformis OX=1665)	Enzyme classification
SLA-04_05576-RA	6.05	AMO Peroxisomal primary amine oxidase (Pichia angusta OX=870730)	not assigned
SLA-04_08124-RA	6.1	CYP55A1 NADP nitrous oxide-forming nitric oxide reductase (Fusarium oxysporum OX=5507)	not assigned
SLA-04_05529-RA	6.1	H3b Histone H3.3 type b (Dictyostelium discoideum OX=44689)	not assigned
SLA-04_04001-RA	6.15	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (Nicotiana tabacum OX=4097)	not assigned
SLA-04_03817-RA	6.2	cdc14ab Dual specificity protein phosphatase CDC14AB (Danio rerio OX=7955)	DNA repair
SLA-04_08183-RA	6.2	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (Nicotiana tabacum OX=4097)	not assigned
SLA-04_05507-RA	6.25	sodA Superoxide dismutase [Mn] (Geobacillus stearothermophilus OX=1422)	not assigned
SLA-04_02931-RA	6.35	NADP-specific glutamate dehydrogenase (Fragment) (Chlorella sorokiniana OX=3076) glutamate to ammonium and Aketoglutarate	Nutrient uptake

SLA-04_07123-RA	6.35	Tpsab1 Tryptase (<i>Rattus norvegicus</i> OX=10116)	protease
SLA-04_05615-RA	6.4	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04390-RA	6.5	At2g45870 UPF0187 protein At2g45870, chloroplastic (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_02153-RA	6.6	AKR4C10 Aldo-keto reductase family 4 member C10 (<i>Arabidopsis thaliana</i> OX=3702)	Enzyme classification
SLA-04_04006-RA	6.65	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04524-RA	6.7	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_05461-RA	6.7	PEPR1 Leucine-rich repeat receptor-like protein kinase PEPR1 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_07724-RB	7.05	CHS2 Chitin synthase 2 (<i>Rhizopus oligosporus</i> OX=4847)	not assigned
SLA-04_02821-RA	7.1	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_01933-RA	7.15	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04004-RA	7.25	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04003-RA	7.3	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_06686-RA	7.35	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_08297-RA	7.55	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04002-RA	7.65	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_01934-RA	7.75	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_03998-RA	7.8	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_05614-RA	8.1	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_08182-RA	8.2	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_05613-RA	8.3	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_01152-RA	8.35	MRS2-7 Magnesium transporter MRS2-7 (<i>Arabidopsis thaliana</i> OX=3702)	Solute transport
SLA-04_08298-RA	8.35	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_04005-RA	8.45	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_02822-RA	8.8	Retrovirus-related Pol polyprotein from transposon TNT 1-94 (<i>Nicotiana tabacum</i> OX=4097)	not assigned
SLA-04_06502-RA	8.9	sll1483 Uncharacterized protein sll1483 (<i>Synechocystis</i> sp. (strain PCC 6803 / Kazusa) OX=1111708)	Cell wall modification
SLA-04_03453-RA	9	pho-4 Phosphate-repressible phosphate permease pho-4 (<i>Neurospora crassa</i> (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=36)	Solute transport
SLA-04_00239-RA	9.4	ACA11 Putative calcium-transporting ATPase 11, plasma membrane-type (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_05577-RA	9.45	AMT1-2 Ammonium transporter 1 member 2 (<i>Arabidopsis thaliana</i> OX=3702)	not assigned
SLA-04_08127-RA	9.45	GLOX Aldehyde oxidase GLOX (<i>Vitis pseudoreticulata</i> OX=231512)	not assigned
SLA-04_03639-RA	9.95	SLT1 Sodium/sulfate cotransporter 1 (<i>Chlamydomonas reinhardtii</i> OX=3055)	not assigned
SLA-04_03358-RA	13	sll1483 Uncharacterized protein sll1483 (<i>Synechocystis</i> sp. (strain PCC 6803 / Kazusa) OX=1111708)	Cell wall modification

References Cited

- Alkhamis, Y., Mathew, R., Ganesan, N., Rahman, S., & Rahman, M. M. (2022, 10/14). pH induced stress enhances lipid accumulation in microalgae grown under mixotrophic and autotrophic condition. *Frontiers in Energy Research*, 10, 1033068. <https://doi.org/10.3389/fenrg.2022.1033068>
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7), 2251-2252. <https://doi.org/10.1093/bioinformatics/btz859>
- Aro, E.-M., & Andersson, B. (2006). *Regulation of photosynthesis* (Vol. 11). Springer Science & Business Media.
- BBDuk. <https://sourceforge.net/projects/bbmap/>
- Beck, A. E., Hunt, K. A., & Carlson, R. P. (2018). Measuring Cellular Biomass Composition for Computational Biology Applications. *Processes*, 6(5), 38. <https://www.mdpi.com/2227-9717/6/5/38>
- Boer, V. M., Crutchfield, C. A., Bradley, P. H., Botstein, D., & Rabinowitz, J. D. (2010). Growth-limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations. *Molecular Biology of the Cell*, 21(1), 198-211. <https://doi.org/10.1091/mbc.e09-07-0597>
- Calhoun, S., Bell, T. A. S., Dahlin, L. R., Kunde, Y., Labutti, K., Louie, K. B., Kuftin, A., Treen, D., Dilworth, D., Mihaltcheva, S., Daum, C., Bowen, B. P., Northen, T. R., Guarnieri, M. T., Starkenburg, S. R., & Grigoriev, I. V. (2021). A multi-omic characterization of temperature stress in a halotolerant *Scenedesmus* strain for algal biotechnology. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-021-01859-y>
- Chen, J., Li J., Dong, W., Zhang, X., Tyagi, R. D., Drogui, P., & Surampalli, R. Y. (2018, 2018/07/01/). The potential of microalgae in biodiesel production. *Renewable and sustainable energy reviews*, 90, 336-346. <https://doi.org/10.1016/j.rser.2018.03.073>
- Chen, W., Feng, C., Guo, W., Shi, D., & Yang, C. (2011). Comparative effects of osmotic-, salt- and alkali stress on growth, photosynthesis, and osmotic adjustment of cotton plants. *Photosynthetica*, 49(3), 417-425. <https://doi.org/10.1007/s11099-011-0050-y>
- Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, 25(3), 294-306. <https://doi.org/10.1016/j.biotechadv.2007.02.001>

- Cooksey, K. E., Guckert, J. B., Williams, S. A., & Callis, P. R. (1987). Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. 6(6), 333-345. [https://doi.org/10.1016/0167-7012\(87\)90019-4](https://doi.org/10.1016/0167-7012(87)90019-4)
- Delrue, F., Setier, P.-A., Sahut, C., Cournac, L., Roubaud, A., Peltier, G., & Froment, A.-K. (2012). An economic, sustainability, and energetic model of biodiesel production from microalgae. *Bioresource Technology*, 111, 191-200.
- Deshpande, R. A., & Shankar, V. (2002, 2002/01/01). Ribonucleases from T2 Family. *Critical Reviews in Microbiology*, 28(2), 79-122. <https://doi.org/10.1080/1040-840291046704>
- Dragone, G., Fernandes, B. D., Vicente, A. A., & Teixeira, J. A. (2010). Third generation biofuels from microalgae.
- Dündar, E., & Bush, D. R. (2009). BAT1, a bidirectional amino acid transporter in Arabidopsis. *Planta*, 229(5), 1047-1056. <https://doi.org/10.1007/s00425-009-0892-8>
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., & Hassabis, D. (2021). *Protein complex prediction with AlphaFold-Multimer*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2021.10.04.463034>
- Gazzarrini, S., Lejay, L., Gojon, A., Ninnemann, O., Frommer, W. B., & von Wirén, N. (1999). Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into Arabidopsis roots. *The Plant Cell*, 11(5), 937-947.
- Gimmler, H., Degenhardt, B. (2001). Alkaliphilic and Alkali-Tolerant Algae. (Algal Adaptations to Environmental Stresses). https://doi.org/10.1007/978-3-642-59491-5_10
- Goemann, C. L. C., Wilkinson, R., Henriques, W., Bui, H., Goemann, H. M., Carlson, R. P., Viamajala, S., Gerlach, R., & Wiedenheft, B. (2023, 2023/01/01/). Genome sequence, phylogenetic analysis, and structure-based annotation reveal metabolic potential of *Chlorella* sp. SLA-04. *Algal Research*, 69, 102943. <https://doi.org/https://doi.org/10.1016/j.algal.2022.102943>
- Hadley, W. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.
- Hartmann, H., & Trumbore, S. (2016). Understanding the roles of nonstructural carbohydrates in forest trees – from what we can measure to what we want to know. *New Phytologist*, 211(2), 386-403. <https://doi.org/10.1111/nph.13955>

- Hohmann-Marriott, M. F., & Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annual Review of Plant Biology*, 62(1), 515-548. <https://doi.org/10.1146/annurev-arplant-042110-103811>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360. <https://doi.org/10.1038/nmeth.3317>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., & Usadel, B. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment*, 37(5), 1250-1258. <https://doi.org/10.1111/pce.12231>
- Losada, M. (1976). REDUCING POWER AND THE REGULATION OF PHOTOSYNTHESIS. In A. Kornberg, B. L. Horecker, L. Cornudella, & J. Oro (Eds.), *Reflections on Biochemistry* (pp. 73-84). Pergamon. <https://doi.org/https://doi.org/10.1016/B978-0-08-021010-0.50014-4>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- Luo, M. B., & Liu, F. (2011, 2011/12/01/). Salinity-induced oxidative stress and regulation of antioxidant defense system in the marine macroalga *Ulva prolifera*. *Journal of Experimental Marine Biology and Ecology*, 409(1), 223-228. <https://doi.org/https://doi.org/10.1016/j.jembe.2011.08.023>
- McKie-Krisberg, Z. M., Laurens, L. M. L., Huang, A., & Polle, J. E. W. (2018, 2018/04/01/). Comparative energetics of carbon storage molecules in green algae. *Algal Research*, 31, 326-333. <https://doi.org/https://doi.org/10.1016/j.algal.2018.01.018>

- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007, Jul). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35(Web Server issue), W182-185. <https://doi.org/10.1093/nar/gkm321>
- Newby, D. T., Mathews, T. J., Pate, R. C., Huesemann, M. H., Lane, T. W., Wahlen, B. D., Mandal, S., Engler, R. K., Feris, K. P., & Shurin, J. B. (2016). Assessing the potential of polyculture to accelerate algal biofuel production. *Algal Research*, 19, 264-277. <https://doi.org/10.1016/j.algal.2016.09.004>
- Paul, M. J., & Foyer, C. H. (2001). Sink regulation of photosynthesis. *Journal of Experimental Botany*, 52(360), 1383-1400. <https://doi.org/10.1093/jexbot/52.360.1383>
- Perera, F., & Nadeau, K. (2022). Climate Change, Fossil-Fuel Pollution, and Children's Health. *New England Journal of Medicine*, 386(24), 2303-2314. <https://doi.org/10.1056/nejmra2117706>
- Rahman, F., Hassan, M., Rosli, R., Almousally, I., Hanano, A., & Murphy, D. J. (2018). Evolutionary and genomic analysis of the caleosin/peroxygenase (CLO/PXG) gene/protein families in the Viridiplantae. *PLoS ONE*, 13(5), e0196669. <https://doi.org/10.1371/journal.pone.0196669>
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014, Jul). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, 42(Web Server issue), W187-191. <https://doi.org/10.1093/nar/gku365>
- Ren, Q., de Roo, G., Ruth, K., Witholt, B., Zinn, M., & Thöny-Meyer, L. (2009, 2009/04/13). Simultaneous Accumulation and Degradation of Polyhydroxyalkanoates: Futile Cycle or Clever Regulation? *Biomacromolecules*, 10(4), 916-922. <https://doi.org/10.1021/bm801431c>
- Rowe, J. M., Dunigan, D. D., Blanc, G., Gurnon, J. R., Xia, Y., & Van Etten, J. L. (2013). Evaluation of higher plant virus resistance genes in the green alga, *Chlorella variabilis* NC64A, during the early phase of infection with *Paramecium bursaria chlorella virus-1*. *Virology*, 442(2), 101-113. <https://doi.org/10.1016/j.virol.2013.04.018>
- Ruperti, F., Papadopoulos, N., Musser, J., & Arendt, D. (2022). *Beyond sequence similarity: cross-phyla protein annotation by structural prediction and alignment*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2022.07.05.498892>

- Sharma, J., Kumar, S. S., Bishnoi, N. R., & Pugazhendhi, A. (2018). Enhancement of lipid production from algal biomass through various growth parameters. *Journal of Molecular Liquids*, 269, 712-720. <https://doi.org/10.1016/j.molliq.2018.08.103>
- Sharma, K. K., Schuhmann, H., & Schenk, P. M. (2012). High Lipid Induction in Microalgae for Biodiesel Production. *Energies*, 5(5), 1532-1553. <https://doi.org/10.3390/en5051532>
- Singh, N. K., Paz, E., Kutsher, Y., Reuveni, M., & Lers, A. (2020). Tomato T2 ribonuclease LE is involved in the response to pathogens. *Molecular Plant Pathology*, 21(7), 895-906. <https://doi.org/10.1111/mpp.12928>
- Vadlamani, A., Pendyala, B., Viamajala, S., & Varanasi, S. (2019). High Productivity Cultivation of Microalgae without Concentrated CO₂ Input. *ACS Sustainable Chemistry & Engineering*, 7(2), 1933-1943. <https://doi.org/10.1021/acssuschemeng.8b04094>
- Vadlamani, A., Viamajala, S., Pendyala, B., & Varanasi, S. (2017). Cultivation of Microalgae at Extreme Alkaline pH Conditions: A Novel Approach for Biofuel Production. *ACS Sustainable Chemistry & Engineering*, 5(8), 7284-7294. <https://doi.org/10.1021/acssuschemeng.7b01534>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2022). Foldseek: fast and accurate protein structure search. *bioRxiv*, 2022.2002.2007.479398. <https://doi.org/10.1101/2022.02.07.479398>
- Vasanth Singan, A. L., James Han and Erika Lindquist. *RNA-Seq Gene Expression Analysis at the Joint Genome Institute*. https://jgi.doe.gov/news-publications/scientific-posters/poster_singan/
- Wilson, W. H., Van Etten, J. L., & Allen, M. J. (2009). The Phycodnaviridae: The Story of How Tiny Giants Rule the World. In (pp. 1-42). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68618-7_1
- Yan, D., Dai, J., & Wu, Q. (2013). Characterization of an ammonium transporter in the oleaginous alga *Chlorella protothecoides*. *Applied Microbiology and Biotechnology*, 97(2), 919-928. <https://doi.org/10.1007/s00253-012-4534-x>
- Yergaliyev, T. M., Nurbekova, Z., Mukiyanova, G., Akbassova, A., Sutula, M., Zhangazin, S., Bari, A., Tleukulova, Z., Shamekova, M., Masalimov, Z. K., & Omarov, R. T. (2016, 2016/12/01/). The involvement of ROS producing aldehyde oxidase in plant response to Tombusvirus infection. *Plant Physiology and Biochemistry*, 109, 36-44. <https://doi.org/https://doi.org/10.1016/j.plaphy.2016.09.001>
- Yuan, L., Loqué, D., Kojima, S., Rauch, S., Ishiyama, K., Inoue, E., Takahashi, H., & von Wirén, N. (2007, Aug). The organization of high-affinity ammonium uptake in

Arabidopsis roots depends on the spatial arrangement and biochemical properties of AMT1-type transporters. *Plant Cell*, 19(8), 2636-2652.
<https://doi.org/10.1105/tpc.107.052134>

CHAPTER FIVE

CONCLUSIONS AND FUTURE STUDIES

Algal biofuels provide a sustainable carbon-neutral alternative to fossil fuels. Optimizing the production and lowering costs of algal biofuels requires a combined understanding of cultivation conditions and algal metabolism. Cultivating algae in high alkalinity conditions is a productive and cost-effective method due to the naturally high levels of CO₂ dissolution and low levels of contamination in outdoor raceway systems. However, prior to the analyses in this dissertation, it was unclear how algae adapt to high alkaline conditions and reshape carbon metabolism to regulate alkalinity-induced stresses. This research provides a foundation for how carbon flows from photosynthesis to lipids in high alkalinity conditions and will be transformative in optimizing algal biomass composition through media optimization, precision stress induction, and genome editing.

Chapter three of this dissertation describes the complete telomere-to-telomere genome sequence of the alkali-tolerant green algae *Chlorella sp.* SLA-04. Annotation was performed using both sequence-based and structural-based protein pipelines, providing a comprehensive representation of SLA-04 metabolism. This genome represents the first complete genome of an alkali-tolerant green algae and can be used to identify adaptations that algae use to survive in alkaline conditions. Genomic analysis of the complete SLA-04 genome identified genes capable of utilizing TCA cycle intermediates to replenish cytosolic acetyl-CoA pools for lipid production. SLA-04 also contains a complete metabolic pathway for sphingolipid anabolism that may allow SLA-04 to readily adapt to changing environmental conditions and facilitate robust cultivation in mass production systems. Overall, this analysis demonstrates the impact that complete algal

genomes can have in improving the accuracy of algal phylogeny and connecting algal genotypes to extreme environmental conditions.

Chapter four of this dissertation explores the gene expression and physiology of *Chlorella* sp. SLA-04 across neutral and alkaline conditions. This analysis identified how carbon metabolism is rewired after nutrient depletion and identifies specific genes within SLA-04 that are likely involved in adaptation to alkaline conditions and mitigating oxidative damage. In response to nitrogen depletion SLA-04 rapidly produces high levels of carbohydrates within 24 hours followed by a slow accumulation of triacylglycerol for the remainder of nitrogen depletion. We propose that SLA-04 utilizes carbon storage compounds as electron sinks to minimize oxidative damage in the cell. Additionally, SLA-04 appears to cycle carbon between TAG storage and the TCA cycle to consume excess photosynthetic reducing power through a proposed futile cycle. Nitrogen limitation also triggers global repression of cellular division suggesting that nitrogen limitation effectively terminates cell growth and forces cells to produce highly reduced carbon storage compounds to balance photosynthetic energy production during stationary phase. In alkaline conditions, SLA-04 appears to maintain high cellular concentrations of carbohydrates, unsaturated lipids, and upregulated ion transporters. We hypothesize that these physiological and metabolic changes are to maintain intracellular homeostasis and mitigate alkaline-induced osmotic stress. We also see an unexpected upregulation of mobile genetic elements and immune genes in alkaline conditions. Together, this analysis confirms that both nitrogen limitation and alkaline growth impact algal metabolism but elicit different stress responses.

These combined analyses provide an in-depth understanding of algal genotypes and potential adaptations required for growth in alkaline conditions. This research has identified

metabolic trends that illuminate resource reallocation in SLA-04 and significantly advance high-alkaline algal cultivation. This analysis provides the foundation for understanding carbon flow in SLA-04, but additional research is necessary to confirm the metabolic trends we see in our transcriptomic data and use this information to streamline lipid metabolism through genome editing. Below are a few additional experiments I propose to explore discrepancies in lipid quantification methods, confirm futile carbon cycles in SLA-04, and expand our analysis beyond transcriptomic expression.

Future Studies

Nile Red Optimization

Quantifying lipid content using GCMS allows researchers to identify the concentration and types of lipids present in a sample with high accuracy, but this protocol requires resources, time, and expensive equipment. Nile red has historically been used as a cheap and immediate way to predict relative lipid content in algae (Cooksey et al., 1987). Using Nile red, researchers can estimate the level of algal lipid production between conditions and time points within a single experiment. While Nile red fluorescence has shown a consistent correlation to lipid content in the past, our experimental results show a discrepancy between fluorescence and lipid content in alkaline conditions. Nile red fluorescence increases significantly after nitrogen depletion within alkaline conditions; however, neither lipid nor carbohydrate concentration increase at a rate that correlates with Nile red fluorescence.

Currently, it is unclear why the predicted levels of lipid from Nile red are so different from the quantified GCMS data within alkaline conditions. Understanding this discrepancy is essential to identify in under what conditions we can confidently use Nile red to predict lipids and what

limitations we need to impose on this fluorescent dye. We hypothesize that increased Nile red fluorescence could be caused either by the synthesis of other non-polar compounds or increased permeability of the cell wall after nitrogen depletion. While lipids are the major non-polar molecule in a cell our GCMS pipeline only quantifies C8 lipids or larger, algae may be synthesizing smaller lipids or other non-polar small molecules that are not detected. Expanding our GCMS pipeline to include smaller molecules or running samples through reverse-phase liquid chromatography would identify if SLA-04 is producing significant amounts of non-polar small molecules in alkaline conditions.

Cell wall permeability could also affect Nile red fluorescence. Decreased cell wall integrity would allow more Nile red to permeate the cell. However, this would only impact fluorescence if Nile red is not fully saturating lipid bodies in samples with lower fluorescence. To test this hypothesis, I propose to do a time-course experiment with increasing concentrations of Nile red. This experiment will identify what incubation times and concentrations of Nile red are needed to fully saturate lipid bodies within SLA-04 in neutral and alkaline conditions. In addition, SLA-04 samples can be thin-sectioned and viewed under the transmission electron microscope to measure cell wall size and integrity between neutral and alkaline conditions. Together, these experiments will identify if cell wall permeability is impacting Nile red fluorescence.

Other lipid dyes including BODIPY 505/515 and DiO-C₁₈ could also be used to observe lipid accumulation in alkaline conditions (Cirulis et al., 2012). BODIPY is a boron-dipyrromethene dye that is insensitive to media pH or polarity and has higher sensitivity and a narrower emission spectrum compared to Nile red (Kacmar et al., 2006; Loudet & Burgess, 2007). DiO-C₁₈ is an amphiphilic and lipophilic dye commonly used to stain membranes (Bradley et al., 2009; Sengupta

et al., 2007). Using these dyes will confirm whether the increased fluorescence in alkaline conditions post-nitrogen depletion is an artifact of Nile red dye or growth in alkaline conditions.

Tracking Carbon Partitioning with ^{13}C Experiment

Physiological data suggests that SLA-04 accumulates significant amounts of TAGs after nitrogen depletion in both neutral and alkaline conditions. Traditionally, carbon flows from photosynthesis to TAGs through chloroplast fatty acid synthesis. In our transcriptomic study, we see a significant downregulation of gene expression within photosynthesis and chloroplast fatty acid synthesis suggesting that these pathways are less functional in nutrient deplete conditions. In contrast, we see an upregulation of genes involved in starch degradation and conversion of TCA intermediated into TAG precursors. TAGs are 50% more reduced than starches and the conversion of starches to TAGs would be an effective way for algae to consume excess photosynthetic energy while also producing high-energy storage molecules. However, while gene expression supports a shift in metabolism, differential regulation of specific pathways does not always directly correlate with levels of metabolism in a cell.

We hypothesize that in nutrient deplete conditions algal metabolism shifts from synthesizing TAG directly from photosynthate to utilizing carbon from starch to synthesize TAGs. To test this hypothesis, I propose to measure carbon flow in SLA-04 after nitrogen depletion to track whether TAG synthesis is using recycled carbon from starch or carbon directly from carbon fixation and plastid fatty acid biosynthesis. Carbon flow can be measured using ^{13}C labeling. Briefly, 24 hours after nitrogen depletion, SLA-04 cultures will be resuspended in media with ^{13}C inorganic carbon. During this time cells will have already accumulated high levels of starch using unlabeled ^{12}C and are in the process of accumulating lipids. Tracking the isotope composition of

lipids using mass spectrometry will identify the ratio of lipids synthesized using carbon from starch (^{12}C) and the ratio of de novo lipid synthesis from photosynthesis (^{13}C). Additional experiments where ^{13}C is added directly after nitrogen depletion can also be performed to identify if the rapid accumulation of starch is coming from photosynthesis or recycled cellular intracellular carbon. Together these experiments will identify if the increase in starch comes directly from photosynthate and whether the slow accumulation of TAGs is coming from TCA intermediates or de novo plastid fatty acid synthesis. Tracking carbon flow will allow us to better understand how SLA-04 deals with excess photosynthetic energy in stressful conditions and learn what pathways to optimize to maximize TAG accumulation in algae.

Identifying Antiviral Genes in Green Algae

Alkaline conditions select for a significant up-regulation of mobile genetic elements within SLA-04. Upregulation of retroelements in SLA-04 cause a genetic burden as they utilize cell resources to replicate and rearrange throughout the genome. In these stressful conditions, we also see an upregulation of dicer, a known immune protein involved in antiviral RNA interference (RNAi) within algae. We hypothesize that the upregulation of retroelements in alkaline conditions causes an upregulation of antiviral genes. In addition to dicer, other upregulated genes in alkaline conditions may play important and previously unappreciated roles in algal antiviral defense. To test this hypothesis, I propose to use protein structural homology to predict the function of upregulated genes with unknown functions. This will give us a better understanding of what these unknown genes may be doing in the cell, and potentially reveal antiviral activity previously unappreciated in green algae. Additionally, genes of interest can be transformed into related species *Chlorella variabilis* NC64A which has a known lytic virus. After transformation, levels of

infection can be measured using plaque assays to determine if genes provide any antiviral immunity in this virus/host interaction.

Characterizing Transcription Factors in SLA-04

In our transcriptomic study, we have found multiple transcription factors that are differentially expressed in SLA-04 grown in either nutrient deplete or alkaline conditions. We hypothesize that these transcription factors function to globally rewire algal metabolism in response to different stresses. To test this hypothesis, I propose to transform constitutively expressed transcription factors in the related algal strain *Chlorella vulgaris* 395 to observe how growth and carbon metabolism are changed. Transcriptional changes can be tracked globally using mRNA sequencing and for specific metabolic pathways using qPCR. Understanding how these transcription factors remodel metabolism and what pathways they regulated will reveal how algae can quickly respond to environmental stresses and maintain cellular homeostasis in stressful conditions. Additionally, these genes can be coopted to develop genetically engineered algal strains with controllable metabolism and optimized biomass composition.

Concluding Remarks

The United Nations has set a science-based goal to limit global warming to 1.5 degrees Celsius by 2030. This will require countries to reduce greenhouse gas emissions by greater than 45% compared to 2010 levels. Failure to meet this goal will have catastrophic and compounding effects on human health, agriculture, and water availability. This task seems daunting and out of reach but I believe that continuous and rigorous scientific advancements can contribute to achieving this goal. If we are to meet this goal, algal biofuels will no doubt play a key role in

replacing petroleum-based fuels with carbon-neutral alternatives. Algal biofuels are still expensive but have dramatically reduced in price from \$19.3 per gallon in 2011 to \$9.84 in 2021 (Energy, 2012; Hoffman et al., 2017; Prieto et al., 2017; Quinn & Davis, 2015; Rafa et al., 2021; Yousuf, 2020). This progress is a visual representation of how science can impact and even save the world. Continued scientific advancements including the research performed in this dissertation will contribute to further decreasing cultivation costs and one-day help make algal biofuels a common feature in day-to-day living. I choose to remain hopeful and trust that all the small contributions of scientists around the world will build a foundation for a sustainable, renewable, and **GREEN** future.

References Cited

- Bradley, J., Luo, R., Otis, T. S., & DiGregorio, D. A. (2009). Submillisecond Optical Reporting of Membrane Potential & In Situ Using a Neuronal Tracer Dye. *The Journal of Neuroscience*, 29(29), 9197. <https://doi.org/10.1523/JNEUROSCI.1240-09.2009>
- Cirulis, J. T., Strasser, B. C., Scott, J. A., & Ross, G. M. (2012). Optimization of staining conditions for microalgae with three lipophilic dyes to reduce precipitation and fluorescence variability. *Cytometry Part A*, 81A(7), 618-626. <https://doi.org/10.1002/cyto.a.22066>
- Cooksey, K. E., Guckert, J. B., Williams, S. A., & Callis, P. R. (1987). Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. 6(6), 333-345. [https://doi.org/10.1016/0167-7012\(87\)90019-4](https://doi.org/10.1016/0167-7012(87)90019-4)
- Energy, D. o. (2012). *algae biofuels, E.E.R.E U.S. Department of Energy in: Growing America's Energy Future.* .
- Hoffman, J., Pate, R. C., Drennen, T., & Quinn, J. C. (2017). Techno-economic assessment of open microalgae production systems. *Algal Research*, 23, 51-57. <https://doi.org/10.1016/j.algal.2017.01.005>
- Kacmar, J., Carlson, R., Balogh, S. J., & Srienc, F. (2006). Staining and quantification of poly-3-hydroxybutyrate in *Saccharomyces cerevisiae* and *Cupriavidus necator* cell populations using automated flow cytometry. *Cytometry Part A*, 69A(1), 27-35. <https://doi.org/10.1002/cyto.a.20197>
- Loudet, A., & Burgess, K. (2007, Nov). BODIPY dyes and their derivatives: syntheses and spectroscopic properties. *Chem Rev*, 107(11), 4891-4932. <https://doi.org/10.1021/cr078381n>
- Prieto, C. V. G., Ramos, F. D., Estrada, V., Villar, M. A., & Diaz, M. S. (2017). Optimization of an integrated algae-based biorefinery for the production of biodiesel, astaxanthin and PHB. *Energy*, 139, 1159-1172.
- Quinn, J. C., & Davis, R. (2015). The potentials and challenges of algae based biofuels: A review of the techno-economic, life cycle, and resource assessment modeling. *Bioresource Technology*, 184, 444-452. <https://doi.org/10.1016/j.biortech.2014.10.075>
- Rafa, N., Ahmed, S., Badruddin, I., Rahman, M. M., & Kamangar, S. (2021, 08/23). Strategies to Produce Cost-Effective Third-Generation Biofuel From Microalgae. *Frontiers in Energy Research*. <https://doi.org/10.3389/fenrg.2021.749968>

Sengupta, P., Holowka, D., & Baird, B. (2007, 2007/05/15/). Fluorescence Resonance Energy Transfer between Lipid Probes Detects Nanoscopic Heterogeneity in the Plasma Membrane of Live Cells. *Biophysical Journal*, 92(10), 3564-3574. <https://doi.org/https://doi.org/10.1529/biophysj.106.094730>

Yousuf, A. (2020). Microalgae Cultivation for Biofuels Production Preface. *Microalgae Cultivation for Biofuels Production*, Xv-Xvi. https://doi.org/Book_Doi 10.1016/C2018-0-01358-4

APPENDICES

APPENDIX A

PEDIGOGICAL RESEARCH

QUANTIFYING MOTIVATION OF MICROBIOLOGY STUDENTS

Introduction

Proper learning requires student motivation to engage with the material (Freeman et al., 2014; Olson, 1997). Without proper student motivation and engagement, students will struggle to develop long-term retention of information (Kaylene C. Williams, 2011; Lieberman & Remedios, 2007). Identifying the motivations of students and shaping a class to foster motivation is a critical component of positive student outcomes. Students from different backgrounds, including major, age, race, and gender approach class material with different experiences and different goals (Ambrose, 2010). This impacts their motivation and willingness to engage with material (Judith M. Harackiewicz, 2002; Trujillo & Tanner, 2014). As a result, a universal teaching style will likely fail as all students and classes are different. It is imperative to identify the specific student characteristics that motivate students within individual classes to provide the best learning environment.

Students taking classes that align with their interests will likely engage deeper with the material while those that are less interested, or only taking the class for credit, will struggle to show similar engagement (Shell & Soh, 2013). This suggests that professors need to modify their teaching styles while instructing major versus non-major students and that the curriculum should be shaped by the individual student characteristics of each class. If professors can make lasting associations of the material with student goals and interests, they can influence student motivation to engage (Svinicki, 2011). Further, teaching methods that cater to underperforming and unengaged students can improve student outcomes, retention, and graduation rates (Judith M. Harackiewicz, 2002; Wallerstedt & Pramling, 2012). Exploring the impacts of application on the

engagement of non-major students is necessary to identify efficient teaching methods for specific classroom settings and will improve student learning, knowledge dissemination throughout a larger community, and faculty teaching skills (Freeman et al., 2014).

Background and Literature Review

Recent trends in higher education have emphasized the importance of student-focused education. Education was previously viewed as providing information to students through traditional lecture-based classes, but now stresses the importance of student learning (Barr & Tagg, 1995). To maximize student learning, students must engage with the material and be willing to critically think about how the new material can be applied to real world applications (Svinicki, 2011). Active learning has been shown to increase student participation and engagement leading to improved student outcomes (Ambrose, 2010). While active learning has provided a platform and environment to encourage student engagement, it is harder to implement these practices in the classroom compared to traditional teaching methods. This is because students need to transition from being passive learners to an active voice within the classroom (Svinicki, 2011). This transition requires students to be motivated and see the value of engaging deeply with the material.

Student motivation is a key component that impacts high level cognition and student engagement (Lumsden, 1994; Nayir, 2017). “Motivation is probably the most important factor that educators can target in order to improve learning” (Olson, 1997). Motivation can be intrinsic where students are driven by their own goals and values, or extrinsic where students are motivated by grades or external rewards (Ambrose, 2010). If students are motivated, they are more likely to see the value of the material and engage in deeper cognition (Lieberman &

Remedios, 2007). Student motivation can be impacted by the teacher, students' mindset, course content, methods, and environment of a class (Kaylene C. Williams, 2011). Certain classroom settings will encourage passive student engagement while others will promote active engagement and motivation. As such, professors must intentionally take extra time to design and establish a classroom environment that is conducive to motivation.

Previous research has identified that students from different backgrounds have different motivational profiles that influence their engagement across different classes (Shell et al., 2005; Shell & Husman, 2008). Further, an individual student's motivational profile can change depending on what class they are taking (Shell & Soh, 2013). These data stress the importance of identifying the characteristics of students and material within individual classes when designing a curriculum that best promotes student success. Of particular interest, motivational differences were identified between students when taking required classes for their major versus non-major electives. STEM students within classes that aligned with their major were more likely to engage with the material while STEM majors taking classes not aligned with their majors had decreased motivation (Shell & Soh, 2013). This analysis was done by surveying a computer science class containing computer science and engineering students and grouping these students into motivational categories using the Students Perception Of Classroom Knowledge building (SPOCK) assessment.

Motivational profiling using SPOCK has identified five different classifications of student motivation (Shell et al., 2005; Shell & Husman, 2008; Shell & Soh, 2013). The two classifications that represent highly motivated students include the strategic profile that represents students that are highly motivated to succeed "by-any-means" and the knowledge-building profile that

represents mastery-oriented students that are goal driven and intrinsically motivated. Other profiles that represent less motivated students include the surface learning profile of students that are uninterested in the material but interested in utility and application, the apathetic profile of a disengaged non-motivated student, and the learned helpless profile of a motivated student that does not have the skills or experience to self-regulate (Shell & Soh, 2013).

Literature has shown that at least for STEM students, students taking classes that are not directly involved within their major tend to fall into the later three profiles of low motivation while students in classes that align with their majors tend to fall into the first two highly motivated categories (Shell & Husman, 2001; Shell & Soh, 2013). Students that are self-motivated and intrinsically driven are more likely to develop a deeper understanding of material and master how to apply new material to real world applications (Judith M. Harackiewicz, 2002). This suggests that students will be more productive in classes within their department, while students will struggle more within non-major classes.

While this analysis has identified differential motivational profiles between different students and classes, its scope is limited to STEM students. Further, this analysis has not investigated any teaching methods that seek to improve the learning outcomes of non-motivated students within non-major classes. I seek to expand upon this literature to identify if similar motivational trends are found within non-STEM students and observe how and if implementing a utility and applications-focused curriculum increases the motivation and engagement of different students. Current literature has identified deficiencies in the curriculum and teaching styles of faculty dealing with students in non-major classes but has provided few methods to overcome these deficiencies. This proposed study will identify if a utility-focused curriculum is

an effective method to improve the engagement and outcomes of students with different motivational profiles.

Statement of Problem, Purpose, and Study Question

The purpose of this study is to identify the motivations and goals of non-major students within an introductory biology course. This study also seeks to identify how focusing on the applications and utility of class information impacts student engagement and motivation. This research aims to identify if non-major and major students have different motivational profiles, and if exploring the application of biology increases non-major interest in the class.

Research questions

1. What are the motivational profiles of non-major students in BIOM103?
2. How does the motivation of students change after experiencing a utility-focused curriculum?

I hypothesize that non-STEM students within an introductory STEM course will tend to classify within lower motivational categories and that a utility-focused curriculum can improve outcomes within these students.

Introduction to Methods

To test my hypothesis, I aim to use a mixed methods approach to measure student motivation. Quantitative based research methods will be used to classify the students within my class into the different motivational groups and then assess through surveys how utility-focused teaching impacts the motivation and goals of students within different motivational profiles.

Qualitative methods will be used to generate personalized feedback from students after the utility-focused intervention.

Positionality and Context

This study will focus on students within an introductory biology course. Most students are non-major freshmen that are taking this course as an elective. This course is designed to focus on fundamental science for the first half of the semester, then transition to focusing on the applications and utility of the covered topics during the second half of the semester. Attention will be paid to presenting the applications that directly apply to the students' interests. This will be determined by observing the students' majors and surveying student interest via muddy questions. This setting was selected based on convenience, and as it fits the criteria to study how application and utility affect non-major student motivation.

Method

Student motivational profiles were assessed using the Student Perceptions of Classroom Knowledge building (SPOCK) assessment distributed at the midpoint of the semester (**Figure 1**) (Shell et al., 2005; Shell & Husman, 2008). SPOCK assessment is a quantitative method that can be used to classify student motivation by asking questions focused on quantifying student self-regulation, knowledge building, question asking, and lack of regulation. These questions were assessed using a Likert scale. In addition to the SPOCK assessment, student motivation, future perspectives, and course affect were measured with similar Likert scales as described by Shell et al 2013 (Shell & Soh, 2013). Numeric data of individual students were submitted to SPSS statistical software where two-way cluster analysis classified students into different motivational

categories based on their responses. This analysis and questionnaire have already been applied to multiple published scientific studies where the reliability and validity have been confirmed. All individual components of this survey have had high coefficient alpha reliability estimates.

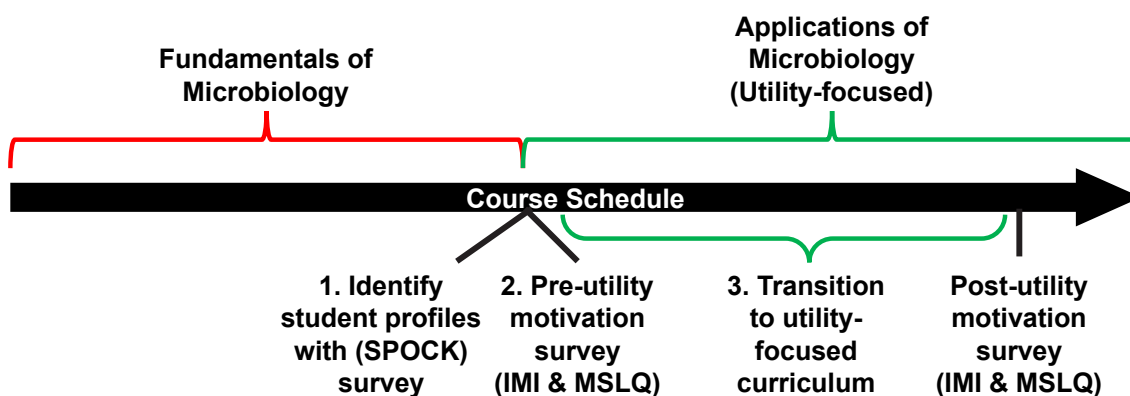


Figure 1: Course schedule. The first half of the semester focused on the fundamentals of microbiology. The second half of the semester transitioned to a utility-focused curriculum that explores the applications of microbiology. SPOCK surveys were distributed at the midpoint of the semester and were used to cluster students into different motivational profiles. IMI and MSLQ pre-surveys were also distributed to students at the semester midpoint. After the utility-focused curriculum IMI and MSLQ post-surveys were given to students. Motivational change was measured by comparing responses to the pre- and post-surveys.

To measure change in motivation students were given a pre- and post-surveys before and after the applications and utility-focused portion of the class (**Figure 1**). This survey was a combination of the Intrinsic Motivation Inventory (IMI) and Motivated Strategies for Learning Questionnaire (MSLQ) (Gibbens, 2019). These surveys focus to assess student intrinsic motivation, self-efficacy, self-value, class-value, and anxiety. The IMI and MSLQ pipelines have been assessed for reliability and shown high reliability coefficients within multiple studies (Gibbens, 2019). Survey questions were gathered using Likert scales and numerical averages were be assessed for the entire class population as well as for individual SPOCK motivational categories. Qualitative data was also gathered from these surveys asking students “What would

they like the professor to change to increase their motivation”. This analysis was used to identify how utility-focused teaching impacts the motivation of non-major students and how change in motivation differs between students in different SPOCK categories.

SPOCK and pre-IMI/MSLQ surveys were distributed to students before the midpoint of the semester and prior to transitioning to utility-focused teaching (**Figure 1**). The post-survey was distributed after students have been exposed to utility-focused intervention.

Instrumentation and Data Collection Procedure

An online questionnaire was used to gather quantitative data from students for both the pre- and post- application-focused stages of the course. BIOM103 is split into two sections, the first half of the semester is focused on the fundamentals of biology while the second section after spring break is focused on the applications. A single survey that combines the peer reviewed reliable questions from the SPOCK, IMI, and MSLQ questionnaires will be given to students during their lab recitation class periods right after spring break prior to beginning the application portion of the course, and after students have engaged in the applications portion of the course for 4 weeks. Data from the SPOCK portion of the questionnaire will be used to group students into different motivational profiles, while the IMI and MSLQ portions will be used to quantify student motivation.

Data Analysis Strategies

The quantitative data from the SPOCK portion of the survey was analyzed for individual students. Individual student data was submitted for two-way cluster analysis using IBM SPSS statistical modeling software. Cluster analysis was used to group student into five motivational

profiles as described in previous literature (Shell & Soh, 2013). Anova statistical analysis was performed to identify significant motivational differences between student groups.

IMI and MSLQ questions are broken down into four groups (intrinsic motivation, self-efficacy, value, anxiety) designed to quantify different aspects of student motivation. IMI and MSLQ data were analyzed by quantifying the mean data of these four question groups for the overall student population, and for students within specific SPOCK motivational profiles. Standard deviation and paired t-tests were used to calculate the significance of motivational changes from pre- and post- questionnaires.

Validity and Reliability

Cronbach's alpha statistics were calculated for these established questionnaires in previous literature (Gibbens, 2019; Shell & Soh, 2013). Alpha coefficients were calculated for each individual question group for the IMI and MSLQ surveys. IMI intrinsic motivation, self-efficacy, value, and anxiety questions has coefficients of 0.918, 0.883, 0.885, 0.843 respectively. MSLQ intrinsic motivation, self-efficacy, value, and anxiety questions has coefficients of 0.803, 0.903, 0.885, 0.717 respectively. All values are within the range of acceptable (0.7-0.8) to excellent (>0.9).

The individual components of the SPOCK questionnaire were assessed for reliability in previous studies (Shell & Soh, 2013). Alpha coefficients for student self-regulation, knowledge building, question asking, lack of regulation, student motivations, future perspectives, and course affect ranged from 0.78-0.91 and are within the range of acceptable to excellent.

Validity of all three questionnaires (SPOCK, IMI, and MSLQ) have been quantified and established in previous literature (Demirhan, 2014; Lin et al., 2014; Markland & Hardy, 1997).

The use of peer reviewed questionnaires increases the confidence that the survey used in this study is reliable and valid.

Findings and Conclusions

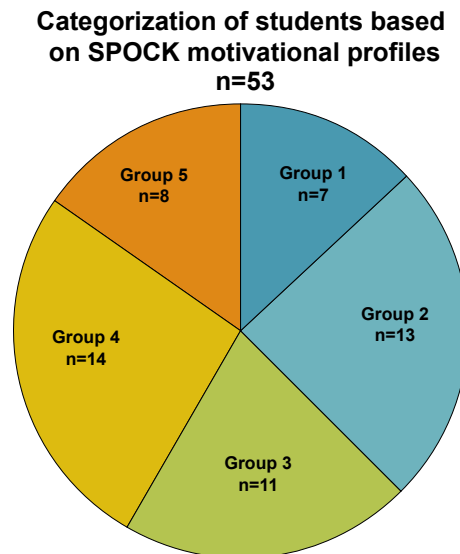


Figure 2: Student clusters based on SPOCK survey. Individual student SPOCK survey results were submitted to SPSS for two-way cluster analysis with a cluster grouping of five.

Total student responses to both SPOCK, Pre-IMI/MSLQ, and Post-IMI/MSLQ was 53. This was 51% of the total class population. SPOCK responses from the individual students (n=53) grouped into 5 clusters (**Figure 2**). The number of students within each cluster ranged from 7-14. Groups one and five had a comparatively low number of students (7 and 8 respectively), compared to groups two, three, and four (13, 11, and 14 respectively).



Figure 3: Student clusters based on SPOCK survey. Relative distribution of student responses separated by student clusters (Y-axis) and separated into individual SPOCK categories (X-axis). Red area plot represents student responses for different SPOCK categories. Distributions that trend to the right of the individual area plots represent higher student responses. Distributions that tend to the left of the individual plots represent lower responses. Individual SPOCK categories are ordered from highest (left) to lowest (right) contribution to cluster prediction.

Clusters were ordered from high to low motivation based on the average responses to individual SPOCK categories (**Figure 3**). Student clusters were labeled as highly motivated students (Group 1) to very amotivated students (Group 5). Group one students had the highest average responses to SPOCK categories that reflect positive aspects of motivation (High level question asking, Performance approach etc.), and the lowest scores in SPOCK categories that reflect negative aspects of motivation (learning avoidance, performance avoidance, and task avoidance). Group five had the lowest average responses to categories reflecting positive aspects of motivation and the highest responses to questions reflecting negative aspects of motivation. The SPOCK categories that contributed greatest to predicting student clusters include high level question asking, endogenous future time perspective, performance approach, and knowledge building. Categories that contributed least to cluster prediction were lack of regulation and affect negative. This suggests that highly motivated students tend to ask more questions, have a better appreciation of how their education will contribute to their future, they are motivated by outperforming other students, and have strong knowledge building skills. These data also suggest that lower motivated students lack these skills.

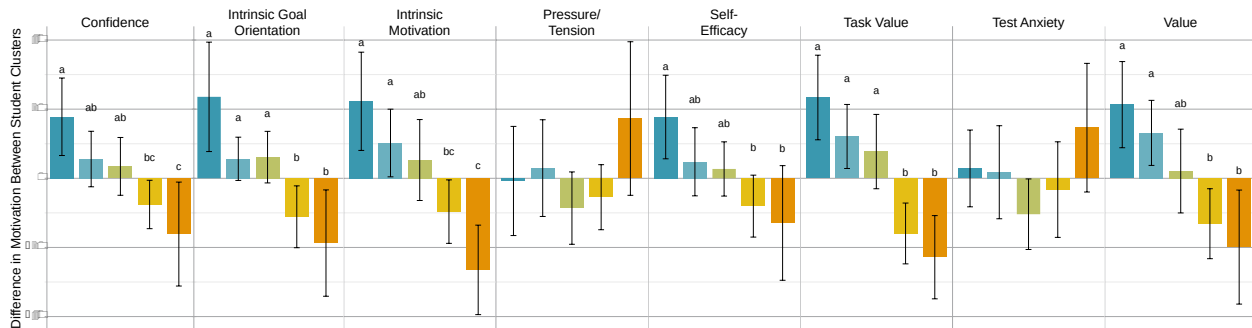


Figure 4: Motivational difference between student clusters. Motivation of student was measured using the IMI and MSLQ surveys. Motivation was broken down into eight sub-categories based on these surveys. Motivational levels of individual student clusters were quantified and compared. A value of zero represents the average motivational level of all analyzed students (n=53). Error bars represent 95% confidence interval. Statistical significance is denoted by letters (a, b, or c). Shared letters denote no statistical difference between groups. Clusters without shared letters are statistically significant. Pressure/Tension and Test Anxiety had no statistically significant differences between student clusters.

While SPOCK analysis identified motivational trends of different student clusters, motivation of student clusters was quantified using IMI and MSLQ survey responses (**Figure 4**). These surveys are designed to quantify student confidence, intrinsic goal orientation, intrinsic motivation, pressure, Self-efficacy, task value, test anxiety, and value. Likert responses for students within each motivational cluster were quantified. Anova statistical analysis was performed to identify significant differences in motivation between student clusters. Statistical differences between student clusters were present in all aspects of motivation excluding pressure/tension and test anxiety. Results suggest that highly motivated students (group 1) have significantly higher levels within all types of motivation compared to amotivated and very amotivated students (groups 4 and 5) excluding pressure/tension and test anxiety (**Figure 4**). Further, motivated students (group 2) have higher intrinsic motivation, and value compared to amotivated and very amotivated students (groups 4 and 5). Lastly, when comparing intrinsic goal orientation and task value all levels of motivated students (Groups 1, 2, and 3) had significantly

higher levels of motivation compared to amotivated and very amotivated students (groups 4 and 5). This data confirms that student clusters have different levels of motivation.

To observe how a utility-focused curriculum impacts the motivation of student within different motivational groups Pre- and Post-IMI/MSLQ survey responses were compared (**Figure 5**). Motivational change was calculated for the total student population (n=53) and for each student cluster. This data showed a significant increase in 5 out of 6 positive indicators of motivation (confidence, intrinsic goal orientation, intrinsic motivation, self-efficacy, and value) within the total student population. Confidence and intrinsic goal orientation showed the greatest increase after utility-focused intervention. Negative indicators of motivation including pressure/tension and test anxiety showed no significant change within the total student population.

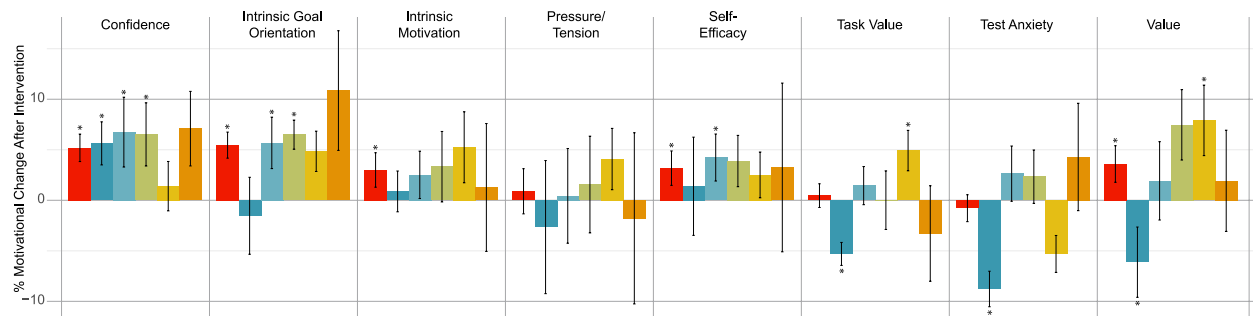


Figure 5: Motivational change of students after utility-focused intervention. Motivational change was calculated by subtracting the pre-survey IMI and MSLQ response values of individual students from their post-survey IMI and MSLQ response values. Data is shown for total student population, and five student clusters determined by SPOCK assessment. A positive increase suggests an increase in the sub-category of motivation. A negative value suggests a decrease in the sub-category of motivation. Error bars represent 95% confidence intervals. Statistically significant values are denoted by an *. Significance suggests that students either increased or decreased in motivation after utility-focused intervention.

Statistical changes in motivation were also observed within individual student clusters. Highly motivated students (n=7) showed the most change with 4 out of 8 motivational categories having significant change. Interestingly, this group showed both positive and negative changes to

motivation after intervention. Confidence increased and test anxiety decreased for this group while task value and value decreased. This may suggest that highly motivated students have low perception of the value of utility-focused application, but benefit from the application, nevertheless.

All other changes in motivation were positive. Motivated and somewhat motivated student clusters showed significant increases in confidence and intrinsic goal orientation. Self-efficacy was increased in motivated student clusters. Finally, amotivated clusters showed an increase in task value and value. This data suggests that a utility-focused curriculum has overwhelmingly positive outcomes in increasing total student motivation and only works to decrease negative predictors of motivation including test anxiety. I hypothesize that increasing the power of this analysis by increasing the sample size of students will identify positive trends in motivational change of all students after utility-focused intervention.

Positive impacts of utility-focused intervention were supported by student feedback and qualitative results (**Table 1**). Prior to the intervention, student reported a desire for “real-life applications” and topics that “applied to life as a non-science major”. Student further reported and believed that a professor “keeping a positive attitude” and “working with our peers” would increase their motivation. Active learning and real-life applications are at the core of utility-focused applications. After intervention, students reported feeling “more motivated” and suggested that the utility-focused curriculum improved their motivation by using “activities during lecture” and “covering topics that apply to life in Montana”. These student goal appear to be in line with utility-focused teaching and may explain why positive changes to student motivation were seen after exploring the application of microbiology.

<u>Pre-intervention</u>	<u>Post-intervention</u>
Students were asked “What would you like the course instructors to do to motivate you this semester?”	
“It motivates me when class topics are related and linked into real life... ”	“ I’m feeling more motivated during the lecture due to the fact that the topics are more applicable to my life.”
“Keep a positive attitude and encourage us.”	“I have greatly appreciated our in-class activities that help us retain more of the information.”
“Provide plenty of real-life applications of the material covered in lecture and lab.”	“I feel like I’m being motivated by the instructors!”
“I like when the topics we cover can be applied to life as a non-science major . That is what I find most interesting...”	“We’ve been covering topics that apply to life in Montana like the microbiology of agriculture and food systems...keep teaching topics that are applicable to most majors at MSU...”
“I think the dynamic of sitting down and watching PowerPoints gets exhausting. I would hope that switching it up and working with our peers could be incorporated.”	“Continue to do activities during lecture . This keeps me engaged...”

Table 1: Qualitative student feedback

Discussion-Recommendations

In this study we explored how a utility-focused curriculum can impact the motivation of non-major students. The goal of this experiment was to identify 1. What are the motivational profiles of non-major students in BIOM103? 2. How does the motivation of students change after experiencing a utility-focused curriculum?

Our initial hypothesis was that non-major students would cluster into lower motivational profiles. However, it appears that students in BIOM103 clustered relatively evenly into all motivational clusters with lower levels of student within highly motivated and highly amotivated groups. Using a mixed methods approach we identified that students desire that the curriculum relates to their interests and report positive motivational outcomes in utility-focused classes. This is further supported by quantitative data that shows utility-focused environments produce an overall increase in most motivational characteristics for all students. While lower motivational profiles (amotivated and very amotivated) did not show a statistically significant positive or negative change in motivation I hypothesize that this is a result of samples size and low power

once students are grouped into individual clusters. Repeating this experiment with a larger samples size would help increase power and identify further trends in motivational change after utility-focused intervention. With an increased samples size, we could also explore how demographics, background, culture, and other aspects impact student motivation.

Literature supports that students need to be motivated to produce long term learning and engage with material. In this study, we show that a utility-focused intervention that explores the applications of microbiology is an effective way to increase the motivation of non-major students. Students come from many different backgrounds and are motivated by different factors. If students goals and motivations do not align with the class material, which is likely with non-major students, faculty need to be aware are ready to shift their curriculum to align with the student's needs. When teaching non-major students, I propose that focusing heavily on the applications of material over the fundamentals will drive students to engage deeper with the material, become active participant in the course, and increase student outcomes.

References Cited

- Ambrose, S., Bridges, M. W., Lovett, M. C., DiPietro, M., Norman, M. K. . (2010). *How Learning Works: 7 Research-Based Principles for Smart Teaching* (Jossey-Bass, Ed.).
- Barr, R. B., & Tagg, J. (1995, 1995/11/01). From Teaching to Learning — A New Paradigm For Undergraduate Education. *Change: The Magazine of Higher Learning*, 27(6), 12-26. <https://doi.org/10.1080/00091383.1995.10544672>
- Demirhan, G. (2014, 01/01). A Validity and Reliability Study of the Motivated Strategies for Learning Questionnaire. *KUYEB*, 14, 9-13.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gibbens, B. (2019). Measuring Student Motivation in an Introductory Biology Class. *The American Biology Teacher*, 81(1), 20-26, 27. <https://doi.org/10.1525/abt.2019.81.1.20>
- Judith M. Harackiewicz, K. E. B., John M. Tauer, Andrew J. Elliot. (2002). Predicting Success in College: A Longitudinal Study of Achievement Goals and Ability Measures as Predictors of Interest and Performance From Freshman Year Through Graduation. *Journal of Educational Psychology*, 94(3), 562-575.
- Kaylene C. Williams, C. C. W. (2011). Five key ingredients for improving student motivation *Research in Higher Education Journal* 11. <https://doi.org/http://aabri.com/manuscripts/11834.pdf>
- Lieberman, D. A., & Remedios, R. (2007). Do undergraduates' motives for studying change as they progress through their degrees? [Article]. *British Journal of Educational Psychology*, 77(2), 379-395. <https://doi.org/10.1348/000709906X157772>
- Lin, K. Y., Hong, H.-Y., & Chai, C. S. (2014, 2014/02/01/). Development and validation of the knowledge-building environment scale. *Learning and Individual Differences*, 30, 124-132. <https://doi.org/https://doi.org/10.1016/j.lindif.2013.10.018>
- Lumsden, L. S. (1994). student motivation. *Research Roundup*, 10(3).
- Markland, D., & Hardy, L. (1997, Mar). On the factorial and construct validity of the Intrinsic Motivation Inventory: conceptual and operational concerns. *Res Q Exerc Sport*, 68(1), 20-32. <https://doi.org/10.1080/02701367.1997.10608863>

- Nayir, F. (2017, 10/20). The Relationship between Student Motivation and Class Engagement Levels. *Eurasian Journal of Educational Research*, 17, 59-78. <https://doi.org/10.14689/ejer.2017.71.4>
- Olson, G. (1997). Motivation, Motivation, Motivation-Secondary School Educators. Retrieved from sysiwyg://934/<http://7-12educators.about...-12educators/library/weekly/aa071897.htm>.
- Shell, D., Husman, J., Turner, J., Cliffel, D., Nath, I., & Sweany, N. (2005, 07/01). The Impact of Computer Supported Collaborative Learning Communities on High School Students' Knowledge Building, Strategic Learning, and Perceptions of the Classroom. *Journal of Educational Computing Research - J EDUC COMPUT RES*, 33, 327-349. <https://doi.org/10.2190/787L-BCBQ-20FN-FW6C>
- Shell, D. F., & Husman, J. (2001). The Multivariate Dimensionality of Personal Control and Future Time Perspective Beliefs in Achievement and Self-Regulation. *Contemporary Educational Psychology*, 26(4), 481-506. <https://doi.org/10.1006/ceps.2000.1073>
- Shell, D. F., & Husman, J. (2008, May 2008 2020-07-02). Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon. *Journal of Educational Psychology*, 100(2), 443-459. <https://doi.org/http://dx.doi.org/10.1037/0022-0663.100.2.443>
- Shell, D. F., & Soh, L.-K. (2013). Profiles of Motivated Self-Regulation in College Computer Science Courses: Differences in Major versus Required Non-Major Courses. *Journal of Science Education and Technology*, 22(6), 899-913. <https://doi.org/10.1007/s10956-013-9437-9>
- Svinicki, M. D., & McKeachie, W. J. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (13 ed.). Cengage Learning.
- Trujillo, G., & Tanner, K. D. (2014). Considering the Role of Affect in Learning: Monitoring Students' Self-Efficacy, Sense of Belonging, and Science Identity. *CBE—Life Sciences Education*, 13(1), 6-15. <https://doi.org/10.1187/cbe.13-12-0241>
- Wallerstedt, C., & Pramling, N. (2012). Learning to play in a goal-directed practice. *Early Years*, 32(1), 5-15. <https://doi.org/10.1080/09575146.2011.593028>

APPENDIX B

NANOPORE SEQUENCING PUBLICATION

COMPLETE GENOME SEQUENCE OF BRUCELLA ABORTUS PHAGE EF4,
DETERMINED USING LONG-READ SEQUENCING

Contribution of Authors and Co-Authors

Manuscript in Appendix B

Author: Calvin Cicha

Contributions: designed the research, performed the research, prepared manuscript figures, wrote the manuscript

Co-Author: Jodi Hedges

Contributions: designed the research, performed the research

Co-Author: Ian Novak

Contributions: performed the research

Co-Author: Deann Snyder

Contributions: performed the research

Co-Author: Mark Jutila

Contributions: designed the research

Co-Author: Blake Wiedenheft

Contributions: designed the research, wrote the manuscript

Manuscript Information

Calvin Cicha, Jodi Hedges, Ian Novak, Deann Snyder, Mark Jutila, Blake Wiedenheft

Microbiology Resource Announcements

Status of Manuscript:

- ☐ Prepared for submission to a peer-reviewed journal
- ☐ Officially submitted to a peer-reviewed journal
- ☐ Accepted by a peer-reviewed journal
- ☒ Published in a peer-reviewed journal

Microbiology Resource Announcements

Volume 9

Issue 11

30 April 2020

DOI: 10.1128/MRA.00212-20



Complete Genome Sequence of *Brucella abortus* Phage EF4, Determined Using Long-Read Sequencing

Calvin Cicha,^a Jodi Hedges,^a Ian Novak,^a Deann Snyder,^a Mark Jutila,^a Blake Wiedenheft^a

^aDepartment of Microbiology and Immunology, Montana State University, Bozeman, Montana, USA

ABSTRACT Brucellaphage EF4 was isolated from elk feces. The 38,321-bp double-stranded DNA genome is predicted to contain 72 coding regions, 38 of which have been assigned predicted functions. This phage displays nucleotide similarity to other brucellaphages of the genus *Perisivirus*.

Brucella species are Gram-negative bacteria and the causative agents of brucellosis, which leads to premature abortions in cattle (1, 2). Brucellosis is a zoonotic disease that can be transmitted from cattle to humans (2). Currently, there is no human vaccine for brucellosis, and *Brucella* species can be cleared only with an extended antibiotic treatment regimen (3). With the need for improved treatment of brucellosis in humans and cattle, phage therapy has been proposed as an alternative treatment (4). Here, we describe the isolation and sequencing of a brucellaphage from the family *Podoviridae* that infects *Brucella abortus* strain S19.

We isolated and sequenced a bacteriophage infecting *Brucella abortus* strain S19 from elk fecal samples collected near the Gravelly Mountains in Montana. Elk fecal samples were homogenized in sterile phosphate-buffered saline and filtered through a 0.2- μ m filter. Fecal filtrates were combined with the S19 vaccine strain of *Brucella abortus* and incubated for 30 min at 37°C. Cultures were suspended in 0.75% potato infusion agar (PIA) and poured onto PIA plates. Plates were incubated for 48 h at 37°C. Eleven fecal samples were screened. One plaque (elk fecal 4 [EF4]) was selected for clonal amplification using the plate lysate protocol (5).

EF4 phage DNA was isolated and purified via phenol-chloroform extraction (6). DNA was prepared for long-read sequencing using the SQK-LSK109 kit provided by Oxford Nanopore Technologies. DNA was sequenced using the MinION platform with an R9.4.1 FLO-MINSP6 flow cell. Sequencing resulted in a total of 417,284 reads, with an average read length of 1 kb. Quality control and base calling were performed with MinkNOW v19.06.8 and Guppy v3.2.4+d9ed22f, respectively (Oxford Nanopore Technologies). Sequences were assembled using Flye v2.5-g0c3de5b (7). This approach resulted in a single viral contig with 995 $\times$ coverage. Prokka v1.14.5 was used to identify protein-coding regions within the viral genome (8). Coding regions were queried against the NCBI nonredundant protein database using BLASTp to assign putative functions (Fig. 1) (9). All tools were run with default parameters for long-read analysis.

Brucellaphages have a core genome of 38 kb, with a subset of phages containing two insertions totaling ~3 kb (10–12). Assembly of the EF4 phage genome produced a single 38,321-bp circular genome (GC content, 48.2%) with neither of the previously described insertions. Since *Perisivirus* genomes are packaged as linear double-stranded DNA but replicate as covalently closed circles, we hypothesize that the circular assembly is a consequence of sequencing both forms of the genome (13). Genome termini were determined via alignment with existing brucellaphages (13). The finding of a single contig with the expected termini and length supports the idea that the whole EF4 phage genome was sequenced. Comparison of the EF4 genome with existing

Citation Cicha C, Hedges J, Novak I, Snyder D, Jutila M, Wiedenheft B. 2020. Complete genome sequence of *Brucella abortus* phage EF4, determined using long-read sequencing. Microbiol Resour Announc 9:e00212–20. <https://doi.org/10.1128/MRA.00212-20>.

Editor John J. Dennehy, Queens College

Copyright © 2020 Cicha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Blake Wiedenheft, bwiedenheft@gmail.com.

Received 2 March 2020

Accepted 7 April 2020

Published 30 April 2020

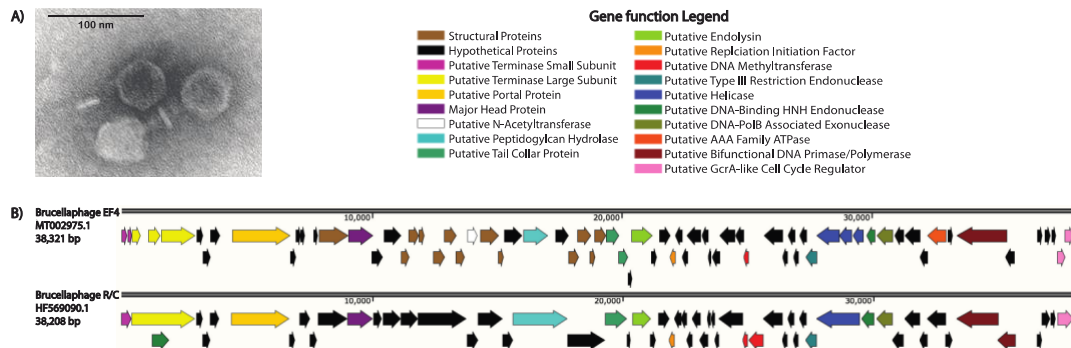


FIG 1 Virion morphology and genome organization of the EF4 phage. (A) Transmission electron microscopy image of brucellaphage EF4, revealing an icosahedral capsid (diameter, 52.5 ± 4.6 nm [$n = 10$]) with a short tail. (B) Genome comparison of brucellaphage EF4 (GenBank accession number [MT002975](#)) and brucellaphage R/C ([HF569090](#)), revealing similar genome organizations. Gene orientations are indicated with arrows, and functional predictions are colored according to the key.

Perisivirus genomes using BLASTn revealed a high level of similarity with the genome of brucellaphage R/C (GenBank accession number [HF569090](#)), with 99.71% nucleotide identity and similar synteny (10–12).

The EF4 genome has a coding density of 87.8%, with 72 coding sequences (CDSs) identified using the Prokka software package (Fig. 1). Of the 72 predicted CDSs, 38 have sequence similarity to proteins of known function, while the other 34 are similar to hypothetical proteins with no known functions. The 38 CDSs with putative functional assignments include major head, portal, and tail collar proteins, as well as major and minor terminases. This genome also contains expected DNA replication machinery, such as a bifunctional DNA polymerase/primase, helicase, exonuclease, and DNA endonuclease. The genes cluster according to predicted function, with structural and lysis proteins being encoded on the positive strand and DNA replication proteins on the negative strand. This orientation may be involved in regulating gene expression.

Data availability. Sequence and related data for the EF4 genome have been deposited under BioProject accession number [PRJNA603786](#), GenBank accession number [MT002975](#), and SRA accession number [SRR10985492](#).

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (grant R21AI130670 to M.J. and grant 1R35GM134867 to B.W.), the U.S. Department of Agriculture (grant MONB00021 to M.J.), and the Montana State University Agricultural Experimental Station (USDA NIFA).

REFERENCES

- Halling SM, Peterson-Burch BD, Bricker BJ, Zuerner RL, Qing Z, Li LL, Kapur V, Alt DP, Olsen SC. 2005. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *J Bacteriol* 187:2715–2726. <https://doi.org/10.1128/JB.187.8.2715-2726.2005>.
- Corbel MJF. 2006. *Brucellosis in humans and animals*. World Health Organization, Geneva, Switzerland.
- Al-Sibai MB, Halim MA, el-Shaker MM, Khan BA, Qadri SM. 1992. Efficacy of ciprofloxacin for treatment of *Brucella-melitensis* infections. *Antimicrob Agents Chemother* 36:150–152. <https://doi.org/10.1128/aac.36.1.150>.
- Corbel MJ, Morris JA. 1980. Investigation of the effect of *Brucella*-phage on the course of experimental infection with *Brucella abortus*. *Br Vet J* 136:278–289. [https://doi.org/10.1016/s0007-1935\(17\)32293-5](https://doi.org/10.1016/s0007-1935(17)32293-5).
- Pickard DJ. 2009. Preparation of bacteriophage lysates and pure DNA. *Methods Mol Biol* 502:3–9. https://doi.org/10.1007/978-1-60327-565-1_1.
- Kramvis A, Bukofzer S, Kew MC. 1996. Comparison of hepatitis B virus DNA extractions from serum by the QIAamp blood kit, GeneReleaser, and the phenol-chloroform method. *J Clin Microbiol* 34:2731–2733. <https://doi.org/10.1128/JCM.34.11.2731-2733.1996>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local

- alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
10. Tevdoradze E, Farlow J, Kotorashvili A, Skhirtladze N, Antadze I, Gunia S, Balarjishvili N, Kvachadze L, Kutateladze M. 2015. Whole genome sequence comparison of ten diagnostic brucellaphages propagated on two *Brucella abortus* hosts. *Virology* 12:66. <https://doi.org/10.1186/s12985-015-0287-3>.
 11. Hammerl JA, Gollner C, Jackel C, Scholz HC, Nockler K, Reetz J, Al Dahouk S, Hertwig S. 2017. Genetic diversity of *Brucella* reference and non-reference phages and its impact on *Brucella*-typing. *Front Microbiol* 8:408. <https://doi.org/10.3389/fmicb.2017.00408>.
 12. Farlow J, Filippov AA, Sergueev KV, Hang J, Kotorashvili A, Nikolich MP. 2014. Comparative whole genome analysis of six diagnostic brucellaphages. *Gene* 541:115–122. <https://doi.org/10.1016/j.gene.2014.01.018>.
 13. Flores V, Lopez-Merino A, Mendoza-Hernandez G, Guarneros G. 2012. Comparative genomic analysis of two brucellaphages of distant origins. *Genomics* 99:233–240. <https://doi.org/10.1016/j.ygeno.2012.01.001>.

APPENDIX C

CO-AUTHOR PUBLICATIONS

TEMPORAL DETECTION AND PHYLOGENETIC ASSESSMENT
OF SARS-COV-2 IN MUNICIPAL WASTEWATER

Contribution of Authors and Co-Authors

Manuscript in Appendix C

Author: Artem Nemudryi

Contributions: designed the research, performed the research, prepared manuscript figures, wrote the manuscript, wrote the manuscript

Co-Author: Anna Nemudraia

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Tanner Wiegand

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Kevin Surya

Contributions: performed the research

Co-Author: Murat Buyukyoruk

Contributions: performed the research

Co-Author: Calvin Cicha

Contributions: performed the research

Co-Author: Karl K. Vanderwood

Contributions: performed the research

Co-Author: Royce Wilkonson

Contributions: performed the research

Co-Author: Blake Wiedenheft

Contributions: designed the research, wrote the manuscript

Manuscript Information

Artem Nemudryi, Anna Nemudraia, Tanner Wiegand, Kevin Surya Murat Buyukyoruk, Calvin Cicha, Karl K. Vanderwood, Royce Wilkinson, and Blake Wiedenheft

Cell Reports Medicine

Status of Manuscript:

☐ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☒ Published in a peer-reviewed journal

Cell Press

Volume 1

Issue 6

22 September 2020

DOI: 10.1016/j.xcrm.2020.100098

TEMPORAL DETECTION AND PHYLOGENETIC ASSESSMENT
OF SARS-COV-2 IN MUNICIPAL WASTEWATER

Artem Nemudryi,¹ Anna Nemudraia,¹ Tanner Wiegand,¹ Kevin Surya,¹ Murat Buyukyoruk,¹

Calvin Cicha,¹ Karl K. Vanderwood,² Royce Wilkinson,¹ and Blake Wiedenheft¹

¹Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA. ²Gallatin City-County Health Department, Bozeman, MT 59715, USA.

Abstract

SARS-CoV-2 has recently been detected in feces, which indicates that wastewater may be used to monitor viral prevalence in the community. Here, we use RT-qPCR to monitor wastewater for SARS-CoV-2 RNA over a 74-day time course. We show that changes in SARS-CoV-2 RNA concentrations follow symptom onset gathered by retrospective interview of patients but precedes clinical test results. In addition, we determine a nearly complete (98.5%) SARS-CoV-2 genome sequence from wastewater and use phylogenetic analysis to infer viral ancestry. Collectively, this work demonstrates how wastewater can be used as a proxy to monitor viral prevalence in the community and how genome sequencing can be used for genotyping viral strains circulating in a community.

Introduction

In late December 2019, authorities from the People's Republic of China (PRC) announced an epidemic of pneumonia.(Organization, 2020) A novel coronavirus (severe acute respiratory syndrome-coronavirus-2 [SARS-CoV-2]) was identified as the etiologic agent and the disease was named coronavirus disease 2019 (COVID-19). The virus spread rapidly, first to Thailand, Japan, Korea, and Europe, and now to >188 countries across all of the continents except Antarctica. The global total of infected individuals now exceeds 20 million (<https://coronavirus.jhu.edu/>).

Public health professionals around the world are working to limit the spread of SARS-CoV-2, and “flatten the curve,” which requires a reduction in cases from one day to the next. However, SARS-CoV-2 containment has been outpaced by viral spread and limited resources for testing. Moreover, mounting evidence suggests that the virus is not only spread by aerosols but may also be transmitted via feces. Both viral RNA and infectious virus have been detected in the stool of COVID-19 patients.(Hindson, 2020; Lodder & de Roda Husman, 2020; Tang et al., 2020; W. Wang et al., 2020; Y. Wu et al., 2020; Xiao et al., 2020; Xu et al., 2020) This has important implications for the spread of the virus and suggests that wastewater may be used to monitor progression or abatement of viral spread at the community level(Ahmed et al., 2020; La Rosa et al., 2020; Mallapaty, 2020; Medema et al., 2020; Peccia et al., 2020; Randazzo et al., 2020; Rimoldi et al., 2020; F. Wu et al., 2020; Wurtzer et al., 2020)

Results

To test whether wastewater could be used for SARS-CoV-2 surveillance, we collected samples from the municipal wastewater treatment plant in Bozeman, Montana (USA). Untreated wastewater samples were collected on 17 different days over the course of a 74-day period using an autosampler that collects a volume proportional to flow for 24 h (Table S6.1). This composite sample reflects the average characteristics of wastewater over the previous day. The samples were filtered and concentrated before RNA extraction. To concentrate SARS-CoV-2, we used ultrafiltration with spin concentrators that efficiently recover viruses from wastewater. (Medema et al., 2020; Ye et al., 2016) Extracted RNA was used as a template for one-step quantitative reverse-transcriptase-polymerase chain reaction (qRT-PCR), performed according to Centers for Disease Control and Prevention (CDC) guidelines (<https://www.fda.gov/media/134922/download>). Each qRT-PCR reaction was performed using 2 primer pairs (N1 and N2), which target distinct regions of the nucleocapsid (N) gene from SARS-CoV-2 (Figures 6.1A and S6.1). Composite samples collected in late March and early April 2020 tested positive for SARS-CoV-2, although concentrations of viral RNA steadily declined and then dropped below the limit of detection. After 1 month of undetectable levels of SARS-CoV-2, the wastewater began testing positive again in late May, which coincided with an increase in COVID-19 cases in the community (Figure 6.1A).

The current methods for tracking the COVID-19 pandemic primarily rely on clinical test results, but this process involves intrinsic delays that preclude real-time tracking of the outbreak. On average, a person develops symptoms 4–5 days after initial exposure, and it is predicted that only 32% of symptomatic individuals are tested. (Lauer et al., 2020; Li et al., 2020) Test results

are typically available 3–9 days after illness onset. (Ahmed et al., 2020; Garg et al., 2020) We hypothesized that wastewater levels of SARS-CoV-2 RNA correlate with COVID-19 incidence rates and that these data could be used as an epidemiological indicator to track the outbreak in real time. To test this hypothesis, we compared our wastewater surveillance data to the frequencies of reported lab-confirmed cases and symptom onset dates that were collected by retrospective interviews. In the initial outbreak (mid-March 2020), the SARS-CoV-2 RNA concentration in the wastewater lagged behind symptom onset data by 8

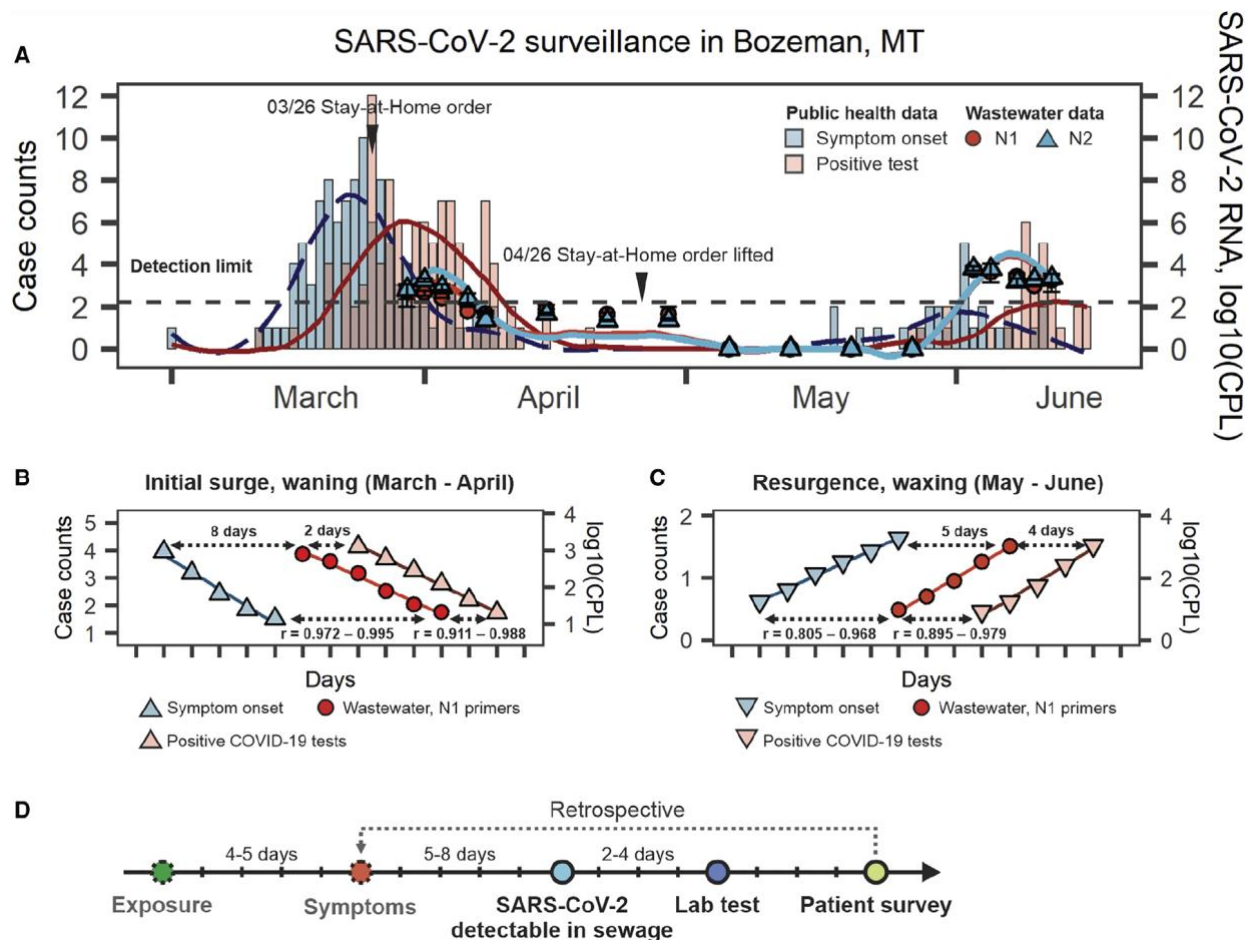


Figure 6.1. Detection and Quantification of SARS-CoV-2 in Wastewater and in the Community (A) Temporal dynamics of SARS-CoV-2 RNA in the municipal wastewater is superimposed on the epidemiological data. Symptom onset data (teal bars) were collected by retrospective interviews of COVID-19 patients who previously tested positive for SARS-CoV-2 (coral bars).

The red circles and blue triangles show SARS-CoV-2 RNA concentration in municipal wastewater (means $\pm$ SDs) measured with qRT-PCR using the N1 and N2 primer pairs, respectively (see Method Details). The lines show curves fitted to qRT-PCR and epidemiological data using local polynomial regression (LOESS, locally estimated scatterplot smoothing). (B and C) Linear regions of the epidemiological and wastewater curves. Curves were displaced relative to each other and Pearson correlation coefficients (r) were calculated. The 95% confidence intervals for the highest Pearson's r values and respective offsets are shown. Data for initial surge (March–April) and resurgence (May) were analyzed separately. Surge boundaries were defined as the earliest reported symptom onset (left boundary) and date with last reported positive test (right boundary). The interval between surges with zero reported cases/symptoms (mid-April–mid-May) was dropped from the analysis. (D) Timeline of the indicators used in the study. Symptom onset is the earliest available estimate of the viral spread. However, these data are collected retrospectively, which preclude its use for real-time tracking of the outbreak. Wastewater correlates with symptom onset and could be used to track progressing outbreak.

days (Pearson's $r = 0.989$; Figure 6.1B), and preceded laboratory test results for individuals by 2 days ($r = 0.969$; Figure 6.1B). When cases resurged in May, wastewater detection trailed symptom onset by 5 days ($r = 0.92$) and foreshadowed the increase in positive tests by 4 days ($r = 0.953$; Figure 6.1C). While wastewater detection trailed symptom onset by 5 days, it is important to note that the retrospective interviews used to collect symptom onset information are available only ~ 10 days after exposure. Our analysis demonstrates that wastewater surveillance is the earliest real-time measure of SARS-CoV-2 prevalence (Figure 6.1D).

To verify that the qRT-PCR results reflect bona fide detection of SARS-CoV-2 rather than priming from an unintended template, we repeated the PCR using 10 primer pairs that tile across the SARS-CoV-2 genome.(Quick, 2020) These primers were designed to target conserved regions of the genome that flank polymorphic sites that have been used to trace viral ancestry and geographic origins (Figure S6.2A).(Bedford et al., 2020; Quick, 2020) RNA isolated from the Bozeman waste stream on March 27, 2020 was used as a template for these RT-PCR reactions, and all 10 primer pairs produced PCR products of the expected sizes (Figure S6.2B; Table S6.2). PCR products were sequenced using the Sanger method and the reads were aligned

to the reference genome using MUSCLE.(Edgar; Fan Wu et al., 2020) We observed no sequence heterogeneity in redundant reads derived from each location of the genome (Figure S6.2C). The same RNA sample was further used to determine a nearly complete (98.5%) SARS-CoV-2 genome sequence using a long-read sequencing platform.(Quick, 2020; Quick et al., 2017)

Mutations that do not confer a fitness defect are preserved in viral progeny, and thus serve as genetic landmarks that can be used to trace viral ancestry. Efforts to understand the origins and evolution of SARS-CoV-2 have resulted in ~82,000 genome sequences from 91 countries as of August 13, 2020 ([https:// www.gisaid.org/](https://www.gisaid.org/)). Phylogenetic analyses of these sequences have enabled molecular tracking of viral spread.(Castillo et al., 2020; Fauver et al., 2020; Gonzalez-Reiche et al., 2020; Stefanelli et al., 2020; van Dorp et al., 2020; Zehender et al., 2020) To determine the ancestry of SARS-CoV-2 strains circulating in Bozeman's wastewater on March 27, we determined the genome sequence using Oxford Nanopore. Approximately 2,000 copies of the viral RNA (estimated with qPCR) were used to generate an amplicon library.(Nathan D. Grubaugh et al., 2019; Quick et al., 2017) Nanopore sequencing on the MinION platform resulted in ~700,000 reads. Quality control and base calling were performed with MinKNOW version 19.06.8 in High Accuracy mode (Oxford Nanopore Technologies), and the sequences were assembled using the bioinformatic pipeline from ARTIC Network (<https://artic.network/ncov-2019>). This approach resulted in a single viral contig with an average sequencing depth of 6,875x that covered 98.5% of the SARS-CoV-2 reference genome (GenBank: MN908947.3). Unsequenced regions of the genome include the 5' and 3' ends and a stretch of 170 bases (22,346–22,515 in the Wuhan-Hu-1 reference genome), which likely had too few reads for basecalling due to PCR bias.(Itokawa et al., 2020)

In total, we found 11 single-nucleotide variants (SNVs) in the assembled genome that distinguish the Bozeman wastewater SARS-CoV-2 sequence from the Wuhan-Hu-1/2019 reference sequence (Figure S6.2D). To verify the authenticity of these SNVs, we examined raw sequencing data, which is available on Mendeley (Mendeley Data: <https://dx.doi.org/10.17632/nfsfvy6xkf.1>). During this analysis, we noticed that one of the variants (A23122T) was introduced by incomplete trimming of a sequencing adaptor (5'-CGTATTGCT) that is partially homologous (underlined) to the reference genome (5'-TACATGCA). This variant has been reported in other genomes sequenced using this protocol,(Lopez-Alvarez et al., 2020) and similar issues have been identified in other regions of the SARS-CoV-2 genome.(Korber et al., 2020) In contrast to A23122T, we found no other evidence for trimming artifacts for the 10 remaining SNVs. It is possible that the consensus genome presented here is derived from a chimeric assembly of distinct genotypes, but >90% of the reads contain the same variant at each of these 10 positions.

The genome was aligned to 14,970 SARS-CoV-2 genomes from 74 different countries (Global Initiative on Sharing All Influenza Data, <https://www.gisaid.org/>). The resulting alignment was used to build a phylogenetic tree (Figure 6.2A), which indicates that the SARS-CoV-2 genome in Bozeman's wastewater is most closely related to genomes from California and

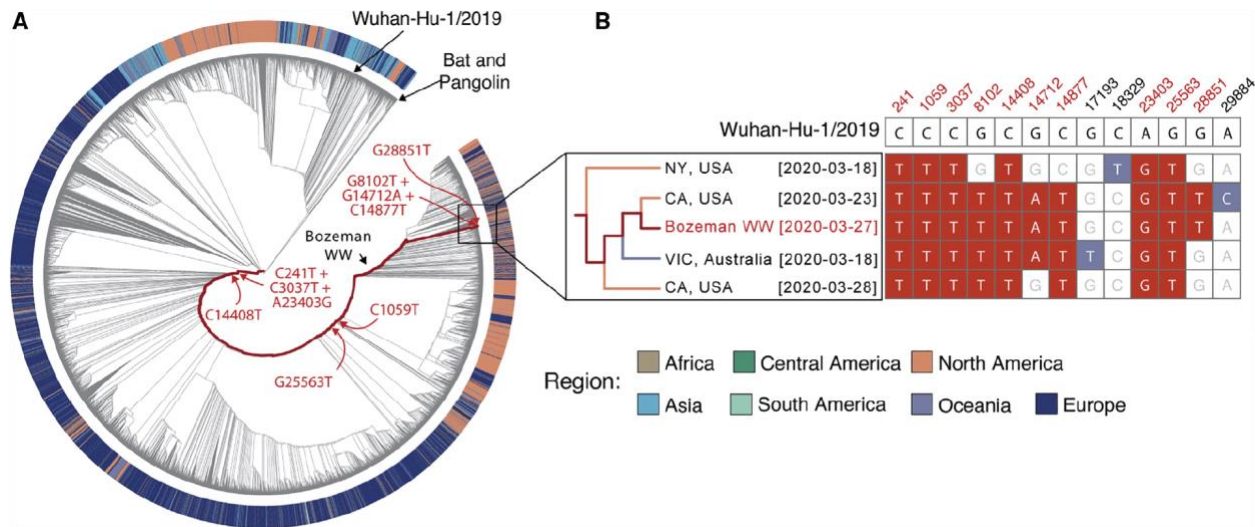


Figure 6.2. Phylogenetic Analysis of SARS-CoV-2 Sequence Isolated from Wastewater. (A) Maximum-likelihood phylogeny of the SARS-CoV-2-related lineage (n = 14,971 sequences). The phylogenetic history of SARS-CoV-2 strain sequenced from Bozeman's wastewater (WW) is shown in crimson. The outer ring is colored according to regions of the world where the samples were isolated. The tree is rooted relative to the RaTG13 genome (a bat coronavirus with 96% sequence similarity to SARS-CoV-2; GenBank: MN996532.1). Mutations that occurred over space and time are shown in red. (B) Sequences isolated from Bozeman WW clade with sequences of US and Australian origin (left). The sequences are named according to the geographic origin and the viral isolation date. A comparison of mutations in sequences is shown in the inset (right). The Wuhan reference sequence for each of the positions where mutations occur is shown across the top. The mutated positions and bases present in Bozeman WW sequence are shown in red, the bases matching Wuhan reference sequence are shown in white, and the mutations not present in the Bozeman WW sequence are shown in blue.

Victoria, Australia). The three mutations that define the Wuhan WA1 lineage (C8782T, C18060T, T28144C) are not present in the Bozeman wastewater (WW) genome, while all 10 mutations in the Bozeman wastewater SARS-CoV-2 sequence co-occur in sequences from California; 9 of these 10 mutations are also present in an isolate from Victoria, Australia (Figures 6.2B, S6.2C, and S6.2D). (Bedford et al., 2020; Deng et al., 2020; Holshue et al., 2020) To determine how these sequence variations may have accumulated over space and time, we mapped each mutation onto the phylogenetic tree of SARS-CoV-2 sequences (Figures 6.2A and S6.2E). This analysis shows that the A28851T mutation has been acquired most recently and confirms that the

assembled genome from the Bozeman wastewater is most closely related to a strain circulating in California. While this sequencing approach reveals the genetic history, it does not measure the fitness of this or any of the mutations associated with distinct geographic locations. We anticipate that temporal genome sequencing from the wastewater will help identify viral strains circulating in a specific community over time.

Discussion

The results presented here demonstrate that wastewater monitoring for SARS-CoV-2 RNA by qRT-PCR provides a real-time measure of viral prevalence in the community (Figure 6.1). Clinical testing for COVID-19 typically occurs 3–9 days after symptom onset and may vary, depending on the availability of tests, care-seeking behavior, workloads in testing facilities, and current testing strategy.(Biggerstaff et al., 2014; Garg et al., 2020; Silverman et al., 2020) In our study, wastewater surveillance for SARS-CoV-2 foreshadowed new case reports by 2–4 days.

The statistics of lab-confirmed COVID-19 cases not only lag behind viral spread but they also underestimate the true scale of the pandemic. A recent analysis of outpatient surveillance data estimates that only 32% of SARS-CoV-2-infected individuals in the United States have sought medical care.(Silverman et al., 2020) These challenges are overcome with wastewater testing, which captures input from all of the individuals in the local community and thus has the potential for estimating the true prevalence of COVID-19 using computational models that account for the median viral load in stool, virus degradation rates, travel time to the treatment facility, and water use per capita.(Ahmed et al., 2020; Hart & Halden, 2020) Furthermore,

wastewater may capture mild and asymptomatic infections that may be used to alert public health officials about emerging undetected transmission events.(Tang et al., 2020)

Monitoring wastewater for SARS-CoV-2 provides a useful epidemiological metric that could help track the outbreak and inform policy. The study presented here complements the rapidly emerging body of work by providing an important link between wastewater surveillance, COVID-19 epidemiology, and tracing SARS-CoV-2 spread patterns with genome sequencing.

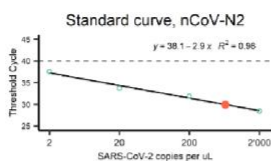
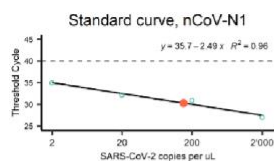
Limitations of Study

Nanopore sequencing has an error rate of ~10%–15%, which precludes the reliable detection of rare genotypes.(Nathan D. Grubaugh et al., 2019; Jain et al., 2018) Therefore, the protocol used here is limited to the detection of genotypes with >10%–15% representation. Due to this limitation of the sequencing approach, we cannot exclude the possibility of chimeric genome assembly. Additional sequencing efforts are required to correlate results from wastewater to clinical isolates and determine how ratios of SARS-CoV-2 variants in wastewater translate to ratios in the population.

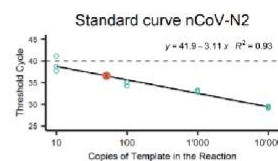
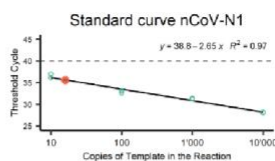
An additional limitation of this study is the use of a single method for the concentration of SARS-CoV-2 from wastewater. While others have shown that centrifugal ultrafiltration is effective, comparative analyses of alternative concentration protocols, RNA extraction methods, and diagnostic primer sets are required to correlate results from different groups and build a standard approach for wastewater-based SARS-CoV-2 surveillance(Ahmed et al., 2020; La Rosa et al., 2020; Mallapaty, 2020; Medema et al., 2020; Peccia et al., 2020; Randazzo et al., 2020; Rimoldi et al., 2020; F. Wu et al., 2020; Wurtzer et al., 2020; Ye et al., 2016).

Supplemental Figures

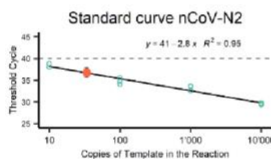
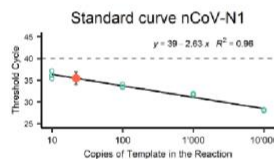
2020/03/27



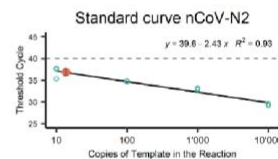
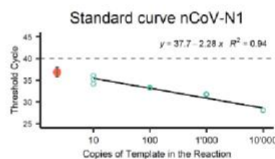
2020/04/03



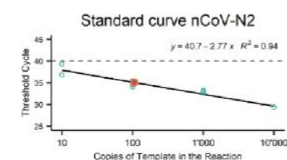
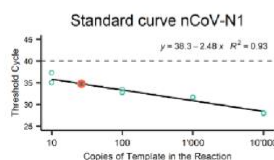
2020/03/30



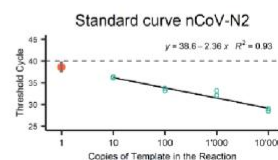
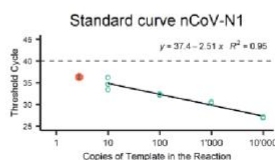
2020/04/06



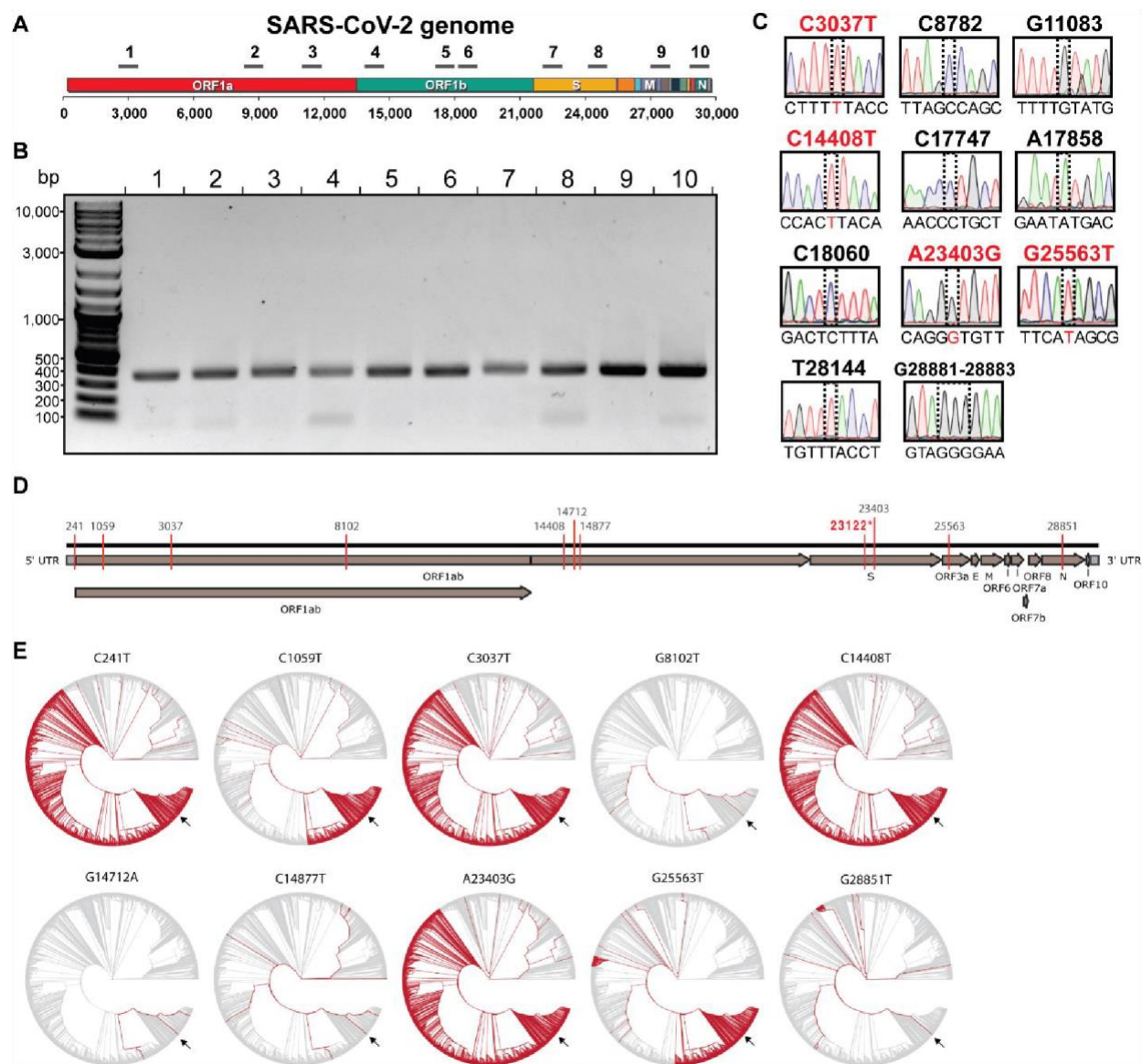
2020/04/01



2020/04/08



Supplemental Figure S6.1, related to Figure 6.1. Standard curves for absolute quantification of SARS-CoV-2 titers. Standard curves were generated for each RT-qPCR run. A 10-fold dilution series of positive control plasmid (IDT#10006625) encoding for SARS-CoV-2 N gene were used, three technical replicates at each dilution (green circles). Data was plotted as Cycle Threshold (Ct) on y-axis versus log₁₀(copy number) on x-axis. Trend lines were fitted to the data by linear regression analysis in RStudio v1.2.1335, Linear equations and R² values are shown for each standard curve. Red dots show mean Ct values for wastewater samples. Error bars represent standard deviation of Ct values.



Supplemental Figure S6.2, related to Figure 6.2. Sequencing and phylogenetics of SARS-CoV-2 in wastewater. (A) Map of the SARS-CoV-2 genome. The solid black lines indicate the approximate location of 10 amplicons spanning phylogenetically informative regions of the genome. (B) Agarose gel of the corresponding PCR products. (C) Ten PCR products shown in B were gel-purified and used to sequence 13 polymorphic sites in SARS-CoV-2 genome 44. Sanger traces (9-bp windows) are shown for each polymorphic site (dotted boxes). Nucleotide numbering is based on the reference genome Wuhan-Hu-1/2019²⁵. Highlighted in red are nucleotide substitutions relative to the reference. (D) Genetic map of SARS-CoV-2 with mutations present in Bozeman wastewater sequence highlighted. *Variant A23122T was further filtered out as an artifact of sequencing library preparation (see explanation in the main text). (E) Phylogenetic tree with each mutation present in Bozeman wastewater mapped in red. The SARS-CoV-2 sequence from Bozeman wastewater is indicated by black arrow.

Supplemental Tables

Sampling date	SARS-CoV-2 RNA concentration estimated with N1, copies / L *	SARS-CoV-2 RNA concentration estimated with N2, copies / L *	Total daily flow volume, L
March 30, 2020	528.9 ± 249	665.6 ± 261.5	22,871,458
April 1, 2020	432.7 ± 66.3	1710.8 ± 179.2	24,097,932
April 3, 2020	254.7 ± 5.4	836.3 ± 144.6	23,287,853
April 6, 2020	57.6 ± 38	257.4 ± 93.8	23,882,163
April 8, 2020	46 ± 9.2	23.5 ± 14.6	23,628,540
April 15, 2020	76 ± 24	46.7 ± 4.2	22,761,681
April 22, 2020	44.6 ± 5.6	22.2 **	22,492,917
April 29, 2020	43.5 ± 34.4	24.6 ± 1.5	22,273,363
May 6, 2020	0 ***	0	22,720,042
May 13, 2020	0	0	22,473,990
May 20, 2020	0	0	22,750,325
May 27, 2020	0	0	22,807,106
June 3, 2020	5400.5 ± 1046.1	6116.2 ± 1034.8	22,121,947
June 5, 2020	4174.6 ± 749.1	5624.6 ± 2509.2	21,754,762
June 8, 2020	2464.2 ± 502.4	1429 ± 322.4	21,016,606
June 10, 2020	971.9 ± 21.5	1947.9 ± 62.4	23,102,368
June 12, 2020	1762.7 ± 733.1	2291.4 ± 680.9	22,591,338
*Data shown as mean ± sem, n = 3			
** Only one of three replicates showed amplification			
*** Samples tested negative and thus were called 0 copies/L			

Supplemental Table S6.1, related to Fig 6.1. Total wastewater flow volumes and SARS-CoV-2 RNA concentration in composite samples.

PCR product	Primer	Sequence (5'-3')
1	nCoV-2019_10_LEFT	TGAGAAGTGCTCTGCCTATACAGT
	nCoV-2019_10_RIGHT	TCATCTAACCAATCTTCTTCTTGCTCT
2	nCoV-2019_29_LEFT	ACTTGTGTTCCCTTTTGTGCTGC
	nCoV-2019_29_RIGHT	AGTGTACTCTATAAGTTTTGATGGTGTGT
3	nCoV-2019_37_LEFT	ACACACCACTGGTTGTTACTCAC
	nCoV-2019_37_RIGHT	GTCCACACTCTCCTAGCACCAT
4	nCoV-2019_48_LEFT	TGTTGACACTGACTTAACAAAGCCT
	nCoV-2019_48_RIGHT	TAGATTACCAGAAGCAGCGTGC
5	nCoV-2019_59_LEFT	TCACGCATGATGTTTCATCTGCA
	nCoV-2019_59_RIGHT	AAGAGTCCTGTTACATTTTCAGCTTG
6	nCoV-2019_60_LEFT	TGATAGAGACCTTTATGACAAGTTGCA
	nCoV-2019_60_RIGHT	GGTACCAACAGCTTCTCTAGTAGC
7	nCoV-2019_77_LEFT	CCAGCAACTGTTTGTGGACCTA
	nCoV-2019_77_RIGHT	CAGCCCCTATTAAACAGCCTGC
8	nCoV-2019_84_LEFT	TGCTGTAGTTGTCTCAAGGGCT
	nCoV-2019_84_RIGHT	AGGTGTGAGTAACTGTTACAAACAAC
9	nCoV-2019_93_LEFT	TGAGGCTGGTTCTAAATCACCCA
	nCoV-2019_93_RIGHT	AGGTCTTCCTTGCCATGTTGAG
10	nCoV-2019_95_LEFT	TGAGGGAGCCTTGAATACACCA
	nCoV-2019_95_RIGHT	CAGTACGTTTTTGCCGAGGCTT

Supplemental Table S6.2, related to Methods. Primer design used for RT-PCR (Supplemental Figure S6.2A and B)

Materials and Methods

Wastewater Samples

Wastewater samples were collected at the Bozeman Water Reclamation Facility (BWRF) that receives and treats domestic, commercial, and industrial wastewater from the City of Bozeman, Montana (USA). Wastewater is sourced from the city limits ($\sim 60 \text{ km}^2$; 49,831 population) with an average flow rate of $\sim 2.31 \times 10^4 \text{ m}^3 / \text{d}$. Composite samples were collected from raw influent with automatic flow proportional sampler Liquistation CSF34 (Endress+Hauser) located at the entrance to the facility downstream of a rock trap. Autosampler was set to collect 150 mL of influent per 150'000 gal of flow ($\sim 5.68 \times 10^5 \text{ L}$) 7 AM to 7 AM. During collection temperature was kept $+2^\circ$ to $+6^\circ\text{C}$, and samples were stored at $+4^\circ\text{C}$ before processing (2-3 h). The composite sample was subsampled in three 500 mL aliquots. No permissions were required for collection of the wastewater.

Symptom Onset Data and Clinical Test Results

Suspected cases of COVID-19 were tested in a CLIA lab and instructed to self-quarantine until notified of the RT-qPCR test results. All laboratory confirmed positive cases of COVID-19 were contacted via telephone by local public health nurses to complete contact tracing. During this interview, the nurses collected recorded symptoms, symptom onset date, travel history, contact with other known laboratory confirmed cases, close contacts and activities on the two days before symptom onset up until notification of a positive test. Data collection was conducted as part of a public health response. Information on COVID-19 patients COVID-19 (gender, age, disease severity, etc) is not available. The study was reviewed by the Montana State University Institutional Review Board (IRB) For the Protection of Human Subjects (FWA 00000165) and

was exempt from IRB oversight in accordance with Code of Federal regulations, Part 46, section 101. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

Wastewater Sample Processing and RNA Extraction

Each wastewater sample (500 mL) was sequentially filtered through 20 mM, 5 mM (Sartorius Biolab Products) and 0.45 mM (Pall Corporation) membrane filters and concentrated down to 150-200 mL using Corning Spin-X UF concentrators with 100 kDa molecular weight cut-off. Total RNA from concentrated samples was extracted with RNeasy Mini Kit (QIAGEN) and eluted with 40 mL of RNase free buffer. This RNA was used as a template for RT-qPCR.

Reverse Transcription Quantitative PCR (RT-qPCR)

RT-qPCR was performed using two primers pairs (N1 and N2) and probes from 2019-nCoV CDC EUA Kit (IDT#10006606). SARS-CoV-2 in wastewater was detected and quantified using one-step RT-qPCR in ABI 7500 Fast Real-Time PCR System according to CDC guidelines and protocols (<https://www.fda.gov/media/134922/download>). 20 mL reactions included 8.5 mL of Nucleasefree Water, 1.5 mL of Primer and Probe mix, 5 mL of TaqPath 1-Step RT-qPCR Master Mix (ThermoFisher, A15299) and 5 mL of the template. Nuclease-free water was used as negative template control (NTC).

Amplification was performed as follows: 25°C for 2 min, 50°C for 15 min, 95°C for 2 min followed by 45 cycles of 95°C for 3 s and 55°C for 30 s. To quantify viral genome copy numbers in the samples, standard curves for N1 and N2 were generated using a dilution series of a positive template control (PTC) plasmid (IDT#10006625) with concentrations ranging from 10 to 10,000 copies per reaction. Three technical replicates were performed at each dilution. The

limit of detection was 10 copies of the control plasmid. The NTC showed no amplification over the 40 cycles of qPCR.

Run data was analyzed in SDS software v1.4 (Applied Biosystems). Threshold cycle (Ct) values were determined by manually adjusting the threshold to fall within exponential phase of the fluorescence curves and above any background signal. Ct values of PTC dilutions were plotted against $\log_{10}(\text{copy number})$ to generate standard curves. Linear regression analysis was performed in RStudio v1.2.1335 and the trend line equation ($Ct = [slope] \cdot 3 [\log_{10}(\text{copy number})] + b$) was used to calculate copy numbers from mean Ct values of technical replicates for each biological replicate. Primer efficiencies calculated as $E = \left(10^{\left(\frac{-1}{[slope]}\right)} - 1\right) \times 100\%$ were $150.36 \pm 12.13\%$ for N1 and $129.45 \pm 25.5\%$ for N2 ($n = 7$ runs, mean $\pm$ sd).

RT-PCR and SARS-CoV-2 Genome Sequencing

Reverse transcription was performed with 10 mL of RNA from SARS-CoV-2 positive wastewater sample using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with 10X RT Random Primers (Applied Biosystems) according to the supplier's protocol. Approximately 2000 viral RNA copies were used as an input for reverse transcription (estimated with qPCR).

The amplicon library for SARS-CoV-2 whole genome sequencing on Oxford Nanopore was generated as described in protocol developed by ARTIC Network (<https://artic.network/ncov-2019>). (Quick, 2020; Quick et al., 2017) Briefly, V3 primer pools containing 110 and 108 primers were used for the multiplex PCR (<https://artic.network/ncov-2019>). PCR reactions were performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs) with the following thermocycling conditions: 98°C for 2min, 35 cycles of 98°C for 15 s

and 65°C for 5 min, 35 cycles. Two resulting amplicon pools were combined and used for library preparation pipeline that included end preparation and Nanopore adaptors ligation. 20 ng of final library DNA was loaded onto the MinION flowcell for sequencing. A total of 304.77 Mb of raw sequencing data was collected.

PCR products used for Sanger sequencing were generated with a subset of primers from ARTIC V3 pools (Table S6.2). PCR reactions were performed as described above. PCR products were analyzed on 1% agarose gels stained with SYBR Safe (Thermo Fisher Scientific), remaining DNA was purified using DNA Clean & Concentrator kit (Zymo Research) and sent to Psomagen for Sanger sequencing. Each PCR product was sequenced with both forward and reverse primers used for PCR.

SARS-CoV-2 Genome Assembly

Nanopore raw reads (304.77 Mb) were base-called with MinKNOW software in high-accuracy mode. Successfully base-called reads (273.8 Mb) were further analyzed using the ARTIC bioinformatic pipeline for COVID-19 (<https://artic.network/ncov-2019>). Consensus sequence was generated with minimap2 and single nucleotide variants were called with nanopolish (both integrated in the pipeline) relative to Wuhan-Hu-1/2019 reference genome (GenBank: MN908947.3). (Li, 2018b; Quick et al., 2016; Fan Wu et al., 2020) The resulting assembly had nearly complete genome coverage (98.51%) with 6,875X average sequencing depth. Regions of the genome that were not captured by this sequencing method include 5' and 3' ends of the genome and a stretch of 170 nucleotides (22,346 – 22,515 nucleotide positions in reference genome), presumably due to amplicon drop-out. Consensus genome sequence was deposited to GISAID: EPI_ISL_437434 (<https://www.gisaid.org/>).

Phylogenetic Analysis

Phylogenetic analysis was performed by aligning the consensus sequence to 14,970 SARS-CoV-2 genomes retrieved from GISAID on 5/5/2020, 8:25:22 AM (<https://www.gisaid.org/>), using the FFT-NS-2 setting in MAFFT v7.429.(Kato et al., 2005; Kato & Standley, 2013) Columns composed of more than 70% gaps were removed with trimAl v1.2rev59.(Capella-Gutierrez et al., 2009) A maximum-likelihood phylogenetic tree was constructed from this alignment using IQTree in the Augur utility of Nextstrain.(Bedford et al., 2020) The APE v5.3 package in R was used to re-root the tree relative to RaTG13 bat coronavirus genome sequence,(Paradis et al., 2004) and the tree was plotted using ggtree v3.10 package in R.(Yu et al., 2016) The subtree, visualized in Figure 6.2B, was rendered in FigTree v1.4.4.(Rambaut & Drummond, 2012)

Position-specific Mutation Analysis

Position specific mutation analysis was conducted in R using the BioStrings package,(Pagès H, 2019) and chromatograms of Sanger sequencing reads were rendered in SnapGene (GSL Biotech; available at snapgene.com).

Quantification and Statistical Analysis

All statistical analyses were performed in RStudio v1.2.1335. Data in figures are shown as mean of three biological replicates (each with two technical replicates) $\pm$ standard deviation (sd). Estimated copy numbers in RT-qPCR reactions were used to calculate titers per liter of wastewater for each biological replicate. Viral RNA concentrations in the composite samples were normalized ($[\text{SARS-CoV-2 concentration}]_{\text{Normalized}} = [\text{SARS-CoV-2 concentration}] \times (\text{Daily flow} / \text{Average flow})$).

Correlation analysis was performed in RStudio using stats, astsa and spatialEco R packages. Symptom onset, positive COVID-19 tests and SARS-CoV-2 RNA concentration in wastewater measured with N1 or N2 primers were fit to a local polynomial regression (LOESS method) using the poly.regression wrapper function from spatialEco package. Resulting models were used to impute missing values (day w/o reported cases, periods between wastewater sampling days). Interpolated data were used for correlation analysis that was separately performed for the surge and resurgence of SARS-Co-V-2. Surge boundaries were determined as earliest reported symptom onset (left boundary) and date with last reported positive test (right boundary). The interval (mid-April – mid-May), when SARS-CoV-2 RNA was not detectable and no reported cases/symptoms were reported, was not included in the analysis. LOESS curves for wastewater data were displaced relative to epidemiological curves in one day increments from 14 to +14 days and Pearson correlation coefficient was calculated for each shift using corr function from stats R package.

Acknowledgements

Research in the Wiedenheft lab is supported by the National Institutes of Health (1R35GM134867), the Montana State University (MSU) Agricultural Experimental Station, the M.J. Murdock Charitable Trust, the Gianforte Family Foundation, and the MSU Office of the Vice President for Research. Wastewater data were collected as part of a SARS-CoV-2 surveillance effort supported by the City of Bozeman. We are grateful to Josh French, Justin Roberts, and the other dedicated wastewater technicians who made this work possible. We thank the GISAID EpiFluDatabase and depositing laboratories. The phylogenetic analysis in this article

would not have been possible without their willingness to share data. We thank the reviewers for helping identify a mutation that was introduced by incomplete trimming of an adaptor.

References Cited

- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *The Science of the total environment*, 728, 138764-138764. <https://doi.org/10.1016/j.scitotenv.2020.138764>
- Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M. L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T. R., Truong, M., Wolf, C. R., Nickerson, D. A., Rieder, M. J., Englund, J. A., Seattle Flu Study, I., Hadfield, J., Hodcroft, E. B., Huddleston, J., Moncla, L. H., Muller, N. F., Neher, R. A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J. S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G. L., Baird, G. S., Chu, H. Y., Shendure, J., & Jerome, K. R. (2020, Oct 30). Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516), 571-575. <https://doi.org/10.1126/science.abc0523>
- Biggerstaff, M., Jhung, M. A., Reed, C., Fry, A. M., Balluz, L., & Finelli, L. (2014). Influenza-like illness, the time to seek healthcare, and influenza antiviral receipt during the 2010-2011 influenza season-United States. *The Journal of infectious diseases*, 210(4), 535-544. <https://doi.org/10.1093/infdis/jiu224>
- Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009, Aug 1). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Castillo, A. E., Parra, B., Tapia, P., Acevedo, A., Lagos, J., Andrade, W., Arata, L., Leal, G., Barra, G., Tambley, C., Tognarelli, J., Bustos, P., Ulloa, S., Fasce, R., & Fernández, J. (2020). Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *Journal of medical virology*, 92(9), 1562-1566. <https://doi.org/10.1002/jmv.25797>
- Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O. G., Faria, N. R., Wang, C., Yu, G., Bushnell, B., Pan, C.-Y., Guevara, H., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Servellita, V., Hsu, E., Miller, S., Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Chu, H. Y., Shendure, J., Jerome, K. R., Anderson, C., Gangavarapu, K., Zeller, M., Spencer, E., Andersen, K. G., MacCannell, D., Paden, C. R., Li, Y., Zhang, J., Tong, S., Armstrong, G., Morrow, S., Willis, M., Matyas, B. T., Mase, S., Kasirye, O., Park, M., Masinde, G., Chan, C., Yu, A. T., Chai, S. J., Villarino, E., Bonin, B., Wadford, D. A., & Chiu, C. Y. (2020). Genomic surveillance reveals

- multiple introductions of SARS-CoV-2 into Northern California. *Science (New York, N.Y.)*, 369(6503), 582-587. <https://doi.org/10.1126/science.abb9263>
- Edgar, R. C. *MUSCLE: multiple sequence alignment with improved accuracy and speed*. Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004., <http://dx.doi.org/10.1109/csb.2004.1332560>
- Fauver, J. R., Petrone, M. E., Hodcroft, E. B., Shioda, K., Ehrlich, H. Y., Watts, A. G., Vogels, C. B. F., Brito, A. F., Alpert, T., Muyombwe, A., Razeq, J., Downing, R., Cheemarla, N. R., Wyllie, A. L., Kalinich, C. C., Ott, I. M., Quick, J., Loman, N. J., Neugebauer, K. M., Greninger, A. L., Jerome, K. R., Roychoudhury, P., Xie, H., Shrestha, L., Huang, M.-L., Pitzer, V. E., Iwasaki, A., Omer, S. B., Khan, K., Bogoch, I. I., Martinello, R. A., Foxman, E. F., Landry, M. L., Neher, R. A., Ko, A. I., & Grubaugh, N. D. (2020). Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell*, 181(5), 990-996.e995. <https://doi.org/10.1016/j.cell.2020.04.021>
- Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., Prill, M., Chai, S. J., Kirley, P. D., Alden, N. B., Kawasaki, B., Yousey-Hindes, K., Niccolai, L., Anderson, E. J., Openo, K. P., Weigel, A., Monroe, M. L., Ryan, P., Henderson, J., Kim, S., Como-Sabetti, K., Lynfield, R., Sosin, D., Torres, S., Muse, A., Bennett, N. M., Billing, L., Sutton, M., West, N., Schaffner, W., Talbot, H. K., Aquino, C., George, A., Budd, A., Brammer, L., Langley, G., Hall, A. J., & Fry, A. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR. Morbidity and mortality weekly report*, 69(15), 458-464. <https://doi.org/10.15585/mmwr.mm6915e3>
- Gonzalez-Reiche, A. S., Hernandez, M. M., Sullivan, M. J., Ciferri, B., Alshammary, H., Obla, A., Fabre, S., Kleiner, G., Polanco, J., Khan, Z., Albuquerque, B., van de Guchte, A., Dutta, J., Francoeur, N., Melo, B. S., Oussenko, I., Deikus, G., Soto, J., Sridhar, S. H., Wang, Y.-C., Twyman, K., Kasarskis, A., Altman, D. R., Smith, M., Sebra, R., Aberg, J., Krammer, F., García-Sastre, A., Luksza, M., Patel, G., Paniz-Mondolfi, A., Gitman, M., Sordillo, E. M., Simon, V., & van Bakel, H. (2020). Introductions and early spread of SARS-CoV-2 in the New York City area. *Science (New York, N.Y.)*, 369(6501), 297-301. <https://doi.org/10.1126/science.abc1917>
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome biology*, 20(1), 8-8. <https://doi.org/10.1186/s13059-018-1618-7>
- Hart, O. E., & Halden, R. U. (2020). Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility,

- economy, opportunities and challenges. *The Science of the total environment*, 730, 138875-138875. <https://doi.org/10.1016/j.scitotenv.2020.138875>
- Hindson, J. (2020). COVID-19: faecal-oral transmission? *Nature reviews. Gastroenterology & hepatology*, 17(5), 259-259. <https://doi.org/10.1038/s41575-020-0295-7>
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M. A., Weldon, W. C., Biggs, H. M., Uyeki, T. M., Pillai, S. K., & Washington State -nCo, V. C. I. T. (2020). First Case of 2019 Novel Coronavirus in the United States. *The New England journal of medicine*, 382(10), 929-936. <https://doi.org/10.1056/NEJMoa2001191>
- Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R., & Kuroda, M. (2020). Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PloS one*, 15(9), e0239403-e0239403. <https://doi.org/10.1371/journal.pone.0239403>
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J., & Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology*, 36(4), 338-345. <https://doi.org/10.1038/nbt.4060>
- Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 33(2), 511-518. <https://doi.org/10.1093/nar/gki198>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Sheffield, C.-G. G., McDanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Whelan, S. P., LaBranche, C. C., Saphire, E. O., & Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e819. <https://doi.org/10.1016/j.cell.2020.06.043>
- La Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L., & Suffredini, E. (2020). First detection of SARS-CoV-2 in untreated wastewaters in Italy. *The Science of the total environment*, 736, 139652-139652. <https://doi.org/10.1016/j.scitotenv.2020.139652>

- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*, 172(9), 577-582. <https://doi.org/10.7326/M20-0504>
- Li, H. (2018b). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., & Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*, 382(13), 1199-1207. <https://doi.org/10.1056/NEJMoa2001316>
- Lodder, W., & de Roda Husman, A. M. (2020). SARS-CoV-2 in wastewater: potential health risk, but also data source. *The lancet. Gastroenterology & hepatology*, 5(6), 533-534. [https://doi.org/10.1016/S2468-1253\(20\)30087-X](https://doi.org/10.1016/S2468-1253(20)30087-X)
- Lopez-Alvarez, D., Parra, B., & Cuellar, W. J. (2020). Genome Sequence of SARS-CoV-2 Isolate Cali-01, from Colombia, Obtained Using Oxford Nanopore MinION Sequencing. *Microbiology resource announcements*, 9(26), e00573-00520. <https://doi.org/10.1128/MRA.00573-20>
- Mallapaty, S. (2020, 2020/04/03). How sewage could reveal true scale of coronavirus outbreak. *Nature*, 580(7802), 176-177. <https://doi.org/10.1038/d41586-020-00973-x>
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020, 2020/05/20). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7), 511-516. <https://doi.org/10.1021/acs.estlett.0c00357>
- [Record #745 is using a reference type undefined in this output style.]
- Pagès H, A. P., Gentleman R, DebRoy S. (2019). Biostrings: Efficient manipulation of biological strings. *R package version 2.52.0*.
- Paradis, E., Claude, J., & Strimmer, K. (2004, Jan 22). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2), 289-290. <https://doi.org/10.1093/bioinformatics/btg412>

Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W., & Omer, S. B. (2020, Oct). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature biotechnology*, 38(10), 1164-1167.
<https://doi.org/10.1038/s41587-020-0684-z>

Quick, J. (2020). *nCoV-2019 sequencing protocol v1* ZappyLab, Inc.
<https://doi.org/10.17504/protocols.io.bbmuik6w>

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Jr., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., & Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols*, 12(6), 1261-1276.
<https://doi.org/10.1038/nprot.2017.066>

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J.-P., Cabeza-Cabrerizo, M., Camino-Sanchez, A., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G. G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallash, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Trina, R., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N. F., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., O'Shea, M. K., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Woelfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keita, S., Rambaut, A., Formenty, P., Gunther, S., & Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228-232. <https://doi.org/10.1038/nature16996>

[Record #404 is using a reference type undefined in this output style.]

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water research*, 181, 115942-115942.
<https://doi.org/10.1016/j.watres.2020.115942>

- Rimoldi, S. G., Stefani, F., Gigantiello, A., Polesello, S., Comandatore, F., Mileto, D., Maresca, M., Longobardi, C., Mancon, A., Romeri, F., Pagani, C., Cappelli, F., Roscioli, C., Moja, L., Gismondo, M. R., & Salerno, F. (2020). Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. *The Science of the total environment*, 744, 140911-140911. <https://doi.org/10.1016/j.scitotenv.2020.140911>
- Silverman, J. D., Hupert, N., & Washburne, A. D. (2020). Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Science translational medicine*, 12(554), eabc1126. <https://doi.org/10.1126/scitranslmed.abc1126>
- Stefanelli, P., Faggioni, G., Lo Presti, A., Fiore, S., Marchi, A., Benedetti, E., Fabiani, C., Anselmo, A., Ciammaruconi, A., Fortunato, A., De Santis, R., Fillo, S., Capobianchi, M. R., Gismondo, M. R., Ciervo, A., Rezza, G., Castrucci, M. R., Lista, F., & On Behalf Of Iss Covid-Study, G. (2020). Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 25(13), 2000305. <https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000305>
- Tang, A., Tong, Z.-D., Wang, H.-L., Dai, Y.-X., Li, K.-F., Liu, J.-N., Wu, W.-J., Yuan, C., Yu, M.-L., Li, P., & Yan, J.-B. (2020). Detection of Novel Coronavirus by RT-PCR in Stool Specimen from Asymptomatic Child, China. *Emerging infectious diseases*, 26(6), 1337-1339. <https://doi.org/10.3201/eid2606.200301>
- van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen, C. J., Pang, J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 83, 104351-104351. <https://doi.org/10.1016/j.meegid.2020.104351>
- Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*, 323(18), 1843-1844. <https://doi.org/10.1001/jama.2020.3786>
- Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A. D., Erickson, T. B., Chai, P. R., Thompson, J., & Alm, E. J. (2020, Jul 21). SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *MSystems*, 5(4). <https://doi.org/10.1128/mSystems.00614-20>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human

- respiratory disease in China. *Nature*, 579(7798), 265-269.
<https://doi.org/10.1038/s41586-020-2008-3>
- Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G., & Huang, X. (2020). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The lancet. Gastroenterology & hepatology*, 5(5), 434-435. [https://doi.org/10.1016/S2468-1253\(20\)30083-2](https://doi.org/10.1016/S2468-1253(20)30083-2)
- Wurtzer, S., Marechal, V., Mouchel, J. M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J. L., & Moulin, L. (2020, Dec). Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveill*, 25(50). <https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000776>
- Xiao, F., Sun, J., Xu, Y., Li, F., Huang, X., Li, H., Zhao, J., Huang, J., & Zhao, J. (2020). Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. *Emerging infectious diseases*, 26(8), 1920-1922. <https://doi.org/10.3201/eid2608.200681>
- Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., Zhang, H., Liu, H., Xia, H., Tang, J., Zhang, K., & Gong, S. (2020). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature medicine*, 26(4), 502-505. <https://doi.org/10.1038/s41591-020-0817-4>
- Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental science & technology*, 50(10), 5077-5085. <https://doi.org/10.1021/acs.est.6b00876>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2016, 2016/09/22). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28-36.
<https://doi.org/10.1111/2041-210x.12628>
- Zehender, G., Lai, A., Bergna, A., Meroni, L., Riva, A., Balotta, C., Tarkowski, M., Gabrieli, A., Bernacchia, D., Rusconi, S., Rizzardini, G., Antinori, S., & Galli, M. (2020). Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. *Journal of medical virology*, 92(9), 1637-1640. <https://doi.org/10.1002/jmv.25794>

SARS-COV-2 GENOMIC SURVEILLANCE IDENTIFIES NATURALLY
OCCURRING TRUNCATION OF ORF7A THAT
LIMITS IMMUNE SUPPRESSION

Contribution of Authors and Co-Authors

Manuscript in Appendix C

Author: Artem Nemudryi

Contributions: designed the research, performed the research, prepared manuscript figures, wrote the manuscript, wrote the manuscript

Co-Author: Anna Nemudraia

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Tanner Wiegand

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Joseph Nichols

Contributions: performed the research

Co-Author: Deann T. Snyder

Contributions: designed the research, performed the research

Co-Author: Jodi F. Hedges

Contributions: designed the research, performed the research

Co-Author: Calvin Cicha

Contributions: performed the research

Co-Author: Helen Lee

Contributions: performed the research

Co-Author: Karl K. Vanderwood

Contributions: performed the research

Co-Author: Diane Bimczok

Contributions: acquired samples

Co-Author: Mark A. Jutila

Contributions: wrote the manuscript

Co-Author: Blake Wiedenheft

Contributions: acquired samples, designed the research, wrote the manuscript

Manuscript Information

Artem Nemudryi, Anna Nemudraia, Tanner Wiegand, Joseph Nichols, Deann T. Snyder, Jodi F. Hedges, Calvin Cicha, Helen Lee, Karl K. Vanderwood, Diane Bimczok, Mark A. Jutila, and Blake Wiedenheft

Cell Reports

Status of Manuscript:

☐ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☒ Published in a peer-reviewed journal

Cell Press

Volume 35

Issue 9

1 June 2021

DOI: 10.1016/j.celrep.2021.109197

SARS-COV-2 GENOMIC SURVEILLANCE IDENTIFIES NATURALLY
OCCURRING TRUNCATION OF ORF7A THAT
LIMITS IMMUNE SUPPRESSION

Artem Nemudryi,¹ Anna Nemudraia,¹ Tanner Wiegand,^{1,3} Joseph Nichols,^{1,3} Deann T. Snyder,¹
Jodi F. Hedges,¹ Calvin Cicha,¹ Helen Lee,¹ Karl K. Vanderwood,² Diane Bimczok,¹ Mark A.
Jutila,¹ and Blake Wiedenheft¹

¹Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA. ²Gallatin City-County Health Department, Bozeman, MT 59715, USA. ³These authors contributed equally.

Abstract

Over 950,000 whole-genome sequences of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been determined for viruses isolated from around the world. These sequences are critical for understanding the spread and evolution of SARS-CoV-2. Using global phylogenomics, we show that mutations frequently occur in the C-terminal end of ORF7a. We isolate one of these mutant viruses from a patient sample and use viral challenge experiments to link this isolate (ORF7a^{Δ115}) to a growth defect. ORF7a is implicated in immune modulation, and we show that the C-terminal truncation negates anti-immune activities of the protein, which results in elevated type I interferon response to the viral infection. Collectively, this work indicates that ORF7a mutations occur frequently, and that these changes affect viral mechanisms responsible for suppressing the immune response.

Introduction

The spillover of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) into the human population has resulted in a global pandemic of coronavirus disease 2019 (COVID-19) with more than 2.8 million deaths worldwide (<https://coronavirus.jhu.edu/>) (Andersen et al., 2020). Although much of the research on this virus is focused on the Spike protein, recent reports demonstrate that accessory proteins of SARS-CoV-2 might be involved in COVID-19 pathogenesis by modulating antiviral host responses (Young et al., 2020; Zhang et al., 2020). SARS-CoV-2 uses multiple strategies to evade host immunity. Accessory proteins ORF3b, ORF6, and ORF7a antagonize various steps of type I interferon (IFN-I) production and signaling, while the proposed function of ORF8 is downregulating antigen presentation (Konno et al., 2020; Lei et al., 2020; Miorin et al., 2020; Tan et al., 2020; Xia et al., 2020). To occlude signal transmission from IFN receptors, ORF7a subverts phosphorylation of STAT2, suppressing transcriptional activation of antiviral IFN-stimulated genes (ISGs) that can otherwise restrict viral replication (Martin-Sancho et al., 2020; Xia et al., 2020). Although the proposed function for ORF7a is intracellular, antibodies against ORF7a are elevated in the serum of COVID-19 patients (Hachim et al., 2020). Work that is currently under review indicates that recombinant ORF7a protein interacts with CD14⁺ monocytes and triggers expression of pro-inflammatory cytokines, including interleukin-6 (IL-6) and tumor necrosis factor alpha (TNF- α). (Zhou et al., 2021) However, it is unclear if these interactions happen in COVID-19 patients. Together, these data suggest that ORF7a plays a dual role in SARS-CoV-2 infection by modulating both the IFN and inflammatory responses.

Results

SARS-CoV-2 Genomic Surveillance Identifies Truncations of ORF7a

Genome sequencing has been used to track the rise and spread of new SARS-CoV-2 lineages over the course of the pandemic (<https://www.gisaid.org/>). (Bedford et al., 2020; Korber et al., 2020) As part of this effort, we sequenced SARS-CoV-2 genomes isolated from patients in Bozeman, Montana (Figure 7.1A). In total, we determined 55 whole-genome sequences of SARS-CoV-2 viruses isolated from patients between April and July 2020 using an amplicon-based Nanopore protocol from the ARTIC network (<https://artic.network/>). (Quick et al., 2017; Tyson et al., 2020) To place the local outbreak in the context of the global SARS-CoV-2 evolutionary trends, we aligned 55 whole-genome sequences isolated from patients in Bozeman to 4,235 genomes subsampled (10 genomes per month per country) from 181,003 SARS-CoV-2 genomes downloaded from the GISAID. The subsampling protocol removes redundancy and bias introduced by uneven distribution of the global SARSCoV-2 sequencing effort. The resulting alignment was used to build a phylogenetic tree (Figure 7.1B). Of the 55 SARS-CoV-2 genomes determined from patients in Bozeman, only 1 was derived from the WA1 lineage (lineage A in Figure 7.1B) that was introduced to Washington state from Wuhan, China, in January 2020. (Bedford et al., 2020; Holshue et al., 2020) The remaining 54 genomes associate with clade B.1 and its offshoots, which are characterized by the D614G mutation in the spike and have prevailed globally since the spring of 2020 (Figure 7.1B). (Rambaut et al., 2020) The genetic variability in SARS-CoV-2 circulating in Bozeman (April to July 2020) represents 14 independent viral lineages, reflecting multiple introductions of the virus to the community from many different sources (Figure S7.1A).

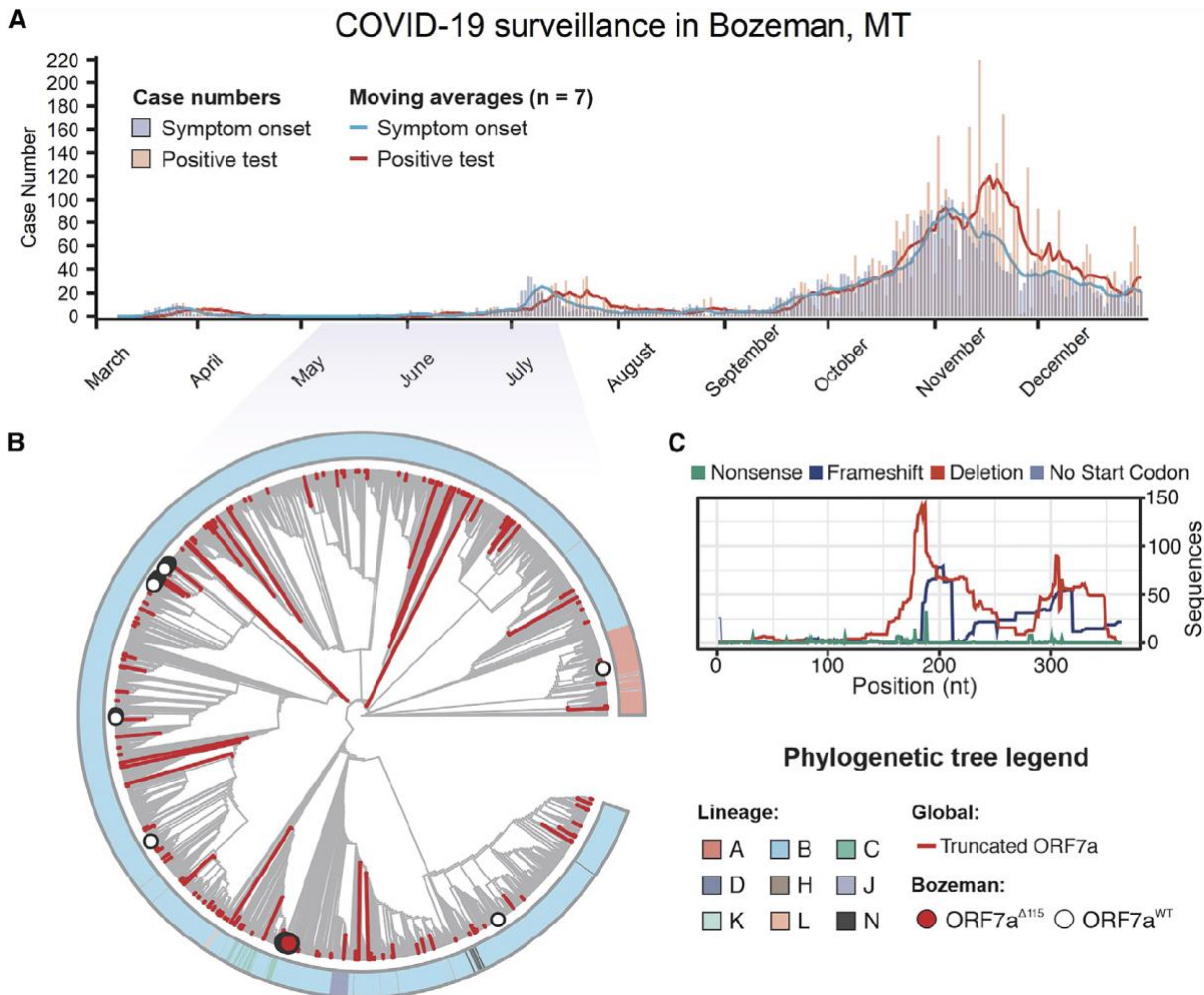


Figure 7.1. SARS-CoV-2 genomic surveillance identifies global reoccurrence of ORF7a truncations. (A) Symptom onset (purple) and PCR-based SARS-CoV-2 test results (coral) for patients in Bozeman, Montana, are shown with vertical bars. Seven-day moving averages, shown with lines, were used to indicate epidemiological trends. (B) Phylogenetic analysis of SARS-CoV-2 genomes sampled in Bozeman and globally. The tree was constructed from an alignment of 55 Bozeman samples and 4,871 genomes subsampled from GISAID. Subsampling was performed using Augur utility (<https://nextstrain.org>) by selecting 10 genomes per country per month since the start of the pandemic. Outer ring shows SARS-CoV-2 lineages assigned to genome sequences (Rambaut et al., 2020a). Major lineages include A (pink) that is associated with initial outbreak in China and B (blue) that emerged later in Europe. Minor lineages (i.e., C-N) are offshoots of lineage B. Red branches identify truncated ORF7a variants ($n = 205$) detected in the global data and merged into the alignment. The red dot highlights 7 of the 55 ORF7a variants that were isolated in Bozeman between April and July (2020). White dots highlight 48 viral genomes isolated in Bozeman that have wild-type ORF7a sequences. (C) Distribution of different mutations that occur along the ORF7a coding sequence.

During the annotation of these genomes, we noticed a reoccurring (7 out of 55 genomes) 115-nt deletion in the gene-encoding accessory protein ORF7a (27,549–27,644 nt). The ORF7a^{Δ115} mutation was found in patient swabs collected over a period of ~1.5 months. RT-PCR and Sanger sequencing verified that all these seven ORF7a^{Δ115} variants are bona fide mutations and not a sequencing artifact (Figures S7.1B and S7.1C). In addition to the ORF7a^{Δ115} deletion, these genomes also share 10 single-nucleotide variants (SNVs) compared with the Wuhan-Hu-1 reference genome. Seven of the 10 SNVs are found frequently worldwide and are a signature of the B.1 lineage of SARS-CoV-2.(Rambaut et al., 2020) The remaining three mutations are rare, do not co-occur in any other genomes on GISAID, and lead to amino acid changes in ORF3a (Q38P, L95F) and N (R195I) proteins. Interestingly, one of the genomes in the ORF7a^{Δ115} cluster has two additional SNVs that likely were acquired during circulation in the community. This virus was sampled from a patient 41 days after sampling the first ORF7a^{Δ115} variant, which agrees well with estimated SARS-CoV-2 mutation rates (2 nt/month).(van Dorp et al., 2020)

To determine if the ORF7a^{Δ115} genotype is unique to Bozeman, we downloaded an alignment of 180,971 SARS-CoV-2 genomes from GISAID, extracted ORF7a sequences, and determined their mutational profiles. In total, we identified 845 unique ORF7a gene variants that are different from the WuhanHu-1 reference sequence. Next, we looked for mutations with major effect on the ORF7a amino acid sequence. We identified 189 unique ORF7a variants (825 total sequences) with frameshifts, deletions, or missense mutations causing premature stop codons (Figure 7.1C). To understand the evolutionary history of these mutations, we integrated these genome sequences into our phylogenetic analysis. Detected ORF7a variants appear to have emerged independently on every continent and were not confined to any single lineage (Figure

7.1B). Next, we aligned translated ORF7a sequences to look for patterns. Most of the identified ORF7a mutations (126 out of 189) truncate the C terminus but preserve the N-terminal half of the protein (Figure 7.1C; Figure S7.1B). A portion (36.5%) of these truncated ORF7a variants appeared in two or more patient samples (116 maximum [max]). Although some of these ORF7a mutants appear to have arisen on multiple independent occasions, others come from genomes that form monophyletic clades, suggesting that viruses with ORF7a truncations are capable of transmission within a host population (Figure S7.1B).

ORF7a Truncation Has a Loss of Function Effect

ORF7a is a type I transmembrane (TM) protein with an N-terminal immunoglobulin-like (Ig-like) ectodomain, stalk, TM domain, and short cytosolic tail (Figure 7.2A). The two β sheets of the ectodomain are held together by two disulfide bonds (i.e., Cys23-Cys58 and Cys35-Cys67) (Figure 7.2B). The cytosolic tail of ORF7a contains a di-lysine (KRKTE) endoplasmic reticulum (ER) retrieval signal (ERRS) that mediates protein trafficking to the ER-Golgi intermediate compartment (ERGIC). (Nelson et al., 2005)

The $\Delta 115$ nt mutation in ORF7a introduces a premature stop codon that eliminates b5, b6, and b7, two of the cystines that form disulfides (Cys58 and Cys67), the TM, and the cytosolic tail (Figures 7.2A and 7.2B). This truncation is expected to destabilize the protein structure and significantly impact protein function and trafficking in the host cell. To determine how loss of TM and sorting signal affects protein localization, we cloned and expressed FLAG-tagged wild-type and truncated ORF7a proteins in HEK293T-hACE2 cells. The wild-type ORF7a accumulated in the perinuclear region of the cell, which is consistent with previously reported ERGIC localization (Figure 2D) (Martin-Sancho et al., 2020; Nelson et al., 2005). In

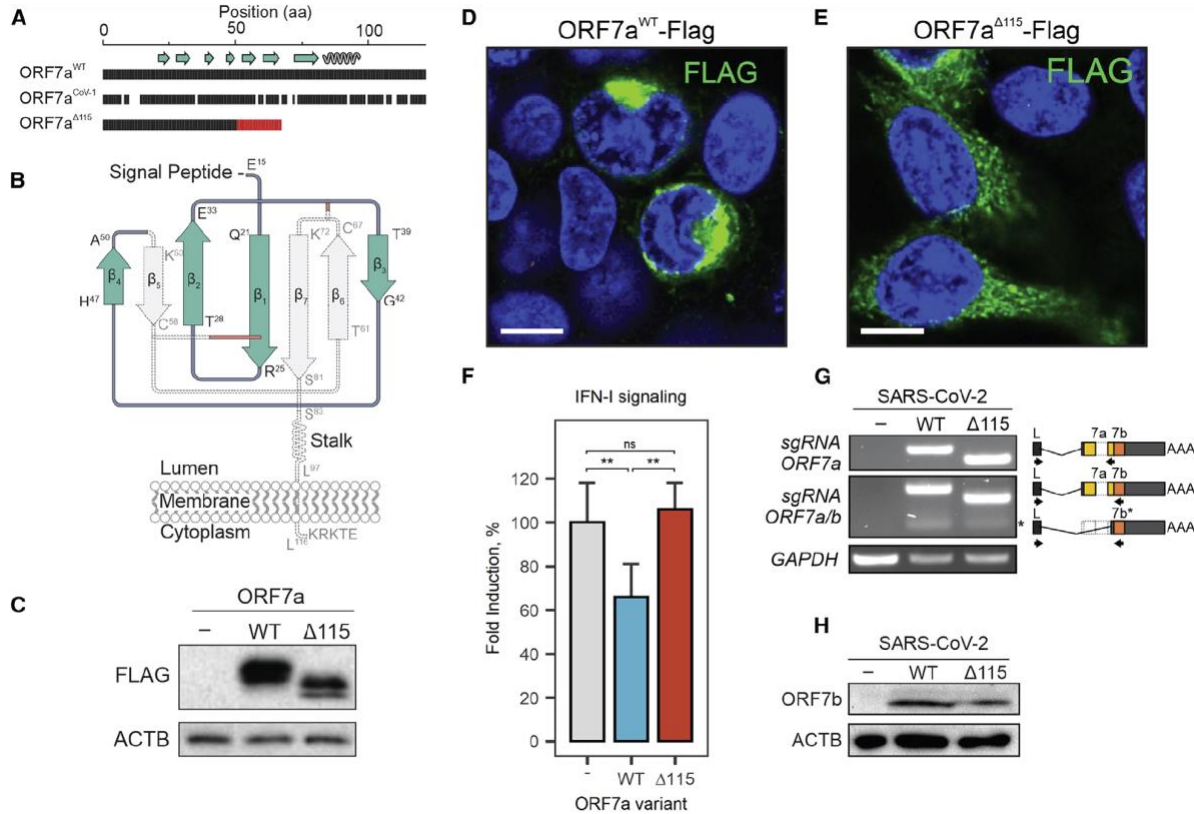


Figure 7.2. ORF7a truncation results in loss of function. (A) Amino acid (aa) sequence alignment of SARS-CoV-2 ORF7a^{WT}, ORF7a^{Δ115}, and SARS-CoV-1 ORF7a. Gaps show non-matching positions; red shows 17-aa sequence resulting from a frameshift in the ORF7a mutant. Beta strands (arrows) and alpha helices (coil) are shown above the alignment.

(B) Diagram of SARS-CoV-2 ORF7a Ig-like fold. Disulfide bonds that stabilize the b sandwich structure are shown with red lines. The portion of the protein eliminated by the deletion is shown in gray. (C) C-terminally FLAG-tagged ORF7a^{WT} and ORF7a^{Δ115} were cloned and overexpressed in HEK293T-hACE2. Protein expression was confirmed with western blot using anti-FLAG antibody. b-Actin (ACTB) was used as a loading control. (D and E) FLAG-tagged ORF7a^{WT} (D) or ORF7a^{Δ115} (E) expressed in HEK293T-hACE2 cells. Immunostaining was performed using an anti-FLAG antibody (green). Cell nuclei were stained with Hoechst 33342 (blue). White scale bar is 10 mm. (F) HEK293T cells with integrated ISRE-luciferase reporter were transfected with pLV-mCherry (“-”), ORF7a^{WT}, or ORF7a^{Δ115} plasmids. Transfected cells were treated with 5 ng/mL human recombinant IFN-α2b for 24 h, and induction of type I IFN signaling was measured with luciferase assay. Fold induction versus non-treated control was determined and normalized to mCherry control. Means (n = 6) were compared with one-way ANOVA (p = 0.00162). Pairwise comparisons were performed using post hoc Tukey’s test. Data are shown as mean ± SD. **p < 0.01; ns p > 0.05. (G) ORF7a and ORF7b sgRNAs were identified by RT-PCR. Lower molecular weight band corresponding to ORF7b sgRNA is indicated with asterisk (*). Diagram on the right shows primer (arrows) positions. Specificity of PCR products was confirmed with Sanger sequencing. GAPDH was used as a control for cDNA synthesis. (H) Lysates from SARS-CoV-2-infected cells were probed with antibodies raised against ORF7b protein. ACTB was used as a loading control.

contrast, the truncated ORF7a is distributed throughout the cytoplasm and does not associate with specific subcellular compartments, which is consistent with the loss of the TM domain and the ERRS signals required for protein targeting (Figure 7.2E).

Extent of the truncation and the defect in intracellular targeting suggests possible loss of ORF7a function. Along with several SARS-CoV-2 accessory proteins, ORF7a has been implicated in suppression of host IFN response to the infection.(Xia et al., 2020) To test the effect of the truncated ORF7a protein on IFN signaling, we generated reporter HEK293T cells with integrated luciferase reporter gene controlled by the interferon-stimulated response element (ISRE). In these cells, activation of interferon response drives reporter expression, which is quantified using a luciferase assay. To test immune suppression, we overexpressed the wild-type and the deletion mutant of ORF7a (ORF7a^{Δ115}) from plasmids and the stimulated cells with IFNα2b. The ORF7a WT protein suppressed IFN response by 34% (p = 0.0074) compared with cells transfected with control plasmid, which is consistent with recent results by Xia et al. (2020).(Xia et al., 2020) In contrast, the ORF7a^{Δ115} protein failed to suppress IFN response and showed no significant difference to a control (mCherry) plasmid (Figure 7.2F). These results indicate that the Δ115 mutation in *ORF7a* has a loss-of-function effect that negates antagonism of IFN signaling.

ORF7a Mutation Has No Collateral Effect on ORF7b

The ORF7a gene overlaps with the downstream ORF7b gene (Figure S7.2). To determine if the Δ115 mutation in ORF7a impacts ORF7b, we first examined whether it eliminates a transcription-regulatory sequence (TRS) that is required for subgenomic RNA (sgRNA) synthesis. None of the TRSs identified via direct RNA sequencing (RNA-seq) of SARS-CoV-2

overlap with the $\Delta 115$ mutation site (Kim et al., 2020),(Kim et al., 2020) suggesting that ORF7b transcription is not affected (Figure S7.2). To confirm this prediction, we used RT-PCR with primers that detect ORF7a and ORF7b sgRNAs (Figure 7.2G). In this assay, we infected 293T-hACE2 cells (MOI = 0.05) with ORF7a^{WT} or ORF7a ^{$\Delta 115$} viral strains and extracted total RNA from cells at 24 hours post-infection (hpi). Both viruses produced specific RT-PCR products corresponding to ORF7a and ORF7b sgRNAs. Finally, to verify that the ORF7a ^{$\Delta 115$} variant does not impact ORF7b translation, we probed cell lysates 24 hpi with an anti-ORF7b antibody (Figure 7.2H). We detected ORF7b protein in both viral strains with band intensities that suggest similar expression levels. Collectively, these data indicate that the ORF7a ^{$\Delta 115$} mutation has no collateral effect on the adjacent gene.

ORF7a ^{$\Delta 115$} Isolate Displays a Replication Defect

To investigate the phenotype of SARS-CoV-2 $\Delta 115$ variant, we infected 293T-hACE2 and Vero E6 cells with ORF7a^{WT} and ORF7a ^{$\Delta 115$} viruses at an MOI of 0.05 and measured viral RNA replication using qRT-PCR (Figures 7.3A and 7.3B). In this assay, we used viral strains sharing the same haplotype (C241T, C3037T, C14408T, A23403G) that defines the SARS-CoV-2 B.1 lineage and its derivatives (Figure S7.3A).(Rambaut et al., 2020) Over the course of 120 h, ORF7a^{WT} viral RNA level increased ~960-fold in supernatants from Vero E6 cells (Figure 7.3A) and ~270-fold in supernatants from 293T-hACE2 cells (Figure 7.3B). The ORF7a ^{$\Delta 115$} virus displayed limited replication that resulted in only 18.6-fold and 7-fold RNA level increase at 120 hpi in VeroE6 and 293T-hACE2, respectively. This growth defect reaches statistical significance starting at 6 hpi in Vero E6 cells ($p = 0.0135$) and 24 hpi in 293T-hACE2 cells ($p = 6.2 \times 10^{-5}$).

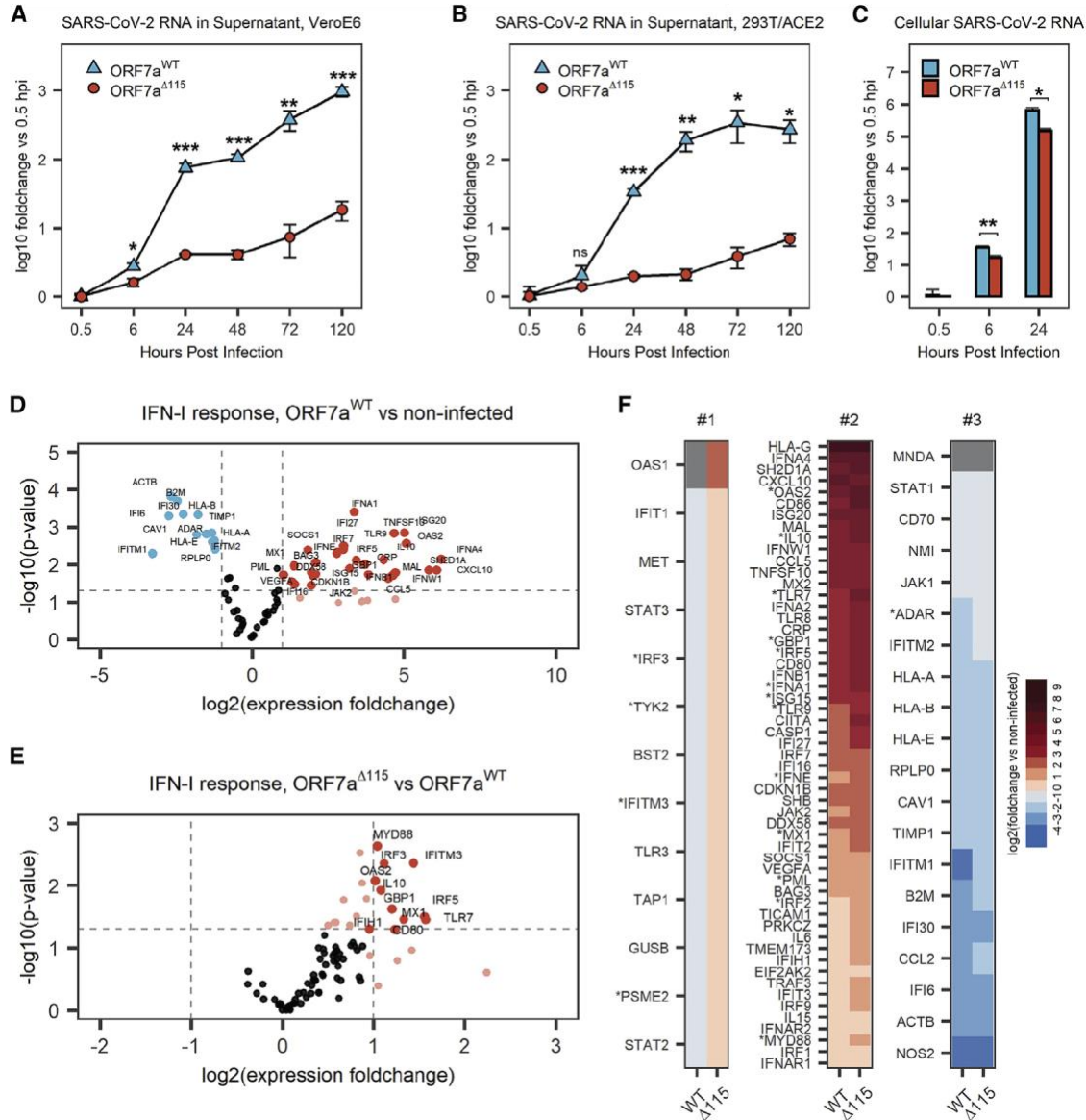


Figure 7.3. The ORF7a^{Δ115} mutation results in replication and immune suppression defects (A and B) VeroE6 (A) or HEK293T-hACE2 (B) cells were infected with ORF7a^{WT} or ORF7a^{Δ115}SARS-CoV-2 strain at MOI = 0.05. Viral RNA was measured in supernatant using qRT-PCR at different time points after infection. Measured RNA levels were normalized (using DCt method) to RNA levels at 0.5 hpi. (C) Total RNA was extracted from infected HEK293T-hACE2 cells, and viral RNA was measured. The DDCt method was used to normalize viral RNA levels to host RNA (*ACTB*) and 0.5 hpi time point. Data in (A)–(C) are presented as mean ± SD of three biological replicates. Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001, or ns p > 0.05 (no significant difference). (D) Volcano plot showing IFN-I response in 293T-hACE2 cells infected with ORF7a^{WT} SARS-CoV-2 strain at MOI = 0.05. Expression of IFN-I response genes was studied 24 hpi using qRT-PCR array targeting 91 human transcripts (88 targets and 3 references). Experiment was performed in three biological replicates. Dashed lines show

regulation (≥ 2 -fold) and statistical significance ($p < 0.05$) thresholds. Each dot represents mean ($n = 3$) normalized expression of a single gene relative to non-infected host. Genes that passed the threshold are labeled. (E) Volcano plot showing IFN-I response in ORF7a Δ^{115} versus ORF7a^{WT} infection. (F) Data shown in (D) and Figure S7.3A were plotted as a heatmap. Genes were classified into three groups. Group #1 genes are oppositely regulated between the two viral strains. Group #2 genes are upregulated by both ORF7a^{WT} and ORF7a Δ^{115} . Group #3 genes are downregulated in both. Genes that have statistically significant difference in expression between two viral strains are marked with asterisk. Gray shows genes with no detectable expression over 40 cycles of qRT-PCR.

To determine if this phenotype results from a defect in viral egress only or if viral RNA replication in the host is also affected, we examined early steps in the infection. We extracted total RNA from the infected 293T-hACE2 cells at 0.5, 6, and 24 hpi and measured SARS-CoV-2 RNA with qRT-PCR. Compared with wild type, ORF7a Δ^{115} RNA level was reduced 2.1-fold inside the host at 6 hpi ($p = 0.002$) and 4.3fold at 24 hpi ($p = 0.02$) (Figure 7.3C). This decrease in viral RNA in the host cell differs from the corresponding decrease in the supernatant RNA at 24 hpi (4.3-fold versus 16.8-fold, respectively; Figures 7.3B and 7.3C). This disparity suggests that observed growth defect in the SARS-CoV-2 Δ^{115} variant might result from defects in genomic RNA replication and transcription, as well as in viral egress.

ORF7a Δ^{115} Virus Induces Elevated IFN-I response *in vitro*

Suppressed or delayed IFN responses characterize SARS-CoV2 infections.(Lei et al., 2020) The C-terminal truncation of ORF7a limits suppression of the type I IFN signaling (Figure 7.2F). Therefore, we hypothesized that ORF7a Δ^{115} virus isolated from the community is deficient in antagonizing IFN response.

To test this hypothesis, we measured expression of type I IFN response genes upon infection with ORF7a^{WT} and ORF7a Δ^{115} strains using a qRT-PCR array targeting 91 human transcripts. The array included three reference genes that we used to normalize expression of 88

target genes. In agreement with recent transcriptomic studies, we detected a type I IFN-mediated antiviral response after infecting HEK293T-hACE2 with SARS-CoV-2 isolates (Figure 7.3D; Figure S7.3B). (Banerjee et al., 2021; Blanco-Melo et al., 2020) Out of 88 type I IFN response genes included in the qRT-PCR array, 42 were differentially expressed (≥ 2 -fold change, $p < 0.05$) in ORF7a^{WT}-infected cells and 55 in ORF7a ^{Δ 115} infection, as compared with the non-infected control (Figure 7.3D; Figure S7.3B). Both viruses stimulated expression of multiple type I interferon genes, antiviral ISGs, ILs, and proinflammatory cytokines.

To understand the effect of ORF7a ^{Δ 115} on the host type I IFN response, we compared expression of genes targeted with qRT-PCR array between the two viruses. In general, the ORF7a ^{Δ 115} variant induces elevated type I IFN response to the infection (Figures S7.3C and S7.3D). The median expression fold change of 88 genes targeted in the qRT-PCR array was 1.72 in ORF7a^{WT} and 2.51 in ORF7a ^{Δ 115} virus ($p = 1.4\text{E-}12$; Figure S7.3C). Analysis of genes differentially expressed between the two viral infections results in an asymmetric volcano plot (Figure 7.3E). A subset of ISGs had expression levels that are significantly different ($p < 0.05$) between the two viral infections (Figures 7.3E and 7.3F). This subset includes sensors (*TLR7*), signal transducers (*MYD88*, *OAS2*), transcriptional regulators (*IRF3*, *IRF5*), and restriction factors (*GBP1*, *IFITM3*, *MX1*) known to combat infections by RNA viruses. (Schoggins, 2019) One of these differentially expressed ISGs (*IFITM3*) restricts SARS-CoV-2 entry. (Shi et al., 2021; Zang et al., 2020) Additionally, polymorphisms in *IFITM3*, *MX1*, and *TLR7* are linked to the severity of COVID-19. (Andolfo et al., 2021; Pati et al., 2021; Y. Zhang, L. Qin, et al., 2020) Upregulation of ISGs indicates that the ORF7a ^{Δ 115} virus is inept at antagonizing the IFN-I

response to the infection, which is consistent with results from our ISRE reporter assay (Figure 7.2F).

Discussion

In this study, we use genomic surveillance and phylogenetics to identify ORF7a variants that have emerged globally throughout the SARS-CoV-2 pandemic. Local occurrence of ORF7a mutations has been reported by others.(Addetia et al., 2020; Holland et al., 2020; Rosenthal et al., 2020) However, the effect of these mutations on viral fitness and host immune responses has not been investigated. Here, we isolated a virus with a deletion mutation in ORF7a (ORF7a^{Δ115}) and show that this genotype results in an *in vitro* growth defect. This growth defect is associated with elevated IFN response to SARS-CoV-2.

Interferon systems are a frontline of host defense that signals infection and elicits an antiviral state.(McNab et al., 2015) Pretreatment with type I, type II, and type III IFNs restricts SARSCoV-2 infection and replication in cell culture.(Mantlo et al., 2020; Miorin et al., 2020) SARS-CoV-2 deploys multiple proteins (e.g., nsp6, nsp13, ORF6, ORF7a, ORF7b, and ORF9b) to shut down IFN signaling and dampen innate immune responses,(Jiang et al., 2020; Lei et al., 2020; Xia et al., 2020) in particular, when overexpressed ORF7a subverts STAT2 phosphorylation, blocking IFN-dependent transcriptional activation of antiviral ISGs.(Xia et al., 2020) Both type I and III IFNs signal through the JAK-STAT pathway and overlap substantially in the transcriptional responses they induce.(Kotenko & Durbin, 2017) Therefore, ORF7a likely suppresses both IFN types; however, its role in suppressing the type III IFN response has yet to be verified.(Xia et al., 2020) Using the ISRE reporter system, we show that C-terminal truncation negates the ability of ORF7a to suppress IFN signaling (Figure 7.2). The ORF7a^{Δ115} variant shows

limited replication in HEK293T-hACE2, as well as in Vero E6, the latter of which does not express type I IFN genes.(Emeny & Morgan, 1979; Osada et al., 2014) Although type I IFN genes are deleted in Vero E6 cells, the type III IFNs are intact and are activated upon infection with RNA viruses, which results in ISG.(Stoltz & Klingstrom, 2010; C. Wang et al., 2020) ISGs exert anti-viral activities that combat the infection.(Schoggins, 2019) We hypothesize that the replication defect of the ORF7a^{Δ115} strain in Vero E6 and 293T-hACE2 cells results from an IFN-dependent response that the virus fails to suppress.

Several groups have replaced ORF7a with a reporter gene.(Hou et al., 2020; Thi Nhu Thao et al., 2020; Xie et al., 2020) Thi Nhu Thao et al. (2020) show that complete deletion of ORF7a reduces replication of the synthetic virus, which agrees well with growth defect in the ORF7a^{Δ115} strain.(Thi Nhu Thao et al., 2020) However, two other studies have not observed this effect. We believe that our results with a naturally occurring ORF7a deletion provide valuable insight to this discussion. Finally, it should be noted that different mutations (i.e., complete deletion versus partial) might have a different outcome for the viral phenotype.

Cytokine profiling and RNA-seq of clinical samples reveal suppressed or delayed IFN responses in COVID-19 patients.(Arunachalam et al., 2020; Hadjadj et al., 2020; Lucas et al., 2020) Dysregulated immune responses coupled with exacerbated inflammatory cytokines drive pathogenesis of the disease and correlate with its severity.(Tay et al., 2020; B. Zhang et al., 2020; Zhou et al., 2020) Here, we detected enhanced IFN signaling in cells infected with a naturally occurring ORF7a^{Δ115} variant of SARS-CoV-2.

Given the importance of the IFN response in COVID-19, we anticipate that ORF7a mutations might affect SARS-CoV-2 pathogenicity. Clinical comparisons will be required to examine this potential effect on COVID-19 features.

Deletions in ORF7a, ORF7b, and ORF8 have previously been reported.(Addetia et al., 2020; Holland et al., 2020; Rosenthal et al., 2020; Su et al., 2020) Similar deletions emerged in Middle East Respiratory Syndrome (MERS) and SARS-CoV, which were linked to viral attenuati.(Chinese, 2004; Muth et al., 2018) A 382-nt deletion in ORF8 is associated with milder SARS-CoV-2 infection in patients.(Young et al., 2020) Although this mutation affects clinical features of COVID-19, it has no evident effect on viral replication or host transcriptional responses in human cell culture.(Gamage et al., 2020)

Several studies illustrate how persistent SARS-CoV-2 infections in immunocompromised hosts associate with accelerated viral evolution.(Avanzato et al., 2020; Choi et al., 2020) In these hosts there is less selective pressure on viral anti-immune proteins, which might permit loss-of-function mutations in accessory genes. The ORF7a^{Δ115} variant that we isolated was present in only seven genomes over a course of ~1.5 months in a local community and then disappeared. We hypothesize that deletions in accessory genes, such as ORF7a^{Δ115}, might emerge in immunocompromised patients but do not persist in the immunocompetent population. Similarly, a deletion mutation in ORF8 (i.e., D382), which was associated with milder COVID-19 disease, disappeared soon after it was discovered.(Young et al., 2020) However, founder effect or mutations that co-occur with other mutations that increase transmissibility (e.g., spike mutations) can result in fixation of an otherwise attenuating mutation. A new more transmissible lineage of SARS-CoV-2 (i.e., B.1.1.7) has recently.(Andrew Rambaut, 2020; Davies et al., 2021) Aside

from changes in the Spike protein, this lineage also includes a premature stop codon in the accessory ORF8 that truncates most of the protein.(Volz et al., 2021) This lineage might illustrate the second scenario for fixing mutations in accessory genes.

Limitations of Study

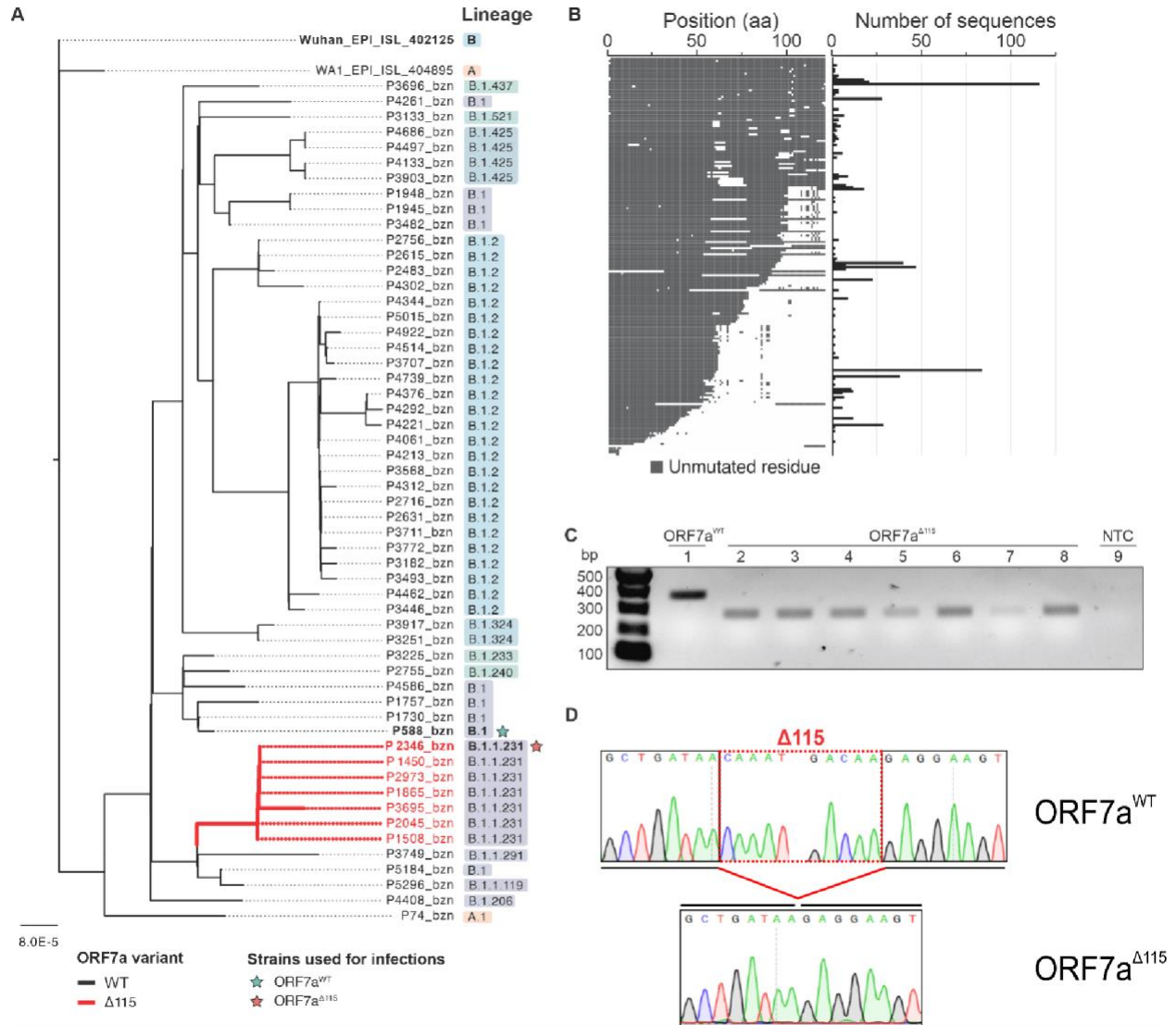
The viral replication defect and cellular immune response assay presented here were performed using HEK293T-hACE2 and Vero E6 cells that are not the natural targets of SARS-CoV-2. However, both cell lines are permissive for SARS-CoV-2 infection and are used extensively to study efficacy of antiviral compounds,(Riva et al., 2020) virion assembly,(Klein et al., 2020) SARS-CoV-2 entry,(Hoffmann et al., 2020) replication,(Thi Nhu Thao et al., 2020) and anti-immune activities of SARS-CoV-2 proteins.(Xia et al., 2020) Although we acknowledge that our conclusions are limited to Vero E6 and 293T-hACE2 cells, we anticipate that the results presented here will provide important context for clinical comparisons, similar to those recently published for ORF8.(Young et al., 2020)

Studies of naturally occurring SARS-CoV-2 variants have limitations, including unavailability of isogenic controls. Here, we compared two viral isolates, ORF7a^{WT} and ORF7a^{Δ115}, that in addition to Δ115 differ at several other nucleotide positions (Figure S7.3A). Although ISRE reporter assay unambiguously links the immunosuppression defect to the truncation of ORF7a, it is not possible to rule out a role for these mutations on the viral life cycle.

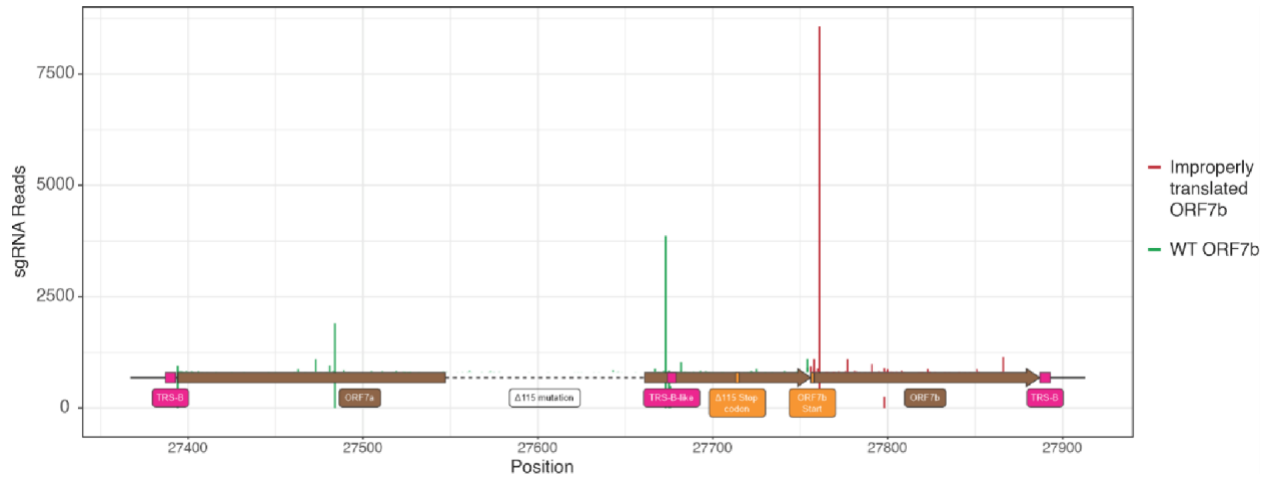
Acknowledgements

We are grateful to members of Bozeman Health that provided de-identified patient samples, specifically C. Nero, D. Smoot, W. Wallace, C. Faurot-Daniels, V. Lawrence, and M. Blauvelt. We are also grateful to M. Flenniken, K. Daughenbaugh, and other members of the COVID task force at MSU for assistance establishing the COVID testing center. We thank the Lefcort lab for generous use of their fluorescent confocal microscope and Dr. Pincus for providing the HEK 293T-hACE2 cells. Research in the Wiedenheft lab is supported by the National Institutes of Health (1R35GM134867), the Montana State University Agricultural Experimental Station, the MJ Murdock Charitable Trust, the Gianforte Foundation, and the MSU Office of the Vice President for Research. We thank the GISAID's EpiFlu Database and contributing laboratories. The phylogenetic analysis in this paper would not have been possible without their willingness to share data. The graphical abstract was created with BioRender. B.W. is the founder of SurGene LLC and VIRIS Detection Systems Inc. B.W., A. Nemudryi, and A. Nemudraia are inventors on patents related to CRISPR-Cas systems and applications thereof.

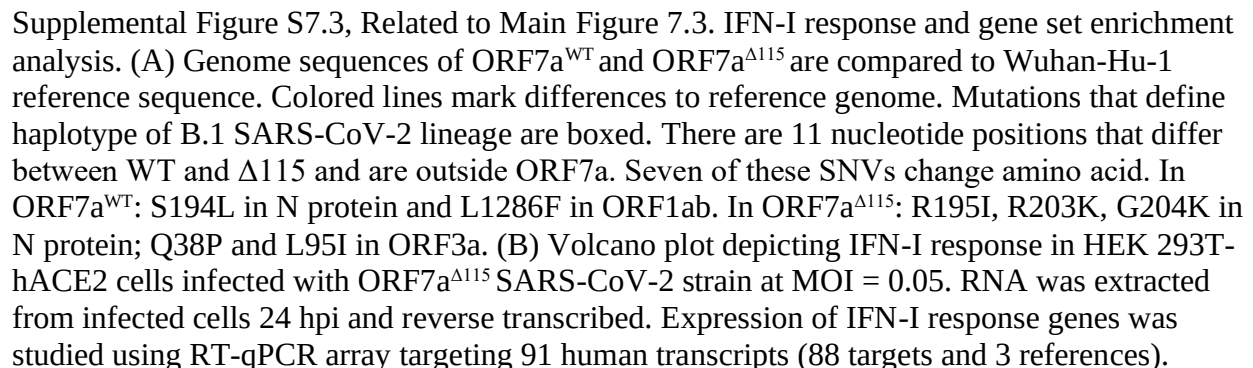
Supplemental Figures



Supplemental Figure S7.1. Phylogenetic and mutational analysis of ORF7a gene sequences. Related to main Figure 7.1. (A) A maximum-likelihood phylogenetic tree indicates that ORF7a^{Δ115} strains form a monophyletic clade (red branches). Viral isolate P588 (teal star) was chosen as a "wildtype" control in our experiments. This strain shares a recent common ancestor with the predominant ORF7a^{Δ115} clade. SARS-CoV-2 lineages were assigned using pangolin utility (cov-lineages.org). (Rambaut et al., 2020) (B) Alignment of unique ORF7a protein sequences with five or more mutations as compared to the SARS-CoV-2 reference strain (Wuhan-Hu-1). Conserved residues are depicted in gray, while mutations and deletions are shown in white. The frequency of each genotype is shown on the right. (C) RT-PCR with primers targeting regions that flank the ORF7a Δ115 deletion sampled in Bozeman, MT. (D) All PCR products in (C) were Sanger sequenced. Representative sequencing chromatograms are shown. Deleted region is shown with dotted red box.



Supplemental Figure S7.2, Mapping TRS-B sites required for ORF7a and ORF7b discontinuous sgRNA transcription. Related to Main Figure 7.2F-H. Direct RNA sequencing data published by Kim et al was used to map TRS-B sites in SARS-CoV-2 ORF7a and ORF7b genes.(Kim et al., 2020) Vertical lines indicate number of reads spanning the junction between TRS-L and “body” of the gene. Dashed line indicate region deleted in ORF7a^{Δ115} strain. The deletion does not remove any of the mapped TRS-B sites suggesting that it has not affected ORF7b sgRNA transcription.



Experiment was performed in 3 biological replicates. Dashed lines show regulation (≥ 2 -fold) and statistical significance thresholds (p-value < 0.05). Each dot represents mean (n = 3) normalized expression of a single gene relative to non-infected host. Genes that passed the threshold are labeled. (C) Plot showing distribution of $\log_2(\text{expression fold change})$ values in ORF7a^{WT} and ORF7a ^{$\Delta 115$} infection vs non-infected control. Each dot represents relative normalized expression for one transcript targeted with the RT-qPCR array. Gray lines connect values for the same transcript in the infection with ORF7a^{WT} and ORF7a ^{$\Delta 115$} variants. Horizontal red bar shows median value. Median ISG activation between two infections was compared using Wilcoxon signed-rank test (***, p-value < 0.001) (D) List of IFN-I response genes that were significantly regulated (p-value < 0.05 , ≥ 2 -fold change in expression) upon infection with ORF7a ^{$\Delta 115$} and ORF7a^{WT} (vs non-infected control) were analyzed for enrichment of associated Reactome Pathways.(Jassal et al., 2020; Zhou et al., 2021) Gene set enrichment analysis (GSEA) was performed using clusterProfiler package in RStudio.(Yu et al., 2012) *Down.WT* – genes downregulated during infection with ORF7a^{WT}; *Up.WT* – genes upregulated during infection with ORF7a^{WT}; *Down. $\Delta 115$* – Reactome pathways enriched in genes downregulated during infection with ORF7a ^{$\Delta 115$} ; *Up. $\Delta 115$* – genes upregulated during infection with ORF7a ^{$\Delta 115$} . *Up. $\Delta 115$ _vs_WT* – Reactome pathways enriched in genes upregulated during infection with ORF7a ^{$\Delta 115$} vs ORF7a^{WT} (Figure 7.3E).

Materials and Methods

Cell Cultures

HEK293T-hACE2 cells (BEI, NR-52511) were generously provided by Dr. Seth Pincus (Montana State University) and maintained at 37°C and 5% CO₂ in Dulbecco's Modified Eagle Medium (DMEM, GIBCO, cat. #12100-061) supplemented with 10% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1), sodium bicarbonate (3.7 g/L), 50 I.U./mL penicillin and 50 µg/mL streptomycin. Vero E6 cells (ATCC® cat. #CRL-1586) were maintained at 37°C and 5% CO₂ in Eagle's Modified Eagle Medium (EMEM, ATCC®, cat. #30-2003) supplemented with 10% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1) and 50 I.U./mL penicillin and 50 µg/mL streptomycin. All cell lines tested negative for mycoplasma.

Plasmids

Human-codon optimized dsDNA gene fragments (gBlocks) encoding for SARS-CoV-2 C-terminally FLAG-tagged ORF7a^{WT} and ORF7a^{Δ115} proteins were synthesized by Integrated DNA Technologies (IDT). Sequence of ORF7a from Wuhan-Hu-1 reference (GenBank MN908947.3) was used for ORF7a^{WT} gBlock. The gene fragments were cloned into the pEGFP-N1 (Clontech cat# 6085-1) backbone to replace the eGFP gene. pLV-mCherry was a gift from Pantelis Tsoulfas (Addgene plasmid # 36084; <http://www.addgene.org/36084/>; RRID:Addgene_36084).

Antibodies

Western blotting: Rabbit anti-ACTB (Cat: AC026, 1:20,000) antibodies were from ABclonal, Mouse anti-FLAG antibodies (Cat: MA1-91878-1MG, 1:1000) were from

ThermoFisher Scientific, sheep anti-ORF7b antibodies (1:120) were from MRC I PPU, College of Life Sciences, University of Dundee, Scotland; Goat anti-mouse IgG peroxidase-conjugated (Cat: 115-035-003, 1:10,000), Goat anti-rabbit IgG peroxidase-conjugated (Cat: 111-035-003, 1:10,000) and Donkey Anti-Sheep IgG peroxidase-conjugated (Cat: 713-035-147, 1:10,000) antibodies were from Jackson ImmunoResearch. Immunocytochemistry: Mouse anti-FLAG antibodies (Cat: MA1-91878-1MG, 1:200), Goat anti-mouse-AlexaFluor-488 (Cat: A11001, 1:2,000) antibodies were from Invitrogen.

Human Clinical Sample Collection and Preparation

Clinical samples were obtained with local IRB approval (protocol #DB033020) and informed consent from patients undergoing testing for SARS-CoV-2 at Bozeman Health Deaconess Hospital. Patients tested for SARS-CoV-2 included both in-patients and out-patients. The latter included individuals who developed symptoms and sought medical care and asymptomatic (at the time of testing) individuals who were exposed to known COVID-19 case and therefore were tested. Nasopharyngeal swabs from patients that tested positive for SARS-CoV-2 were collected in viral transport media. RNA was extracted from all patient samples using QIAamp Viral RNA Mini Kit (QIAGEN) a biosafety level 3 (BSL3) laboratory. All samples were heat-inactivated before removing from BSL3.

Production and Titration of Coronavirus Stocks

The nasopharyngeal swabs in the viral transport media were used to generate viral stocks as previously described.(Harcourt et al., 2020) Briefly, 100 ul of the viral transport media was two-fold serially diluted and applied to Vero E6 cells When infected cells showed extensive cytopathic effect, the media was collected and utilized to generate a greater volume second

passage in Vero E6 cells. When CPE was apparent, supernatants were centrifuged 1,000 RCF for 5 minutes to remove cellular debris. Virus titer in clarified supernatants was determined with plaque assay in Vero E6 cells as described.(Loveday et al., 2021) For plaque assay, Vero E6 cells were incubated with viral inoculum at limiting dilutions. Inoculated cells were overlaid with either 0.75% methylcellulose, DMEM supplemented with 2% FBS and 1% pen-strep and incubated for 4 days. Cells were fixed and stained with 0.5% methylene blue in 70% ethanol. Plaques were counted and the overall titer was calculated. Viral stock's identity was confirmed via whole genome sequencing on Oxford Nanopore. All SARS-CoV-2 experiments were performed in a BSL3 laboratory.

Symptom Onset Data and Clinical Test Results

Suspect cases of COVID-19 were tested in a CLIA lab and instructed to self-quarantine until notified of the RT-qPCR test results. All laboratory confirmed positive cases of COVID-19 were contacted via telephone by local public health nurses to complete contact tracing. During this interview, the nurses collected recorded symptoms, symptom onset date, travel history, contact with other known laboratory confirmed cases, close contacts and activities on the two days before symptom onset up until notification of a positive test. Data collection was conducted as part of a public health response. The study was reviewed by the Montana State University Institutional Review Board (IRB) For the Protection of Human Subjects (FWA 00000165) and was exempt from IRB oversight in accordance with Code of Federal regulations, Part 46, section 101. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

Quantitative Reverse Transcription PCR (qRT-PCR)

qRT-PCR was performed using CDC primers (N1 and N2) and probes from the 2019-nCoV RUO Kit (IDT# 10006713). SARS-CoV-2 RNA was quantified using one-step qRT-PCR in ABI 7500 Fast Real-Time PCR System according to CDC protocol (<https://www.fda.gov/media/134922/download>). In brief, 20 μ L reactions included 8.5 μ L of Nuclease-free Water, 1.5 μ L of Primer and Probe mix (IDT, 10006713), 5 μ L of TaqPath 1-Step RT-qPCR Master Mix (ThermoFisher, A15299) and 5 μ L of the template. Nuclease-free water was used as negative template control (NTC). Amplification was performed using following program: 25°C for 2 min, 50°C for 15 min, 95°C for 2 min followed by 45 cycles of 95°C for 3 s and 55°C for 30 s. Run data was analyzed in SDS software v1.4 (Applied Biosystems). The NTC showed no amplification throughout the 40 cycles of qPCR.

RT-PCR and SARS-CoV-2 Genome Sequencing

For sequencing 10 μ L of RNA from SARS-CoV-2 positive patient sample was reverse transcribed using SuperScript IV (Thermo Fisher Scientific) according to the supplier's protocol. The protocol developed by ARTIC Network was used to generate and sequence amplicon library that covers whole SARS-CoV-2 genome on Oxford Nanopore using ligation sequencing kit (SQK-LSK109) (<https://artic.network/ncov-2019>). Briefly, two multiplex PCR reactions were performed with primer pools from ARTIC nCoV-2019 V3 Panel (IDT; Table S3) using Q5 High-Fidelity DNA Polymerase (New England Biolabs). Reactions were performed with the following thermocycling conditions: 98°C for 2min, 30 cycles of 98°C for 15 s and 65°C for 5 min. Two resulting amplicon pools for each patient sample were combined and used for library preparation. After end repair (NEB E7546) samples were barcoded using Native Barcoding Expansion Kits EXP-NBD104 and EXP-NBD114 from Oxford Nanopore. A total of 24

barcoded samples were pooled together and Nanopore adaptors were ligated. After clean-up 20 ng of multiplexed library DNA was loaded onto the MinION flowcell for sequencing. A total of 0.3 Gb of raw sequencing data was collected per patient sample. Nuclease-free water was used as a negative control for library preparation and sequencing. Non-specific amplification in negative control was additionally checked using an agarose gel. SARS-CoV-2 Isolate USA-WA1/2020 (BEI, NR-52281) was used as a positive control for genome sequencing.

SARS-CoV-2 Genome Assembly

MinKNOW software was used to basecall raw Nanopore reads in high-accuracy mode. ARTIC bioinformatic pipeline for COVID-19 was used to analyze reads (<https://artic.network/ncov-2019>). Pipeline included demultiplexing with guppy barcoder and generating consensus sequences with minimap2 and calling single nucleotide variants with nanopolish relative to Wuhan-Hu-1/2019 reference genome.(Li, 2018b; Quick, 2020; Fan Wu et al., 2020) 20X coverage was used as a threshold both for consensus and variant calling. Nucleotide positions with less than 20X coverage were masked with Ns in the final consensus. Consensus sequences were uploaded to GISAID (<https://www.gisaid.org/>), accession IDs are available in the online supplementary material associated with this manuscript (<https://doi.org/10.1016/j.celrep.2021.109197>).

RT-PCR and Sanger Sequencing

To verify the ORF7aΔ115 mutation we used RT-PCR with primers that flank the deletion region. Reactions were performed with Q5 High-Fidelity DNA Polymerase in 25 µL volume (New England Biolabs) using following program: 98°C for 2min, 35 cycles of 98°C for 15 s and 65°C for 5 min, 35 cycles. PCR products were analyzed on 1% agarose gels stained with SYBR

Safe (Thermo Fisher Scientific). The remaining volume of the reaction was purified using DNA Clean & Concentrator kit (Zymo Research) and sent to Psomagen for Sanger sequencing. Each PCR product was sequenced with both forward and reverse primers used for PCR. To detect sgRNAs of ORF7a and ORF7b primers annealing to Leader region of SARS-CoV-2 and inside the ORF were used. PCR products were cut out from 1% agarose gel and purified with Zymoclean Gel DNA Recovery Kit. Purified products were sequenced with forward and reverse primers. Sequence was verified by aligning to SARS-CoV-2 genome in UGENE software (Unipro).

Phylogenetic and ORF7a Mutational Analysis

An alignment of 181,003 SARS-CoV-2 genomes was downloaded from the GISAID database at 9:11 AM on 2020-11-11, and sequences without corresponding metadata entries were removed. The nucleotide positions encoding ORF7a in the reference sequence (Wuhan-Hu-1; EP_ISL_402125) were extracted for the remaining 180,971 sequences, and ORF7a sequences with 100% sequence identity were clustered using CD-HIT (with settings: -c 1 -aL 1). One representative was randomly selected from each of the 846 ORF7a sequence clusters and mutations in each nucleotide and translated sequence were determined using the Biostrings package in R. Graphs of these mutations were rendered with the ggplot2 package. To construct a phylogenetic tree of global and Bozeman SARS-CoV-2 sequences, up to ten sequences from each country in the world were selected for each month of the pandemic using the Filter utility in the Augur pipeline (`–group-by country year month –sequences-per-group 10`). (Bedford et al., 2020; Lemieux et al., 2021) This resulted in an alignment of 4,235 GISAID sequences. Then, the Nextclade online tool was used to determine the quality of 73 SARS-CoV-2 genomes (%)

genome covered) sequenced from Bozeman patients. 55 Bozeman sequences that were determined to be of “Good” quality were merged with the alignment of GISAID sequences, as well as an alignment of 669 representative SARS-CoV-2 genomes that had non-synonymous mutations in ORF7a. The bat coronavirus RaTG13 (MN996532.2) and the Wuhan- Hu-1 SARS-CoV-2 sequences were also merged, resulting in an alignment of 4,959 full genome sequences, and positions that are hypervariable or prone to sequencing artifacts were masked with a VCF file downloaded from EMBL at 12:45PM on 2020-05- 29. The alignment was trimmed of columns composed of 90% or more gaps using Trimal (-gt 0.9), and IqTree was used to construct a maximum likelihood phylogeny (-m GTR -ninit 2 -n 2 -me 0.05). Branch lengths and internal nodes were rescaled on the resulting tree using TreeTime, and tree was re-rooted to the RaTG13 sequence using the APE package before being visualized with the ggTree package in R. Clades were assigned to SARS-CoV-2 genomes using Nextclade and pangolin (<https://cov-lineages.org/>) utilities.

Western Blot

293T-hACE2 cells were transfected using Lipofectamine 3000 (Invitrogen, L3000-015) with plasmids encoding for ORF7aWT, ORF7a Δ 115 or control pEGFP-N1 plasmid. To detect ORF7b expression cells were infected with ORF7aWT or ORF7a Δ 115 SARS-CoV-2 strains at MOI = 0.05. After 24 hours, cells were washed two times with PBS and lysed in RIPA buffer (150 mM sodium chloride, 1% NP-40, 0.5% sodium deoxycholate, 0.2% SDS, 50 mM Tris, pH 8.0) at 4°C for 30 min. The lysates were clarified by centrifugation (10'000 g, 20 min) and stored at -80°C. For western blot lysates were mixed with 6xLaemmli SDS-PAGE buffer and heated at 98°C for 5 min, resolved in 12% SDS-PAGE gel and transferred onto a PVDF (ORF7b) or

nitrocellulose (ORF7a) membrane using Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, #1703930). Membranes were blocked and probed with indicated antibodies. The proteins were visualized using Pierce ECL Western Blotting Substrate (ThermoFisher Scientific, #32106) and exposed to X-Ray film (sc-201696, Santa Cruz Biotech).

Immunocytochemistry

HEK293T-hACE2 were transfected with plasmid DNA using Lipofectamine 3000 (Invitrogen REF: L3000-015) according to manufacturer's instructions. Cells were either transfected with expression vectors for SARS-CoV-2 ORF7aWT, SARS-CoV-2 ORF7a Δ 115, or mock transfected. 24 hours after transfection, cells were detached and seeded at 50% confluency on glass coverslips (Fisherbrand, #12-545-80) pre-coated with poly-D-lysine (Cultrex, #3439-100-01) in a 24 well plate. The next day, media was removed, and cells were washed with PBS and fixed with ice-cold 100% methanol on ice for 10 minutes. Half of the methanol was replaced with PBS three times before the methanol solution was discarded and cells were washed three times in PBS (5 minutes each wash). Cells were blocked with 1% BSA (Fisher Scientific BP9703-100) in PBS + 0.1% Tween-20 (PBST) for 30 minutes at room temperature. After blocking, cells were stained with mouse monoclonal anti-Flag antibody (Invitrogen, MA1-91878) (1:200 in 1% BSA in PBST) overnight at 4°C. The next day, cells were washed 3X in PBS (5 min each) and stained with secondary antibody diluted 1:2000 in 1% BSA in PBST for 1 h at RT in the dark (Goat anti-mouse AlexaFluor488; Invitrogen, A11001). After 3x PBS washes nuclei were stained with Hoechst 33342 in PBS for 5 minutes at RT. Coverslips were then mounted with ProLong Gold antifade mounting media (Invitrogen, P36934) on Superfrost

Plus microscope slides (#22-037-246). Immunostained cells were imaged using a Leica SP8 confocal microscope.

IFN Inhibition Assay

To generate ISRE reporter system, HEK293T cells were transduced with lentiviruses carrying pGreenFire1-ISRE (Cat. TR016PA-1, SBI) construct and selected with puromycin (1 μ g/ml) for 2 days. After selection cells (1×10^5 cells/well, 48-well plate) were transfected with expression plasmids (250 ng) for SARS-CoV-2 ORF7aWT, SARS-CoV-2 ORF7a Δ 115, or control plasmid pLV-mCherry. 16 hours post transfection, cells were treated for 24 h with 5 ng/ml of human IFN- α 2b (Cat. rcyc-hifna2b, InvivoGen) according to manufacturer's recommendations. After treatment activation of IFN signaling was measured using Luciferase Assay System (Cat. E1500, Promega) on Cytation 5 (BioTek) plate reader. Assay was performed in six replicates for each condition.

SARS-CoV-2 Replication Assays

HEK293T-hACE2 and Vero E6 cells were seeded in the 48-well plate and infected with ORF7aWT or ORF7a Δ 115 SARS-CoV-2 strains at MOI = 0.05 in 50 μ L of the cell culture media. Infections were performed in triplicates. Next, 0.5 h post infection (hpi) 500 μ L of the fresh cell culture media was added to each well. The supernatants (140 μ L) from the infected cells were harvested at the following time points 0.5, 6, 24, 48, 72 and 120 hpi. The RNAs from the supernatants were extracted using QIAamp Viral RNA Mini Kit (QIAGEN) and used as an input in qRT-PCR with CDC N1 and N2 primers and probes as described above. Relative quantification was used (Δ Ct method) to calculate viral replication versus 0.5 hpi time point as $2^{-\Delta$ Ct.

The infected cells were harvested 0.5, 6 and 24 hpi and washed with PBS. Total RNA from cells was extracted using RNeasy kit (QIAGEN, Cat No./ID: 74104) with on-column DNase digestion step (QIAGEN, 79254) according to the manufacturers protocol. Viral RNA was quantified using qRT-PCR CDC N1 primers and probe. Human ACTB endogenous control was quantified with TaqMan assay (ThermoFisher, 4333762T). $\Delta\Delta C_t$ method was used to normalize viral RNA level to host RNA and 0.5 hpi. Viral replication was calculated for each replicate as $2^{-\Delta\Delta C_t}$.

Type I Interferon Response Assay

HEK293T-hACE2 were seeded in the 48-well plate and infected with ORF7aWT or ORF7a Δ 115 SARS-CoV-2 strains at MOI = 0.05 in 50 μ L of the media. After 30 min on infection the 500 μ L of the cell culture media was added to each well. At 24 hpi the cells were washed once with PBS, detached using Trypsin and the total RNA was extracted using RNeasy kit (QIAGEN, Cat No./ID: 74104) with on-column DNase digestion (QIAGEN, Cat No./ID: 79254) according to the manufacturer's protocol. Non-infected cells were processed using the same protocol with mock infection. After extraction, 0.5-1 μ g of total RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) according to the supplier's protocol. The cDNA was diluted 10-fold in water and analyzed with Type I interferon response (SAB Target List) H96 qPCR array. Each reaction contained 2 μ L of cDNA, 10 μ L of 2X SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and 8 μ L of water. Reactions were performed on QuantStudio 3 Real-Time PCR System instrument with following thermocycling conditions: 95°C for 2min, 40 cycles of 95°C for 5 s and 60°C for 30 s. After amplification, melt curve analysis was performed to examine product specificity. Assay was

performed in triplicates for non-infected cells and both viruses. Run data was analyzed in PrimePCR Analysis software (Bio-Rad). Expression of IFN-I response genes was normalized to geometric mean of 3 reference transcripts (TBP, GAPDH, HPRT1) and non-infected control ($\Delta\Delta C_t$ method).

Quantification and Statistical Analysis

All statistical analyses were performed in RStudio v1.2.1335. Moving averages ($n = 7$) for symptom onset and positive COVID-19 tests data were calculated using `geom_ma` function from `tidyquant` package and plotted with `ggplot2` in RStudio. qRT-PCR data is shown as mean of three biological replicates (each with three technical replicates) $\pm$ standard deviation (SD). Data in ISRE reporter assay was analyzed with one-way ANOVA with Tukey's post hoc pairwise comparisons. Replication of viruses and expression of IFN-I response genes was compared using Student's *t* test. Medians of IFN-I response between two viruses were compared with Wilcoxon signed-rank test. Significance levels in figures: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ or ns (no significant difference, $p > 0.05$).

References Cited

- Addetia, A., Xie, H., Roychoudhury, P., Shrestha, L., Loprieno, M., Huang, M. L., Jerome, K. R., & Greninger, A. L. (2020, Aug). Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol*, 129, 104523. <https://doi.org/10.1016/j.jcv.2020.104523>
- Andolfo, I., Russo, R., Lasorsa, V. A., Cantalupo, S., Rosato, B. E., Bonfiglio, F., Frisso, G., Abete, P., Cassese, G. M., Servillo, G., Esposito, G., Gentile, I., Piscopo, C., Villani, R., Fiorentino, G., Cerino, P., Buonerba, C., Pierri, B., Zollo, M., Iolascon, A., & Capasso, M. (2021, Apr 23). Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. *iScience*, 24(4), 102322. <https://doi.org/10.1016/j.isci.2021.102322>
- Andrew Rambaut, N. L., Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson, Erik Volz, on behalf of COVID-19 Genomics Consortium UK (CoG-UK). (2020). *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*.
- Arunachalam, P. S., Wimmers, F., Mok, C. K. P., Perera, R., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., Wagh, D., Collier, J., Pellegrini, K. L., Kazmin, D., Alaaeddine, G., Leung, W. S., Chan, J. M. C., Chik, T. S. H., Choi, C. Y. C., Huerta, C., Paine McCullough, M., Lv, H., Anderson, E., Edupuganti, S., Upadhyay, A. A., Bosinger, S. E., Maecker, H. T., Khatri, P., Roupheal, N., Peiris, M., & Pulendran, B. (2020, Sep 4). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science*, 369(6508), 1210-1220. <https://doi.org/10.1126/science.abc6261>
- Avanzato, V. A., Matson, M. J., Seifert, S. N., Pryce, R., Williamson, B. N., Anzick, S. L., Barbian, K., Judson, S. D., Fischer, E. R., Martens, C., Bowden, T. A., de Wit, E., Riedo, F. X., & Munster, V. J. (2020, Dec 23). Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell*, 183(7), 1901-1912 e1909. <https://doi.org/10.1016/j.cell.2020.10.049>
- Banerjee, A., El-Sayes, N., Budykowski, P., Jacob, R. A., Richard, D., Maan, H., Aguiar, J. A., Demian, W. L., Baid, K., D'Agostino, M. R., Ang, J. C., Murdza, T., Tremblay, B. J., Afkhami, S., Karimzadeh, M., Irving, A. T., Yip, L., Ostrowski, M., Hirota, J. A., Kozak, R., Capellini, T. D., Miller, M. S., Wang, B., Mubareka, S., McGeer, A. J., McArthur, A. G., Doxey, A. C., & Mossman, K. (2021, May 21). Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience*, 24(5), 102477. <https://doi.org/10.1016/j.isci.2021.102477>
- Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M. L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A.,

- Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T. R., Truong, M., Wolf, C. R., Nickerson, D. A., Rieder, M. J., Englund, J. A., Seattle Flu Study, I., Hadfield, J., Hodcroft, E. B., Huddleston, J., Moncla, L. H., Muller, N. F., Neher, R. A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J. S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G. L., Baird, G. S., Chu, H. Y., Shendure, J., & Jerome, K. R. (2020, Oct 30). Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516), 571-575.
<https://doi.org/10.1126/science.abc0523>
- Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Moller, R., Jordan, T. X., Oishi, K., Panis, M., Sachs, D., Wang, T. T., Schwartz, R. E., Lim, J. K., Albrecht, R. A., & tenOever, B. R. (2020, May 28). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, 181(5), 1036-1045 e1039.
<https://doi.org/10.1016/j.cell.2020.04.026>
- Chinese, S. M. E. C. (2004, Mar 12). Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science*, 303(5664), 1666-1669.
<https://doi.org/10.1126/science.1092002>
- Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., Solomon, I. H., Kuo, H. H., Boucau, J., Bowman, K., Adhikari, U. D., Winkler, M. L., Mueller, A. A., Hsu, T. Y., Desjardins, M., Baden, L. R., Chan, B. T., Walker, B. D., Lichterfeld, M., Brigl, M., Kwon, D. S., Kanjilal, S., Richardson, E. T., Jonsson, A. H., Alter, G., Barczak, A. K., Hanage, W. P., Yu, X. G., Gaiha, G. D., Seaman, M. S., Cernadas, M., & Li, J. Z. (2020, Dec 3). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *The New England journal of medicine*, 383(23), 2291-2293.
<https://doi.org/10.1056/NEJMc2031364>
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., Group, C. C.-W., Consortium, C.-G. U., Diaz-Ordaz, K., Keogh, R., Eggo, R. M., Funk, S., Jit, M., Atkins, K. E., & Edmunds, W. J. (2021, Apr 9). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538).
<https://doi.org/10.1126/science.abg3055>
- Emeny, J. M., & Morgan, M. J. (1979, Apr). Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *J Gen Virol*, 43(1), 247-252.
<https://doi.org/10.1099/0022-1317-43-1-247>
- Gamage, A. M., Tan, K. S., Chan, W. O. Y., Liu, J., Tan, C. W., Ong, Y. K., Thong, M., Andiappan, A. K., Anderson, D. E., Wang, Y., & Wang, L. F. (2020, Dec). Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking

- ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLoS Pathog*, 16(12), e1009130. <https://doi.org/10.1371/journal.ppat.1009130>
- Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., Breillat, P., Carlier, N., Gauzit, R., Morbieu, C., Pene, F., Marin, N., Roche, N., Szwed, T. A., Merkl, S. H., Treluyer, J. M., Veyer, D., Mouthon, L., Blanc, C., Tharaux, P. L., Rozenberg, F., Fischer, A., Duffy, D., Rieux-Laucat, F., Kerneis, S., & Terrier, B. (2020, Aug 7). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, 369(6504), 718-724. <https://doi.org/10.1126/science.abc6027>
- Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., Queen, K., Tao, Y., Paden, C. R., Zhang, J., Li, Y., Uehara, A., Wang, H., Goldsmith, C., Bullock, H. A., Wang, L., Whitaker, B., Lynch, B., Gautam, R., Schindewolf, C., Lokugamage, K. G., Scharton, D., Plante, J. A., Mirchandani, D., Widen, S. G., Narayanan, K., Makino, S., Ksiazek, T. G., Plante, K. S., Weaver, S. C., Lindstrom, S., Tong, S., Menachery, V. D., & Thornburg, N. J. (2020, Jun). Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerging infectious diseases*, 26(6), 1266-1273. <https://doi.org/10.3201/eid2606.200516>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., & Pohlmann, S. (2020, Apr 16). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271-280 e278. <https://doi.org/10.1016/j.cell.2020.02.052>
- Holland, L. A., Kaelin, E. A., Maqsood, R., Estifanos, B., Wu, L. I., Varsani, A., Halden, R. U., Hogue, B. G., Scotch, M., & Lim, E. S. (2020, Jul 1). An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *J Virol*, 94(14). <https://doi.org/10.1128/JVI.00711-20>
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M. A., Weldon, W. C., Biggs, H. M., Uyeki, T. M., Pillai, S. K., & Washington State -nCoV, V. C. I. T. (2020). First Case of 2019 Novel Coronavirus in the United States. *The New England journal of medicine*, 382(10), 929-936. <https://doi.org/10.1056/NEJMoa2001191>
- Hou, Y. J., Okuda, K., Edwards, C. E., Martinez, D. R., Asakura, T., Dinno, K. H., 3rd, Kato, T., Lee, R. E., Yount, B. L., Mascenik, T. M., Chen, G., Olivier, K. N., Ghio, A., Tse, L. V., Leist, S. R., Gralinski, L. E., Schafer, A., Dang, H., Gilmore, R., Nakano, S., Sun, L., Fulcher, M. L., Livraghi-Butrico, A., Nicely, N. I., Cameron, M., Cameron, C., Kelvin, D. J., de Silva, A., Margolis, D. M., Markmann, A., Bartelt, L., Zumwalt, R., Martinez, F. J., Salvatore, S. P., Borczuk, A., Tata, P. R., Sontake, V., Kimple, A., Jaspers, I., O'Neal,

- W. K., Randell, S. H., Boucher, R. C., & Baric, R. S. (2020, Jul 23). SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell*, 182(2), 429-446 e414. <https://doi.org/10.1016/j.cell.2020.05.042>
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., & D'Eustachio, P. (2020, Jan 8). The reactome pathway knowledgebase. *Nucleic Acids Res*, 48(D1), D498-D503. <https://doi.org/10.1093/nar/gkz1031>
- Jiang, H. W., Zhang, H. N., Meng, Q. F., Xie, J., Li, Y., Chen, H., Zheng, Y. X., Wang, X. N., Qi, H., Zhang, J., Wang, P. H., Han, Z. G., & Tao, S. C. (2020, Sep). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol*, 17(9), 998-1000. <https://doi.org/10.1038/s41423-020-0514-8>
- Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020, May 14). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, 181(4), 914-921 e910. <https://doi.org/10.1016/j.cell.2020.04.011>
- Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., & Chlanda, P. (2020, Nov 18). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun*, 11(1), 5885. <https://doi.org/10.1038/s41467-020-19619-7>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Sheffield, C.-G. G., McDanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Whelan, S. P., LaBranche, C. C., Saphire, E. O., & Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e819. <https://doi.org/10.1016/j.cell.2020.06.043>
- Kotenko, S. V., & Durbin, J. E. (2017, May 5). Contribution of type III interferons to antiviral immunity: location, location, location. *J Biol Chem*, 292(18), 7295-7303. <https://doi.org/10.1074/jbc.R117.777102>
- Lei, X., Dong, X., Ma, R., Wang, W., Xiao, X., Tian, Z., Wang, C., Wang, Y., Li, L., Ren, L., Guo, F., Zhao, Z., Zhou, Z., Xiang, Z., & Wang, J. (2020, Jul 30). Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun*, 11(1), 3810. <https://doi.org/10.1038/s41467-020-17665-9>
- Lemieux, J. E., Siddle, K. J., Shaw, B. M., Loreth, C., Schaffner, S. F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C. H., Krasilnikova, L. A., DeRuff, K. C., Rudy, M., Bauer, M. R., Lagerborg, K. A., Normandin, E., Chapman, S. B., Reilly, S. K.,

- Anahtar, M. N., Lin, A. E., Carter, A., Myhrvold, C., Kembball, M. E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J. B., Branda, J. A., Hooper, D., Gaeta, J. M., Baggett, T. P., O'Connell, J., Gnirke, A., Lieberman, T. D., Philippakis, A., Burns, M., Brown, C. M., Luban, J., Ryan, E. T., Turbett, S. E., LaRocque, R. C., Hanage, W. P., Gallagher, G. R., Madoff, L. C., Smole, S., Pierce, V. M., Rosenberg, E., Sabeti, P. C., Park, D. J., & MacInnis, B. L. (2021, Feb 5). Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*, 371(6529). <https://doi.org/10.1126/science.abe3261>
- Li, H. (2018b). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Loveday, E. K., Hain, K. S., Kochetkova, I., Hedges, J. F., Robison, A., Snyder, D. T., Brumfield, S. K., Young, M. J., Jutila, M. A., Chang, C. B., & Taylor, M. P. (2021, Mar 26). Effect of Inactivation Methods on SARS-CoV-2 Virion Protein and Structure. *Viruses*, 13(4). <https://doi.org/10.3390/v13040562>
- Lucas, C., Wong, P., Klein, J., Castro, T. B. R., Silva, J., Sundaram, M., Ellingson, M. K., Mao, T., Oh, J. E., Israelow, B., Takahashi, T., Tokuyama, M., Lu, P., Venkataraman, A., Park, A., Mohanty, S., Wang, H., Wyllie, A. L., Vogels, C. B. F., Earnest, R., Lapidus, S., Ott, I. M., Moore, A. J., Muenker, M. C., Fournier, J. B., Campbell, M., Odio, C. D., Casanovas-Massana, A., Yale, I. T., Herbst, R., Shaw, A. C., Medzhitov, R., Schulz, W. L., Grubaugh, N. D., Dela Cruz, C., Farhadian, S., Ko, A. I., Omer, S. B., & Iwasaki, A. (2020, Aug). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*, 584(7821), 463-469. <https://doi.org/10.1038/s41586-020-2588-y>
- Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S., & Huang, C. (2020, Jul). Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Res*, 179, 104811. <https://doi.org/10.1016/j.antiviral.2020.104811>
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A., & O'Garra, A. (2015, Feb). Type I interferons in infectious disease. *Nat Rev Immunol*, 15(2), 87-103. <https://doi.org/10.1038/nri3787>
- Miorin, L., Kehrer, T., Sanchez-Aparicio, M. T., Zhang, K., Cohen, P., Patel, R. S., Cupic, A., Makio, T., Mei, M., Moreno, E., Danziger, O., White, K. M., Rathnasinghe, R., Uccellini, M., Gao, S., Aydillo, T., Mena, I., Yin, X., Martin-Sancho, L., Krogan, N. J., Chanda, S. K., Schotsaert, M., Wozniak, R. W., Ren, Y., Rosenberg, B. R., Fontoura, B. M. A., & Garcia-Sastre, A. (2020, Nov 10). SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc Natl Acad Sci U S A*, 117(45), 28344-28354. <https://doi.org/10.1073/pnas.2016650117>
- Muth, D., Corman, V. M., Roth, H., Binger, T., Dijkman, R., Gottula, L. T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtaric, D., Toplak, I., Ameneiros, R. S., Pfeifer, A., Thiel, V., Drexler, J. F., Muller, M. A., & Drosten, C. (2018, Oct 11). Attenuation of replication

- by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*, 8(1), 15177. <https://doi.org/10.1038/s41598-018-33487-8>
- Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S., & Fremont, D. H. (2005, Jan). Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure*, 13(1), 75-85. <https://doi.org/10.1016/j.str.2004.10.010>
- Osada, N., Kohara, A., Yamaji, T., Hirayama, N., Kasai, F., Sekizuka, T., Kuroda, M., & Hanada, K. (2014, Dec). The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res*, 21(6), 673-683. <https://doi.org/10.1093/dnares/dsu029>
- Pati, A., Padhi, S., Suvankar, S., & Panda, A. K. (2021, Jan 4). Minor Allele of Interferon-Induced Transmembrane Protein 3 Polymorphism (rs12252) Is Covered Against Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Mortality: A Worldwide Epidemiological Investigation. *The Journal of infectious diseases*, 223(1), 175-178. <https://doi.org/10.1093/infdis/jiaa630>
- Quick, J. (2020). *nCoV-2019 sequencing protocol v1* ZappyLab, Inc. <https://doi.org/10.17504/protocols.io.bbmuik6w>
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Jr., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., & Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols*, 12(6), 1261-1276. <https://doi.org/10.1038/nprot.2017.066>
- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020, Nov). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*, 5(11), 1403-1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., Burgstaller-Muehlbacher, S., De Jesus, P. D., Teriete, P., Hull, M. V., Chang, M. W., Chan, J. F., Cao, J., Poon, V. K., Herbert, K. M., Cheng, K., Nguyen, T. H., Rubanov, A., Pu, Y., Nguyen, C., Choi, A., Rathnasinghe, R., Schotsaert, M., Miorin, L., Dejoze, M., Zwaka, T. P., Sit, K. Y., Martinez-Sobrido, L., Liu, W. C., White, K. M., Chapman, M. E., Lendy, E. K., Glynn, R. J., Albrecht, R., Rupp, E., Mesecar, A. D., Johnson, J. R., Benner, C., Sun, R., Schultz, P. G., Su, A. I., Garcia-Sastre, A., Chatterjee, A. K., Yuen, K. Y., & Chanda, S. K. (2020, Oct). Discovery of SARS-CoV-2 antiviral drugs through

- large-scale compound repurposing. *Nature*, 586(7827), 113-119.
<https://doi.org/10.1038/s41586-020-2577-1>
- Rosenthal, S. H., Kagan, R. M., Gerasimova, A., Anderson, B., Grover, D., Hua, M., Liu, Y., Owen, R., & Lachawan, F. (2020). Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. *BioRxiv*, 2020.2012.2010.418855.
<https://doi.org/10.1101/2020.12.10.418855>
- Schoggins, J. W. (2019, Sep 29). Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol*, 6(1), 567-584. <https://doi.org/10.1146/annurev-virology-092818-015756>
- Shi, G., Kenney, A. D., Kudryashova, E., Zani, A., Zhang, L., Lai, K. K., Hall-Stoodley, L., Robinson, R. T., Kudryashov, D. S., Compton, A. A., & Yount, J. S. (2021, Feb 1). Opposing activities of IFITM proteins in SARS-CoV-2 infection. *EMBO J*, 40(3), e106501. <https://doi.org/10.15252/embj.2020106501>
- Stoltz, M., & Klingstrom, J. (2010, Sep). Alpha/beta interferon (IFN-alpha/beta)-independent induction of IFN-lambda1 (interleukin-29) in response to Hantaan virus infection. *J Virol*, 84(18), 9140-9148. <https://doi.org/10.1128/JVI.00717-10>
- Su, Y. C. F., Anderson, D. E., Young, B. E., Linster, M., Zhu, F., Jayakumar, J., Zhuang, Y., Kalimuddin, S., Low, J. G. H., Tan, C. W., Chia, W. N., Mak, T. M., Octavia, S., Chavatte, J. M., Lee, R. T. C., Pada, S., Tan, S. Y., Sun, L., Yan, G. Z., Maurer-Stroh, S., Mendenhall, I. H., Leo, Y. S., Lye, D. C., Wang, L. F., & Smith, G. J. D. (2020, Jul 21). Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *MBio*, 11(4).
<https://doi.org/10.1128/mBio.01610-20>
- Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A., & Ng, L. F. P. (2020, Jun). The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*, 20(6), 363-374.
<https://doi.org/10.1038/s41577-020-0311-8>
- Thi Nhu Thao, T., Labroussaa, F., Ebert, N., V'Kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., Kratzel, A., Gultom, M., Schmied, K., Laloli, L., Husser, L., Wider, M., Pfaender, S., Hirt, D., Cippa, V., Crespo-Pomar, S., Schroder, S., Muth, D., Niemeyer, D., Corman, V. M., Muller, M. A., Drosten, C., Dijkman, R., Jores, J., & Thiel, V. (2020, Jun). Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, 582(7813), 561-565. <https://doi.org/10.1038/s41586-020-2294-9>
- Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajeky, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020, Sep 4). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRxiv*. <https://doi.org/10.1101/2020.09.04.283077>

- van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen, C. J., Pang, J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 83, 104351-104351. <https://doi.org/10.1016/j.meegid.2020.104351>
- Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D. P., consortium, T. C.-G. U., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., & Ferguson, N. M. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.2012.2030.20249034. <https://doi.org/10.1101/2020.12.30.20249034>
- Wang, C., Shan, L., Qu, S., Xue, M., Wang, K., Fu, F., Wang, L., Wang, Z., Feng, L., Xu, W., & Liu, P. (2020). The Coronavirus PEDV Evades Type III Interferon Response Through the miR-30c-5p/SOCS1 Axis. *Front Microbiol*, 11, 1180. <https://doi.org/10.3389/fmicb.2020.01180>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269. <https://doi.org/10.1038/s41586-020-2008-3>
- Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y., Wang, H., Menachery, V. D., Rajsbaum, R., & Shi, P. Y. (2020, Oct 6). Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep*, 33(1), 108234. <https://doi.org/10.1016/j.celrep.2020.108234>
- Xie, X., Muruato, A., Lokugamage, K. G., Narayanan, K., Zhang, X., Zou, J., Liu, J., Schindewolf, C., Bopp, N. E., Aguilar, P. V., Plante, K. S., Weaver, S. C., Makino, S., LeDuc, J. W., Menachery, V. D., & Shi, P. Y. (2020, May 13). An Infectious cDNA Clone of SARS-CoV-2. *Cell Host Microbe*, 27(5), 841-848 e843. <https://doi.org/10.1016/j.chom.2020.04.004>
- Young, B. E., Fong, S. W., Chan, Y. H., Mak, T. M., Ang, L. W., Anderson, D. E., Lee, C. Y., Amrun, S. N., Lee, B., Goh, Y. S., Su, Y. C. F., Wei, W. E., Kalimuddin, S., Chai, L. Y. A., Pada, S., Tan, S. Y., Sun, L., Parthasarathy, P., Chen, Y. Y. C., Barkham, T., Lin, R. T. P., Maurer-Stroh, S., Leo, Y. S., Wang, L. F., Renia, L., Lee, V. J., Smith, G. J. D., Lye, D. C., & Ng, L. F. P. (2020, Aug 29). Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational

- cohort study. *Lancet*, 396(10251), 603-611. [https://doi.org/10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8)
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012, May). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 16(5), 284-287. <https://doi.org/10.1089/omi.2011.0118>
- Zang, R., Case, J. B., Yutuc, E., Ma, X., Shen, S., Gomez Castro, M. F., Liu, Z., Zeng, Q., Zhao, H., Son, J., Rothlauf, P. W., Kreutzberger, A. J. B., Hou, G., Zhang, H., Bose, S., Wang, X., Vahey, M. D., Mani, K., Griffiths, W. J., Kirchhausen, T., Fremont, D. H., Guo, H., Diwan, A., Wang, Y., Diamond, M. S., Whelan, S. P. J., & Ding, S. (2020, Dec 15). Cholesterol 25-hydroxylase suppresses SARS-CoV-2 replication by blocking membrane fusion. *Proc Natl Acad Sci U S A*, 117(50), 32105-32113. <https://doi.org/10.1073/pnas.2012197117>
- Zhang, B., Zhou, X., Qiu, Y., Song, Y., Feng, F., Feng, J., Song, Q., Jia, Q., & Wang, J. (2020). Clinical characteristics of 82 cases of death from COVID-19. *PloS one*, 15(7), e0235458. <https://doi.org/10.1371/journal.pone.0235458>
- Zhang, Y., Qin, L., Zhao, Y., Zhang, P., Xu, B., Li, K., Liang, L., Zhang, C., Dai, Y., Feng, Y., Sun, J., Hu, Z., Xiang, H., Knight, J. C., Dong, T., & Jin, R. (2020, Jun 16). Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. *The Journal of infectious diseases*, 222(1), 34-37. <https://doi.org/10.1093/infdis/jiaa224>
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020, Mar 28). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*, 395(10229), 1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- Zhou, Z., Huang, C., Zhou, Z., Huang, Z., Su, L., Kang, S., Chen, X., Chen, Q., He, S., Rong, X., Xiao, F., Chen, J., & Chen, S. (2021, Mar 19). Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14(+) monocytes. *iScience*, 24(3), 102187. <https://doi.org/10.1016/j.isci.2021.102187>

INTRINSIC SIGNAL AMPLIFICATION BY TYPE III CRISPR-CAS
SYSTEM PROVIDES A SEQUENCE-SPECIFIC
SARS-COV-2 DIAGNOSTIC

Contribution of Authors and Co-Authors

Manuscript in Appendix C

Author: Andrew Santiago-Frangos

Contributions: designed the research, performed the research, prepared manuscript figures, wrote the manuscript, wrote the manuscript

Co-Author: Laina N. Hall

Contributions: performed the research

Co-Author: Anna Nemudraia

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Artem Nemudryi

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Pushya Krishna

Contributions: performed the research

Co-Author: Tanner Wiegand

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Royce A. Wilkinson

Contributions: performed the research

Co-Author: Deann T. Snyder

Contributions: acquired samples, performed the research

Co-Author: Jodi F. Hedges

Contributions: acquired samples, performed the research

Co-Author: Calvin Cicha

Contributions: performed the research

Co-Author: Helen Lee

Contributions: performed the research

Co-Author: Ava Graham

Contributions: performed the research

Co-Author: Mark A. Jutila

Contributions: acquired samples, wrote the manuscript

Co-Author: Matthew P. Taylor

Contributions: acquired samples, wrote the manuscript

Co-Author: Blake Wiedenheft

Contributions: designed the research, wrote the manuscript

Manuscript Information

Andrew Santiago-Frangos, Laina N. Hall, Anna Nemudraia, Artem Nemudryi, Pushya Krishna, Tanner Wiegand, Royce A. Wilkinson, Deann T. Snyder, Jodi F. Hedges, Calvin Cicha, Helen H. Lee, Ava Graham, Mark A. Jutila, Matthew P. Taylor, and Blake Wiedenheft

Cell Reports Medicine

Status of Manuscript:

- ☐ Prepared for submission to a peer-reviewed journal
- ☐ Officially submitted to a peer-reviewed journal
- ☐ Accepted by a peer-reviewed journal
- ☒ Published in a peer-reviewed journal

Cell Press

Volume 2

Issue 6

15 June 2021

DOI: 10.1016/j.xcrm.2021.100319

INTRINSIC SIGNAL AMPLIFICATION BY TYPE III CRISPR-CAS
SYSTEM PROVIDES A SEQUENCE-SPECIFIC
SARS-COV-2 DIAGNOSTIC

Andrew Santiago-Frangos,¹ Laina N. Hall,¹ Anna Nemudraia,¹ Artem Nemudryi,¹ Pushya
Krishna,¹ Tanner Wiegand,¹ Royce A. Wilkinson,¹ Deann T. Snyder,¹ Jodi F. Hedges,¹ Calvin
Cicha,¹ Helen H. Lee,¹ Ava Graham,¹ Mark A. Jutila,¹ Matthew P. Taylor,¹ and Blake
Wiedenheft¹

¹Department of Microbiology and Immunology, Montana State University, Bozeman,
MT 59717, USA.

Abstract

There is an urgent need for inexpensive new technologies that enable fast, reliable, and scalable detection of viruses. Here, we repurpose the type III CRISPR-Cas system for sensitive and sequence-specific detection of SARS-CoV-2. RNA recognition by the type III CRISPR complex triggers Cas10-mediated polymerase activity, which simultaneously generates pyrophosphates, protons, and cyclic oligonucleotides. We show that all three Cas10-polymerase products are detectable using colorimetric or fluorometric readouts. We design ten guide RNAs that target conserved regions of SARS-CoV-2 genomes. Multiplexing improves the sensitivity of amplification-free RNA detection from 10^7 copies/ μL for a single guide RNA to 10^6 copies/ μL for ten guides. To decrease the limit of detection to levels that are clinically relevant, we developed a two-pot reaction consisting of RT-LAMP followed by T7-transcription and type III CRISPR-based detection. The two-pot reaction has a sensitivity of 200 copies/ μL and is completed using patient samples in less than 30 min.

Introduction

Frequent tests and quick results are critical for limiting the spread of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) and ending the current coronavirus disease 2019 (COVID-19) pandemic.(Larremore et al., 2021; Paltiel et al., 2020) qRT-PCR (quantitative reverse transcriptase-polymerase chain reaction) has been the gold standard for viral diagnostics, but this method is slow and requires sophisticated equipment that is expensive to purchase and operate. A national survey of individuals tested using nasal swab-based qRT-PCR found that the average wait times for results was 4.1 days, with 10% of tests taking 310 days.(Lazer, 2020) Thus, there is an urgent need for inexpensive new technologies that enable fast, reliable, and scalable detection of viruses.

Recently, loop-mediated isothermal amplification (LAMP)(Notomi et al., 2000) was developed as a sensitive (1–100 copies/mL) point-of-care diagnostic.(Dao Thi et al., 2020; Y. Zhang, G. Ren, et al., 2020) However, LAMP and reverse-transcription LAMP (RT-LAMP) are prone to generating mis-primed or template-independent amplicons that cause false positive readings if they are not coupled to sequence-specific verification of the amplicon.(Gao et al., 2019; Hardinge & Murray, 2019; Meagher et al., 2018; Ooi et al., 2021; Rolando et al., 2020; Zou et al., 2020) For example, the POP7-targeting primers used in the DETECTR CRISPR-based diagnostic lead to excessive nonspecific amplification; however, the sequence-specific verification of the RT-LAMP amplicon by Cas12 yields sensitive and specific detection of POP7 RNA.(Broughton et al., 2020; Ooi et al., 2021) Therefore, platforms that can further verify nucleic acid sequences increase the utility of LAMP and RT-LAMP applications. Type V (Cas12-based) and type VI (Cas13-based) CRISPR systems have been coupled to LAMP or RPA

(recombinase polymerase amplification) for sensitive and reliable detection of viral nucleic acids.(Broughton et al., 2020; J. S. Chen et al., 2018; East-Seletsky et al., 2016; Gootenberg et al., 2018; Gootenberg et al., 2017; Joung et al., 2020) Following isothermal amplification, the RNA-guided Cas12 or Cas13 proteins bind to the amplified target and trigger a non-sequencespecific nuclease activity that cleaves a fluorophore and quencher-labeled DNA or RNA.(J. S. Chen et al., 2018; East-Seletsky et al., 2016; Gootenberg et al., 2018) Cleavage of the tether results in an increase in fluorescence that can be detected in 45 min. Furthermore, the collateral nuclease activity of Cas13 has previously been used to cleave and mature an RNA oligo into a linear pseudo-ligand for the RNase Csm6, resulting in a 3.5-fold increase signal sensitivity.(Gootenberg et al., 2018) While Cas12 and Cas13 detection methods have been optimized over several iterations to be compatible with the isothermal amplification of viral RNA, the ultimate goal is to develop CRISPR-based technologies that are sensitive enough to detect the viral RNA directly, without prior amplification. Recently, Fozouni et al. reported that type IV (Cas13a-based) CRISPR systems can be used for amplification-free detection of SARS-CoV-2 RNA in ~30 min and with the sensitivity of ~100 copies/mL.(Fozouni et al., 2021)

Like the type VI (Cas13-based) systems, type III systems also target RNA.(Hale et al., 2009; Kazlauskienė et al., 2016; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014) However, type III systems rely on a unique intrinsic signal amplification mechanism (Figure 8.1). Here, we set out to determine whether intrinsic amplification by type III systems could be used to detect viral RNA without prior amplification. The proof-of-concept presented here demonstrates that the Csm complex from *Thermus thermophilus* (TtCsm) can be programmed to specifically recognize the SARS-CoV-2 genome. SARS-CoV-2, but not SARS-CoV-1 or a panel

of other respiratory pathogens, activates the Cas10 polymerase, which generates ~1,000 cyclic nucleotides (e.g., cA₄) after binding an RNA target.(Jia, Jones, Sukenick, et al., 2019; Kazlauskienė et al., 2017; Niewoehner et al., 2017; Rouillon et al., 2018) Like all polymerases, nucleotide polymerization by Cas10 also generates protons (H⁺) and pyrophosphates (PPi). We demonstrate that each of these three products can be used to detect SARS-CoV-2 RNA using either colorimetric, fluorometric, or both methods simultaneously. The assay can be performed in 1–30 min, depending on the detection method and the concentration of the RNA. When coupled with RTLAMP, the assay can be performed in <30 min and has a limit of detection of ~200 copies/mL. Collectively, this work indicates that the type III systems can be repurposed for fast and sensitive diagnostics.

Results

Sequence-specific Activation of Cas10 Polymerase Yields Three Detectable Products

Sequence-specific recognition of RNA by type III CRISPR systems initiates a signaling cascade (Figure 8.1A).(Kazlauskienė et al., 2017; Niewoehner et al., 2017; Rouillon et al., 2018) RNA binding by the *Thermus thermophilus* (TtCsm) complex triggers a conformational change that activates the Palm domain of the Cas10 subunit, which amplifies the RNA binding signal by converting ATP into ~1,000 cyclic oligoadenylates (e.g., cA₄). (Jia, Jones, Sukenick, et al., 2019; Kazlauskienė et al., 2017; Niewoehner et al., 2017; Rouillon et al., 2018) We hypothesized that the intrinsic signal amplification unique to type III CRISPR systems would boost the sensitivity of direct RNA detection, while maintaining specificity. To test this hypothesis, we expressed and purified the type III-A CRISPR RNA (crRNA)guided surveillance complex from TtCsm with a

A WT Csm complex Csm complex Csm^{Csm3-D34A} complex

4ATP → cA₄ + 4PPI + 4H⁺

Target binding → Target release

Target binding → SARS-CoV-2 RNA

4ATP → cA₄ + 4H⁺ + 4PPI

B

Alignment of 45,625 SARS-CoV-2 Genomes

N1 gRNA: 3'-ACCGCGCUUUUACGUGUUAUAAACGGGGUUCGCGAAGUCGCA-5'

SARS-CoV-2: 5'...UGGCGCGCAAAUUGGCACAAUUUGCCCCACGCGCUUACGAGCU...3'

SARS-CoV-1: UGGCGCGCAAAUUGGCACAAUUUGCUCAACUGCGCUCUGCAU

C 1-5 min.

Fluorimetric detection

Fluorescence (x10³)

WT TiCsm TiCsm^{Csm3-D34A}

SARS-CoV-2 CoV-1

RNA Copies: 10⁰ 10¹ 10² 10³ 10⁴ 10⁵ 10⁶ 10⁷ 10⁸ 10⁹

D 30 min.

Colorimetric detection

TiCsm: - 200 nM

SARS-CoV: 2 2 1

RNA Copies: 10¹¹ 10¹⁰ 10⁹ 10⁸ 10⁷ 10⁶ 10⁵ 10⁴ 10³ 10² 10¹ 10⁰

E 10 min.

Visible Fluorimetric detection

SARS-CoV: - 2 1

RNA Copies: - 10¹⁰ 10⁹ 10⁸ 10⁷ 10⁶ 10⁵ 10⁴ 10³ 10² 10¹ 10⁰

Figure 8.1. Detection of SARS-CoV-2 using the type III CRISPR-Cas system. (A) Schematic of the type III Csm complex from *Thermus thermophilus* (TtCsm). TtCsm complex consists of a crRNA (red) and an unequal stoichiometry of 5 different proteins (Cas10₁, pink; Csm4₁, blue;

Csm3₆, gray: Csm2₄, green: Csm5₁, white). RNA binding activates the Cas10-polymerase, Cas10 DNase, and Csm3 RNase (left, transparent). Target RNA cleavage by Csm3 subunits (blue stars) causes target dissociation and inactivates Cas10. A mutation in the Csm3 subunit (TtCsm^{Csm3-D34A}) renders the complex RNase-dead (right). (B) Schematic of SARS-CoV-2 genome. The N1 CRISPR RNA (crRNA_{N1}) targets a sequence on the 3' end of genomic and subgenomic RNAs, which is conserved in 45,625 SARS-CoV-2 genomes. Mismatches between the crRNA-guide and SARS-CoV-1 are highlighted (navy blue). (C) Fluorometric detection of *in vitro* transcribed SARS-CoV-2 (red squares) and SARS-CoV-1 N-gene (black squares). Cyclic tetra-adenylate (cA₄) activates TtCsm6, which is a non-sequence-specific ancillary nuclease (Figure S8.2). Activated TtCsm6 cleaves an RNA tether, which links a fluorophore (star) to a quencher (gray hexagon). TtCsm^{Csm3-D34A} N1 (right graph) exhibits a LoD 3-fold lower than wild-type TtCsm N1 (left graph) and retains specificity for SARS-CoV-2 RNA. The mean of 3 technical replicates are shown; error bars represent ± 1 SD. (D) Colorimetric detection of SARS-CoV-2 RNA by TtCsm^{Csm3-D34A} N1 complex using a pH-sensitive dye (Phenol Red). Reactions were incubated for 30 min at 60°C. Technical replicates are shown in Figure S8.3. (E) Visible fluorometric detection of SARS-CoV-2 RNA by TtCsm^{Csm3-D34A} N1 complex using calcein. Reactions were incubated for up to 1 h at 60°C. Technical replicates and kinetics are shown in Figure S8.4.

Csm3-mediated cleavage of the target RNA (e.g., SARS-SARCoV-2 RNA) would reduce the target RNA concentration over time and thus limit the sensitivity of the assay. Therefore, we mutated residues in the Csm3 subunit responsible for target RNA cleavage (D34A),(Liu et al., 2017; Tamulaitis et al., 2014) and purified the RNase-dead complex (TtCsm^{Csm3-D34A}) (Figure S8.1). To measure the limit of detection (LoD), we added the mutant or wild-type Csm complex to a reaction containing the TtCsm6 nuclease, a fluorescent reporter (i.e., FAM-RNA-Iowa Black FQ), and a titration of RNA corresponding to the N-gene of either SARS-CoV-2 or SARS-CoV-1 (Figures 8.1C and S8.2; Tables S8.2 and S8.3). Single mismatches in the 3' end of target RNAs have been shown to reduce cyclic oligoadenylate production by SepCsm (type III-A) by up to 30-fold as compared to the fully complementary target,³⁰ while double-nucleotide mutations in the 3' end of target RNAs ablated cyclic oligoadenylate production by SsoCsm (type III-D).(Rouillon et al., 2018) Furthermore, single- and double-nucleotide complementarity between the crRNA 5' handle and the target 3' anti-tag significantly decreased the allosteric activation of

Cas10 and cyclic oligoadenylate production by SthCsm (type III-A).(You et al., 2019) Single mismatches in the target RNA have been shown to result in 10-fold lower amounts of cyclic oligoadenylate production by other Csm complexes.(Nasef et al., 2019) Using fluorometric detection, both the mutant and the wild-type Csm complex could detect the SARS-CoV-2 RNA at concentrations $>10^8$ copies per reaction, and neither complex cross-reacted with the SARS-CoV-1 RNA at the highest concentrations tested. The RNase-dead TtCsm complex was roughly 3-fold more sensitive than the wild type, with an LoD of $\sim 10^7$ copies per reaction.

In addition to fluorometric detection, we developed a colorimetric RNA detection method that uses a pH change that occurs during nucleotide polymerization (Figures 8.1D and S8.3). Specific recognition of SARS-CoV-2 by RNase-dead TtCsm complex activates Cas10. Cas10 polymerizes ATP,(Jia, Jones, Sukenick, et al., 2019; Kazlauskienė et al., 2017; Niewoehner et al., 2017; Rouillon et al., 2018) releasing one proton per incorporated nucleotide. Cas10-generated protons acidify the solution and change the color of a pH indicator (i.e., Phenol Red) from fuchsia through orange (10^{10} RNA copies) to yellow (10^{11} RNA copies). Similarly, we developed a visible fluorometric detection method that relies on the sequestration of metallic ions by pyrophosphate. The metal indicator calcein is initially quenched by bound Mn^{2+} ions.(Tomita et al., 2008) In addition to the cyclic oligoadenylates and protons, Cas10 polymerase generates one pyrophosphate per ATP polymerized. Pyrophosphate forms an insoluble precipitate with Mn^{2+} , which unquenches calcein. Free calcein is then bound by excess Mg^{2+} , forming a fluorescent complex that can be seen by eye or with a UV lamp in <10 min (Figures 8.1E and S8.4).

Csm-based Direct Detection of SARS-CoV-2 RNA in Patient Samples

The LoD using crRNA_{N1} is between 10^7 and 10^8 copies of IVT RNA per microliter, which is insufficient to be clinically relevant (Figures 8.1C and 8.2C). To identify other guides that may outperform or complement the activity of crRNA_{N1}, we aligned 45,641 SARS-CoV-2 genomes available from GISAID.(Elbe & Buckland-Merrett, 2017) These alignments were used to select guides based on four key criteria. First, each target sequence had to be >99% identical among the available SARS-CoV-2 genomes. Second, complementarity between the target and the crRNA was not allowed to extend beyond the spacer sequence (guide) and into repeat derived portions of the crRNA that have been shown to suppress Cas10 activity.(Kazlauskienė et al., 2017) Third, we targeted regions of SARS-CoV-2 that were different by at least 2 nt in SARS-CoV-1 and MERS-CoV. Fourth, the list of target sequences was pruned to remove guides with similarity to human mRNA sequences or common oral and respiratory pathogen sequences (E value <1,000). Finally, we focused on target sequences located 3' of the ORF3a gene, which are present on both the viral genome and subgenomic RNAs generated during infection. In total, we designed crRNAs targeting 10 different locations on the SARS-CoV-2 genome (Figures 8.2A and S8.1; Table S8.1). To determine how each of these guides perform, we measured sequence-specific detection of RNA

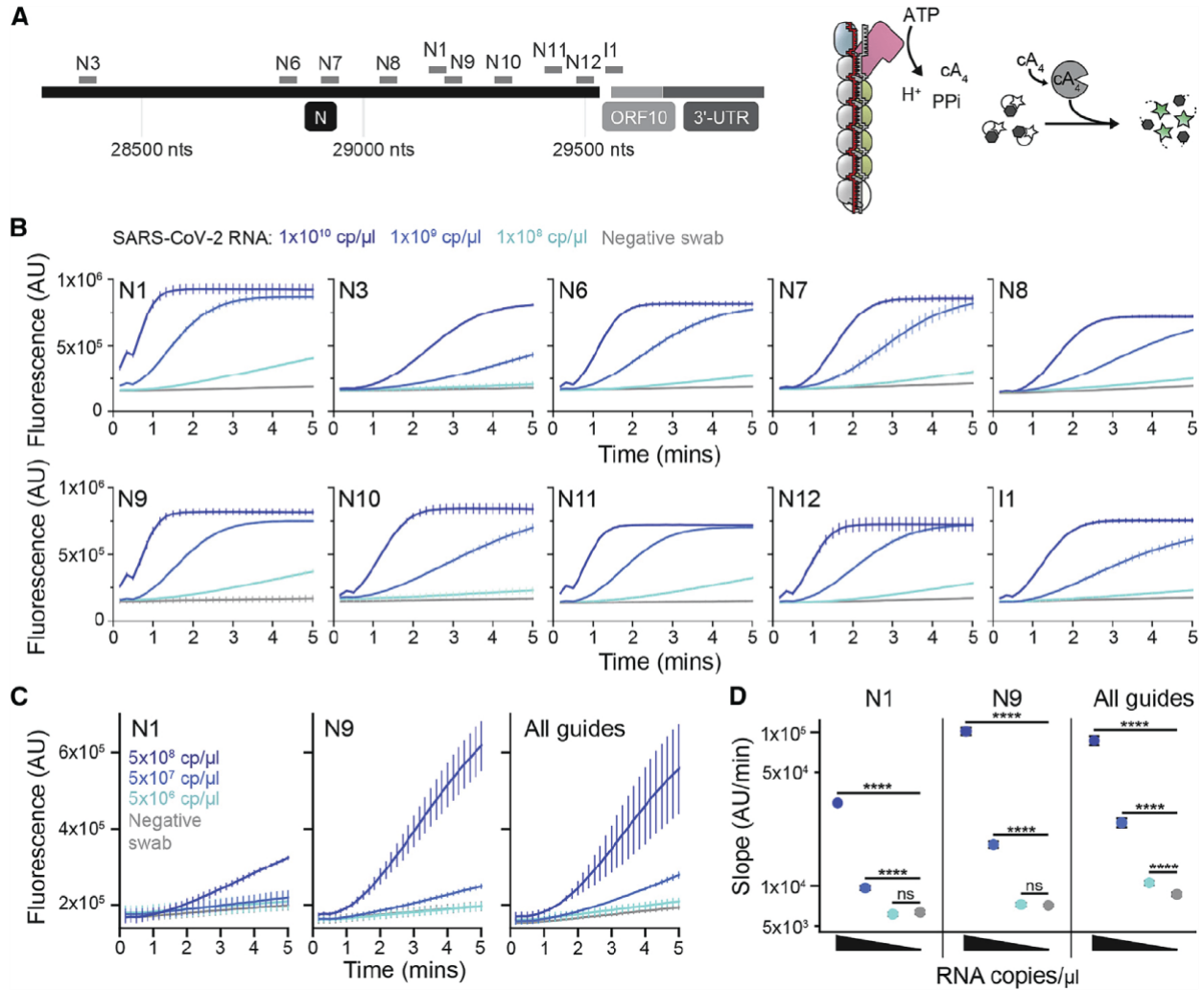


Figure 8.2. Screening and multiplexing crRNA guides. (A) Regions of SARS-CoV-2 genome targeted by each of the 10 guides, and schematic of RNA reporter-based assay (right) used to test guides. (B) Detection of SARS-CoV-2 IVT RNA, spiked into RNA extracted from patients lacking a SARS-CoV-2 infection, by 10 different TtCsm^{Csm3-D34A} complexes (25 nM), via a reporter RNA-based assay (Figure 8.1C; Tables S8.1 and S8.3). Means and SDs of 2 technical replicates are shown. (C) Direct detection of SARS-CoV-2 genomes in RNA extracted from patient samples by 25 nM TtCsm^{Csm3-D34A} N1 or N9, or a mixture of 10 different TtCsm^{Csm3-D34A} complexes, each at 2.5 nM, via a reporter RNA-based assay. RNA extracted from a patient with a high viral load (5.3 × 10⁸ copies/mL as determined by qRT-PCR) was diluted into RNA extracted from patients lacking a SARS-CoV-2 infection. Means and SDs of 3 technical replicates are shown. (D) Slopes of increasing fluorescence, measured in (C), were calculated by simple linear regression. The calculated slope and ±95% confidence intervals are shown. Positive RNA slopes were compared to the negative swab RNA slope by an F-test: ****p < 0.0001, ns, not significantly higher than negative swab RNA control.

using a fluorometric reporter assay (i.e., FAM-RNA-Iowa Black FQ) (Figure 8.2B). Most of the crRNAs provide similar sensitivity; however, crRNA_{N1} and crRNA_{N9} generated significantly more signal than the next best complex tested ($p < 0.0001$). We then tested crRNA_{N1} and crRNA_{N9} on RNA isolated from the nasal swab of an infected patient (Figure 8.2C). Both crRNA_{N1}- and crRNA_{N9}-guided complexes generate a similar signal for either IVT RNA or RNA isolated from a SARS-CoV-2⁺ patient (i.e., 2- to 3-fold increase in signal by 5 min, relative to the first time point). The LoD for crRNA_{N1} and crRNA_{N9} is 10^7 copies/mL (Figure 8.2D) ($p < 0.0001$).

Fozouni et al. recently showed that multiplexing Cas13 (i.e., combining multiple guides into a single reaction) improves the sensitivity of SARS CoV-2 detection.(Fozouni et al., 2021) We reasoned that similar benefits may be possible for Csm-based detection. To test this idea, we combined 10 of the guides (2.5 nM each) into a single multiplexed reaction. Multiplexing 10 guides improves the sensitivity of TtCsm-mediated detection of SARS-CoV-2 RNA isolated from the nasal swab of a positive patient by ~10 times (Figures 8.2C and 8.2D). However, the sensitivity of direct detection appears to increase additively with the number of TtCsm complexes, which precludes the direct detection of RNA at concentrations that are clinically relevant.

Testing Clinical Samples for SARS-CoV-2 using RT-LAMP and T7-Csm

Csm-based detection is currently not sensitive enough to directly detect SARS-CoV-2 in all of the patients capable of spreading the infection, which requires an LoD of 10^3 RNA copies/mL (Figure 8.2C).(La Scola et al., 2020; Larremore et al., 2021; Paltiel et al., 2020; Wolfel et al., 2020) To decrease the LoD of a type III CRISPR-based diagnostic to $\sim 10^3$ RNA copies/mL, we

incorporated an upstream nucleic acid amplification technique (Figure 8.3A). SARS-CoV-2 genomic RNA is reverse transcribed (RT) into DNA, which is then amplified by

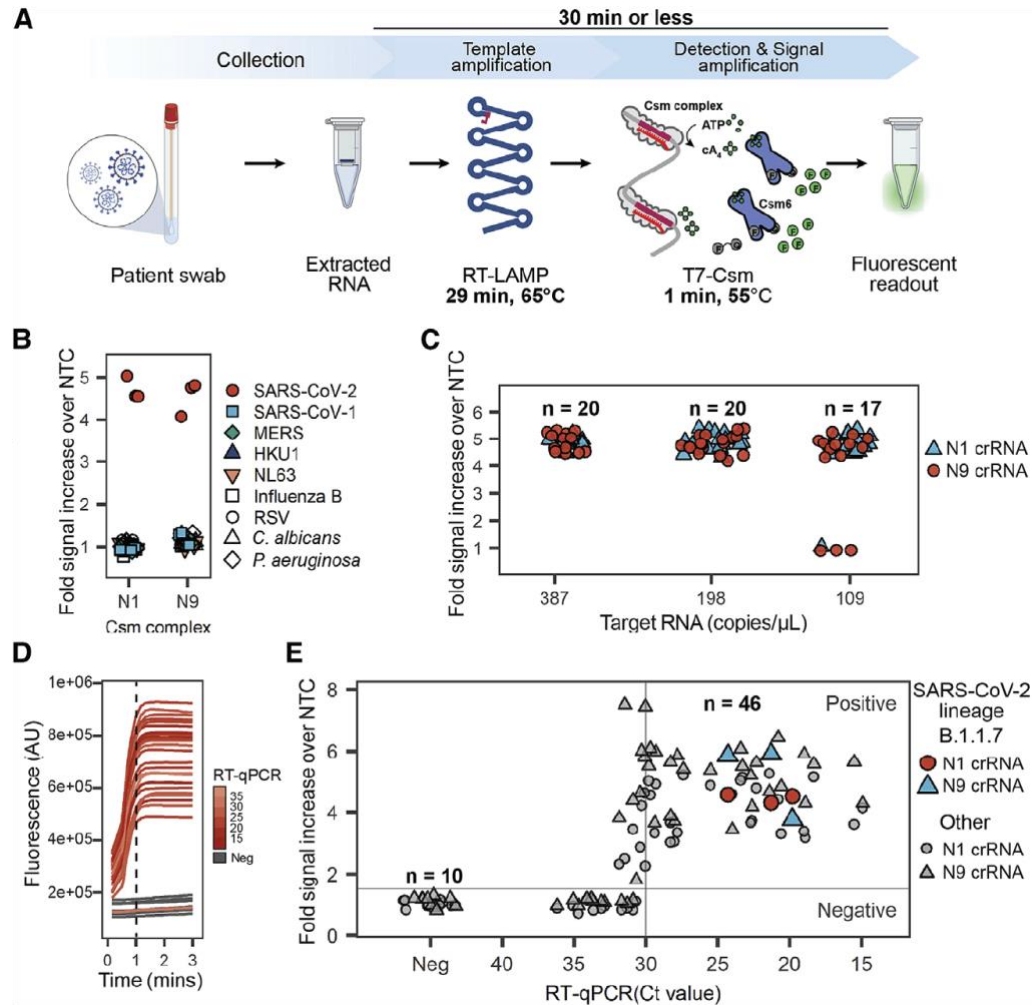


Figure 8.3. CRISPR-Csm-based detection of SARS-CoV-2 in clinical nasopharyngeal swab samples. (A) Schematic of RT-LAMP-T7-Csm based detection. The viral RNA is reverse transcribed, and the resulting DNA is amplified in an RT-LAMP reaction to produce transcription templates for T7 RNA polymerase, in one pot. An aliquot of the RT-LAMP reaction (29 min) is then mixed with the T7-Csm reaction (1 min). (B) RT-LAMP-T7-Csm is specific. Neither crRNA_{N1} nor crRNA_{N9} cross-react with other coronaviruses, or other common human oral pathogens or flora. Detection of SARS-CoV-2 by both crRNAs is rapid, 1 min, and robust, 4- to 5-fold increase in signal over no-template control (NTC). Three technical replicates are shown. (C) RT-LAMP-T7-Csm is sensitive. TtCsm^{Csm3-D34A} complexes loaded with either crRNA_{N1} or crRNA_{N9} have LoDs of 198 copies/mL (20/20 technical replicates); 20 technical replicates are shown. (D) Kinetics of fluorescence signal increase in T7-Csm reactions. SARS-CoV-2⁺ patient samples are observed to have a ~2-fold increase in signal over NTC by 10 s

(crRNA_{N1} median = 1.8, crRNA_{N9} median = 2.3), which rapidly increases to a 4-5-fold increase in signal over an NTC reaction by 1 min (crRNA_{N1} median = 4.2, crRNA_{N9} median = 5.5). A subset of traces is shown for clarity (Figure S8.5). (E) Nasopharyngeal swabs from 56 individuals were tested with qRT-PCR (x-axis) and RT- LAMP-T7-Csm (y-axis). Swabs with Ct values <40 for both N1 and N2 CDC diagnostic primers are considered positive for SARS-CoV-2 RNA. RT-LAMP-T7-Csm reliably identifies patient samples with a Ct < 30.7 (200–100 RNA copies/mL) as positive for SARS-CoV-2 (Figure S8.6). The B.1.1.7 variant is positively identified by both crRNA_{N1} and crRNA_{N9}. Data are shown as fold change in fluorescence as compared to NTC reaction. Single technical replicates are shown for qRT-PCR and RT-LAMP-T7- Csm experiments with either N1 or N9 crRNAs.

LAMP using primers that flank regions of the SARS-CoV-2 genome targeted by crRNA_{N1} and crRNA_{N9}. One of the LAMP primers incorporates a T7 promoter into the amplified DNA, which is then used for *in vitro* transcription (T7) and detected by TtCsm (Figures 8.3A and S8.5; Table S8.4).

To confirm the specificity of TtCsm-based detection, we tested SARS-CoV-2 alongside a panel of eight other oral and respiratory pathogens, including coronaviruses SARS-CoV-1, Middle East respiratory syndrome coronavirus (MERS-CoV), human coronavirus HKU1, and human coronavirus NL63 (Figure 8.3B). These samples resulted in a background signal similar to the no-template control (NTC) (Figure 8.3B). In contrast, SARS-CoV-2 RNA results in a 4- to 5-fold increase in signal.

To determine the LoD of RT-LAMP-T7-Csm, we tested 20 replicates of 2-fold serial dilutions ranging from ~100-400 copies/mL SARS-CoV-2 RNA (Figure 8.3C). The LoD of RT-LAMP-T7-Csm is 198 copies/mL SARS-CoV-2 RNA (20/20 replicates), in an assay that relies on a 29-min RT-LAMP step, followed by a 1-min T7Csm fluorometric detection reaction (Figure 8.3D; Table 8.1).

To further validate this method, we next tested RNA extracted from 56 nasopharyngeal swab samples taken from patients who had previously been tested using qRT-PCR. Of the 56

samples tested, 46 were positive for SARS-CoV-2 and 10 were negative by qRT-PCR (Figure 8.3E). Using two different crRNA guides, we demonstrate that the type III CRISPR system has a

Test	Time to result (min)	LoD (copies/ μ L)	Positive agreement (%)		Negative agreement (%)
			2x LoD	All Samples Tested	All Samples Tested
DETECTR	45	20	95.0%	95%	100.0%
SHERLOCK	50	6.75	100.0%	100.0%	100.0%
RT-LAMP-T7-Csm	30	100	100%	74%	100.0%

Table 8.1. Comparison of clinical parameters of RT-LAMP-T7-Csm and FDA EUA-approved protocols of DETECTR and SHERLOCK diagnostics.(M. Biosciences, 2020; S. Biosciences, 2020)

specificity (negative predictive agreement) of 100%, as well as a positive predictive agreement of 100% for nasopharyngeal swab samples with 100–200 copies/mL SARS-CoV-2 RNA as determined by qRT-PCR (Figures 8.3E and S8.5). Whole-genome sequencing revealed that three of the patient samples used here belonged to the B.1.1.7. lineage. These genome sequences have been deposited in GISAID accession IDs: EPI_ISL_1081321, EPI-ISL_1081322, EPI_ISL_1081323).(Elbe & Buckland-Merrett, 2017) Importantly, the B.1.1.7. variants were positively identified by RT-LAMP-T7-Csm with both N1 and N9 crRNA guides (Figure 8.3E; N1, red circles; N9, blue triangles).

Discussion

Collectively, this work demonstrates that sequence-specific detection of viral RNA by the type III CRISPR-Cas complex triggers the synthesis of cA4, pyrophosphate, and protons, each of which are detectable within 1–30 min, using colorimetric or fluorometric methods (Figure 8.1). Coupling RT-LAMP to T7 transcription and Csm-based detection creates a rapid (<30 min) testing protocol with attomolar sensitivity and high specificity. Moreover, we show that this protocol is capable of detecting the B.1.1.7 SARS-CoV-2 variant, and we anticipate that the

guide design criteria described here could be applied to the specific detection of other pathogens. Furthermore, the RT-LAMP and T7-Csm reactions occur in similar buffers and at similar temperatures; therefore, it is possible that additional modifications that either increase the efficiency of RT-LAMP at 55°C (i.e., redesigning primer sets), or that increase the efficiency of T7-Csm at 65°C (i.e., a more thermostable RNA polymerase), could enable an integrated one-pot diagnostic.

While the LoD for RT-LAMP-T7-Csm is 1–1.5 orders of magnitude higher than FDA Emergency Use Authorization (EUA)-approved DETECTR and SHERLOCK assays (20 and 6.75 copies/mL, respectively), the time to result is ~30% faster, which would translate to a higher throughput of samples analyzed per unit of machine run time. Sensitivity was strategically sacrificed for time in this proof-of-concept application for optimal detection of infectious SARS-CoV-2 patients, which requires an LoD of 1,000 copies/mL.(La Scola et al., 2020; Larremore et al., 2021; Paltiel et al., 2020; Wolfel et al., 2020) However, an increase in the length of the RT-LAMP step (i.e., to 40 min, as is used in SHERLOCK), would likely increase the sensitivity. Additional modifications that could further increase sensitivity include the optimization of RT-LAMP primer sets, a screen for more active type III Csm or Cmr surveillance complexes, or a screen for CARF-fusion effectors that do not cleave their cyclic oligoadenylate ligands.(Jia, Jones, Yang, et al., 2019) Alternatively, more sensitive methods of measuring Cas10-generated protons, such as ion-sensitive field-effector transistors used in label-free sequencing by synthesis technologies, could decrease the LoD.

Type V (Cas12-based) and type VI (Cas13-based) CRISPRbased diagnostics hinge completely on their collateral nuclease activity.(Broughton et al., 2020; J. S. Chen et al., 2018;

Gootenberg et al., 2018; Gootenberg et al., 2017; Joung et al., 2020; Yan et al., 2019) The Cas10 subunits of Csm complexes possess an analogous nuclease activity triggered upon target binding.(Elmore et al., 2016; Kazlauskienė et al., 2016; Liu et al., 2017; Sofos et al., 2020; Zhang et al., 2016) However, here, we focused on the NTP polymerase activity of the activated Csm complex, which enables readouts based on several different chemistries (Figure 8.1). Furthermore, there is a rich resource of naturally occurring downstream effector proteins that have evolved to be activated by the Cas10 polymerized NTP products, which possess a wide range of enzymatic activities. These effectors include RNases, DNases, putative proteases, nitrilases, adenosine deaminases, and adenylate cyclases.(Makarova et al., 2020) Future efforts are aimed at incorporating complementary effectors with the goal of reducing the LoD for the direct detection of RNA in clinical samples.

Limitations of Study

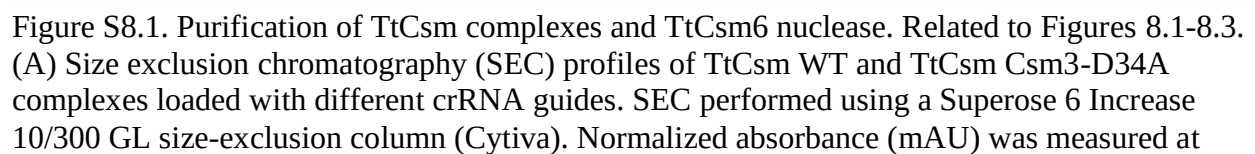
This study demonstrates a proof-of-concept for a sensitive, specific, and rapid diagnostic based on the type III CRISPR system. The current implementation requires pre-amplification using RT-LAMP, followed by T7-transcription and type III detection (T3D). Clinical implementation of this approach will benefit from changes that eliminate the need for RT-LAMP and T7 amplification, improvements that limit sample handling, as well as a better understanding of the base pairing requirements that activate the Cas10 polymerase.(Fozouni et al., 2021; Joung et al., 2020; Nasef et al., 2019; Rouillon et al., 2018; You et al., 2019) Improvements will benefit from screening additional guides, testing other type III complexes for accelerated polymerase activity, and incorporating other ancillary nucleases or other effectors in a way that boosts the sensitivity and reduces the time to result.

Acknowledgements

We are grateful to members of Bozeman Health who provided deidentified patient samples, specifically Christopher Nero, Douglas Smoot, Winfield Wallace, Cayley Faurot-Daniels, Verena Lawrence, and Melissa Blauvelt. We are also grateful to Michelle Flenniken, Katie Daughenbaugh, Diane Bimczok, and other members of the COVID task force at Montana State University (MSU) for assistance establishing the COVID testing center. A.S.-F. is a postdoctoral fellow of the Life Science Research Foundation, which is supported by the Simons Foundation. A.S.-F. is supported by the Postdoctoral Enrichment Program Award from the Burroughs Wellcome Fund. Research in the Wiedenheft lab and the biosafety level 3 facility is supported by the Office of the Vice President for Research at Montana State University, a generous donation from the Rosolowsky family, and a sponsored research agreement from VIRIS Detection Systems. Some graphical assets were created with BioRender.com.

B.W. is the founder of SurGene and VIRIS Detection Systems. B.W., A.S.-F., A. Nemudraia, and A. Nemudryi are inventors on patent applications related to CRISPR-Cas systems and applications thereof.

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.



260nm (red) and 280nm (blue). Fractions 9 up to 16 of each SEC purification were collected, concentrated, and stored at -80°C . (B) The fractions were combined, concentrated and run on an SDS-PAGE. All five Csm proteins are present and the intensities of each band correspond with our understanding of the protein stoichiometry of the assembled TtCsm complex. (C) RNA was isolated from the pooled and concentrated SEC fractions. Denaturing urea polyacrylamide gel of nucleic acids associated with each TtCsm complex. The full-length crRNA intermediate is expected to be 76 nucleotides (nts) long. (D) SEC profile of TtCsmCsm3-D34A N1 complex. Six successive fractions, representing the entire peak, were collected concentrated and stored separately. (E) Sequence-specific activation of Cas10 was estimated for each of the six fractions in panel D. ^{32}P -ATP polymerization was measured using thin-layer chromatography (TLC). 500nM TtCsm^{Csm3-D34A} N1 complex was incubated with 10^{10} copies of target RNA, 50 μM ATP and 10 nM α - ^{32}P -ATP, at 60°C for 1 hour. Nucleic acids were phenol-chloroform extracted from each reaction and spotted on a silica gel TLC plate coated with fluorescent indicator F254, developed in solvent (0.2 M ammonium bicarbonate pH 9.3, 70% ethanol, 30% water). An unlabeled cA4 standard (Axxora) was run on the same TLC plate, in a parallel lane, and was visualized by illumination with a handheld shortwave (254 nm) UV lamp (Analytik Jena) and a Galaxy S9 phone (Samsung). One of the two major ^{32}P -labelled products generated by target RNA-bound TtCsm complex migrates similarly to the cA4 standard. All TtCsm complexes polymerize similar amounts of α - ^{32}P -ATP (bottom band) into similar ratios of similarly migrating products (top bands). (F) SEC profile of TtCsm6 ancillary nuclease purified on a Superdex 200 26/600 size-exclusion column (Cytiva). (G) SDS-PAGE gel image of purified TtCsm6.

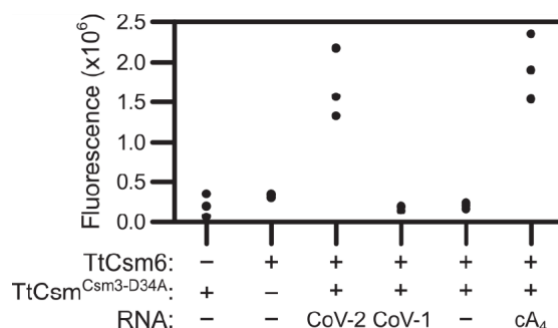


Figure S8.2. Csm-mediated synthesis of cA₄ is necessary for activation of the non-sequence specific nuclease Csm6. Related to Figure 8.1. Detection of invitro transcribed SARS-CoV-2 RNA using the type III surveillance complex TtCsm^{Csm3-D34A} and the ancillary nuclease, TtCsm6. TtCsm^{Csm3-D34A}, TtCsm6 and SARS-CoV-2 RNA, but not SARS-CoV-1 RNA, are all required to trigger cleavage of the fluorescent RNA reporter. Addition of cA₄, bypasses the signaling pathway and activates TtCsm6 directly, cleaving the reporter RNA in the absence of SARS-CoV-2 RNA (last lane). These reactions were performed using 10^8 copies of *in vitro* transcribed SARS-CoV-1 or -2 RNA spiked into nasopharyngeal swab clinical matrix. Results from three technical replicates are shown. Results presented in the final column (cA₄) demonstrate that the fluorescent signal is generated by cA₄-dependent activation of Csm6.

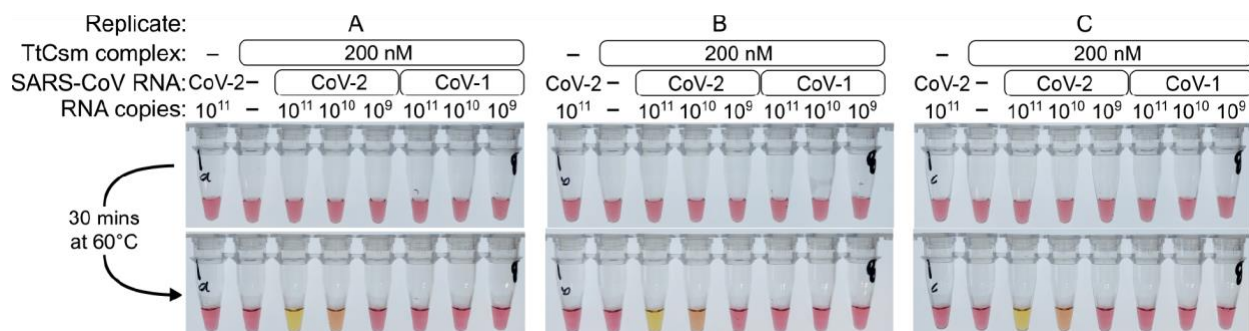


Figure S8.3. Colorimetric detection of a specific RNA sequence. Related to Figure 8.1. TtCsm^{Csm3-D34A} was incubated with either RNA target, in the presence of ATP, for 30 minutes at 60°C. Specific RNA recognition, and associated Cas10-mediated ATP polymerization, causes acidification of the solution, and changes the color of Phenol Red pH indicator from fuchsia, through orange, to yellow. Three technical replicates (A, B and C) are shown.

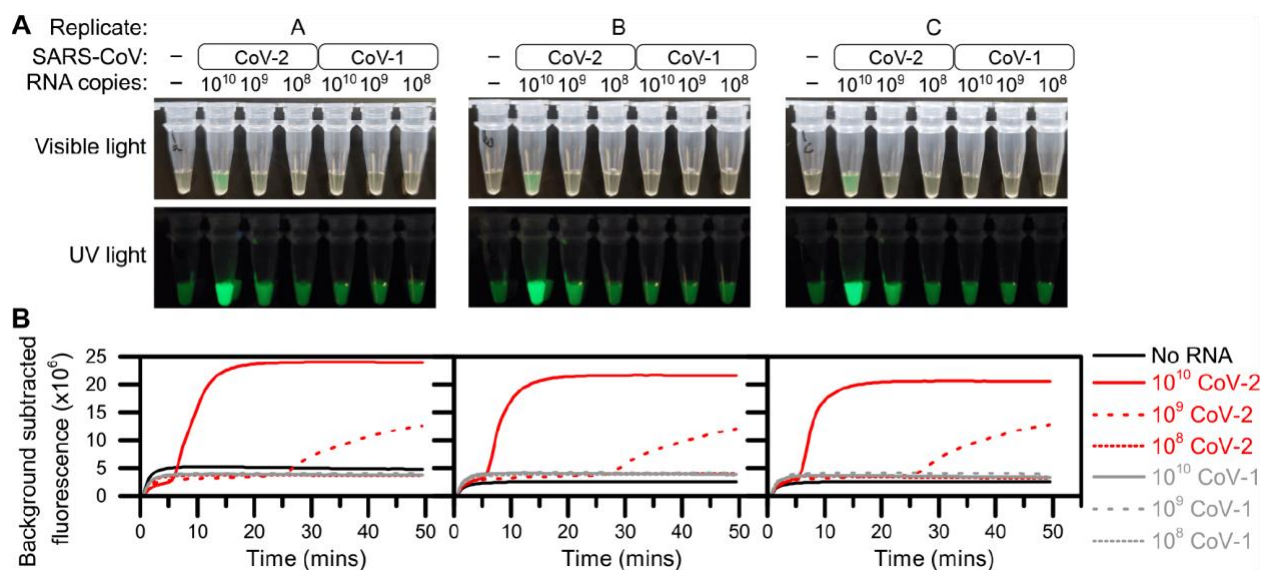


Figure S8.4. Visible fluorometric detection of a specific RNA sequence. Related to Figure 8.1. (A) TtCsm^{Csm3-D34A} was incubated with no RNA, a specific (i.e., SARS-CoV-2) or non-specific (i.e., SARS-CoV-1) RNA target, in the presence of ATP for 50 minutes at 60°C. The reactions contain an excess of Mg^{2+} and Mn^{2+} ions, and the metal indicator Calcein. Calcein preferentially binds Mn^{2+} forming a quenched complex. Specific RNA recognition, and associated Cas10-mediated ATP polymerization, generates pyrophosphate which precipitates with Mg^{2+} and Mn^{2+} ions. Excess Mg^{2+} ions bind Calcein to form a highly fluorescent complex that can be seen by eye in visible light or with UV light. (B) Kinetics of Calcein fluorescence for corresponding reactions shown in Panel A. 10^{10} SARS-CoV-2 RNAs can be detected in 10 minutes, and 10^9 copies are detected in 40 minutes. No increase in fluorescence is seen beyond the first five minutes for either a no RNA control, or for SARS-CoV-21 RNA containing samples.

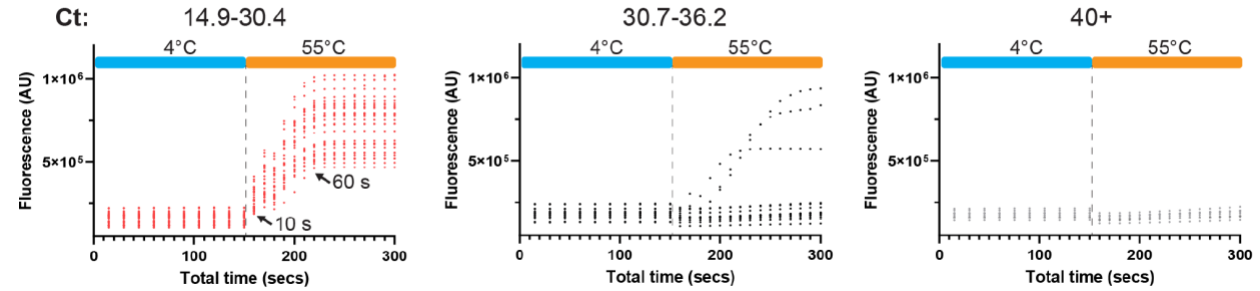
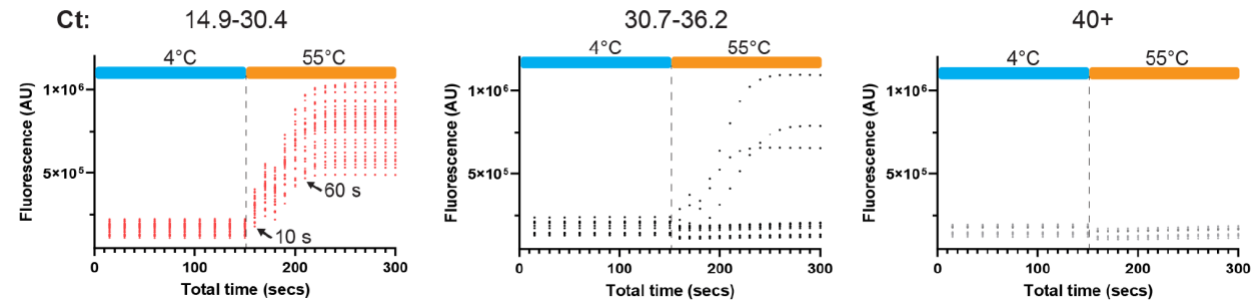
A TtCsm^{Csm3-D34A} N1 complex**B** TtCsm^{Csm3-D34A} N9 complex

Figure S8.5. Rapid and specific detection of SARS-CoV-2 by RT-LAMP-T7-Csm. Related to Figure 8.3. Raw fluorescence kinetics for the T7-Csm stage of RT-LAMP-T7-Csm detection of SARS-CoV-2 RNA from patient samples with Ct values of 14.9-30.4 (left), 30.7-36.2 (middle), and 40+ (right), as detected by the RNase dead TtCsm complex (dTtCsm) loaded with (A) crRNA_{N1} or (B) crRNA_{N9}. Reactions were first incubated in an RT-PCR machine at 4°C and fluorescence was measured every 15 seconds for 150 seconds. Fluorescence readings in both positive and negative reactions are low until the T7-Csm reactions are heated to 55°C, upon which fluorescence rapidly increases within 10 seconds and continues to increase for 60 seconds in most samples.

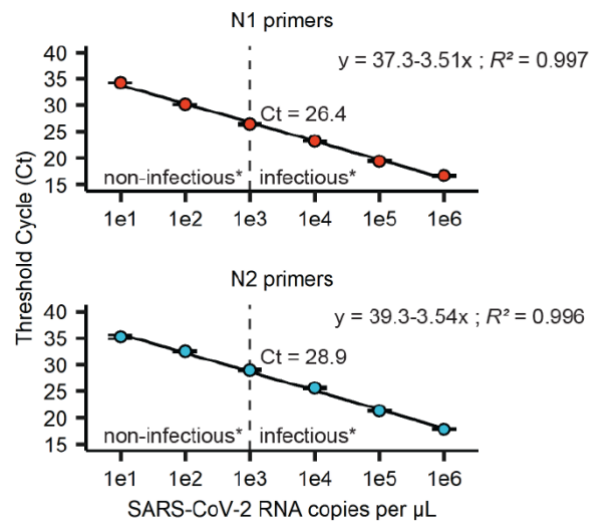


Figure S8.6. Standard curves for absolute quantification of SARS-CoV-2 titers. Related to Figures 8.2 and 8.3. A 10-fold dilution series of SARS-CoV-2 synthetic RNA fragment (RTGM 10169, National Institute of Standards and Technology) containing the Nucleocapsid gene was used. Data was plotted as Cycle Threshold (Ct) on y-axis versus log₁₀ (copies per ml) on x-axis. The mean Ct of three technical replicates are shown, error bars represent ± 1 standard deviation. Trend lines were fit to the data using the `geom_smooth` function of the `ggplot2` R package; linear equations and R^2 values are shown. Boundaries corresponding to non-infectious and infectious Ct values refer to observations that patients with viral titers below 10^6 copies/ml are rarely infectious.

Supplementary Tables

Name	Sequence (5'-3') [Target binding region underlined]
N1 (76 nt guide)	AUUGCGAC <u>ACGCUGAAGCGCUGGGGGCAA</u> UUUGUGCAAUU UGCGGCCAGUUGCAAGGGAUUGAGCCCCGUAAGGGG
N3 (76 nt guide)	AUUGCGAC <u>GGCCGACGUUGUUUUGAUCGCGCCCCACUGCG</u> UUCUCCAUGUUGCAAGGGAUUGAGCCCCGUAAGGGG
N6 (76 nt guide)	AUUGCGAC <u>GUUGCGACUACGUGAUGAGGAACGAGAAGAG</u> GCUUGACUGGUUGCAAGGGAUUGAGCCCCGUAAGGGG
N7 (76 nt guide)	AUUGCGAC <u>AGCAGCAGCAAAGCAAGAGCAGCAUCACCGCC</u> <u>AUUGCCAGGUUGCAAGGGAUUGAGCCCCGUAAGGGG</u>
N8 (76 nt guide)	AUUGCGAC <u>AUGCUUUAGUGGCAGUACGUUUUUGCCGAGGC</u> <u>UUCUUAGA</u> GUUGCAAGGGAUUGAGCCCCGUAAGGGG
N9 (76 nt guide)	AUUGCGAC <u>UCCGAAGAACGCUGAAGCGCUGGGGGCAA</u> U <u>UGUGCAAU</u> GUUGCAAGGGAUUGAGCCCCGUAAGGGG
N10 (76 nt guide)	AUUGCGAC <u>AUUCAGCAAAAUGACUUGAUCUUUGAAAUUU</u> GGAUCUUUGGUUGCAAGGGAUUGAGCCCCGUAAGGGG
N11 (76 nt guide)	AUUGCGAC <u>CAGUUUGCUGUUUCUUCUGUCUCUGCGGUAAG</u> GCUUGAGUGUUGCAAGGGAUUGAGCCCCGUAAGGGG
N12 (76 nt guide)	AUUGCGAC <u>GUCAGCACUGCUCAUGGAUUGUUGCAAUUGUU</u> <u>UGGAGAAAGUUGCAAGGGAUUGAGCCCCGUAAGGGG</u>
I1 (76 nt guide)	AUUGCGAC <u>AAAAGCGAAAACGUUUUAUAUAGCCCAUCUGCC</u> <u>UUGUGUGGGUUGCAAGGGAUUGAGCCCCGUAAGGGG</u>

Table S8.1. TtCsm guide RNAs. Related to Figures 8.1-8.3 and S8.1-S8.5.

Name	Sequence (5'-3')	Associated Figure
RNA Reporter A	/56-FAM/rCrUrCrUrCrU/3IABkFQ/	1
RNA Reporter B	/56-FAM/rArUrCrUrUrCrUrUrArU/3IABkFQ/	2 and 3

Table S8.2. Fluorescent reporter RNAs. Related to Figures 8.1-8.3 and S8.5.

Primer name	Sequence (5'-3')
SARS-CoV-2 N1 T7 template Forward	GATAATACGACTCACTATAGGGAAGTATTACAAACATTG GCCGCAAATTGCACAATT
SARS-CoV-2 N1 T7 template Reverse	GCGCGACATTCCGAAGAACGCTGAAGCGCTGGGGGCAAA TTGTGCAATTTGCGGCC
SARS-CoV-1 N1 T7 template Forward	GATAATACGACTCACTATAGGGAAGTATTACAAACATTG GCCGCAAATTGCACAATT
SARS-CoV-1 N1 T7 template Reverse	GCGTGACATTCCAAAGAATGCAGAGGCACTTGGAGCAAA TTGTGCAATTTGCGGCC
SARS-CoV-2 target all ten complexes_F	CTAGAGCTCGATAATACGACTCACTATAGGGCGTGTGTTT TAGATTTTCATCTAAACG
SARS-CoV-2 target all ten complexes_R	ATCCTGCAGGCACACTGATTAAAGATTGCTATGTG

Table S8.3. Primers to generate IVT RNA templates. Related to Figures 8.1, 8.2, S8.2, S8.3, and S8.4

Primer name	Sequence (5'-3')
F3	GCTGCTGAGGCTTCTAAG
B3	GCGTCAATATGCTTATTACAGC
BIP	TCAGCGTTCTTCGGAATGTCGCTGTGTAGGTCAACCACG
FIP	GCGGCCAATGTTTGTAATCAGTAGACGTGGTCCAGAACAA
Loop Forward	CCTTGTCTGATTAGTTCCTGGT
Loop Reverse	TGGCATGGAAGTCACACC
T7-FIP	TAATACGACTCACTATAGGGAGACGTGGTCCAGAACAA

Table S8.4. Primers for RT-LAMP. Related to Figures 8.3 and S8.5.

Name	Source
SARS-CoV-2	The National Institute of Standards and Technology(RGTM 10169)
SARS-CoV-1	American Type Culture Collection (ATCC) (VR-3280SD)
MERS-CoV	ATCC (VR-3248SD)
Human coronavirus HKU1	ATCC (VR-3262SD)
Influenza B	ATCC (VR-1885DQ)
Human coronavirus NL63	ATCC (VR-3263SD)
Human respiratory syncytial	ATCC (VR-1580DQ)
<i>Pseudomonas aeruginosa</i>	ATCC (27853D-5)
<i>Candida albicans</i>	ATCC (10231D-5)

Table S8.5. Purified genomic nucleic acids. Related to Figure 8.3

Materials and Methods

Bacterial Strains

Escherichia coli DH5a (Thermo Fisher Scientific) cells were used to amplify plasmids used in this paper. *E. coli* BL21 DE3 (NEB) cells were used to express proteins used in this paper. *E. coli* were grown in LB (Lennox) media at either 37°C, or 16°C after induction of protein expression with 0.5 mM IPTG (isopropyl-b-D-thiogalactoside).

Nucleic Acid Preparation

Previously published LAMP primers (Eurofins) were designed to amplify the SARS-CoV-2 N-gene.(Broughton et al., 2020) Target SARS-CoV-2 and SARS-CoV-1 RNAs were *in vitro* transcribed with MEGAscript T7 (Thermo Fisher Scientific) from PCR products generated from pairs of synthesized overlapping DNA oligos or using SARS-CoV-2 genome as a template (Table S8.3) (Eurofins). Previously designed primer pools (IDT) were used for RT-PCR and sequencing of SARS-CoV-2 genomes (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). Transcribed RNAs were purified by denaturing PAGE. Fluorescent reporter RNA A and fluorescent reporter RNA B purified by RNase-free HPLC (Table S8.2) (IDT). Purified genomes of viral, bacterial and fungal pathogens were used as is, or resuspended in 1x TE (10 mM Tris-HCl pH 7.5, 1 mM Ethylenediaminetetraacetic acid (EDTA)) to ~1x10⁶ genomes/mL (Table S8.5).

Plasmids

Expression vectors for *Thermus thermophilus* type III-A *csm1-csm5* genes, pCDF-5xT7-TtCsm were purchased from Addgene (plasmid # 128572).(Liu et al., 2019) pCDF-5xT7-TtCsm

was used as a template for site-directed mutagenesis to mutate the Csm3 residue D33 to alanine (D33A) to inactivate Csm3-mediated cleavage of target RNA (pCDF-5xT7-TtCsmCsm3-D34A).(Liu et al., 2017) The CRISPR array in pACYC-TtCas6-4xcrRNA4.5 (Addgene plasmid # 127764)(Liu et al., 2019) was replaced with a synthetic CRISPR array (GeneArt) containing five repeats and four identical spacers, designed to target the N-gene of SARS-CoV-2 (i.e., pACYC-TtCas6-4xgCoV2N1). TtCas6 was PCR amplified from the pACYC-TtCas6-4xcrRNA4.5 plasmid and cloned between the NcoI and XhoI sites of pRSF-1b (MilliporeSigma) (pRSF-TtCas6). The CARF-HEPN nuclease TtCsm6 was expressed from pC0075 TtCsm6 His6-TwinStrep-SUMO-BsaI (Addgene plasmid # 115270).(Gootenberg et al., 2018)

Protein Purifications

Expression and purification of the TtCsm complex was performed as previously described with minor modifications.(Liu et al., 2019) Briefly, the crRNA plasmid (e.g., pACYC-TtCas6-4xgCoV2N1) was co-transformed with pRSF-TtCas6 and either pCDF-5xT7-TtCsm or pCDF-5xT7-TtCsmCsm3-D34A into *Escherichia coli* BL21(DE3) cells and grown in LB Broth (Lennox) (Thermo Fisher Scientific) at 37°C to an OD₆₀₀ of 0.5. Cultures were then induced with 0.5 mM IPTG (isopropyl-b-D-thiogalactoside) for expression overnight at 25°C. Cells were pelleted (3,000 x *g* for 25 mins at 4°C) and lysed via sonication in Lysis buffer (25 mM HEPES pH 7.5, 150 mM KCl, 10 mM imidazole, 1 mM TCEP, 0.01 % Triton X-100, 5 % glycerol, 1 mM PMSF). Lysate was clarified by centrifugation at 10,000 x *g* for 25 mins at 4°C. The lysate was then heat-treated at 55°C for 45 minutes and further clarified by centrifugation at 10,000 x *g* for 25 mins at 4°C. His-tagged Csm1 and TtCsm complex were bound to HisTrap HP resin (Cytiva) and washed with Wash buffer (50 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP, 5 %

glycerol, 20 mM imidazole). Protein was eluted in Lysis buffer supplemented with 300 mM imidazole. Eluted protein was concentrated (100k MWCO Corning Spin-X concentrators) at 4°C before further purification over HiLoad Superdex 200 26/600 or Superose 6 Increase 10/300 GL size-exclusion columns (Cytiva) in 25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM TCEP. Fractions containing the TtCsm complex were pooled, concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Expression and purification of TtCsm6 was performed as previously described with minor modifications.(Gootenberg et al., 2018) pTtCsm6 was transformed into *Escherichia coli* BL21(DE3) cells and grown in LB Broth (Lennox) (Thermo Fisher Scientific) at 37°C to an OD₆₀₀ of 0.5. Cultures were then incubated on ice for 1 hour, and then induced with 0.5 mM IPTG for expression overnight at 16°C. Cells were lysed via sonication in TtCsm6 Lysis buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM TCEP) and lysate was clarified by centrifugation at 10,000 x *g* for 25 mins at 4°C. The lysate was heat-treated at 55°C for 45 minutes and clarified by centrifugation at 10,000 x *g* for 25 mins at 4°C. His6-TwinStrep-tagged TtCsm6 was bound to StrepTrap HP resin (Cytiva) and washed in TtCsm6 Lysis buffer. The protein was eluted with TtCsm6 Lysis buffer supplemented with 2.5 mM desthiobiotin and concentrated (10k MWCO Corning Spin-X concentrators) at 4°C. Affinity tags were removed from TtCsm6 using SUMO protease (100 mL of 2.5 mg/ ml protease per 20 mg of TtCsm6 substrate) during dialysis against SUMO digest buffer (30 mM Tris-HCl pH 8, 500 mM NaCl 1 mM DTT, 0.15% Igepal) at 4°C overnight. Cleaved His6-TwinStrep tag and uncleaved His6-TwinStrep-TtCsm6 were removed by binding to HisTrap HP resin (Cytiva), and the flow-through was concentrated using Corning Spin-X concentrators at 4°C. Finally, TtCsm6 was purified using a HiLoad Superdex 200 26/600

size-exclusion column (Cytiva) in 20 mM Tris-HCl pH 7.5, 1 mM DTT, 400 mM monopotassium glutamate, 5 % glycerol. Fractions containing TtCsm6 were pooled, concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at 80°C.

To screen guide RNAs in a high throughput format, ten TtCsm complexes were first crudely purified. 8 mL cultures of *E. coli* BL21DE3 cells transformed with pTtCsm and pT7-5xCRISPR-Cas6 were grown at 37°C and 250 RPM in LB media with selective antibiotics until they reached an OD600 reading of 0.4. Protein expression was then induced with the addition of 0.5 mM IPTG to the media, and cells were grown overnight at 16°C. Cells were collected by centrifugation at 4000 RPM, and cell pellets were resuspended in 250 mL of Ni-NTA Equilibration buffer (PBS; 100mM sodium phosphate, 600mM sodium chloride), 0.05% Tween-20 Detergent, 30mM imidazole; pH 8.0). Resuspended cells were sonicated twice for twenty seconds, then clarified by centrifugation at 15,000 rpm for 20 minutes at 4°C to remove cellular debris. The lysate was then heat-treated at 55°C for 45 minutes, and re-clarified by centrifugation at 15,000 rpm, for 30 mins at 4°C. TtCsm was then purified using HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific) according to the manufacturers recommendations, but with modified wash (25 mM HEPES pH 7.5, 150 mM NaCl, 0.05% Tween-20, 1 mM TCEP) and equilibration (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP) and elution buffers (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, 300 mM Imidazole). TtCsm complex concentration was quantified on a Nanodrop (Thermo Fisher Scientific).

Type III CRISPR-Csm Based Detection

For experiments shown in Figure 8.1C, RNA was extracted from nasopharyngeal swabs derived from patients that tested negative for SARS-CoV-2 as determined by RT-qPCR. This

RNA was used as is or spiked with *in vitro* transcribed SARS-CoV-2 or SARS-CoV-1 RNA. These RNA samples were mixed with 250 mM ATP, 500 nM fluorescent reporter RNA A, 500 nM TtCsm complex, and 2500 nM of TtCsm6 in reaction buffer (20 mM Tris-HCl pH 7.9, 200 mM Monopotassium glutamate, 10 mM Ammonium sulfate, 5 mM Magnesium sulfate and 1 mM TCEP (tris(2-carboxyethyl)phosphine)) in a 30 mL reaction. Reactions were incubated at 60°C (CRISPR-Csm alone), and fluorescence was measured over time in an ABI 7500 Fast Real-Time PCR System (Applied Biosystems), using the manufacturers default filter settings for FAM dye. Fluorescence measurements at an incubation time of 45 minutes are reported.

For experiments shown in Figure 8.2C, an RNA detection mixture was made containing either 25 nM TtCsmCsm3-D34A N1, or 25 nM TtCsmCsm3-D34A N9, or 2.5 nM each of ten complexes (TtCsmCsm3-D34A N1, N3, N6, N7, N8, N9, N10, N11, N12 and I1), mixed with 250 mM ATP, 150 nM fluorescent reporter RNA B, 300 nM TtCsm6, in reaction buffer (20 mM Tris-HCl pH 7.8, 250 mM Monopotassium glutamate, 10 mM Ammonium sulfate, 5 mM Magnesium sulfate and 1 mM TCEP). 3 mL of RNA extracted from a patient nasopharyngeal swab with high SARS-CoV-2 viral load ($\sim 5 \times 10^8$ copies/mL) was added to 27 mL of the above RNA detection mixture. Alternatively, RNA from this positive patient sample was first diluted 10- or 100-fold into RNA extracted from a patient negative for SARS-CoV-2 (CT > 40), and 3 mL of these dilutions was added to 27 mL of the above RNA detection mixture. Reactions were incubated at 60°C and fluorescence was measured every 10 s for up to 20 minutes, in a QuantStudio 3 Real-Time PCR system (ThermoFisher), using the manufacturers default filter settings for FAM dye.

Colorimetric CRISPR-Csm Based Detection

TtCsm^{Csm3-D34A} stocks were buffer exchanged into a low buffering capacity buffer (0.5 mM Tris-HCl pH 8.8, 50 mM Potassium chloride, 10 mM Ammonium sulfate, 8 mM Magnesium sulfate) using Microspin G25 columns (Cytiva) as per the manufacturer's instructions. TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) or *in vitro* transcribed SARS-CoV-2 or SARS-CoV-1 RNA were incubated with 200 nM TtCsm^{Csm3-D34A} in 1x WarmStart Colorimetric LAMP Master Mix (NEB), supplemented with an additional 1 mM ATP, in a 25 mL reaction. The volume of buffer-exchanged TtCsm used contributed approximately 40 mM Tris-HCl pH 8.8 buffer to the final reaction. Reactions were assembled on ice and imaged on an LED tracing pad with a Galaxy S9 phone (Samsung). Then reactions were incubated at 60°C for 30 minutes, rapidly cooled, and imaged again.

Visible Fluorometric CRISPR-Csm Based Detection

TE buffer or *in vitro* transcribed SARS-CoV-2 or SARS-CoV-1 RNA were incubated with 500 nM TtCsm^{Csm3-D34A} in reaction buffer (20 mM Tris-HCl pH 8.8, 100 mM Potassium chloride, 10 mM Ammonium sulfate, 6 mM Magnesium sulfate, 0.5 mM Manganese chloride, 1 mM TCEP, 1 mM ATP and 25 mM Calcein), in a 30 mL reaction. Reactions were incubated at 60°C, and fluorescence was measured over time in an ABI 7500 Fast Real-Time PCR System (Applied Biosystems), using the manufacturers default filter settings for FAM dye. After incubating at 60°C for 50 minutes, the same reactions were then imaged under visible light, and under UV light (365 nm) with a Galaxy S9 phone (Samsung). To screen guide RNAs in a high throughput format (Figure 8.2B), 200 nM crude purified TtCsm complex was incubated with

10^{12} copies of IVT SARS-CoV-2 RNA in the above buffers and fluorescence was recorded in a ABI 7500 Fast Real-Time PCR System (Applied Biosystems) machine as above.

RT-LAMP-T7-Csm

Isothermal amplification of nucleic acids in swab samples was performed by RT-LAMP. In brief, 25 μ L reactions contained 8 units (U) of WarmStart Bst 2.0 (NEB), and 7.5 U of WarmStart RTx Reverse Transcriptase (NEB), 1.4 mM dNTPs, LAMP primers, 25 U of Murine RNase Inhibitor (NEB) in reaction buffer (20 mM Tris-HCl pH 7.8, 8 mM Magnesium sulfate, 10 mM Ammonium sulfate, 50 mM potassium chloride, 0.1% Tween-20). LAMP primers designed to amplify the SARS-CoV-2 N-gene,(Broughton et al., 2020) were added at an optimized final concentration of 0.2 mM F3 and B3, 0.4 mM LoopF and LoopB, 1.6 mM BIP, 0.53 mM FIP, and 1.07 mM of T7-FIP (Table S8.4). The T7-FIP primer consists of a T7 promoter fused to the 5' end of the FIP primer, and allows for the generation of T7 transcription templates during the second step of T7-Csm reaction. RT-LAMP reactions were performed using 5 μ L of input RNA at 65°C for 29 minutes. 3 μ L of RT-LAMP reactions were mixed with 27 μ L of a modified T7-Csm fluorescent detection reaction containing 0.5 mM rNTPs, 300 nM TtCsm6, 5.5 units of Hi-T7 RNA Polymerase (NEB), 150 nM fluorescent reporter RNA B, and 20 nM of either TtCsm^{Csm3-D34A} N1 or N9, in reaction buffer (40 mM Tris-HCl pH 7.5, 4 mM Magnesium chloride, 50 mM Sodium chloride, 2 mM spermidine, 1 mM DTT). Reactions were incubated at 55°C for up to 20 min and fluorescence kinetics was monitored in a QuantStudio 3 Real-Time PCR system (ThermoFisher) as described above.

LoD standards were prepared by diluting SARS-CoV-2 RNA into RNA extracted from COVID-19-negative patient nasopharyngeal swabs. Concentrations were determined with RT-

qPCR using a standard curve generated from 10-fold dilution series (1×10^6 - 1×10^0) of IVT fragment.

Human Clinical Sample Collection and Preparation

Nasopharyngeal swabs from patients that either tested negative or positive for SARS-CoV-2 were collected in viral transport media. RNA was extracted from all patient samples using QIAamp Viral RNA Mini Kit (QIAGEN).

RT-qPCR

RT-qPCR was performed using two primers pairs (N1 and N2) and probes from the 2019-nCoV CDC EUA Kit (IDT#10006606). SARS-CoV-2 in RNA-extracted, nasopharyngeal patient samples was detected and quantified using one-step RT-qPCR in ABI 7500 Fast Real-Time PCR System according to CDC guidelines and protocols (<https://www.fda.gov/media/134922/download>). In brief, 20 μ L reactions included 8.5 μ L of Nuclease-free Water, 1.5 μ L of Primer and Probe mix (IDT, 10006713), 5 μ L of TaqPath 1-Step RT-qPCR Master Mix (ThermoFisher, A15299) and 5 μ L of the template. Nuclease-free water was used as negative template control (NTC). Amplification was performed as follows: 25°C for 2 min, 50°C for 15 min, 95°C for 2 min followed by 45 cycles of 95°C for 3 s and 55°C for 30 s. To quantify viral genome copy numbers in the samples, standard curves for N1 and N2 were generated using a dilution series of a SARS-CoV-2 synthetic RNA fragment (RTGM 10169, National Institute of Standards and Technology) spanning N gene with concentrations ranging from 10 to 10^6 copies per μ L. Three technical replicates were performed at each dilution. The NTC showed no amplification throughout the 45 cycles of qPCR.

Bioinformatic Design of TtCsm crRNA Guides Targeting SARS-CoV-2

An alignment of 45,641 SARS-CoV-2 genomes was downloaded from the GISAID database (Global Initiative for Sharing All Influenza Data; <https://www.gisaid.org/> on 6-23-2020.(Elbe & Buckland-Merrett, 2017; Katoh et al., 2005) The alignment was scanned for conservation with a 40-nucleotide sliding window, and 40-nucleotide segments with strong conservation were saved for downstream analysis. Next, four nucleotides flanking the above 40-nucleotide candidate viral target sequences were checked for base pairing to the first four nucleotides of the prospective 5'-crRNA handle (underlined; 5'-AUUGCGAC-3'), only candidates lacking handle complementarity were considered further. Candidate sites with less than two mismatches to SARS-CoV (NC_004718.3) and MERS-CoV (NC_019843.3) in the first 18 nucleotides of the target sequence were discarded. Next, candidate crRNAs targeting the above sites were screened for potential cross-reactivity with human mRNAs and a list of human pathogens and common respiratory flora downloaded from the FDA's Emergency Use Authorization requirements (downloaded on 7-29-2020) using BLAST (E-value 1000). The remaining 6,229 crRNA sequences were then sorted by genomic location and only guides that were located 3' of the SARS-CoV-2 ORF3a gene (positions 25,393 to 29,903) were considered further. Finally, 76 guides were selected from the remaining pool that had the greatest conservation among SARS-CoV-2 sequences and the largest number of mismatches to SARS-CoV and MERS-CoV sequences.

Sequencing of SARS-CoV-2 RNA Isolated from Patient Samples

SARS-CoV-2 genomic RNA isolated from patient samples was sequenced as previously described.(Nemudryi et al., 2021) In brief, 10 mL of SARS-CoV-2 genomic RNA extracted from

nasopharyngeal patient swabs was first reverse transcribed with SuperScript IV (ThermoFisher) according to the manufacturer's instructions. The ARTIC Network protocol was followed to generate a sequence amplicon library covering the whole SARS-CoV-2 genome on Oxford Nanopore using a ligation sequencing kit (Oxford Nanopore, SQK-LSK109) (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). (N. D. Grubaugh et al., 2019; Tyson et al., 2020) Two multiplex PCR reactions were performed with primer pools described in the ARTIC nCoV-2019 V3 Panel, amplified with Q5 DNA Polymerase (NEB). The two resulting amplicon pools for each patient sample were then combined and used for library preparation. Samples were end repaired (NEB, E7546) and then barcoded using Native Barcoding Expansion Kits (Oxford Nanopore, EXP-NBD104 and EXP-NBD114). Barcoded samples were pooled together and then Nanopore adaptors were ligated.

The multiplexed library was loaded onto the MinION flowcell, and a total of 0.3 Gb of raw sequencing data was collected per patient sample. Raw Nanopore reads were base-called in high-accuracy mode (Oxford Nanopore, MinKNOW), and further analyzed using the ARTIC bioinformatic pipeline for COVID-19 (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). (Loman, 2020) Consensus sequences were uploaded to GISAID (<https://www.gisaid.org/>), IDs: EPI_ISL_1081321, EPI-ISL_1081322, EPI_ISL_1081323. (Elbe & Buckland-Merrett, 2017) These three SARS-CoV-2 genome sequences were identified as members of the B.1.1.7 lineage by an automated lineage assigner (<https://github.com/cov-lineages/pangolin>). (Rambaut et al., 2020)

Quantification and Statistical Analysis

All experiments were performed in triplicate or duplicate and error is reported as ± 1 standard deviation. The merged datasets of replicates of fluorescence kinetics of direct Csm-based detection of SARS-CoV-2 RNA in patient samples was fit to a simple linear regression, in Prism 9 (Graphpad). The fitted slopes of SARS-CoV-2 RNA-containing patient samples were compared pairwise to the negative swab RNA control by an F-test, **** $p < 0.0001$.

References Cited

- Athukoralage, J. S., Graham, S., Rouillon, C., Gruschow, S., Czekster, C. M., & White, M. F. (2020, Apr 27). The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *Elife*, 9. <https://doi.org/10.7554/eLife.55852>
- Biosciences, M. (2020). *SARS-CoV-2 RNA DETECTR Reagent Kit*.
- Biosciences, S. (2020). *Instructions for use: Sherlock™ CRISPR SARS-CoV-2 kit*.
- Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., Miao, X., Streithorst, J. A., Granados, A., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Hsu, E., Gu, W., Miller, S., Pan, C. Y., Guevara, H., Wadford, D. A., Chen, J. S., & Chiu, C. Y. (2020, Jul). CRISPR-Cas12-based detection of SARS-CoV-2. *Nature biotechnology*, 38(7), 870-874. <https://doi.org/10.1038/s41587-020-0513-4>
- Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018, Apr 27). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, 360(6387), 436-439. <https://doi.org/10.1126/science.aar6245>
- Dao Thi, V. L., Herbst, K., Boerner, K., Meurer, M., Kremer, L. P., Kirrmaier, D., Freistaedter, A., Papagiannidis, D., Galmozzi, C., Stanifer, M. L., Boulant, S., Klein, S., Chlanda, P., Khalid, D., Barreto Miranda, I., Schnitzler, P., Krausslich, H. G., Knop, M., & Anders, S. (2020, Aug 12). A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science translational medicine*, 12(556). <https://doi.org/10.1126/scitranslmed.abc7075>
- East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., & Doudna, J. A. (2016, Oct 13). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, 538(7624), 270-273. <https://doi.org/10.1038/nature19802>
- Elbe, S., & Buckland-Merrett, G. (2017, Jan). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*, 1(1), 33-46. <https://doi.org/10.1002/gch2.1018>
- Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016, Feb 15). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev*, 30(4), 447-459. <https://doi.org/10.1101/gad.272153.115>
- Fozouni, P., Son, S., Diaz de Leon Derby, M., Knott, G. J., Gray, C. N., D'Ambrosio, M. V., Zhao, C., Switz, N. A., Kumar, G. R., Stephens, S. I., Boehm, D., Tsou, C. L., Shu, J., Bhuiya, A., Armstrong, M., Harris, A. R., Chen, P. Y., Osterloh, J. M., Meyer-Franke, A., Joehnk, B., Walcott, K., Sil, A., Langelier, C., Pollard, K. S., Crawford, E. D.,

- Puschnik, A. S., Phelps, M., Kistler, A., DeRisi, J. L., Doudna, J. A., Fletcher, D. A., & Ott, M. (2021, Jan 21). Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell*, 184(2), 323-333 e329. <https://doi.org/10.1016/j.cell.2020.12.001>
- Gao, X., Sun, B., & Guan, Y. (2019, Feb). Pullulan reduces the non-specific amplification of loop-mediated isothermal amplification (LAMP). *Anal Bioanal Chem*, 411(6), 1211-1218. <https://doi.org/10.1007/s00216-018-1552-2>
- Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., & Zhang, F. (2018, Apr 27). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, 360(6387), 439-444. <https://doi.org/10.1126/science.aag0179>
- Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A., Myhrvold, C., Bhattacharyya, R. P., Livny, J., Regev, A., Koonin, E. V., Hung, D. T., Sabeti, P. C., Collins, J. J., & Zhang, F. (2017, Apr 28). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, 356(6336), 438-442. <https://doi.org/10.1126/science.aam9321>
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019, Jan 8). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*, 20(1), 8. <https://doi.org/10.1186/s13059-018-1618-7>
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009, Nov 25). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), 945-956. <https://doi.org/10.1016/j.cell.2009.07.040>
- Hardinge, P., & Murray, J. A. H. (2019, May 14). Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification using Quenched Fluorescent Primers. *Sci Rep*, 9(1), 7400. <https://doi.org/10.1038/s41598-019-43817-z>
- Jia, N., Jones, R., Sukenick, G., & Patel, D. J. (2019, Sep 5). Second Messenger cA4 Formation within the Composite Csm1 Palm Pocket of Type III-A CRISPR-Cas Csm Complex and Its Release Path. *Mol Cell*, 75(5), 933-943 e936. <https://doi.org/10.1016/j.molcel.2019.06.013>
- Jia, N., Jones, R., Yang, G., Ouerfelli, O., & Patel, D. J. (2019, Sep 5). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Mol Cell*, 75(5), 944-956 e946. <https://doi.org/10.1016/j.molcel.2019.06.014>

- Joung, J., Ladha, A., Saito, M., Kim, N. G., Woolley, A. E., Segel, M., Barretto, R. P. J., Ranu, A., Macrae, R. K., Faure, G., Ioannidi, E. I., Krajewski, R. N., Bruneau, R., Huang, M. W., Yu, X. G., Li, J. Z., Walker, B. D., Hung, D. T., Greninger, A. L., Jerome, K. R., Gootenberg, J. S., Abudayyeh, O. O., & Zhang, F. (2020, Oct 8). Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. *The New England journal of medicine*, 383(15), 1492-1494. <https://doi.org/10.1056/NEJMc2026172>
- Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 33(2), 511-518. <https://doi.org/10.1093/nar/gki198>
- Kazlauskienė, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., & Siksnys, V. (2017, Aug 11). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, 357(6351), 605-609. <https://doi.org/10.1126/science.aao0100>
- Kazlauskienė, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., & Siksnys, V. (2016, Apr 21). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol Cell*, 62(2), 295-306. <https://doi.org/10.1016/j.molcel.2016.03.024>
- La Scola, B., Le Bideau, M., Andreani, J., Hoang, V. T., Grimaldier, C., Colson, P., Gautret, P., & Raoult, D. (2020, Jun). Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *Eur J Clin Microbiol Infect Dis*, 39(6), 1059-1061. <https://doi.org/10.1007/s10096-020-03913-9>
- Larremore, D. B., Wilder, B., Lester, E., Shehata, S., Burke, J. M., Hay, J. A., Tambe, M., Mina, M. J., & Parker, R. (2021, Jan). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv*, 7(1). <https://doi.org/10.1126/sciadv.abd5393>
- Lazer, D. S., M.; Perlis, R.H.; Ognyanova, K.; Baum, M.A.; Quintana, A.; Druckman, J.; Volpe, J.D.; Chwe, H.; Simonson, M. (2020). *State of the Nation: A 50-State COVID-19 Survey: Report #8: Failing the Test: Waiting Times for COVID Diagnostic Tests Across the U.S.* <https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/en/grc-740424>
- Liu, T. Y., Iavarone, A. T., & Doudna, J. A. (2017). RNA and DNA Targeting by a Reconstituted *Thermus thermophilus* Type III-A CRISPR-Cas System. *PloS one*, 12(1), e0170552. <https://doi.org/10.1371/journal.pone.0170552>
- Liu, T. Y., Liu, J. J., Aditham, A. J., Nogales, E., & Doudna, J. A. (2019, Jul 5). Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. *Nat Commun*, 10(1), 3001. <https://doi.org/10.1038/s41467-019-10780-2>

- Loman, N. J. R., W.; Rambaut, A. (2020). *NCoV-2019 Novel Coronavirus Bioinformatics Protocol*. <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>
- Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksny, V., Venclovas, C., & Koonin, E. V. (2020, Sep 18). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antiviral defense. *Nucleic Acids Res*, 48(16), 8828-8847. <https://doi.org/10.1093/nar/gkaa635>
- Meagher, R. J., Priye, A., Light, Y. K., Huang, C., & Wang, E. (2018, Apr 16). Impact of primer dimers and self-amplifying hairpins on reverse transcription loop-mediated isothermal amplification detection of viral RNA. *Analyst*, 143(8), 1924-1933. <https://doi.org/10.1039/c7an01897e>
- Nasef, M., Muffly, M. C., Beckman, A. B., Rowe, S. J., Walker, F. C., Hatoum-Aslan, A., & Dunkle, J. A. (2019, Aug). Regulation of cyclic oligoadenylate synthesis by the *Staphylococcus epidermidis* Cas10-Csm complex. *RNA*, 25(8), 948-962. <https://doi.org/10.1261/rna.070417.119>
- Nemudryi, A., Nemudraia, A., Wiegand, T., Nichols, J., Snyder, D. T., Hedges, J. F., Cicha, C., Lee, H., Vanderwood, K. K., Bimczok, D., Jutila, M. A., & Wiedenheft, B. (2021, Jun 1). SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Rep*, 35(9), 109197. <https://doi.org/10.1016/j.celrep.2021.109197>
- Niewoehner, O., Garcia-Doval, C., Rostol, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017, Aug 31). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, 548(7669), 543-548. <https://doi.org/10.1038/nature23467>
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000, Jun 15). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*, 28(12), E63. <https://doi.org/10.1093/nar/28.12.e63>
- Ooi, K. H., Liu, M. M., Tay, J. W. D., Teo, S. Y., Kaewsapsak, P., Jin, S., Lee, C. K., Hou, J., Maurer-Stroh, S., Lin, W., Yan, B., Yan, G., Gao, Y. G., & Tan, M. H. (2021, Mar 19). An engineered CRISPR-Cas12a variant and DNA-RNA hybrid guides enable robust and rapid COVID-19 testing. *Nat Commun*, 12(1), 1739. <https://doi.org/10.1038/s41467-021-21996-6>
- Paltiel, A. D., Zheng, A., & Walensky, R. P. (2020, Jul 1). Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. *JAMA Netw Open*, 3(7), e2016818. <https://doi.org/10.1001/jamanetworkopen.2020.16818>

- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020, Nov). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*, 5(11), 1403-1407.
<https://doi.org/10.1038/s41564-020-0770-5>
- Rolando, J. C., Jue, E., Barlow, J. T., & Ismagilov, R. F. (2020, Apr 17). Real-time kinetics and high-resolution melt curves in single-molecule digital LAMP to differentiate and study specific and non-specific amplification. *Nucleic Acids Res*, 48(7), e42.
<https://doi.org/10.1093/nar/gkaa099>
- Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S., & White, M. F. (2018, Jul 2). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife*, 7.
<https://doi.org/10.7554/eLife.36734>
- Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015, May 21). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, 161(5), 1164-1174. <https://doi.org/10.1016/j.cell.2015.04.027>
- Sofos, N., Feng, M., Stella, S., Pape, T., Fuglsang, A., Lin, J., Huang, Q., Li, Y., She, Q., & Montoya, G. (2020, Sep 3). Structures of the Cmr-beta Complex Reveal the Regulation of the Immunity Mechanism of Type III-B CRISPR-Cas. *Mol Cell*, 79(5), 741-757 e747.
<https://doi.org/10.1016/j.molcel.2020.07.008>
- Staals, R. H., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., Koehorst, J. J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P. J., Urlaub, H., Heck, A. J., Nogales, E., Doudna, J. A., Shinkai, A., & van der Oost, J. (2014, Nov 20). RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell*, 56(4), 518-530.
<https://doi.org/10.1016/j.molcel.2014.10.005>
- Tamulaitis, G., Kazlauskienė, M., Manakova, E., Venclovas, C., Nwokeoji, A. O., Dickman, M. J., Horvath, P., & Siksnys, V. (2014, Nov 20). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol Cell*, 56(4), 506-517.
<https://doi.org/10.1016/j.molcel.2014.09.027>
- Tomita, N., Mori, Y., Kanda, H., & Notomi, T. (2008). Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat Protoc*, 3(5), 877-882. <https://doi.org/10.1038/nprot.2008.57>
- Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajek, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020, Sep 4). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRxiv*. <https://doi.org/10.1101/2020.09.04.283077>

- Wolfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Muller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brunink, S., Schneider, J., Ehmann, R., Zwirgmaier, K., Drosten, C., & Wendtner, C. (2020, May). Virological assessment of hospitalized patients with COVID-2019. *Nature*, 581(7809), 465-469. <https://doi.org/10.1038/s41586-020-2196-x>
- Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019, Jan 4). Functionally diverse type V CRISPR-Cas systems. *Science*, 363(6422), 88-91. <https://doi.org/10.1126/science.aav7271>
- You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X., & Wang, Y. (2019, Jan 10). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell*, 176(1-2), 239-253 e216. <https://doi.org/10.1016/j.cell.2018.10.052>
- Zhang, J., Graham, S., Tello, A., Liu, H., & White, M. F. (2016, Feb 29). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res*, 44(4), 1789-1799. <https://doi.org/10.1093/nar/gkw020>
- Zhang, Y., Ren, G., Buss, J., Barry, A. J., Patton, G. C., & Tanner, N. A. (2020, Sep). Enhancing colorimetric loop-mediated isothermal amplification speed and sensitivity with guanidine chloride. *Biotechniques*, 69(3), 178-185. <https://doi.org/10.2144/btn-2020-0078>
- Zou, Y., Mason, M. G., & Botella, J. R. (2020). Evaluation and improvement of isothermal amplification methods for point-of-need plant disease diagnostics. *PloS one*, 15(6), e0235216. <https://doi.org/10.1371/journal.pone.0235216>

REFERENCES CITED

REFERENCES CITED

- Addetia, A., Xie, H., Roychoudhury, P., Shrestha, L., Loprieno, M., Huang, M. L., Jerome, K. R., & Greninger, A. L. (2020, Aug). Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol*, 129, 104523. <https://doi.org/10.1016/j.jcv.2020.104523>
- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *The Science of the total environment*, 728, 138764-138764. <https://doi.org/10.1016/j.scitotenv.2020.138764>
- Aid, F. (2020). Plant Lipid Metabolism. In. IntechOpen. <https://doi.org/10.5772/intechopen.81355>
- Alkhamis, Y., Mathew, R., Ganesan, N., Rahman, S., & Rahman, M. M. (2022, 10/14). pH induced stress enhances lipid accumulation in microalgae grown under mixotrophic and autotrophic condition. *Frontiers in Energy Research*, 10, 1033068. <https://doi.org/10.3389/fenrg.2022.1033068>
- Álvarez, I., & Wendel, J. F. (2003, 2003/12/01/). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, 29(3), 417-434. [https://doi.org/https://doi.org/10.1016/S1055-7903\(03\)00208-2](https://doi.org/https://doi.org/10.1016/S1055-7903(03)00208-2)
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1). <https://doi.org/10.1186/s13059-020-1935-5>
- Ambrose, S., Bridges, M. W., Lovett, M. C., DiPietro, M., Norman, M. K. . (2010). *How Learning Works: 7 Research-Based Principles for Smart Teaching* (Jossey-Bass, Ed.).
- Andolfo, I., Russo, R., Lasorsa, V. A., Cantalupo, S., Rosato, B. E., Bonfiglio, F., Frisso, G., Abete, P., Cassese, G. M., Servillo, G., Esposito, G., Gentile, I., Piscopo, C., Villani, R., Fiorentino, G., Cerino, P., Buonerba, C., Pierri, B., Zollo, M., Iolascon, A., & Capasso, M. (2021, Apr 23). Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. *iScience*, 24(4), 102322. <https://doi.org/10.1016/j.isci.2021.102322>
- Andrew Rambaut, N. L., Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson, Erik Volz, on behalf of COVID-19 Genomics Consortium UK (CoG-UK). (2020). *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.*

- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7), 2251-2252. <https://doi.org/10.1093/bioinformatics/btz859>
- Arisz, S. A., van Himbergen, J. A., Musgrave, A., van den Ende, H., & Munnik, T. (2000, Jan). Polar glycerolipids of *Chlamydomonas moewusii*. *Phytochemistry*, 53(2), 265-270. [https://doi.org/10.1016/s0031-9422\(99\)00505-1](https://doi.org/10.1016/s0031-9422(99)00505-1)
- Aro, E.-M., & Andersson, B. (2006). *Regulation of photosynthesis* (Vol. 11). Springer Science & Business Media.
- Arriola, M. B., Velmurugan, N., Zhang, Y., Plunkett, M. H., Hondzo, H., & Barney, B. M. (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. <https://doi.org/10.1111/tpj.13789>
- Arunachalam, P. S., Wimmers, F., Mok, C. K. P., Perera, R., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., Wagh, D., Coller, J., Pellegrini, K. L., Kazmin, D., Alaaeddine, G., Leung, W. S., Chan, J. M. C., Chik, T. S. H., Choi, C. Y. C., Huerta, C., Paine McCullough, M., Lv, H., Anderson, E., Edupuganti, S., Upadhyay, A. A., Bosinger, S. E., Maecker, H. T., Khatri, P., Roupahel, N., Peiris, M., & Pulendran, B. (2020, Sep 4). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science*, 369(6508), 1210-1220. <https://doi.org/10.1126/science.abc6261>
- Athukoralage, J. S., Graham, S., Rouillon, C., Gruschow, S., Czekster, C. M., & White, M. F. (2020, Apr 27). The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *Elife*, 9. <https://doi.org/10.7554/eLife.55852>
- Avanzato, V. A., Matson, M. J., Seifert, S. N., Pryce, R., Williamson, B. N., Anzick, S. L., Barbian, K., Judson, S. D., Fischer, E. R., Martens, C., Bowden, T. A., de Wit, E., Riedo, F. X., & Munster, V. J. (2020, Dec 23). Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell*, 183(7), 1901-1912 e1909. <https://doi.org/10.1016/j.cell.2020.10.049>
- Avidan, O., & Pick, U. (2015). Acetyl-CoA synthetase is activated as part of the PDH-bypass in the oleaginous green alga *Chlorella desiccata*. *Journal of Experimental Botany*, 66(22), 7287-7298. <https://doi.org/10.1093/jxb/erv424>
- Banerjee, A., El-Sayes, N., Budylowski, P., Jacob, R. A., Richard, D., Maan, H., Aguiar, J. A., Demian, W. L., Baid, K., D'Agostino, M. R., Ang, J. C., Murdza, T., Tremblay, B. J., Afkhami, S., Karimzadeh, M., Irving, A. T., Yip, L., Ostrowski, M., Hirota, J. A., Kozak,

- R., Capellini, T. D., Miller, M. S., Wang, B., Mubareka, S., McGeer, A. J., McArthur, A. G., Doxey, A. C., & Mossman, K. (2021, May 21). Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience*, 24(5), 102477. <https://doi.org/10.1016/j.isci.2021.102477>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1). <https://doi.org/10.1186/s13100-015-0041-9>
- Barr, R. B., & Tagg, J. (1995, 1995/11/01). From Teaching to Learning — A New Paradigm For Undergraduate Education. *Change: The Magazine of Higher Learning*, 27(6), 12-26. <https://doi.org/10.1080/00091383.1995.10544672>
- Bauer, D. E., Hatzivassiliou, G., Zhao, F., Andreadis, C., & Thompson, C. B. (2005). ATP citrate lyase is an important component of cell growth and transformation. *Oncogene*, 24(41), 6314-6322. <https://doi.org/10.1038/sj.onc.1208773>
- BBDuk. <https://sourceforge.net/projects/bbmap/>
- Beck, A. E., Hunt, K. A., & Carlson, R. P. (2018). Measuring Cellular Biomass Composition for Computational Biology Applications. *Processes*, 6(5), 38. <https://www.mdpi.com/2227-9717/6/5/38>
- Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M. L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T. R., Truong, M., Wolf, C. R., Nickerson, D. A., Rieder, M. J., Englund, J. A., Seattle Flu Study, I., Hadfield, J., Hodcroft, E. B., Huddleston, J., Moncla, L. H., Muller, N. F., Neher, R. A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J. S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G. L., Baird, G. S., Chu, H. Y., Shendure, J., & Jerome, K. R. (2020, Oct 30). Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516), 571-575. <https://doi.org/10.1126/science.abc0523>
- Behrenfeld, M. J. (2001). Biospheric Primary Production During an ENSO Transition. *Science*, 291(5513), 2594-2597. <https://doi.org/10.1126/science.1055071>
- Bellou, S., & Aggelis, G. (2013). Biochemical activities in *Chlorella* sp. and *Nannochloropsis salina* during lipid and sugar synthesis in a lab-scale open pond simulating reactor. *Journal of Biotechnology*, 164(2), 318-329. <https://doi.org/10.1016/j.jbiotec.2013.01.010>
- Biggerstaff, M., Jhung, M. A., Reed, C., Fry, A. M., Balluz, L., & Finelli, L. (2014). Influenza-like illness, the time to seek healthcare, and influenza antiviral receipt during the 2010-

- 2011 influenza season-United States. *The Journal of infectious diseases*, 210(4), 535-544. <https://doi.org/10.1093/infdis/jiu224>
- Biosciences, M. (2020). *SARS-CoV-2 RNA DETECTR Reagent Kit*.
- Biosciences, S. (2020). *Instructions for use: Sherlock™ CRISPR SARS-CoV-2 kit*.
- Blanc-Mathieu, R., Verhelst, B., Derelle, E., Rombauts, S., Bouget, F.-Y., Carré, I., Château, A., Eyre-Walker, A., Grimsley, N., Moreau, H., Piégu, B., Rivals, E., Schackwitz, W., Van De Peer, Y., & Piganeau, G. (2014). An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics*, 15(1), 1103. <https://doi.org/10.1186/1471-2164-15-1103>
- Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Moller, R., Jordan, T. X., Oishi, K., Panis, M., Sachs, D., Wang, T. T., Schwartz, R. E., Lim, J. K., Albrecht, R. A., & tenOever, B. R. (2020, May 28). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, 181(5), 1036-1045 e1039. <https://doi.org/10.1016/j.cell.2020.04.026>
- Boer, V. M., Crutchfield, C. A., Bradley, P. H., Botstein, D., & Rabinowitz, J. D. (2010). Growth-limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations. *Molecular Biology of the Cell*, 21(1), 198-211. <https://doi.org/10.1091/mbc.e09-07-0597>
- Bold, H. C. (1949). The Morphology of *Chlamydomonas chlamydogama*, Sp. Nov. *Bulletin of the Torrey Botanical Club*, 76(2), 101. <https://doi.org/10.2307/2482218>
- Bradley, J., Luo, R., Otis, T. S., & DiGregorio, D. A. (2009). Submillisecond Optical Reporting of Membrane Potential & In Situ Using a Neuronal Tracer Dye. *The Journal of Neuroscience*, 29(29), 9197. <https://doi.org/10.1523/JNEUROSCI.1240-09.2009>
- Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., Miao, X., Streithorst, J. A., Granados, A., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Hsu, E., Gu, W., Miller, S., Pan, C. Y., Guevara, H., Wadford, D. A., Chen, J. S., & Chiu, C. Y. (2020, Jul). CRISPR-Cas12-based detection of SARS-CoV-2. *Nature biotechnology*, 38(7), 870-874. <https://doi.org/10.1038/s41587-020-0513-4>
- Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth's lakes. *Geophysical Research Letters*, 44(1), 209-218. <https://doi.org/10.1002/2016gl071378>
- Cai, H., Li, Q., Fang, X., Li, J., Curtis, N. E., Altenburger, A., Shibata, T., Feng, M., Maeda, T., Schwartz, J. A., Shigenobu, S., Lundholm, N., Nishiyama, T., Yang, H., Hasebe, M., Li,

- S., Pierce, S. K., & Wang, J. (2019). A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Scientific Data*, 6(1), 190022. <https://doi.org/10.1038/sdata.2019.22>
- Calhoun, S., Bell, T. A. S., Dahlin, L. R., Kunde, Y., Labutti, K., Louie, K. B., Kuftin, A., Treen, D., Dilworth, D., Mihaltcheva, S., Daum, C., Bowen, B. P., Northen, T. R., Guarnieri, M. T., Starkenburg, S. R., & Grigoriev, I. V. (2021). A multi-omic characterization of temperature stress in a halotolerant *Scenedesmus* strain for algal biotechnology. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-021-01859-y>
- Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48(1). <https://doi.org/10.1002/0471250953.bi0411s48>
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., & Gutell, R. R. (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3(1), 2. <https://doi.org/10.1186/1471-2105-3-2>
- Canon-Rubio, K. A., Sharp, C. E., Bergerson, J., Strous, M., & De la Hoz Siegler, H. (2016, Feb). Use of highly alkaline conditions to improve cost-effectiveness of algal biotechnology. *Applied Microbiology and Biotechnology*, 100(4), 1611-1622. <https://doi.org/10.1007/s00253-015-7208-7>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825-5829. <https://doi.org/10.1093/molbev/msab293>
- Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., & Yandell, M. (2007). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18(1), 188-196. <https://doi.org/10.1101/gr.6743907>
- Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009, Aug 1). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Castillo, A. E., Parra, B., Tapia, P., Acevedo, A., Lagos, J., Andrade, W., Arata, L., Leal, G., Barra, G., Tambley, C., Tognarelli, J., Bustos, P., Ulloa, S., Fasce, R., & Fernández, J. (2020). Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *Journal of medical virology*, 92(9), 1562-1566. <https://doi.org/10.1002/jmv.25797>

- Chapman, R. L. (2013). Algae: the world's most important "plants"—an introduction. *Mitigation and Adaptation Strategies for Global Change*, 18(1), 5-12.
<https://doi.org/10.1007/s11027-010-9255-9>
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In (pp. 207-232). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35306-5_10
- Chen, J., Li J., Dong, W., Zhang, X., Tyagi, R. D., Drogui, P., & Surampalli, R. Y. (2018, 2018/07/01/). The potential of microalgae in biodiesel production. *Renewable and sustainable energy reviews*, 90, 336-346.
<https://doi.org/https://doi.org/10.1016/j.rser.2018.03.073>
- Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018, Apr 27). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, 360(6387), 436-439. <https://doi.org/10.1126/science.aar6245>
- Chen, W., Feng, C., Guo, W., Shi, D., & Yang, C. (2011). Comparative effects of osmotic-, salt- and alkali stress on growth, photosynthesis, and osmotic adjustment of cotton plants. *Photosynthetica*, 49(3), 417-425. <https://doi.org/10.1007/s11099-011-0050-y>
- Chinese, S. M. E. C. (2004, Mar 12). Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science*, 303(5664), 1666-1669.
<https://doi.org/10.1126/science.1092002>
- Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, 25(3), 294-306.
<https://doi.org/10.1016/j.biotechadv.2007.02.001>
- Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., Solomon, I. H., Kuo, H. H., Boucau, J., Bowman, K., Adhikari, U. D., Winkler, M. L., Mueller, A. A., Hsu, T. Y., Desjardins, M., Baden, L. R., Chan, B. T., Walker, B. D., Lichterfeld, M., Brigl, M., Kwon, D. S., Kanjilal, S., Richardson, E. T., Jonsson, A. H., Alter, G., Barczak, A. K., Hanage, W. P., Yu, X. G., Gaiha, G. D., Seaman, M. S., Cernadas, M., & Li, J. Z. (2020, Dec 3). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *The New England journal of medicine*, 383(23), 2291-2293.
<https://doi.org/10.1056/NEJMc2031364>
- Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C., & Jones, C. L. Q., R.B. Myneni, S. Piao and P. Thornton., (2013). *Carbon and Other Biogeochemical Cycles*. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

- Cirulis, J. T., Strasser, B. C., Scott, J. A., & Ross, G. M. (2012). Optimization of staining conditions for microalgae with three lipophilic dyes to reduce precipitation and fluorescence variability. *Cytometry Part A*, 81A(7), 618-626. <https://doi.org/10.1002/cyto.a.22066>
- Cooksey, K. E., Guckert, J. B., Williams, S. A., & Callis, P. R. (1987). Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. 6(6), 333-345. [https://doi.org/10.1016/0167-7012\(87\)90019-4](https://doi.org/10.1016/0167-7012(87)90019-4)
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- Dao Thi, V. L., Herbst, K., Boerner, K., Meurer, M., Kremer, L. P., Kirrmaier, D., Freistaedter, A., Papagiannidis, D., Galmozzi, C., Stanifer, M. L., Boulant, S., Klein, S., Chlanda, P., Khalid, D., Barreto Miranda, I., Schnitzler, P., Krausslich, H. G., Knop, M., & Anders, S. (2020, Aug 12). A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science translational medicine*, 12(556). <https://doi.org/10.1126/scitranslmed.abc7075>
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., Group, C. C.-W., Consortium, C.-G. U., Diaz-Ordaz, K., Keogh, R., Eggo, R. M., Funk, S., Jit, M., Atkins, K. E., & Edmunds, W. J. (2021, Apr 9). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538). <https://doi.org/10.1126/science.abg3055>
- De Bigault Du Granrut, A., & Cacas, J.-L. (2016). How Very-Long-Chain Fatty Acids Could Signal Stressful Conditions in Plants? *Frontiers in Plant Science*, 7, 1490-1490. <https://doi.org/10.3389/fpls.2016.01490>
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518-524. <https://doi.org/10.1038/nbt.3423>
- Delrue, F., Setier, P.-A., Sahut, C., Cournac, L., Roubaud, A., Peltier, G., & Froment, A.-K. (2012). An economic, sustainability, and energetic model of biodiesel production from microalgae. *Bioresource Technology*, 111, 191-200.
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361-375. <https://doi.org/10.1038/nrg1603>
- Demirhan, G. (2014, 01/01). A Validity and Reliability Study of the Motivated Strategies for Learning Questionnaire. *KUYEB*, 14, 9-13.

- Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O. G., Faria, N. R., Wang, C., Yu, G., Bushnell, B., Pan, C.-Y., Guevara, H., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Servellita, V., Hsu, E., Miller, S., Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Chu, H. Y., Shendure, J., Jerome, K. R., Anderson, C., Gangavarapu, K., Zeller, M., Spencer, E., Andersen, K. G., MacCannell, D., Paden, C. R., Li, Y., Zhang, J., Tong, S., Armstrong, G., Morrow, S., Willis, M., Matyas, B. T., Mase, S., Kasirye, O., Park, M., Masinde, G., Chan, C., Yu, A. T., Chai, S. J., Villarino, E., Bonin, B., Wadford, D. A., & Chiu, C. Y. (2020). Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science (New York, N.Y.)*, 369(6503), 582-587. <https://doi.org/10.1126/science.abb9263>
- Deshpande, R. A., & Shankar, V. (2002, 2002/01/01). Ribonucleases from T2 Family. *Critical Reviews in Microbiology*, 28(2), 79-122. <https://doi.org/10.1080/1040-840291046704>
- Dohm, J. C., Peters, P., Stralis-Pavese, N., & Himmelbauer, H. (2020). Benchmarking of long-read correction methods. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/nargab/lqaa037>
- Doucha, J., Straka, F., & Lívanský, K. (2005). Utilization of flue gas for cultivation of microalgae *Chlorella* sp.) in an outdoor open thin-layer photobioreactor. *Journal of Applied Phycology*, 17(5), 403-412. <https://doi.org/10.1007/s10811-005-8701-7>
- Dragone, G., Fernandes, B. D., Vicente, A. A., & Teixeira, J. A. (2010). Third generation biofuels from microalgae.
- Drost, H.-G., Gabel, A., Grosse, I., & Quint, M. (2015). Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Molecular Biology and Evolution*, 32(5), 1221-1231. <https://doi.org/10.1093/molbev/msv012>
- Dündar, E., & Bush, D. R. (2009). BAT1, a bidirectional amino acid transporter in Arabidopsis. *Planta*, 229(5), 1047-1056. <https://doi.org/10.1007/s00425-009-0892-8>
- Eakins, B. W., G.F. Sharman. (2010). *Volumes of the World's Oceans from ETOPO1* NOAA National Geophysical Data Center, Boulder, CO.
- East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., & Doudna, J. A. (2016, Oct 13). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, 538(7624), 270-273. <https://doi.org/10.1038/nature19802>
- Eddy, S. R., & Nawrocki, E. P. (2009). Structural rna homology search and alignment using covariance models.

- Edgar, R. C. *MUSCLE: multiple sequence alignment with improved accuracy and speed* Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004., <http://dx.doi.org/10.1109/csb.2004.1332560>
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Erik, L., Hirsh, L., Paladin, L., Piovesan, D., Silvio, C., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427-D432. <https://doi.org/10.1093/nar/gky995>
- El-Sheekh, M., & Rady, A. (1995, 01/01). Effect of phosphorus starvation on growth, photosynthesis and some metabolic processes in the unicellular green alga *Chlorella kessleri*. *Phyton*, 35, 139-151.
- Elbe, S., & Buckland-Merrett, G. (2017, Jan). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*, 1(1), 33-46. <https://doi.org/10.1002/gch2.1018>
- Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9(1), 18. <https://doi.org/10.1186/1471-2105-9-18>
- Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016, Feb 15). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev*, 30(4), 447-459. <https://doi.org/10.1101/gad.272153.115>
- Emeny, J. M., & Morgan, M. J. (1979, Apr). Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *J Gen Virol*, 43(1), 247-252. <https://doi.org/10.1099/0022-1317-43-1-247>
- Energy, D. o. (2012). *algae biofuels, E.E.R.E U.S. Department of Energy in: Growing America's Energy Future*. .
- Energy, U. S. D. o., Energy, O. o. E. E. a. R., & Office, B. T. (2017). *Algae Cultivation for Carbon Capture and Utilization Workshop Summary Report*.
- Eustance, E., Lai, Y.-J. S., Shesh, T., & Rittmann, B. E. (2020, 2020/10/01/). Improved CO₂ utilization efficiency using membrane carbonation in outdoor raceways. *Algal Research*, 51, 102070. <https://doi.org/https://doi.org/10.1016/j.algal.2020.102070>
- Evans, J. H. (1958). The Survival of Freshwater Algae During Dry Periods: Part I. An Investigation of the Algae of Five Small Ponds. *The Journal of Ecology*, 46(1), 149. <https://doi.org/10.2307/2256910>

- Evans, K. M., Wortley, A. H., & Mann, D. G. (2007, 2007/07/18/). An Assessment of Potential Diatom “Barcode” Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). *Protist*, 158(3), 349-364. <https://doi.org/https://doi.org/10.1016/j.protis.2007.04.001>
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., & Hassabis, D. (2021). *Protein complex prediction with AlphaFold-Multimer*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2021.10.04.463034>
- Fan, J., Ning, K., Zeng, X., Luo, Y., Wang, D., Hu, J., Li, J., Xu, H., Huang, J., Wan, M., Wang, W., Zhang, D., Shen, G., Run, C., Liao, J., Fang, L., Huang, S., Jing, X., Su, X., Wang, A., Bai, L., Hu, Z. M., Xu, J., & Li, Y. (2015). Genomic Foundation of Starch to Lipid Switch in Oleaginous Chlorella. *Plant Physiology*, pp.01174.02015. <https://doi.org/10.1104/pp.15.01174>
- Fauver, J. R., Petrone, M. E., Hodcroft, E. B., Shioda, K., Ehrlich, H. Y., Watts, A. G., Vogels, C. B. F., Brito, A. F., Alpert, T., Muyombwe, A., Razeq, J., Downing, R., Cheemarla, N. R., Wyllie, A. L., Kalinich, C. C., Ott, I. M., Quick, J., Loman, N. J., Neugebauer, K. M., Greninger, A. L., Jerome, K. R., Roychoudhury, P., Xie, H., Shrestha, L., Huang, M.-L., Pitzer, V. E., Iwasaki, A., Omer, S. B., Khan, K., Bogoch, I. I., Martinello, R. A., Foxman, E. F., Landry, M. L., Neher, R. A., Ko, A. I., & Grubaugh, N. D. (2020). Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell*, 181(5), 990-996.e995. <https://doi.org/10.1016/j.cell.2020.04.021>
- Field, C. B. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374), 237-240. <https://doi.org/10.1126/science.281.5374.237>
- Fozouni, P., Son, S., Diaz de Leon Derby, M., Knott, G. J., Gray, C. N., D'Ambrosio, M. V., Zhao, C., Switz, N. A., Kumar, G. R., Stephens, S. I., Boehm, D., Tsou, C. L., Shu, J., Bhuiya, A., Armstrong, M., Harris, A. R., Chen, P. Y., Osterloh, J. M., Meyer-Franke, A., Joehnk, B., Walcott, K., Sil, A., Langelier, C., Pollard, K. S., Crawford, E. D., Puschnik, A. S., Phelps, M., Kistler, A., DeRisi, J. L., Doudna, J. A., Fletcher, D. A., & Ott, M. (2021, Jan 21). Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell*, 184(2), 323-333 e329. <https://doi.org/10.1016/j.cell.2020.12.001>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>

- Fulnečková, J., Hasíková, T., Fajkus, J., Lukešová, A., Eliáš, M., & Sýkorová, E. (2012). Dynamic Evolution of Telomeric Sequences in the Green Algal Order Chlamydomonadales. *Genome Biology and Evolution*, 4(3), 248-264. <https://doi.org/10.1093/gbe/evs007>
- Gamage, A. M., Tan, K. S., Chan, W. O. Y., Liu, J., Tan, C. W., Ong, Y. K., Thong, M., Andiappan, A. K., Anderson, D. E., Wang, Y., & Wang, L. F. (2020, Dec). Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLoS Pathog*, 16(12), e1009130. <https://doi.org/10.1371/journal.ppat.1009130>
- Gao, X., Sun, B., & Guan, Y. (2019, Feb). Pullulan reduces the non-specific amplification of loop-mediated isothermal amplification (LAMP). *Anal Bioanal Chem*, 411(6), 1211-1218. <https://doi.org/10.1007/s00216-018-1552-2>
- Gardner, R. D., Cooksey, K. E., Mus, F., Macur, R., Moll, K., Eustance, E., Carlson, R. P., Gerlach, R., Fields, M. W., & Peyton, B. M. (2012). Use of sodium bicarbonate to stimulate triacylglycerol accumulation in the chlorophyte *Scenedesmus* sp. and the diatom *Phaeodactylum tricornutum*. *Journal of Applied Phycology*, 24(5), 1311-1320. <https://doi.org/10.1007/s10811-011-9782-0>
- Gardner, R. D., Lohman, E., Gerlach, R., Cooksey, K. E., & Peyton, B. M. (2013). Comparison of CO₂ and bicarbonate as inorganic carbon sources for triacylglycerol and starch accumulation in *Chlamydomonas reinhardtii*. *Biotechnology and Bioengineering*, 110(1), 87-96. <https://doi.org/10.1002/bit.24592>
- Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., Prill, M., Chai, S. J., Kirley, P. D., Alden, N. B., Kawasaki, B., Yousey-Hindes, K., Niccolai, L., Anderson, E. J., Openo, K. P., Weigel, A., Monroe, M. L., Ryan, P., Henderson, J., Kim, S., Como-Sabetti, K., Lynfield, R., Sosin, D., Torres, S., Muse, A., Bennett, N. M., Billing, L., Sutton, M., West, N., Schaffner, W., Talbot, H. K., Aquino, C., George, A., Budd, A., Brammer, L., Langley, G., Hall, A. J., & Fry, A. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR. Morbidity and mortality weekly report*, 69(15), 458-464. <https://doi.org/10.15585/mmwr.mm6915e3>
- Gazzarrini, S., Lejay, L., Gojon, A., Ninnemann, O., Frommer, W. B., & von Wirén, N. (1999). Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into *Arabidopsis* roots. *The Plant Cell*, 11(5), 937-947.
- Gibbens, B. (2019). Measuring Student Motivation in an Introductory Biology Class. *The American Biology Teacher*, 81(1), 20-26, 27. <https://doi.org/10.1525/abt.2019.81.1.20>

- Gimmler, H., Degenhardt, B. (2001). Alkaliphilic and Alkali-Tolerant Algae. (Algal Adaptations to Environmental Stresses). https://doi.org/10.1007/978-3-642-59491-5_10
- Goemann, C. L. C., Wilkinson, R., Henriques, W., Bui, H., Goemann, H. M., Carlson, R. P., Viamajala, S., Gerlach, R., & Wiedenheft, B. (2023, 2023/01/01/). Genome sequence, phylogenetic analysis, and structure-based annotation reveal metabolic potential of *Chlorella* sp. SLA-04. *Algal Research*, 69, 102943. <https://doi.org/https://doi.org/10.1016/j.algal.2022.102943>
- Goemann, H. M., Gay, J. D., Mueller, R. C., Brookshire, E. N. J., Miller, P., Poulter, B., & Peyton, B. M. (2021, Dec 2021 2021-11-08). Aboveground and belowground responses to cyanobacterial biofertilizer supplement in a semi-arid, perennial bioenergy cropping system. *Global Change Biology. Bioenergy*, 13(12), 1908-1923. <https://doi.org/https://doi.org/10.1111/gcbb.12892>
- Goncalves, E. C., Wilkie, A. C., Kirst, M., & Rathinasabapathi, B. (2016). Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnology Journal*, 14(8), 1649-1660. <https://doi.org/10.1111/pbi.12523>
- Gonzalez-Reiche, A. S., Hernandez, M. M., Sullivan, M. J., Ciferri, B., Alshammary, H., Obla, A., Fabre, S., Kleiner, G., Polanco, J., Khan, Z., Albuquerque, B., van de Guchte, A., Dutta, J., Francoeur, N., Melo, B. S., Oussenko, I., Deikus, G., Soto, J., Sridhar, S. H., Wang, Y.-C., Twyman, K., Kasarskis, A., Altman, D. R., Smith, M., Sebra, R., Aberg, J., Krammer, F., García-Sastre, A., Luksza, M., Patel, G., Paniz-Mondolfi, A., Gitman, M., Sordillo, E. M., Simon, V., & van Bakel, H. (2020). Introductions and early spread of SARS-CoV-2 in the New York City area. *Science (New York, N.Y.)*, 369(6501), 297-301. <https://doi.org/10.1126/science.abc1917>
- Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., & Zhang, F. (2018, Apr 27). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, 360(6387), 439-444. <https://doi.org/10.1126/science.aag0179>
- Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A., Myhrvold, C., Bhattacharyya, R. P., Livny, J., Regev, A., Koonin, E. V., Hung, D. T., Sabeti, P. C., Collins, J. J., & Zhang, F. (2017, Apr 28). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, 356(6336), 438-442. <https://doi.org/10.1126/science.aam9321>
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019, Jan 8). An amplicon-based sequencing framework for accurately measuring intrahost virus

- diversity using PrimalSeq and iVar. *Genome Biol*, 20(1), 8. <https://doi.org/10.1186/s13059-018-1618-7>
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome biology*, 20(1), 8-8. <https://doi.org/10.1186/s13059-018-1618-7>
- Guiry, M. D. (2012). HOW MANY SPECIES OF ALGAE ARE THERE? *Journal of Phycology*, 48(5), 1057-1063. <https://doi.org/10.1111/j.1529-8817.2012.01222.x>
- Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., Breillat, P., Carlier, N., Gauzit, R., Morbieu, C., Pene, F., Marin, N., Roche, N., Szwebel, T. A., Merklings, S. H., Treluyer, J. M., Veyer, D., Mouthon, L., Blanc, C., Tharaux, P. L., Rozenberg, F., Fischer, A., Duffy, D., Rieux-Laucat, F., Kerneis, S., & Terrier, B. (2020, Aug 7). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, 369(6504), 718-724. <https://doi.org/10.1126/science.abc6027>
- Hadley, W. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009, Nov 25). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), 945-956. <https://doi.org/10.1016/j.cell.2009.07.040>
- Hall, D. O., & Rao, K. (1999). *Photosynthesis*. Cambridge University Press.
- Hanifzadeh, M., Garcia, E. C., & Viamajala, S. (2018). Production of lipid and carbohydrate from microalgae without compromising biomass productivities: Role of Ca and Mg. *Renewable Energy*, 127, 989-997. <https://doi.org/10.1016/j.renene.2018.05.012>
- Hanifzadeh, M.-M. (2018). *Approaches for Sustainable Production of Microalgae with High Productivity and Flexible Composition* [The University of Toledo].
- Hanschen, E. R., Hovde, B. T., & Starkenburg, S. R. (2020, 2020/10/01/). An evaluation of methodology to determine algal genome completeness. *Algal Research*, 51, 102019. <https://doi.org/https://doi.org/10.1016/j.algal.2020.102019>
- Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., Queen, K., Tao, Y., Paden, C. R., Zhang, J., Li, Y., Uehara, A., Wang, H., Goldsmith, C., Bullock, H. A., Wang, L., Whitaker, B., Lynch, B., Gautam, R., Schindewolf, C., Lokugamage, K. G., Scharton, D., Plante, J. A., Mirchandani, D., Widen, S. G., Narayanan, K., Makino, S.,

- Ksiazek, T. G., Plante, K. S., Weaver, S. C., Lindstrom, S., Tong, S., Menachery, V. D., & Thornburg, N. J. (2020, Jun). Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerging infectious diseases*, 26(6), 1266-1273. <https://doi.org/10.3201/eid2606.200516>
- Hardinge, P., & Murray, J. A. H. (2019, May 14). Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification using Quenched Fluorescent Primers. *Sci Rep*, 9(1), 7400. <https://doi.org/10.1038/s41598-019-43817-z>
- Hart, O. E., & Halden, R. U. (2020). Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *The Science of the total environment*, 730, 138875-138875. <https://doi.org/10.1016/j.scitotenv.2020.138875>
- Hartmann, H., & Trumbore, S. (2016). Understanding the roles of nonstructural carbohydrates in forest trees – from what we can measure to what we want to know. *New Phytologist*, 211(2), 386-403. <https://doi.org/10.1111/nph.13955>
- Higashiyama, T., Maki, S., & Yamada, T. (1995). Molecular organization of *Chlorella vulgaris* chromosome I: presence of telomeric repeats that are conserved in higher plants. *Molecular and General Genetics MGG*, 246(1), 29-36. <https://doi.org/10.1007/bf00290130>
- Hindson, J. (2020). COVID-19: faecal-oral transmission? *Nature reviews. Gastroenterology & hepatology*, 17(5), 259-259. <https://doi.org/10.1038/s41575-020-0295-7>
- Hoek, C., Mann, D., Jahns, H. M., & Jahns, M. (1995). *Algae: an introduction to phycology*. Cambridge university press.
- Hoffman, J., Pate, R. C., Drennen, T., & Quinn, J. C. (2017). Techno-economic assessment of open microalgae production systems. *Algal Research*, 23, 51-57. <https://doi.org/10.1016/j.algal.2017.01.005>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., & Pohlmann, S. (2020, Apr 16). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271-280 e278. <https://doi.org/10.1016/j.cell.2020.02.052>
- Hohmann-Marriott, M. F., & Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annual Review of Plant Biology*, 62(1), 515-548. <https://doi.org/10.1146/annurev-arplant-042110-103811>

- Holland, L. A., Kaelin, E. A., Maqsood, R., Estifanos, B., Wu, L. I., Varsani, A., Halden, R. U., Hogue, B. G., Scotch, M., & Lim, E. S. (2020, Jul 1). An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *J Virol*, 94(14). <https://doi.org/10.1128/JVI.00711-20>
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M. A., Weldon, W. C., Biggs, H. M., Uyeki, T. M., Pillai, S. K., & Washington State -nCo, V. C. I. T. (2020). First Case of 2019 Novel Coronavirus in the United States. *The New England journal of medicine*, 382(10), 929-936. <https://doi.org/10.1056/NEJMoa2001191>
- Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12(1), 491. <https://doi.org/10.1186/1471-2105-12-491>
- Hou, Y. J., Okuda, K., Edwards, C. E., Martinez, D. R., Asakura, T., Dinno, K. H., 3rd, Kato, T., Lee, R. E., Yount, B. L., Mascenik, T. M., Chen, G., Olivier, K. N., Ghio, A., Tse, L. V., Leist, S. R., Gralinski, L. E., Schafer, A., Dang, H., Gilmore, R., Nakano, S., Sun, L., Fulcher, M. L., Livraghi-Butrico, A., Nicely, N. I., Cameron, M., Cameron, C., Kelvin, D. J., de Silva, A., Margolis, D. M., Markmann, A., Bartelt, L., Zumwalt, R., Martinez, F. J., Salvatore, S. P., Borczuk, A., Tata, P. R., Sontake, V., Kimple, A., Jaspers, I., O'Neal, W. K., Randell, S. H., Boucher, R. C., & Baric, R. S. (2020, Jul 23). SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell*, 182(2), 429-446 e414. <https://doi.org/10.1016/j.cell.2020.05.042>
- Hovde, B. T., Hanschen, E. R., Steadman Tyler, C. R., Lo, C.-C., Kunde, Y., Davenport, K., Daligault, H., Msanne, J., Canny, S., Eyun, S.-I., Riethoven, J.-J. M., Polle, J., & Starkenburg, S. R. (2018). Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Research*, 35, 449-461. <https://doi.org/10.1016/j.algal.2018.09.012>
- Hsieh, C.-H., & Wu, W.-T. (2009, 2009/09/01/). Cultivation of microalgae for oil production with a cultivation strategy of urea limitation. *Bioresource Technology*, 100(17), 3921-3926. <https://doi.org/https://doi.org/10.1016/j.biortech.2009.03.019>
- Hynes, M. J., & Murray, S. L. (2010). ATP-Citrate Lyase Is Required for Production of Cytosolic Acetyl Coenzyme A and Development in *Aspergillus nidulans*. *Eukaryotic Cell*, 9(7), 1039-1048. <https://doi.org/10.1128/ec.00080-10>
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009, Nov 15). Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins*, 77(3), 499-508. <https://doi.org/10.1002/prot.22458>

- Illman, A. M., Scragg, A. H., & Shales, S. W. (2000, Nov 1). Increase in *Chlorella* strains calorific values when grown in low nitrogen medium. *Enzyme Microb Technol*, 27(8), 631-635. [https://doi.org/10.1016/s0141-0229\(00\)00266-0](https://doi.org/10.1016/s0141-0229(00)00266-0)
- Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R., & Kuroda, M. (2020). Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PloS one*, 15(9), e0239403-e0239403. <https://doi.org/10.1371/journal.pone.0239403>
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J., & Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology*, 36(4), 338-345. <https://doi.org/10.1038/nbt.4060>
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., & D'Eustachio, P. (2020, Jan 8). The reactome pathway knowledgebase. *Nucleic Acids Res*, 48(D1), D498-D503. <https://doi.org/10.1093/nar/gkz1031>
- Jia, N., Jones, R., Sukenick, G., & Patel, D. J. (2019, Sep 5). Second Messenger cA4 Formation within the Composite Csm1 Palm Pocket of Type III-A CRISPR-Cas Csm Complex and Its Release Path. *Mol Cell*, 75(5), 933-943 e936. <https://doi.org/10.1016/j.molcel.2019.06.013>
- Jia, N., Jones, R., Yang, G., Ouerfelli, O., & Patel, D. J. (2019, Sep 5). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Mol Cell*, 75(5), 944-956 e946. <https://doi.org/10.1016/j.molcel.2019.06.014>
- Jiang, H. W., Zhang, H. N., Meng, Q. F., Xie, J., Li, Y., Chen, H., Zheng, Y. X., Wang, X. N., Qi, H., Zhang, J., Wang, P. H., Han, Z. G., & Tao, S. C. (2020, Sep). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol*, 17(9), 998-1000. <https://doi.org/10.1038/s41423-020-0514-8>
- Joung, J., Ladha, A., Saito, M., Kim, N. G., Woolley, A. E., Segel, M., Barretto, R. P. J., Ranu, A., Macrae, R. K., Faure, G., Ioannidi, E. I., Krajewski, R. N., Bruneau, R., Huang, M. W., Yu, X. G., Li, J. Z., Walker, B. D., Hung, D. T., Greninger, A. L., Jerome, K. R., Gootenberg, J. S., Abudayyeh, O. O., & Zhang, F. (2020, Oct 8). Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. *The New England journal of medicine*, 383(15), 1492-1494. <https://doi.org/10.1056/NEJMc2026172>

- Judith M. Harackiewicz, K. E. B., John M. Tauer, Andrew J. Elliot. (2002). Predicting Success in College: A Longitudinal Study of Achievement Goals and Ability Measures as Predictors of Interest and Performance From Freshman Year Through Graduation. *Journal of Educational Psychology*, 94(3), 562-575.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kacmar, J., Carlson, R., Balogh, S. J., & Srienc, F. (2006). Staining and quantification of poly-3-hydroxybutyrate in *Saccharomyces cerevisiae* and *Cupriavidus necator* cell populations using automated flow cytometry. *Cytometry Part A*, 69A(1), 27-35. <https://doi.org/10.1002/cyto.a.20197>
- Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 33(2), 511-518. <https://doi.org/10.1093/nar/gki198>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Kaylene C. Williams, C. C. W. (2011). Five key ingredients for improving student motivation *Research in Higher Education Journal* 11. <https://doi.org/http://aabri.com/manuscripts/11834.pdf>
- Kazlauskienė, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., & Siksnys, V. (2017, Aug 11). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, 357(6351), 605-609. <https://doi.org/10.1126/science.aao0100>
- Kazlauskienė, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., & Siksnys, V. (2016, Apr 21). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol Cell*, 62(2), 295-306. <https://doi.org/10.1016/j.molcel.2016.03.024>
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, 181(4610), 662-666. <https://doi.org/10.1038/181662a0>

- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020, May 14). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, 181(4), 914-921 e910. <https://doi.org/10.1016/j.cell.2020.04.011>
- Kingsford, C., Schatz, M. C., & Pop, M. (2010). Assembly complexity of prokaryotic genomes using short reads. *BMC Bioinformatics*, 11(1), 21. <https://doi.org/10.1186/1471-2105-11-21>
- Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., & Chlanda, P. (2020, Nov 18). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun*, 11(1), 5885. <https://doi.org/10.1038/s41467-020-19619-7>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540-546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kong, J. N., Hardin, K., Dinkins, M., Wang, G., He, Q., Mujadzic, T., Zhu, G., Bielawski, J., Spassieva, S., & Bieberich, E. (2015). Regulation of Chlamydomonas flagella and ependymal cell motile cilia by ceramide-mediated translocation of GSK3. *Molecular Biology of the Cell*, 26(24), 4451-4465. <https://doi.org/10.1091/mbc.e15-06-0371>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Sheffield, C.-G. G., McDanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Whelan, S. P., LaBranche, C. C., Saphire, E. O., & Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e819. <https://doi.org/10.1016/j.cell.2020.06.043>
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., & Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*, 30(7), 693-700. <https://doi.org/10.1038/nbt.2280>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research*, 27(5), 722-736. <https://doi.org/10.1101/gr.215087.116>

- Kotenko, S. V., & Durbin, J. E. (2017, May 5). Contribution of type III interferons to antiviral immunity: location, location, location. *J Biol Chem*, 292(18), 7295-7303.
<https://doi.org/10.1074/jbc.R117.777102>
- Kothari, R., Pandey, A., Ahmad, S., Kumar, A., Pathak, V. V., & Tyagi, V. V. (2017). Microalgal cultivation for value-added products: a critical enviro-economical assessment. *3 Biotech*, 7(4). <https://doi.org/10.1007/s13205-017-0812-8>
- Kou, Y., Liu, M., Sun, P., Dong, Z., & Liu, J. (2020, 2020/09/01/). High light boosts salinity stress-induced biosynthesis of astaxanthin and lipids in the green alga *Chromochloris zofingiensis*. *Algal Research*, 50, 101976.
<https://doi.org/https://doi.org/10.1016/j.algal.2020.101976>
- Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, 16(5), 263-276.
<https://doi.org/10.1038/nrmicro.2018.9>
- La Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L., & Suffredini, E. (2020). First detection of SARS-CoV-2 in untreated wastewaters in Italy. *The Science of the total environment*, 736, 139652-139652.
<https://doi.org/10.1016/j.scitotenv.2020.139652>
- La Scola, B., Le Bideau, M., Andreani, J., Hoang, V. T., Grimaldier, C., Colson, P., Gautret, P., & Raoult, D. (2020, Jun). Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *Eur J Clin Microbiol Infect Dis*, 39(6), 1059-1061. <https://doi.org/10.1007/s10096-020-03913-9>
- Lara-Gil, J. A., Alvarez, M. M., & Pacheco, A. (2014, Feb). Toxicity of flue gas components from cement plants in microalgae CO₂ mitigation systems. *Journal of Applied Phycology*, 26(1), 357-368. <https://doi.org/10.1007/s10811-013-0136-y>
- Larremore, D. B., Wilder, B., Lester, E., Shehata, S., Burke, J. M., Hay, J. A., Tambe, M., Mina, M. J., & Parker, R. (2021, Jan). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv*, 7(1). <https://doi.org/10.1126/sciadv.abd5393>
- Latsos, C., Wassenaar, E., Moerdijk, T., Coleman, B., Robbens, J., Van Roy, S., Bastiaens, L., Van Houcke, J., & Timmermans, K. R. (2022). Effect of pH on *Rhodomonas salina* growth, biochemical composition, and taste, produced in semi-large scale under sunlight conditions. *Journal of Applied Phycology*, 34(3), 1215-1226.
<https://doi.org/10.1007/s10811-022-02730-1>
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019

- (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*, 172(9), 577-582. <https://doi.org/10.7326/M20-0504>
- Lazer, D. S., M.; Perlis, R.H.; Ognyanova, K.; Baum, M.A.; Quintana, A.; Druckman, J.; Volpe, J.D.; Chwe, H.; Simonson, M. (2020). *State of the Nation: A 50-State COVID-19 Survey: Report #8: Failing the Test: Waiting Times for COVID Diagnostic Tests Across the U.S.* <https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/en/grc-740424>
- Lei, X., Dong, X., Ma, R., Wang, W., Xiao, X., Tian, Z., Wang, C., Wang, Y., Li, L., Ren, L., Guo, F., Zhao, Z., Zhou, Z., Xiang, Z., & Wang, J. (2020, Jul 30). Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun*, 11(1), 3810. <https://doi.org/10.1038/s41467-020-17665-9>
- Lemieux, J. E., Siddle, K. J., Shaw, B. M., Loreth, C., Schaffner, S. F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C. H., Krasilnikova, L. A., DeRuff, K. C., Rudy, M., Bauer, M. R., Lagerborg, K. A., Normandin, E., Chapman, S. B., Reilly, S. K., Anahtar, M. N., Lin, A. E., Carter, A., Myhrvold, C., Kembell, M. E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J. B., Branda, J. A., Hooper, D., Gaeta, J. M., Baggett, T. P., O'Connell, J., Gnirke, A., Lieberman, T. D., Philippakis, A., Burns, M., Brown, C. M., Luban, J., Ryan, E. T., Turbett, S. E., LaRocque, R. C., Hanage, W. P., Gallagher, G. R., Madoff, L. C., Smole, S., Pierce, V. M., Rosenberg, E., Sabeti, P. C., Park, D. J., & MacInnis, B. L. (2021, Feb 5). Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*, 371(6529). <https://doi.org/10.1126/science.abe3261>
- Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14), 2103-2110. <https://doi.org/10.1093/bioinformatics/btw152>
- Li, H. (2018a). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H. (2018b). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., & Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*, 382(13), 1199-1207. <https://doi.org/10.1056/NEJMoa2001316>

- Li, Y., Lou, Y., Mu, T., Ke, A., Ran, Z., Xu, J., Chen, J., Zhou, C., Yan, X., Xu, Q., & Tan, Y. (2017). Sphingolipids in marine microalgae: Development and application of a mass spectrometric method for global structural characterization of ceramides and glycosphingolipids in three major phyla. *Analytica Chimica Acta*, 986, 82-94. <https://doi.org/10.1016/j.aca.2017.07.039>
- Liang, M.-H., & Jiang, J.-G. (2016, 2016/12/01/). Characterization and nitrogen deficiency response of ATP-citrate lyase from unicellular alga *Dunaliella tertiolecta*. *Algal Research*, 20, 77-86. <https://doi.org/https://doi.org/10.1016/j.algal.2016.09.025>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lieberman, D. A., & Remedios, R. (2007). Do undergraduates' motives for studying change as they progress through their degrees? [Article]. *British Journal of Educational Psychology*, 77(2), 379-395. <https://doi.org/10.1348/000709906X157772>
- Lin, K. Y., Hong, H.-Y., & Chai, C. S. (2014, 2014/02/01/). Development and validation of the knowledge-building environment scale. *Learning and Individual Differences*, 30, 124-132. <https://doi.org/https://doi.org/10.1016/j.lindif.2013.10.018>
- Liu, T. Y., Iavarone, A. T., & Doudna, J. A. (2017). RNA and DNA Targeting by a Reconstituted *Thermus thermophilus* Type III-A CRISPR-Cas System. *PloS one*, 12(1), e0170552. <https://doi.org/10.1371/journal.pone.0170552>
- Liu, T. Y., Liu, J. J., Aditham, A. J., Nogales, E., & Doudna, J. A. (2019, Jul 5). Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. *Nat Commun*, 10(1), 3001. <https://doi.org/10.1038/s41467-019-10780-2>
- Lodder, W., & de Roda Husman, A. M. (2020). SARS-CoV-2 in wastewater: potential health risk, but also data source. *The lancet. Gastroenterology & hepatology*, 5(6), 533-534. [https://doi.org/10.1016/S2468-1253\(20\)30087-X](https://doi.org/10.1016/S2468-1253(20)30087-X)
- Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., & Usadel, B. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment*, 37(5), 1250-1258. <https://doi.org/10.1111/pce.12231>
- Loman, N. J. R., W.; Rambaut, A. (2020). NCoV-2019 Novel Coronavirus Bioinformatics Protocol. <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>

- Lopez-Alvarez, D., Parra, B., & Cuellar, W. J. (2020). Genome Sequence of SARS-CoV-2 Isolate Cali-01, from Colombia, Obtained Using Oxford Nanopore MinION Sequencing. *Microbiology resource announcements*, 9(26), e00573-00520. <https://doi.org/10.1128/MRA.00573-20>
- Losada, M. (1976). REDUCING POWER AND THE REGULATION OF PHOTOSYNTHESIS. In A. Kornberg, B. L. Horecker, L. Cornudella, & J. Oro (Eds.), *Reflections on Biochemistry* (pp. 73-84). Pergamon. <https://doi.org/https://doi.org/10.1016/B978-0-08-021010-0.50014-4>
- Loudet, A., & Burgess, K. (2007, Nov). BODIPY dyes and their derivatives: syntheses and spectroscopic properties. *Chem Rev*, 107(11), 4891-4932. <https://doi.org/10.1021/cr078381n>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- Loveday, E. K., Hain, K. S., Kochetkova, I., Hedges, J. F., Robison, A., Snyder, D. T., Brumfield, S. K., Young, M. J., Jutila, M. A., Chang, C. B., & Taylor, M. P. (2021, Mar 26). Effect of Inactivation Methods on SARS-CoV-2 Virion Protein and Structure. *Viruses*, 13(4). <https://doi.org/10.3390/v13040562>
- Lucas, C., Wong, P., Klein, J., Castro, T. B. R., Silva, J., Sundaram, M., Ellingson, M. K., Mao, T., Oh, J. E., Israelow, B., Takahashi, T., Tokuyama, M., Lu, P., Venkataraman, A., Park, A., Mohanty, S., Wang, H., Wyllie, A. L., Vogels, C. B. F., Earnest, R., Lapidus, S., Ott, I. M., Moore, A. J., Muenker, M. C., Fournier, J. B., Campbell, M., Odio, C. D., Casanovas-Massana, A., Yale, I. T., Herbst, R., Shaw, A. C., Medzhitov, R., Schulz, W. L., Grubaugh, N. D., Dela Cruz, C., Farhadian, S., Ko, A. I., Omer, S. B., & Iwasaki, A. (2020, Aug). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*, 584(7821), 463-469. <https://doi.org/10.1038/s41586-020-2588-y>
- Lumsden, L. S. (1994). student motivation. *Research Roundup*, 10(3).
- Luo, M. B., & Liu, F. (2011, 2011/12/01/). Salinity-induced oxidative stress and regulation of antioxidant defense system in the marine macroalga *Ulva prolifera*. *Journal of Experimental Marine Biology and Ecology*, 409(1), 223-228. <https://doi.org/https://doi.org/10.1016/j.jembe.2011.08.023>
- M. H. Langholtz, B. J. S., and L. M. Eaton (2016). 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy. *U.S. Department of Energy*.

- Ma, Z., Tian, M., Tan, Y., Cui, G., Feng, Y., Cui, Q., & Song, X. (2017). Response mechanism of the docosahexaenoic acid producer *Aurantiochytrium* under cold stress. *Algal Research*, 25, 191-199.
- Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksny, V., Venclovas, C., & Koonin, E. V. (2020, Sep 18). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antiviral defense. *Nucleic Acids Res*, 48(16), 8828-8847. <https://doi.org/10.1093/nar/gkaa635>
- Mallapaty, S. (2020, 2020/04/03). How sewage could reveal true scale of coronavirus outbreak. *Nature*, 580(7802), 176-177. <https://doi.org/10.1038/d41586-020-00973-x>
- Manni, M. e., Matthew, Seppey, M., Felipe, & Evgeny. (2021, 2021-06-22). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *arXiv pre-print server*. <https://doi.org/None>
arxiv:2106.11799
- Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S., & Huang, C. (2020, Jul). Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Res*, 179, 104811. <https://doi.org/10.1016/j.antiviral.2020.104811>
- Markland, D., & Hardy, L. (1997, Mar). On the factorial and construct validity of the Intrinsic Motivation Inventory: conceptual and operational concerns. *Res Q Exerc Sport*, 68(1), 20-32. <https://doi.org/10.1080/02701367.1997.10608863>
- McBride, R. C., Lopez, S., Meenach, C., Burnett, M., Lee, P. A., Nohilly, F., & Behnke, C. (2014). Contamination Management in Low Cost Open Algae Ponds for Biofuels Production. *10*(3), 221-227. <https://doi.org/10.1089/ind.2013.0036>
- McCree, K. J. (1981). Photosynthetically Active Radiation. In O. L. Lange, P. S. Nobel, C. B. Osmond, & H. Ziegler (Eds.), *Physiological Plant Ecology I: Responses to the Physical Environment* (pp. 41-55). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68090-8_3
- McKie-Krisberg, Z. M., Laurens, L. M. L., Huang, A., & Polle, J. E. W. (2018, 2018/04/01/). Comparative energetics of carbon storage molecules in green algae. *Algal Research*, 31, 326-333. <https://doi.org/https://doi.org/10.1016/j.algal.2018.01.018>
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A., & O'Garra, A. (2015, Feb). Type I interferons in infectious disease. *Nat Rev Immunol*, 15(2), 87-103. <https://doi.org/10.1038/nri3787>
- Meagher, R. J., Priye, A., Light, Y. K., Huang, C., & Wang, E. (2018, Apr 16). Impact of primer dimers and self-amplifying hairpins on reverse transcription loop-mediated isothermal

- amplification detection of viral RNA. *Analyst*, 143(8), 1924-1933.
<https://doi.org/10.1039/c7an01897e>
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020, 2020/05/20). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7), 511-516.
<https://doi.org/10.1021/acs.estlett.0c00357>
- Mehrotra, S., & Goyal, V. (2014). Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics, proteomics & bioinformatics*, 12(4), 164-171. <https://doi.org/10.1016/j.gpb.2014.07.003>
- Mehta, P., Sahil, K., Sarao, L. K., Jangra, M. S., & Bhardwaj, S. K. (2023). Algal Biofuels: Clean Energy to Combat the Climate Change. In (pp. 187-210). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6810-5_7
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Marechal-Drouard, L., Marshall, W. F., Qu, L. H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C. L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I. T., Pazour, G., Purton, S., Ral, J. P., Riano-Pachon, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C. J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W. C. A., Otiillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S., & Grossman, A. R. (2007). The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science*, 318(5848), 245-250. <https://doi.org/10.1126/science.1143609>
- Michael Huesemann, J. B. (2017). Microalgae Biofuels Production on CO2 from Air.
- Miorin, L., Kehrer, T., Sanchez-Aparicio, M. T., Zhang, K., Cohen, P., Patel, R. S., Cupic, A., Makio, T., Mei, M., Moreno, E., Danziger, O., White, K. M., Rathnasinghe, R., Uccellini, M., Gao, S., Aydillo, T., Mena, I., Yin, X., Martin-Sancho, L., Krogan, N. J., Chanda, S.

- K., Schotsaert, M., Wozniak, R. W., Ren, Y., Rosenberg, B. R., Fontoura, B. M. A., & Garcia-Sastre, A. (2020, Nov 10). SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc Natl Acad Sci U S A*, 117(45), 28344-28354. <https://doi.org/10.1073/pnas.2016650117>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007, Jul). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35(Web Server issue), W182-185. <https://doi.org/10.1093/nar/gkm321>
- Mustapa, M., Sallehudin, N., Mohamed, M., Mohammad Noor, N., & Raus, R. (2016, 01/01). Decontamination of *Chlorella* sp. Culture using antibiotics and antifungal cocktail treatment. *ARNP Journal of Engineering and Applied Sciences*, 11.
- Muth, D., Corman, V. M., Roth, H., Binger, T., Dijkman, R., Gottula, L. T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtaric, D., Toplak, I., Ameneiros, R. S., Pfeifer, A., Thiel, V., Drexler, J. F., Muller, M. A., & Drosten, C. (2018, Oct 11). Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*, 8(1), 15177. <https://doi.org/10.1038/s41598-018-33487-8>
- Nalley, J. O., O'Donnell, D. R., & Litchman, E. (2018). Temperature effects on growth rates and fatty acid content in freshwater algae and cyanobacteria. *Algal Research*, 35, 500-507.
- Nasef, M., Muffly, M. C., Beckman, A. B., Rowe, S. J., Walker, F. C., Hatoum-Aslan, A., & Dunkle, J. A. (2019, Aug). Regulation of cyclic oligoadenylate synthesis by the *Staphylococcus epidermidis* Cas10-Csm complex. *RNA*, 25(8), 948-962. <https://doi.org/10.1261/rna.070417.119>
- Nayir, F. (2017, 10/20). The Relationship between Student Motivation and Class Engagement Levels. *Eurasian Journal of Educational Research*, 17, 59-78. <https://doi.org/10.14689/ejer.2017.71.4>
- Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S., & Fremont, D. H. (2005, Jan). Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure*, 13(1), 75-85. <https://doi.org/10.1016/j.str.2004.10.010>
- Nemudryi, A., Nemudraia, A., Wiegand, T., Nichols, J., Snyder, D. T., Hedges, J. F., Cicha, C., Lee, H., Vanderwood, K. K., Bimczok, D., Jutila, M. A., & Wiedenheft, B. (2021, Jun 1). SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a

- that limits immune suppression. *Cell Rep*, 35(9), 109197.
<https://doi.org/10.1016/j.celrep.2021.109197>
- Newby, D. T., Mathews, T. J., Pate, R. C., Huesemann, M. H., Lane, T. W., Wahlen, B. D., Mandal, S., Engler, R. K., Feris, K. P., & Shurin, J. B. (2016). Assessing the potential of polyculture to accelerate algal biofuel production. *Algal Research*, 19, 264-277.
<https://doi.org/10.1016/j.algal.2016.09.004>
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.
<https://doi.org/10.1093/molbev/msu300>
- Niewoehner, O., Garcia-Doval, C., Rostol, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017, Aug 31). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, 548(7669), 543-548.
<https://doi.org/10.1038/nature23467>
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000, Jun 15). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*, 28(12), E63. <https://doi.org/10.1093/nar/28.12.e63>
- Olson, G. (1997). Motivation, Motivation, Motivation-Secondary School Educators. Retrieved from sysiwyg://934/<http://7-12educators.about...-12educators/library/weekly/aa071897.htm>.
- Ooi, K. H., Liu, M. M., Tay, J. W. D., Teo, S. Y., Kaewsapsak, P., Jin, S., Lee, C. K., Hou, J., Maurer-Stroh, S., Lin, W., Yan, B., Yan, G., Gao, Y. G., & Tan, M. H. (2021, Mar 19). An engineered CRISPR-Cas12a variant and DNA-RNA hybrid guides enable robust and rapid COVID-19 testing. *Nat Commun*, 12(1), 1739. <https://doi.org/10.1038/s41467-021-21996-6>
- [Record #745 is using a reference type undefined in this output style.]
- Osada, N., Kohara, A., Yamaji, T., Hirayama, N., Kasai, F., Sekizuka, T., Kuroda, M., & Hanada, K. (2014, Dec). The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res*, 21(6), 673-683. <https://doi.org/10.1093/dnares/dsu029>
- Pagès H, A. P., Gentleman R, DebRoy S. (2019). Biostrings: Efficient manipulation of biological strings. *R package version 2.52.0*.
- Paltiel, A. D., Zheng, A., & Walensky, R. P. (2020, Jul 1). Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United

- States. *JAMA Netw Open*, 3(7), e2016818.
<https://doi.org/10.1001/jamanetworkopen.2020.16818>
- Paradis, E., Claude, J., & Strimmer, K. (2004, Jan 22). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2), 289-290.
<https://doi.org/10.1093/bioinformatics/btg412>
- Pati, A., Padhi, S., Suvankar, S., & Panda, A. K. (2021, Jan 4). Minor Allele of Interferon-Induced Transmembrane Protein 3 Polymorphism (rs12252) Is Covered Against Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Mortality: A Worldwide Epidemiological Investigation. *The Journal of infectious diseases*, 223(1), 175-178.
<https://doi.org/10.1093/infdis/jiaa630>
- Paul, M. J., & Foyer, C. H. (2001). Sink regulation of photosynthesis. *Journal of Experimental Botany*, 52(360), 1383-1400. <https://doi.org/10.1093/jexbot/52.360.1383>
- Pawlita-Posmyk, M., Wzorek, M., & Płaczek, M. (2018). The influence of temperature on algal biomass growth for biogas production. MATEC Web of Conferences,
- Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2018). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, 35(13), 2193-2198.
<https://doi.org/10.1093/bioinformatics/bty841>
- Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W., & Omer, S. B. (2020, Oct). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature biotechnology*, 38(10), 1164-1167.
<https://doi.org/10.1038/s41587-020-0684-z>
- Perera, F., & Nadeau, K. (2022). Climate Change, Fossil-Fuel Pollution, and Children's Health. *New England Journal of Medicine*, 386(24), 2303-2314.
<https://doi.org/10.1056/nejmra2117706>
- Pérez-Pérez, M. E., Lemaire, S. D., & Crespo, J. L. (2012). Reactive oxygen species and autophagy in plants and algae. *Plant Physiology*, 160(1), 156-164.
- Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, 9, 304. <https://doi.org/10.12688/f1000research.23297.2>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290-295. <https://doi.org/10.1038/nbt.3122>

- Praveenkumar, R., Shameera, K., Mahalakshmi, G., Akbarsha, M. A., & Thajuddin, N. (2012, 2012/02/01/). Influence of nutrient deprivations on lipid accumulation in a dominant indigenous microalga *Chlorella* sp., BUM11008: Evaluation for biodiesel production. *Biomass and Bioenergy*, 37, 60-66.
<https://doi.org/https://doi.org/10.1016/j.biombioe.2011.12.035>
- Prieto, C. V. G., Ramos, F. D., Estrada, V., Villar, M. A., & Diaz, M. S. (2017). Optimization of an integrated algae-based biorefinery for the production of biodiesel, astaxanthin and PHB. *Energy*, 139, 1159-1172.
- Quick, J. (2020). *nCoV-2019 sequencing protocol v1* ZappyLab, Inc.
<https://doi.org/10.17504/protocols.io.bbmuik6w>
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Jr., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., & Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols*, 12(6), 1261-1276.
<https://doi.org/10.1038/nprot.2017.066>
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J.-P., Cabeza-Cabrero, M., Camino-Sanchez, A., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G. G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallash, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Trina, R., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N. F., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., O'Shea, M. K., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Woelfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keita, S., Rambaut, A., Formenty, P., Gunther, S., & Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228-232. <https://doi.org/10.1038/nature16996>
- Quinn, J. C., & Davis, R. (2015). The potentials and challenges of algae based biofuels: A review of the techno-economic, life cycle, and resource assessment modeling. *Bioresource Technology*, 184, 444-452. <https://doi.org/10.1016/j.biortech.2014.10.075>

- Rafa, N., Ahmed, S., Badruddin, I., Rahman, M. M., & Kamangar, S. (2021, 08/23). Strategies to Produce Cost-Effective Third-Generation Biofuel From Microalgae. *Frontiers in Energy Research*. <https://doi.org/10.3389/fenrg.2021.749968>
- Rahman, F., Hassan, M., Rosli, R., Almously, I., Hanano, A., & Murphy, D. J. (2018). Evolutionary and genomic analysis of the caleosin/peroxygenase (CLO/PXG) gene/protein families in the Viridiplantae. *PLoS ONE*, 13(5), e0196669. <https://doi.org/10.1371/journal.pone.0196669>
- [Record #404 is using a reference type undefined in this output style.]
- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020, Nov). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*, 5(11), 1403-1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014, Jul). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, 42(Web Server issue), W187-191. <https://doi.org/10.1093/nar/gku365>
- Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water research*, 181, 115942-115942. <https://doi.org/10.1016/j.watres.2020.115942>
- Raven, J. A., & Geider, R. J. (1988). Temperature and algal growth. *New Phytologist*, 110(4), 441-461.
- Reitan, K. I., Rainuzzo, J. R., & Olsen, Y. (1994). EFFECT OF NUTRIENT LIMITATION ON FATTY ACID AND LIPID CONTENT OF MARINE MICROALGAE1. *Journal of Phycology*, 30(6), 972-979. <https://doi.org/10.1111/j.0022-3646.1994.00972.x>
- Ren, Q., de Roo, G., Ruth, K., Witholt, B., Zinn, M., & Thöny-Meyer, L. (2009, 2009/04/13). Simultaneous Accumulation and Degradation of Polyhydroxyalkanoates: Futile Cycle or Clever Regulation? *Biomacromolecules*, 10(4), 916-922. <https://doi.org/10.1021/bm801431c>
- Renaud, S. M., Thinh, L.-V., Lambrinidis, G., & Parry, D. L. (2002). Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures. *Aquaculture*, 211(1-4), 195-214.
- Resemann, H. C., Herrfurth, C., Feussner, K., Hornung, E., Ostendorf, A. K., Gömann, J., Mittag, J., Van Gessel, N., Vries, J. D., Ludwig-Müller, J., Markham, J., Reski, R., &

- Feussner, I. (2021). Convergence of sphingolipid desaturation across over 500 million years of plant evolution. *Nature Plants*, 7(2), 219-232. <https://doi.org/10.1038/s41477-020-00844-3>
- Rice, P., Longden, I., & Bleasby, A. (2000, Jun). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16(6), 276-277. [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2)
- Rimoldi, S. G., Stefani, F., Gigantiello, A., Polesello, S., Comandatore, F., Mileto, D., Maresca, M., Longobardi, C., Mancon, A., Romeri, F., Pagani, C., Cappelli, F., Roscioli, C., Moja, L., Gismondo, M. R., & Salerno, F. (2020). Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. *The Science of the total environment*, 744, 140911-140911. <https://doi.org/10.1016/j.scitotenv.2020.140911>
- Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., Burgstaller-Muehlbacher, S., De Jesus, P. D., Teriete, P., Hull, M. V., Chang, M. W., Chan, J. F., Cao, J., Poon, V. K., Herbert, K. M., Cheng, K., Nguyen, T. H., Rubanov, A., Pu, Y., Nguyen, C., Choi, A., Rathnasinghe, R., Schotsaert, M., Miorin, L., Dejosez, M., Zwaka, T. P., Sit, K. Y., Martinez-Sobrido, L., Liu, W. C., White, K. M., Chapman, M. E., Lendy, E. K., Glynne, R. J., Albrecht, R., Rupp, E., Mesecar, A. D., Johnson, J. R., Benner, C., Sun, R., Schultz, P. G., Su, A. I., Garcia-Sastre, A., Chatterjee, A. K., Yuen, K. Y., & Chanda, S. K. (2020, Oct). Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature*, 586(7827), 113-119. <https://doi.org/10.1038/s41586-020-2577-1>
- Rolando, J. C., Jue, E., Barlow, J. T., & Ismagilov, R. F. (2020, Apr 17). Real-time kinetics and high-resolution melt curves in single-molecule digital LAMP to differentiate and study specific and non-specific amplification. *Nucleic Acids Res*, 48(7), e42. <https://doi.org/10.1093/nar/gkaa099>
- Rosenthal, S. H., Kagan, R. M., Gerasimova, A., Anderson, B., Grover, D., Hua, M., Liu, Y., Owen, R., & Lacbawan, F. (2020). Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. *BioRxiv*, 2020.2012.2010.418855. <https://doi.org/10.1101/2020.12.10.418855>
- Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S., & White, M. F. (2018, Jul 2). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife*, 7. <https://doi.org/10.7554/eLife.36734>
- Rowe, J. M., Dunigan, D. D., Blanc, G., Gurnon, J. R., Xia, Y., & Van Etten, J. L. (2013). Evaluation of higher plant virus resistance genes in the green alga, *Chlorella variabilis* NC64A, during the early phase of infection with *Paramecium bursaria chlorella virus-1*. *Virology*, 442(2), 101-113. <https://doi.org/10.1016/j.virol.2013.04.018>

- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019, Jul 2). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*, 47(W1), W5-w10. <https://doi.org/10.1093/nar/gkz342>
- Ruperti, F., Papadopoulos, N., Musser, J., & Arendt, D. (2022). *Beyond sequence similarity: cross-phyla protein annotation by structural prediction and alignment*. Cold Spring Harbor Laboratory. <https://dx.doi.org/10.1101/2022.07.05.498892>
- Ryan Davis, J. M., Christopher Kinchin, Nicholas Grundl, Eric Tan, David Hymbird. (2016). Process Design and Economics for the Production of Algal Biomass: Algal Biomass Production in Open Pond Systems and Processing Through Dewatering for Downstream Conversion.
- Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015, May 21). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, 161(5), 1164-1174. <https://doi.org/10.1016/j.cell.2015.04.027>
- Schoggins, J. W. (2019, Sep 29). Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol*, 6(1), 567-584. <https://doi.org/10.1146/annurev-virology-092818-015756>
- Sengupta, P., Holowka, D., & Baird, B. (2007, 2007/05/15/). Fluorescence Resonance Energy Transfer between Lipid Probes Detects Nanoscopic Heterogeneity in the Plasma Membrane of Live Cells. *Biophysical Journal*, 92(10), 3564-3574. <https://doi.org/https://doi.org/10.1529/biophysj.106.094730>
- Sharma, J., Kumar, S. S., Bishnoi, N. R., & Pugazhendhi, A. (2018). Enhancement of lipid production from algal biomass through various growth parameters. *Journal of Molecular Liquids*, 269, 712-720. <https://doi.org/10.1016/j.molliq.2018.08.103>
- Sharma, K. K., Schuhmann, H., & Schenk, P. M. (2012). High Lipid Induction in Microalgae for Biodiesel Production. *Energies*, 5(5), 1532-1553. <https://doi.org/10.3390/en5051532>
- Shell, D., Husman, J., Turner, J., Cliffl, D., Nath, I., & Sweany, N. (2005, 07/01). The Impact of Computer Supported Collaborative Learning Communities on High School Students' Knowledge Building, Strategic Learning, and Perceptions of the Classroom. *Journal of Educational Computing Research - J EDUC COMPUT RES*, 33, 327-349. <https://doi.org/10.2190/787L-BCBQ-20FN-FW6C>
- Shell, D. F., & Husman, J. (2001). The Multivariate Dimensionality of Personal Control and Future Time Perspective Beliefs in Achievement and Self-Regulation. *Contemporary Educational Psychology*, 26(4), 481-506. <https://doi.org/10.1006/ceps.2000.1073>
- Shell, D. F., & Husman, J. (2008, May 2008)

- 2020-07-02). Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon. *Journal of Educational Psychology*, 100(2), 443-459. <https://doi.org/http://dx.doi.org/10.1037/0022-0663.100.2.443>
- Shell, D. F., & Soh, L.-K. (2013). Profiles of Motivated Self-Regulation in College Computer Science Courses: Differences in Major versus Required Non-Major Courses. *Journal of Science Education and Technology*, 22(6), 899-913. <https://doi.org/10.1007/s10956-013-9437-9>
- Shi, G., Kenney, A. D., Kudryashova, E., Zani, A., Zhang, L., Lai, K. K., Hall-Stoodley, L., Robinson, R. T., Kudryashov, D. S., Compton, A. A., & Yount, J. S. (2021, Feb 1). Opposing activities of IFITM proteins in SARS-CoV-2 infection. *EMBO J*, 40(3), e106501. <https://doi.org/10.15252/emboj.2020106501>
- Shi, T. Q., Wang, L. R., Zhang, Z. X., Sun, X. M., & Huang, H. (2020). Stresses as First-Line Tools for Enhancing Lipid and Carotenoid Production in Microalgae. *Front Bioeng Biotechnol*, 8, 610. <https://doi.org/10.3389/fbioe.2020.00610>
- Silverman, J. D., Hupert, N., & Washburne, A. D. (2020). Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Science translational medicine*, 12(554), eabc1126. <https://doi.org/10.1126/scitranslmed.abc1126>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Singh, N. K., Paz, E., Kutsher, Y., Reuveni, M., & Lers, A. (2020). Tomato T2 ribonuclease LE is involved in the response to pathogens. *Molecular Plant Pathology*, 21(7), 895-906. <https://doi.org/10.1111/mpp.12928>
- Singh, S., & Singh, P. (2015). Effect of temperature and light on the growth of algae species: A review. *Renewable and sustainable energy reviews*, 50, 431-444.
- [Record #564 is using a reference type undefined in this output style.]
- Sofos, N., Feng, M., Stella, S., Pape, T., Fuglsang, A., Lin, J., Huang, Q., Li, Y., She, Q., & Montoya, G. (2020, Sep 3). Structures of the Cmr-beta Complex Reveal the Regulation of the Immunity Mechanism of Type III-B CRISPR-Cas. *Mol Cell*, 79(5), 741-757 e747. <https://doi.org/10.1016/j.molcel.2020.07.008>
- Song, X., Liu, B.-F., Kong, F., Ren, N.-Q., & Ren, H.-Y. (2022, 2022/08/01/). Overview on stress-induced strategies for enhanced microalgae lipid production: Application,

- mechanisms and challenges. *Resources, Conservation and Recycling*, 183, 106355. <https://doi.org/https://doi.org/10.1016/j.resconrec.2022.106355>
- Staals, R. H., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., Koehorst, J. J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P. J., Urlaub, H., Heck, A. J., Nogales, E., Doudna, J. A., Shinkai, A., & van der Oost, J. (2014, Nov 20). RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell*, 56(4), 518-530. <https://doi.org/10.1016/j.molcel.2014.10.005>
- Stefanelli, P., Faggioni, G., Lo Presti, A., Fiore, S., Marchi, A., Benedetti, E., Fabiani, C., Anselmo, A., Ciammaruconi, A., Fortunato, A., De Santis, R., Fillo, S., Capobianchi, M. R., Gismondo, M. R., Ciervo, A., Rezza, G., Castrucci, M. R., Lista, F., & On Behalf Of Iss Covid-Study, G. (2020). Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 25(13), 2000305. <https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000305>
- Stephens, E., Ross, I. L., Mussgnug, J. H., Wagner, L. D., Borowitzka, M. A., Posten, C., Kruse, O., & Hankamer, B. (2010). Future prospects of microalgal biofuel production systems. *Trends in Plant Science*, 15(10), 554-564. <https://doi.org/10.1016/j.tplants.2010.06.003>
- Stoltz, M., & Klingstrom, J. (2010, Sep). Alpha/beta interferon (IFN-alpha/beta)-independent induction of IFN-lambda1 (interleukin-29) in response to Hantaan virus infection. *J Virol*, 84(18), 9140-9148. <https://doi.org/10.1128/JVI.00717-10>
- Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, 12(1). <https://doi.org/10.1186/s13100-020-00230-y>
- Su, Y. C. F., Anderson, D. E., Young, B. E., Linster, M., Zhu, F., Jayakumar, J., Zhuang, Y., Kalimuddin, S., Low, J. G. H., Tan, C. W., Chia, W. N., Mak, T. M., Octavia, S., Chavatte, J. M., Lee, R. T. C., Pada, S., Tan, S. Y., Sun, L., Yan, G. Z., Maurer-Stroh, S., Mendenhall, I. H., Leo, Y. S., Lye, D. C., Wang, L. F., & Smith, G. J. D. (2020, Jul 21). Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *MBio*, 11(4). <https://doi.org/10.1128/mBio.01610-20>
- Supuran, C. T. (2008). Carbonic anhydrases-an overview. *Current pharmaceutical design*, 14(7), 603-614.

- Sushchik, N., Kalacheva, G., Zhila, N., Gladyshev, M., & Volova, T. (2003). A Temperature Dependence of the Intra-and Extracellular Fatty-Acid Composition of Green Algae and Cyanobacterium. *Russian journal of plant physiology*, 50(3).
- Sutton, J. M., Millwood, J. D., Case McCormack, A., & Fierst, J. L. (2021). Optimizing experimental design for genome sequencing and assembly with Oxford Nanopore Technologies. *Gigabyte*, 2021, 1-26. <https://doi.org/10.46471/gigabyte.27>
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server), W609-W612. <https://doi.org/10.1093/nar/gkl315>
- Svinicki, M. D., & McKeachie, W. J. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (13 ed.). Cengage Learning.
- Tamulaitis, G., Kazlauskienė, M., Manakova, E., Venclovas, C., Nwokeoji, A. O., Dickman, M. J., Horvath, P., & Siksnys, V. (2014, Nov 20). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol Cell*, 56(4), 506-517. <https://doi.org/10.1016/j.molcel.2014.09.027>
- Tan, K. W. M., & Lee, Y. K. (2016). The dilemma for lipid productivity in green microalgae: importance of substrate provision in improving oil yield without sacrificing growth. *Biotechnology for Biofuels*, 9(1). <https://doi.org/10.1186/s13068-016-0671-2>
- Tang, A., Tong, Z.-D., Wang, H.-L., Dai, Y.-X., Li, K.-F., Liu, J.-N., Wu, W.-J., Yuan, C., Yu, M.-L., Li, P., & Yan, J.-B. (2020). Detection of Novel Coronavirus by RT-PCR in Stool Specimen from Asymptomatic Child, China. *Emerging infectious diseases*, 26(6), 1337-1339. <https://doi.org/10.3201/eid2606.200301>
- Tarailo-Graovac, M., & Chen, N. (2009, Mar). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*, Chapter 4, Unit 4.10. <https://doi.org/10.1002/0471250953.bi0410s25>
- Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A., & Ng, L. F. P. (2020, Jun). The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*, 20(6), 363-374. <https://doi.org/10.1038/s41577-020-0311-8>
- Teresa Lopes da Silva, A. R. (2015). scale-up problems for the large scale production of algae. https://doi.org/10.1007/978-3-319-22813-6_6
- Thi Nhu Thao, T., Labroussaa, F., Ebert, N., V'Kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., Kratzel, A., Gultom, M., Schmied, K., Laloli, L., Husser, L., Wider, M., Pfaender, S., Hirt, D., Cippa, V., Crespo-Pomar, S., Schroder, S., Muth, D., Niemeyer, D., Corman, V. M., Muller, M. A., Drosten, C., Dijkman, R., Jores, J., &

- Thiel, V. (2020, Jun). Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, 582(7813), 561-565. <https://doi.org/10.1038/s41586-020-2294-9>
- Tomita, N., Mori, Y., Kanda, H., & Notomi, T. (2008). Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat Protoc*, 3(5), 877-882. <https://doi.org/10.1038/nprot.2008.57>
- Trujillo, G., & Tanner, K. D. (2014). Considering the Role of Affect in Learning: Monitoring Students' Self-Efficacy, Sense of Belonging, and Science Identity. *CBE—Life Sciences Education*, 13(1), 6-15. <https://doi.org/10.1187/cbe.13-12-0241>
- Tsai, D. D.-W., Chen, P. H., Chou, C. M.-j., Hsu, C.-F., & Ramaraj, R. (2015, 2015/11/01/). Carbon sequestration by alga ecosystems. *Ecological Engineering*, 84, 386-389. <https://doi.org/https://doi.org/10.1016/j.ecoleng.2015.09.024>
- Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajek, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020, Sep 4). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRxiv*. <https://doi.org/10.1101/2020.09.04.283077>
- Vadlamani, A., Pendyala, B., Viamajala, S., & Varanasi, S. (2019). High Productivity Cultivation of Microalgae without Concentrated CO₂ Input. *ACS Sustainable Chemistry & Engineering*, 7(2), 1933-1943. <https://doi.org/10.1021/acssuschemeng.8b04094>
- Vadlamani, A., Viamajala, S., Pendyala, B., & Varanasi, S. (2017). Cultivation of Microalgae at Extreme Alkaline pH Conditions: A Novel Approach for Biofuel Production. *ACS Sustainable Chemistry & Engineering*, 5(8), 7284-7294. <https://doi.org/10.1021/acssuschemeng.7b01534>
- van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen, C. J., Pang, J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 83, 104351-104351. <https://doi.org/10.1016/j.meegid.2020.104351>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2022). Foldseek: fast and accurate protein structure search. *biorxiv*, 2022.2002.2007.479398. <https://doi.org/10.1101/2022.02.07.479398>
- Varshney, P., Mikulic, P., Vonshak, A., Beardall, J., & Wangikar, P. P. (2015, May). Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour Technol*, 184, 363-372. <https://doi.org/10.1016/j.biortech.2014.11.040>

- Vasanth Singan, A. L., James Han and Erika Lindquist. *RNA-Seq Gene Expression Analysis at the Joint Genome Institute*. https://jgi.doe.gov/news-publications/scientific-posters/poster_singan/
- Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D. P., consortium, T. C.-G. U., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., & Ferguson, N. M. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.2012.2030.20249034. <https://doi.org/10.1101/2020.12.30.20249034>
- Walker, J. C. G. (1980). The Oxygen Cycle. In (pp. 87-104). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-24940-6_5
- Wallerstedt, C., & Pramling, N. (2012). Learning to play in a goal-directed practice. *Early Years*, 32(1), 5-15. <https://doi.org/10.1080/09575146.2011.593028>
- Wang, C., Shan, L., Qu, S., Xue, M., Wang, K., Fu, F., Wang, L., Wang, Z., Feng, L., Xu, W., & Liu, P. (2020). The Coronavirus PEDV Evades Type III Interferon Response Through the miR-30c-5p/SOCS1 Axis. *Front Microbiol*, 11, 1180. <https://doi.org/10.3389/fmicb.2020.01180>
- Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*, 323(18), 1843-1844. <https://doi.org/10.1001/jama.2020.3786>
- Wang, Y., He, B., Sun, Z., & Chen, Y.-F. (2016). Chemically enhanced lipid production from microalgae under low sub-optimal temperature. *Algal Research*, 16, 20-27.
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Weyrich, L. S., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., Handsley-Davis, M., Adler, C. J., Breen, J., & Cooper, A. (2019). Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources*, 19(4), 982-996. <https://doi.org/10.1111/1755-0998.13011>
- Widjaja, A., Chien, C.-C., & Ju, Y.-H. (2009, 2009/01/01/). Study of increasing lipid production from fresh water microalgae *Chlorella vulgaris*. *Journal of the Taiwan Institute of Chemical Engineers*, 40(1), 13-20. <https://doi.org/https://doi.org/10.1016/j.jtice.2008.07.007>

- Wilson, W. H., Van Etten, J. L., & Allen, M. J. (2009). The Phycodnaviridae: The Story of How Tiny Giants Rule the World. In (pp. 1-42). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-540-68618-7_1
- Wolfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Muller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brunink, S., Schneider, J., Ehmann, R., Zwirgmaier, K., Drosten, C., & Wendtner, C. (2020, May). Virological assessment of hospitalized patients with COVID-2019. *Nature*, 581(7809), 465-469. <https://doi.org/10.1038/s41586-020-2196-x>
- Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A. D., Erickson, T. B., Chai, P. R., Thompson, J., & Alm, E. J. (2020, Jul 21). SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *MSystems*, 5(4). <https://doi.org/10.1128/mSystems.00614-20>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269.
<https://doi.org/10.1038/s41586-020-2008-3>
- Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G., & Huang, X. (2020). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The lancet. Gastroenterology & hepatology*, 5(5), 434-435. [https://doi.org/10.1016/S2468-1253\(20\)30083-2](https://doi.org/10.1016/S2468-1253(20)30083-2)
- Wurtzer, S., Marechal, V., Mouchel, J. M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J. L., & Moulin, L. (2020, Dec). Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveill*, 25(50). <https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000776>
- Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y., Wang, H., Menachery, V. D., Rajsbaum, R., & Shi, P. Y. (2020, Oct 6). Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep*, 33(1), 108234. <https://doi.org/10.1016/j.celrep.2020.108234>
- Xiao, F., Sun, J., Xu, Y., Li, F., Huang, X., Li, H., Zhao, J., Huang, J., & Zhao, J. (2020). Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. *Emerging infectious diseases*, 26(8), 1920-1922. <https://doi.org/10.3201/eid2608.200681>

- Xie, X., Muruato, A., Lokugamage, K. G., Narayanan, K., Zhang, X., Zou, J., Liu, J., Schindewolf, C., Bopp, N. E., Aguilar, P. V., Plante, K. S., Weaver, S. C., Makino, S., LeDuc, J. W., Menachery, V. D., & Shi, P. Y. (2020, May 13). An Infectious cDNA Clone of SARS-CoV-2. *Cell Host Microbe*, 27(5), 841-848 e843. <https://doi.org/10.1016/j.chom.2020.04.004>
- Xin, C., Addy, M. M., Zhao, J., Cheng, Y., Cheng, S., Mu, D., Liu, Y., Ding, R., Chen, P., & Ruan, R. (2016). Comprehensive techno-economic analysis of wastewater-based algal biofuel production: A case study. *Bioresource Technology*, 211, 584-593. <https://doi.org/10.1016/j.biortech.2016.03.102>
- Xin, L., Hong-ying, H., Ke, G., & Ying-xue, S. (2010, 2010/07/01/). Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga *Scenedesmus* sp. *Bioresource Technology*, 101(14), 5494-5500. <https://doi.org/https://doi.org/10.1016/j.biortech.2010.02.016>
- Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., Zhang, H., Liu, H., Xia, H., Tang, J., Zhang, K., & Gong, S. (2020). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature medicine*, 26(4), 502-505. <https://doi.org/10.1038/s41591-020-0817-4>
- Yan, D., Dai, J., & Wu, Q. (2013). Characterization of an ammonium transporter in the oleaginous alga *Chlorella protothecoides*. *Applied Microbiology and Biotechnology*, 97(2), 919-928. <https://doi.org/10.1007/s00253-012-4534-x>
- Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019, Jan 4). Functionally diverse type V CRISPR-Cas systems. *Science*, 363(6422), 88-91. <https://doi.org/10.1126/science.aav7271>
- Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental science & technology*, 50(10), 5077-5085. <https://doi.org/10.1021/acs.est.6b00876>
- Yeh, K. L., & Chang, J. S. (2011, Nov). Nitrogen starvation strategies and photobioreactor design for enhancing lipid content and lipid production of a newly isolated microalga *Chlorella vulgaris* ESP-31: implications for biofuels. *Biotechnol J*, 6(11), 1358-1366. <https://doi.org/10.1002/biot.201000433>
- Yergaliyev, T. M., Nurbekova, Z., Mukiyanova, G., Akbassova, A., Sutula, M., Zhangazin, S., Bari, A., Tleukulova, Z., Shamekova, M., Masalimov, Z. K., & Omarov, R. T. (2016, 2016/12/01/). The involvement of ROS producing aldehyde oxidase in plant response to Tombusvirus infection. *Plant Physiology and Biochemistry*, 109, 36-44. <https://doi.org/https://doi.org/10.1016/j.plaphy.2016.09.001>

- Yodsuwan, N., Sawayama, S., & Sirisansaneeyakul, S. (2017). Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom *Phaeodactylum tricornutum*. *Agriculture and Natural Resources*, 51(3), 190-197.
<https://doi.org/10.1016/j.anres.2017.02.004>
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004, May). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol*, 21(5), 809-818.
<https://doi.org/10.1093/molbev/msh075>
- You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X., & Wang, Y. (2019, Jan 10). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell*, 176(1-2), 239-253 e216.
<https://doi.org/10.1016/j.cell.2018.10.052>
- Young, B. E., Fong, S. W., Chan, Y. H., Mak, T. M., Ang, L. W., Anderson, D. E., Lee, C. Y., Amrun, S. N., Lee, B., Goh, Y. S., Su, Y. C. F., Wei, W. E., Kalimuddin, S., Chai, L. Y. A., Pada, S., Tan, S. Y., Sun, L., Parthasarathy, P., Chen, Y. Y. C., Barkham, T., Lin, R. T. P., Maurer-Stroh, S., Leo, Y. S., Wang, L. F., Renia, L., Lee, V. J., Smith, G. J. D., Lye, D. C., & Ng, L. F. P. (2020, Aug 29). Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet*, 396(10251), 603-611. [https://doi.org/10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8)
- Yousuf, A. (2020). Microalgae Cultivation for Biofuels Production Preface. *Microalgae Cultivation for Biofuels Production*, Xv-Xvi. https://doi.org/Book_Doi 10.1016/C2018-0-01358-4
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2016, 2016/09/22). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28-36.
<https://doi.org/10.1111/2041-210x.12628>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28-36.
<https://doi.org/10.1111/2041-210x.12628>
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012, May). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 16(5), 284-287.
<https://doi.org/10.1089/omi.2011.0118>
- Yuan, L., Loqué, D., Kojima, S., Rauch, S., Ishiyama, K., Inoue, E., Takahashi, H., & von Wirén, N. (2007, Aug). The organization of high-affinity ammonium uptake in

- Arabidopsis roots depends on the spatial arrangement and biochemical properties of AMT1-type transporters. *Plant Cell*, 19(8), 2636-2652.
<https://doi.org/10.1105/tpc.107.052134>
- Yuvraj, & Padmanabhan, P. (2017). Technical insight on the requirements for CO₂-saturated growth of microalgae in photobioreactors. *3 Biotech*, 7(2).
<https://doi.org/10.1007/s13205-017-0778-6>
- Zang, R., Case, J. B., Yutuc, E., Ma, X., Shen, S., Gomez Castro, M. F., Liu, Z., Zeng, Q., Zhao, H., Son, J., Rothlauf, P. W., Kreutzberger, A. J. B., Hou, G., Zhang, H., Bose, S., Wang, X., Vahey, M. D., Mani, K., Griffiths, W. J., Kirchhausen, T., Fremont, D. H., Guo, H., Diwan, A., Wang, Y., Diamond, M. S., Whelan, S. P. J., & Ding, S. (2020, Dec 15). Cholesterol 25-hydroxylase suppresses SARS-CoV-2 replication by blocking membrane fusion. *Proc Natl Acad Sci U S A*, 117(50), 32105-32113.
<https://doi.org/10.1073/pnas.2012197117>
- Zehender, G., Lai, A., Bergna, A., Meroni, L., Riva, A., Balotta, C., Tarkowski, M., Gabrieli, A., Bernacchia, D., Rusconi, S., Rizzardini, G., Antinori, S., & Galli, M. (2020). Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. *Journal of medical virology*, 92(9), 1637-1640. <https://doi.org/10.1002/jmv.25794>
- Zhang, B., Zhou, X., Qiu, Y., Song, Y., Feng, F., Feng, J., Song, Q., Jia, Q., & Wang, J. (2020). Clinical characteristics of 82 cases of death from COVID-19. *PloS one*, 15(7), e0235458.
<https://doi.org/10.1371/journal.pone.0235458>
- Zhang, J., Graham, S., Tello, A., Liu, H., & White, M. F. (2016, Feb 29). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res*, 44(4), 1789-1799. <https://doi.org/10.1093/nar/gkw020>
- Zhang, L., Pei, H., Chen, S., Jiang, L., Hou, Q., Yang, Z., & Yu, Z. (2018, 2018/02/01/). Salinity-induced cellular cross-talk in carbon partitioning reveals starch-to-lipid biosynthesis switching in low-starch freshwater algae. *Bioresource Technology*, 250, 449-456. <https://doi.org/https://doi.org/10.1016/j.biortech.2017.11.067>
- Zhang, Y., Qin, L., Zhao, Y., Zhang, P., Xu, B., Li, K., Liang, L., Zhang, C., Dai, Y., Feng, Y., Sun, J., Hu, Z., Xiang, H., Knight, J. C., Dong, T., & Jin, R. (2020, Jun 16). Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. *The Journal of infectious diseases*, 222(1), 34-37.
<https://doi.org/10.1093/infdis/jiaa224>
- Zhang, Y., Ren, G., Buss, J., Barry, A. J., Patton, G. C., & Tanner, N. A. (2020, Sep). Enhancing colorimetric loop-mediated isothermal amplification speed and sensitivity with guanidine chloride. *Biotechniques*, 69(3), 178-185. <https://doi.org/10.2144/btn-2020-0078>

- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020, Mar 28). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*, 395(10229), 1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- Zhou, Z., Huang, C., Zhou, Z., Huang, Z., Su, L., Kang, S., Chen, X., Chen, Q., He, S., Rong, X., Xiao, F., Chen, J., & Chen, S. (2021, Mar 19). Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14(+) monocytes. *iScience*, 24(3), 102187. <https://doi.org/10.1016/j.isci.2021.102187>
- Zou, Y., Mason, M. G., & Botella, J. R. (2020). Evaluation and improvement of isothermal amplification methods for point-of-need plant disease diagnostics. *PloS one*, 15(6), e0235216. <https://doi.org/10.1371/journal.pone.0235216>