GENOMIC ANALYSIS OF MICROBIAL ADAPTATIONS AND INTERACTIONS WITHIN PLANT HOSTS

by

SHUMENG ZHANG

(Under the Direction of Jeffrey L. Bennetzen)

ABSTRACT

Sarracenia rosea is a carnivorous plant that uses pitchers (formed by modified leaves) to trap and digest small animals as a source of nutrients. The pitcher fluids are a microecosystem that contains diverse invertebrates and microorganisms. Six bacterial isolates from pitcher fluids of *S. rosea* were identified as *Hafnia paralvei*, marking the first discovery of *H. paralvei* from plants. Pan and core genome comparisons between pitcher strains and 16 non-pitcher strains of *H. paralvei* identified 276 genes that were in all pitcher strains but not in any other strains. Environment-impacted pathways including flagellar assembly and the type VI secretion system were found to be consistently different between pitcher and non-pitcher strains. A total of 148 genes were found to be present in some but not all pitcher strains. Diverse bacterial activities closely associated with the pitcher environment, including nitrogen metabolism and anaerobic respiration, were discovered in this intraspecific variable gene set. The study of *H. paralvei* strains from pitcher plants provides insights for the exploration of adaption of microbial communities within the pitcher fluids.

Puccinia emaculata is a biotrophic and parasitic fungus that causes rust disease on the leaves of the bioenergy crop switchgrass via urediniospores, resulting in decreased biomass production. *P. emaculata* infects switchgrass leaves by penetration. Numerous different microbes were found to be associated with urediniospores, which is not a commonly found phenomenon. The microbial communities of switchgrass inoculated and uninoculated with urediniospores were explored by plating of homogenized leaf tissues and metagenomic analysis. Microbes inside urediniospores-inoculated switchgrass leaves were more abundant than uninoculated leaves, confirming that multiple microbial species were transferred into host plant leaves during the infection of *P. emaculata*. Bacteria, none of which has been found to be able to penetrate the epidermis of plant leaves, from different genera were found to be transferred into switchgrass leaves during this rust infection. Interestingly, bacteria abundant on the outside of the spores were not transferred into host plant tissues.

INDEX WORDS: pitcher plant, *Hafnia paralvei*, plant-associated microbes, comparative genome analysis, fungal infection, genome assembly, *Puccinia emaculata*, switchgrass

GENOMIC ANALYSIS OF
MICROBIAL ADAPTATIONS AND INTERACTIONS WITHIN PLANT HOSTS

by

Shumeng Zhang

MS, Huazhong Agricultural University, China, 2015

BS, Huazhong Agricultural University, China, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

GENOMIC ANALYSIS OF

MICROBIAL ADAPTATIONS AND INTERACTIONS WITHIN PLANT HOSTS

by

SHUMENG ZHANG

Major Professor: Jeffrey L. Bennetzen

Committee: Jan Mrazek
Katrien M. Devos
Liang Liu
Shavannor Smith

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

DEDICATION

I dedicate this work to my parents, Guangdong Zhang and Lijuan Su, in thanks for their encouragement and financial support.

ACKNOWLEDGEMENTS

I want to thank my advisor, Dr. Jeffrey Bennetzen, for his invaluable guidance with experimental design, data analysis, results interpretation and writing in the past five years. I was not able to complete my dissertation without his help.

I want to thank my committee members Dr. Katrien Devos, Dr. Shavannor Smith, Dr. Jan Mrazek and Dr. Liang Liu who provided invaluable suggestion on my dissertation in and out of committee meetings.

I want to thank the previous postdoc fellow Dr. Xuewen Wang for his help with genome sequencing and suggestions about genomic analysis. I want to thank the research scientist Dr. Srinivasa Chaluvadi who patiently lead me through wet lab work. I want to thank Lorna Nissen in Dr. Shavannor Smith's lab for her guidance to handle fungal spores.

I want to thank my loving lab mates, alumni and undergraduates, Dr. Cristina Vives Cobo, Dr. Daniel Frailey, Dr. Robert Arthur, Dr. Guomin Han, Dr. Domitille Chalopin, Dr. Hao Wang, Dr. Minkyu Park, Hongye Zhou, Rocky Giarratano, Aye Htun, Josue Fernandez, Ahmad Nisar, Dustin Bailey and many more, for academic help, encouragement , support and fun time together.

I want to thank my friends here in Athens who are funny, helpful and considerable. I want to thank my parents and my family who provide unlimited love and support. I want to thank my boyfriend Daniel Busby and his family who provide tremendous care and help. I want to thank myself for not giving up in the journey of Ph.D..

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

SECTION 1.1

MICROECOSYSTEMS IN *SARRACENIA* PITCHER FLUIDS

Carnivorous pitcher plants

Pitcher plants are a group of carnivorous plants that inhabit sunny, aquatic, nutrient-poor places such as bogs (NACZI et al.,1999). Pitcher plant environments contain very limited nutrients and are acidic, so that most plants are not able to survive (Joye, 1989). Pitchers are special organs that have deep cavity features formed by modified leaves, working as pitfalls to prey on small organisms ranging from bacteria to small vertebrates (Adlassnig et al., 2010). The purpose of pitchers in trapping prey is to absorb their soluble nutrients to facilitate the survival of pitcher plants in sterile habitats (Bradshaw & Creelman, 1984). The uptake of inorganic nutrients such as nitrogen and phosphorus are of particular importance, while the uptake of organic nutrients is less important because pitcher plants are photosynthetically productive (Adamec, 1997; Schulze et al.,1997; Schulze et al., 2001). More than 100 species of pitcher plants have been classified and/or collected, including all

members of the *Sarraceniaceae* and *Nepenthaceae* families, plus some members of the *Bromeliaceae* and *Cephalotaceae* families (Juniper et al., 1989).

Pitcher structures differ among species of pitcher plants. A pitcher can be as colorful as a flower, which may be helpful in capturing prey. The outer surface on the inside of a pitcher is rough and hairy. Typically, a pitcher is comprised of five zones (Figure 1.1). Zone 1 consists of an open or unopen hood and pitcher peristome. The hood protects the pitcher fluids from being washed away or overly diluted by rainwater (Bauer et al., 2009). The pitcher peristome is wet and slippery, and thus causes prey to fall into the pitcher, while inward-pointing hairs inside the pitcher prevent the prey from escaping. There are attraction glands on the hood that produce sugar nectar or volatiles that attract prey, which can then fall into the pitcher hollow and drown (Juniper, 1989; Joel, 1986; Jürgens et al., 2009). Attraction glands may exist on the peristome as well. Zone 2 is called the prey-retention zone, and has an inner surface that is waxy and slippery. There are hairs that point downward in Zone 2. The features of the prey-retention zone keep trapped prey from climbing or flying out. Zone 3 is the glandular zone. The glands secrete digestive enzymes in at least some species, including species in the genus *Sarracenia* (Robinson, 1908; Glenn & Bodri, 2012; Gallie & Chang, 1997). Zone 4 is filled with pitcher fluids. These fluids contain abundant microbes and zone 4 is the location where the prey are digested and their nutrients mobilized. This is also the location where the mobilized nutrients are taken up by the plant via foliar absorption (Plummer & Kethley, 1964). Zone 5 has unknown function(s). The volumes of pitchers range from 0.2 mL to 1.5 L, depending on both pitcher genotypes and environmental factors (Adlassnig et al., 2010).

Figure 1.1 Longitudinal section of a *Sarracenia purpurea* pitcher (Glenn & Bodri, 2012). The numbers from 1 to 5 represent the zones that are described in the text.

Pitcher fluids and digestive enzymes

At the bottom of pitchers, pitcher fluids digest prey to provide nutrients for the pitcher plant. Fluid solvents can be secreted by pitcher plants or collected from rainfall. Both *Nepenthes* and *Cephalotus* pitcher plants secrete fluids that have low pH and high saturation with oxygen. Fluids are composed differently in *Sarracenia* pitcher plants. *Sarracenia* pitchers senesce in the fall on these perennial plants, but new pitchers form

each year in the spring. When they first develop in the spring, *Sarracenia* pitchers are closed and empty. Later, the pitchers' hoods open to collect water from rainfall (Arber, 1941; Hepburn et al., 1927; Peterson et al., 2008; Plummer & Jackson, 1963). Microbial communities are assembled from various inocula (wind, rain splash, bog flooding, prey arrival, etc.). The community is sustained by continual re-inoculations by these same processes and by the availability of nutrients from the digesting prey. The prey-free, initial fluids of *Sarracenia* pitcher plants are slightly acidic and have a low ion content, in agreement with their primary origin from rainfall and/or bog water (Adlassnig et al., 2010).

The digestive enzymes solubilized in pitcher fluids can be produced by either the host plant and/or the pitcher microbes (Luciano et al., 2017). The fluids secreted by *Nepenthes* and *Cephalotus* pitcher plants contain abundant digestive enzymes, but the pitcher fluids contain a limited diversity of microbes. The opposite is true for *Sarracenia* pitchers. Although *Sarracenia purpurea* was reported to secrete a few droplets of digestive fluids, they are believed to be insufficient for the digestion of prey (Batalin, 1880; Hepburn et al., 1927; Smith, 1893).

Nucleases, RNase, proteases and phosphatases are found in the pitcher fluid of *S. purpurea* sterilized with ampicillin, carbenicillin and cefotaxime, but enzymes were secreted only in the first weeks after the pitcher opened (Gallie and Chang, 1997). However, *S. purpurea* does not secrete chitinases (which degrade chitin, a major component of prey exoskeletons (Hepburn and Jones, 1927; Gallie and Chang, 1997)). In *Sarracenia* pitcher plants, microbial communities appear to be very important in the digestion of prey. Due to the activity of hydrolytic enzymes generated by microbes, prey are able to be broken down

into their basic constituents, and then some are mineralized into soluble limiting nutrients such as N, P and K sources (Butler & Ellison, 2007; Butler et al., 2008). Chitinases are also generated and secreted by pitcher microbes (Bay, 1893; Lindquist, 1975).

Traditionally, Indonesian and Malagasy peoples use *Nepenthes* pitcher fluids to wash eyes and treat headaches, asthma and burns (Schoenwetter et al., 2006). Pharmaceutically, several compounds in *Nepenthes* have been used as digestive enzymes to remove antigens from red blood cells (Adlassnig et al., 2010). Proteins with distinctive activities against microbes are found in *Nepenthes* as well (Hatano and Hamada, 2008)

Microorganism communities and interactions in a pitcher

A pitcher provides the environment for a microecosystem that consists of diverse inquilines (arthropods, protozoa, insect larvae, etc.) and microbes (archaea, fungi, phage, protists and bacteria). The complexity of the microecosystem builds up after pitchers open and fill with rainwater, but the initial microecosystem present before the hood opens has not been characterized. In at least one studied environment, the diversity of the microecosystem increases until four weeks after the pitcher opens (Zander, 2016). Both compositions of pitcher fluids and types of prey affect the composition of inquiline communities (Miller et al., 1994; Trzcinski et al., 2005). Protists and bacteria are more abundant in pitcher fluids that contain more organic nutrients (Higashi et al., 1993). The number of published inquiline species ranges from 3 in *S. minor* to 165 in *S. purpurea* (Adlassnig et al., 2010). The majority of microorganisms in pitchers are aerobic and heterotrophic (Juniper et al., 1989; Williams, 2006; Butler et al., 2008). Oxygen dissolved from the surface of pitcher fluids is believed to be insufficient because the surface area of

pitcher fluids is small (Bradshaw & Creelman, 1984). Pitcher plants also diffuse oxygen to pitcher fluids through cuticular gaps on pitcher walls (Bradshaw & Creelman, 1984; Juniper et al., 1989). Besides, pitcher plants have epidermal chloroplasts that consume the carbon dioxide generated from microorganismal catabolism and release oxygen to pitcher fluids (Adlassnig et al., 2010). The exchange of gas by pitcher epidermal chloroplasts is not only good for the stabilization of microorganism communities but also benefits the plant. However, too much prey can lead to an oxygen shortage inside pitchers, and thus anaerobic microorganisms accumulate (Bradshaw & Creelman, 1984). However, the degradation of prey under anaerobic conditions is not well understood.

The establishment of a pitcher microecosystem requires five prerequisites (Adlassnig et al., 2010). First, the microbes must share a habitat with a pitcher plant because many microbes are limited to certain areas. Second, organisms should be able to find a pitcher. Organisms can enter the pitcher through rainfall, flying spores or pitcher traps. Third, organisms must be able to survive in the pitcher fluids. Fourth, there should be enough energy resources in the pitcher. Prey serves as the main energy resources of a pitcher. Fifth, the survival of any resilient pitcher organism should be allowed by predators, competitors, the life span of a pitcher and abiotic stresses (e.g., heat, drought, cold) on the pitcher fluids (Adlassnig et al., 2010).

The interactions between a pitcher plant and its pitcher community can be classified into three types. First, some organisms are harmful to plants. For example, some caterpillars bite the inner walls of pitchers in both *Nepenthes* and *Sarracenia* (Juniper et al., 1989; Atwater et al., 2006). Second, some microorganisms are beneficial to the plant,

which is of the most interest to me. For example, microbes help the plant with the digestion of prey. This type of interaction is common in *Sarracenia* pitcher plants because they are not able to secrete sufficient digestive enzymes (Peroutka et al.,2008). Bacterial and fungal microorganisms are found to secrete proteases and chitinases that degrade prey in *S. purpurea* (Bay, 1893; Lindquist, 1975). The third type of relationship is no interaction between the plant and the pitcher community. This usually happens to pitcher visitors including geckoes and mantises (Adlassnig et al., 2010). These visitors can steal prey from the pitcher.

Studies on the microecosystems of carnivorous pitcher plants

S. purpurea pitchers have served as a model microecosystem for studies of microorganism communities and food web dynamics for more than a century. It has been reported that 75% of the prey of *S. purpurea* are ants that feed on plant nectars (Newell and Natase, 1998). The formic acid brought by ants may be the reason that pitcher fluids are so acidic (Plummer and Jackson, 1963). *S. purpurea* were found to trap newts (*Notophthalmus viridescens*) as well (Butler et al., 2005). The trapping efficiency (number of captures divided by number of visits for both newts and ants) of *S. purpurea* is lower than 1% (Newell and Nastase, 1998). The larvae of the pitcher plant mosquito, *Wyeomyia smithii*, is usually the top predator in the microecosystem. *W. smithii* depends on *S. purpurea* to complete its larval development (Bradshaw, 1983). There is no strong correlation between the volume of fluid and inquiline diversity (Gotelli and Ellison, 2006), but the presence of the top predator affects inquiline composition (Zander, 2016). In *S. purpurea*, microbes play a more important role in nitrogen uptake than large arthropods

(Karagatzides et al., 2009). Among these bacteria, *Azotobacteriaceae* were found to assimilate atmospheric N_2 (Prankevicius and Cameron, 1991). The amount of nitrogen fixed was estimated to exceed the amount of nitrogen that was needed by the plant even without trapping prey (Ellison et al., 2003).

SECTION 1.2

INTERACTIONS BETWEEN PATHOGENS AND HOST PLANTS

Plant pathogen lifestyle

Plants are exposed to many types of microbial pathogens such as fungi, bacteria, and viruses that cause specific plant diseases. However, different types of pathogens infect plant cells in different ways. Plant pathogens can be categorized by their virulence mechanisms. So far, three types of pathogen lifestyles are defined: biotrophic, necrotrophic and hemibiotrophic (Oliver and Ipcho, 2004). Biotrophic pathogens infect living cells to proliferate. They deliver effectors to suppress the innate plant immune system, but these effectors are often recognized as signals in the plant to initiate hypersensitive resistance (Jones and Dangl, 2006). Effectors also manipulate host metabolism to provide sugar, amino acids and other nutrients for the pathogen (Fatima & Senthil-Kumar, 2015). Rust fungi, powdery mildew fungi and oomycete downy mildews are classified as obligate biotrophs that depend on living host cells for growth and reproduction (Voegele and Mendgen, 2003). Necrotrophic pathogens enter the plant through dead tissues, but also derive most of their nutrients from living plant tissues (Laluk and Mengiste, 2010). Some

necrotrophic pathogens that produce Host-Selective Toxins (HST) are called narrow-host-range necrotrophs. The other necrotrophic pathogens that produce Cell Wall Degrading Enzymes (CWDE) are broad-host-range necrotrophs. Hemibiotrophic pathogens pursue biotrophic lifestyles early in the infection and switch to necrotrophic lifestyles at later stages (Oliver and Ipcho, 2004; Glazebrook, 2005).

Biotrophic fungal infections in plants

Many pathogens infect plant cells through pre-penetration and penetration of the plant epidermis (Figure 1.2). For example, biotrophic rust fungi penetrate plant leaves with a specialized structure called an appressorium (Szabo and Bushnell, 2001). After spores are dispersed in the environment and deposited on a surface, including the leaf surfaces of susceptible plants, the spores geminate in the presence of sufficient moisture to produce germ tubes. Appressoria develop when a germ tube contacts a stomate. The formation of appressoria relies on the stomatal shape on the leaf surface (Allen *et al*., 1991; Staples and Hoch, 1987). The appressorium forms a penetration peg to penetrate the plant surface through the stomate in between guard cells. Then, other infection structures (substomatal vesicles, infection hyphae, haustorial mother cells and haustoria) are consecutively developed inside the leaf (Hoch *et al*., 1987). The haustorium is a special organ that invades the host cell to uptake nutrients. The haustorial cells penetrate the cell walls of additional plant cells underlying the stomate, but they do not penetrate the plant cell membranes. A haustorium is comprised of several structures, including (from outside to inside) extrahaustorial membrane, extrahaustorial matrix, haustorial wall, and haustorial plasma membrane (Figure 1.3). Substrate transfer from plant cytoplasm to haustorial cytoplasm

requires substrates to pass through all of these layers. The extrahaustorial membrane is a thickened derivative of the plant plasma membrane that keeps the haustorium from reaching the plant cytoplasm. The extrahaustorial matrix is a gel-like layer enriched in carbohydrates (Szabo and Bushnell, 2001).

The haustorium is formed *in planta*, and it cannot be cultured completely in vitro. This makes it hard to analyze haustoria at a molecular level (Deising *et al*., 1991; Voegele and Mendgen, 2003). The role that haustoria play in infection is controversial. Haustoria are assumed to function in nutrient uptake, suppression of host defense responses, and manipulation of host metabolism (Voegele and Mendgen, 2003). A hexose transporter (HXT1) was found to locate at haustorial membranes but not intercellular hyphal membranes, suggesting that hexose uptake occurs in the haustorium (Voegele *et al*., 2001).

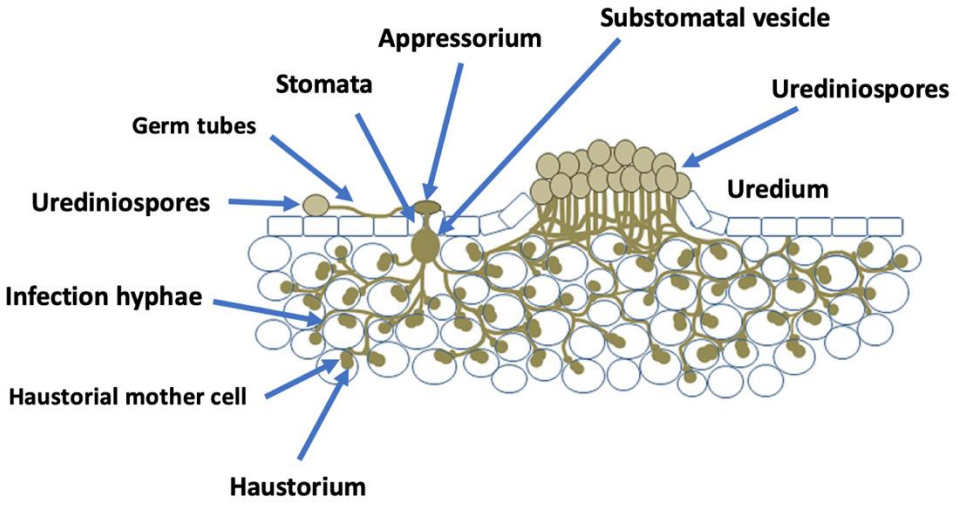


Figure 1.2 Infection process of urediniospores (Edited from Bettgenhaeuser et al. 2014)

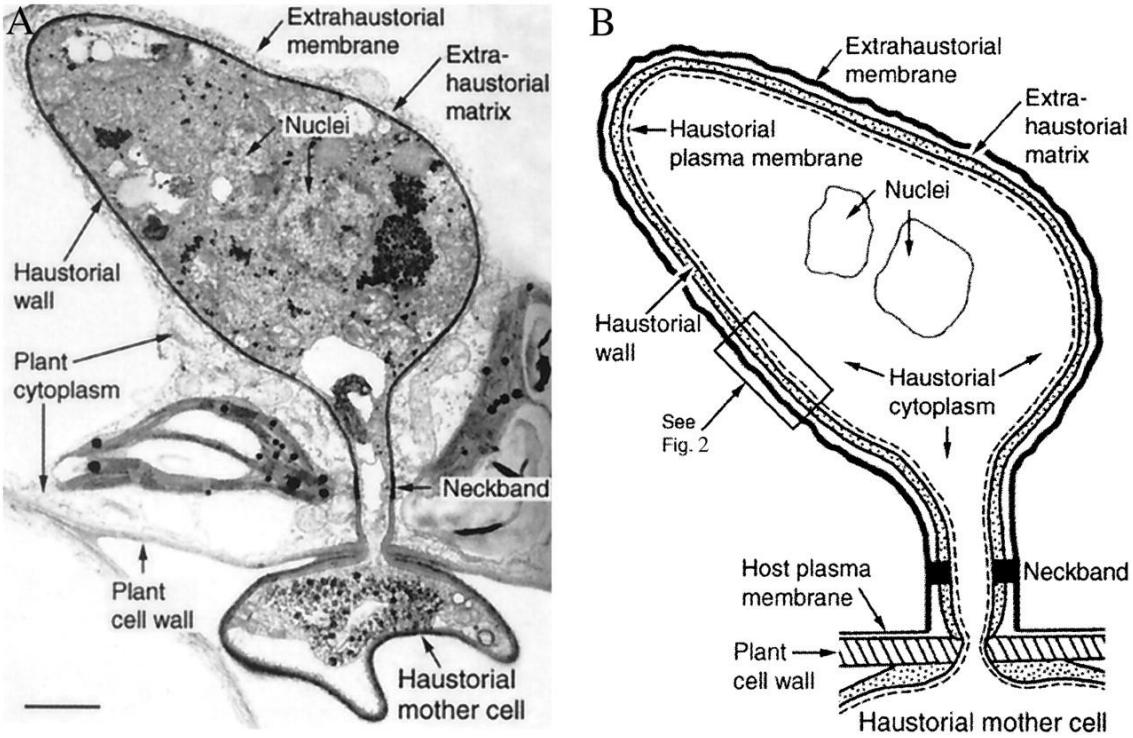


Figure 1.3 Key features of a haustorium (Szabo and Bushnell, 2001).

(A) Transmission electron micrograph of a flax rust haustorium; (B) Drawing that shows key features of the fungal haustorium.

Effector proteins of pathogens

Effectors proteins, also known as virulence factors, are secreted proteins that are produced mostly by biotrophic and hemibiotrophic pathogens. They play a crucial role in host infection, particularly in suppressing innate disease resistance in plants, but how they function within host tissues is only partially understood. Because effector proteins are able to suppress plant defense responses, they then allow plant cells to provide nutrients for the pathogen (Lo Presti *et al*., 2015). Many of the first identified effector proteins were initially called Avr (avirulence) proteins because they were recognized by host resistance genes

involved in the gene-for-gene process of race-specific disease resistance (Flor, 1942). Avr proteins are virulence proteins that are recognized by a cognate resistance gene to trigger a plant immune response that makes the pathogen avirulent on any plant that has that resistance gene allele (Van Der Biezen and Jones, 1998). Another noteworthy feature of effector proteins is that they are functionally redundant. Diverse effector proteins are likely to play similar roles, as a strategy to evade resistance gene recognition. This indicates that losing some effector proteins may not dramatically affect pathogen virulence. For example, no obvious change in infection phenotype is found after silencing the effector protein AvrL567 in the flax rust fungus *Melampsora lini* (Lawrence *et al*., 2010).

Some effectors of fungi function in the zone between fungal hyphae and the plant cell, while others are delivered into the plant cytoplasm (Lo Presti *et al*., 2015). Different from the type III secretion system (T3SS) commonly used by bacterial pathogens to deliver effectors into host cells, rust fungi are thought to deliver effector proteins into host cells via other secretion systems in the haustoria (Rafiqi *et al*., 2012). However, the mechanism by which effectors cross the extrahaustorial membrane is unknown. Some effectors are internalized in the plant cell by signals in their N-terminal regions, but signal features are not conserved between different rust effectors (Rafiqi *et al*., 2010). Only a few secreted effector proteins expressed in haustoria have been identified among rust fungi. AvrM, AvrL567, AvrP123, and AvrP4 were identified in the flax rust fungus *Melampsora lini* (Ellis *et al*., 2007). RTP1 was found in the bean rust fungus *Uromyces fabae* (Kemen *et al*., 2005). PGTAUSPE-10-1 was identified in the wheat stem rust fungus *Puccinia*

graminis f. sp. *tritici* (Upadhyaya *et al*., 2014). However, their biochemical function is not clear. No *P. emaculata* effector proteins have been identified so far.

The first bacterial Avr gene was cloned in 1984 (Staskawicz *et al*., 1984). Later, new effectors of bacteria were identified based on conserved genetic signatures and protein signatures. Genetic signatures include plant-induced promoters, expression from the operon of T3SS chaperone, and pathogenicity islands (Deslandes and Rivas, 2012). Protein signatures include T3SS targeting signals, a T3SS chaperone binding site and (for bacterial effectors) eukaryote-like motifs (Giraldo and Valent, 2013). Also, new bacterial effectors can be confirmed by T3SS delivery (Deslandes and Rivas, 2012). The first *Avr* gene in oomycetes was found in 2004 (Tian *et al*., 2004), and cytoplasmic effector proteins in oomycetes have translocation motifs (Kamoun, 2006). New effectors are found by their appearance at distinct infection stages and by containing a signal peptide. Different from bacteria and oomycetes, fungal effectors have been mainly discovered in recent years via genome analysis and transcriptomics (especially with RNA-Seq from purified haustoria) (Petre *et al*., 2014). Conserved motifs such as RXLR in oomycete effectors (Gasteiger *et al*., 2005) have not been found in fungal effectors. Fungal effectors are considered secreted proteins because they function outside the pathogen.

Plant defense: non-host resistance of plants

Non-host resistance is a powerful and durable disease resistance of plants against a variety of potentially pathogenic microbes, including bacterial and fungal pathogens. The first-line protection of plants from pathogens are waxy cuticles and plant cell walls on the epidermis. They work as a physical barrier against pathogens. Pathogens have to penetrate

the epidermis to start infections. Penetration by a pathogen can trigger rearrangements of plant cell microfilaments and microtubules (Kobayashi et al., 1997; Takemoto et al. 2003), which leads to alterations in the plant epidermis that hinder the invasion of pathogens into non-host plants. Plants also have chemical barriers to defend against pathogens. Antimicrobial chemicals produced by plants are toxic to a variety of microorganisms. Phytoanticipins and phytoalexins are two major groups of antimicrobial chemicals that are produced by plants. Phytoanticipins are produced constitutively (Hölscher et al., 2014). Phytoalexins are produced and accumulate after invasion of pathogens (Lei et al., 2014). There are two possible outcomes of non-host resistance. One outcome is to trigger PTI (described in the next paragraph). Another outcome is to trigger ETI (described in the second paragraph below), which activates a hypersensitive response resulting in cell death (Mysore & Ryu, 2004). Non-host resistance is the most common mode of disease resistance used by plants. Most pathogens are specific to only one species or one lineage of species probably because they have not evolved to suppress the non-host resistance of non-host plants.

Plant defense: PTI and ETI models

The ZigZag model proposed in 2006 describes plant defense against pathogens and the evolutionary arms race between plants and pathogens (Jones and Dangl, 2006). Plant extracellular pattern recognition receptors (PRRs) recognize conserved microbe associated molecular patterns (MAMPs), including cell wall components (e.g. chitin, peptidoglycan) and signature proteins (e.g. flagellin). MAMPs are largely invariant, essential and broadly distributed molecular signatures in microbes that cannot be easily discarded by pathogens

without fitness consequences. The activation of a PRR results in a pattern-triggered immune response (PTI) that confers the first mode of protection against a wide range of pathogens (Boller and Felix, 2009). Formation and phosphorylation of immune receptor complexes happen upon the recognitions of MAMPs (Bigeard et al., 2015). Pathogens have evolved to deliver effector proteins to suppress PTI defense and signaling (Göhre and Robatzek, 2008).

Plants have evolved resistance genes (R genes) for a second mode of defense response. An R gene is a single, usually dominant gene that encodes a single protein that recognizes a specific effector protein or (less commonly) an effector protein product. This recognition activates a signal transduction pathway that results in the effector triggered immune response (ETI). ETI is a powerful defense response that is associated with programmed cell death at the site of infection, known in the field of plant pathology as the hypersensitive response (Tsuda and Katagiri, 2010).

Flor proposed the gene-for-gene model that an effector and a cognate R protein directly interact to trigger the ETI response (Flor, 1942). This mode of resistance is traditionally not durable because a change in the Avr protein or its disappearance (e.g. genetic deletion) can remove resistance. In 1990s, the guard model was proposed because of the discovery of many indirect interactions between effectors and R proteins (Van Der Biezen and Jones, 1998). Another reason for the proposal of the guard hypothesis is that a few R proteins in plants can recognize a large number of different pathogen effectors. The guard model proposes that effectors target a plant internal protein (guardee) rather than the R protein and the R protein (guard) detects the change of the guardee to trigger the ETI

response (Dangl and Jones, 2001). In the guard model, the natural selective pressure is mostly on guardee to match the rapid change of Avr proteins, and the R protein senses the interaction between the guardee and effectors. There is no reason why the gene-for-gene and guard models might not both be true, depending on the particular interaction observed.

In the guard model, high selective pressure on the guardee may cause plant fitness costs that reduce the durability because the guardee may have other plant functions beside pathogen perception. In 2008, the decoy model was proposed, stipulating that sometimes the target of effectors in a plant monitored by the R protein is a decoy that mimics the effector target but does not have any function except effector perception (Van der Hoorn and Kamoun, 2008). Some decoys are found structurally fused to R proteins, which led to the proposal of the integrated decoy model.

SECTION 1.3

INTERACTIONS BETWEEN *PUCCINIA EMACULATA* AND SWITCHGRASS

Switchgrass and rust disease

Switchgrass, *Panicum virgatum*, is a C4 perennial plant that was selected as a "model" herbaceous bioenergy crop by the Department of Energy (DOE) in 1991 (Wright *et al*., 1992). Switchgrass can be cultivated on marginal lands with high and reliable yield (Wright and Turhollow, 2010). In the following years, perennial switchgrass has been grown in monoculture over large acreages in the US. Monoculture of switchgrass intensifies problems with fungal pathogens such as *P. emaculata*, *P. graminis* and *Uromyces*

graminicola that cause foliar rust diseases to be more and more epidemic in switchgrass fields (Gustafson, 2003). *P. emaculata* is the predominant fungal pathogen that infects the leaves of switchgrass in the US. It is a dikaryotic fungus that contains two haplontic nuclei that can be highly polymorphic (Gill *et al*., 2015), which increases the difficulty for genome analysis. It has been reported to infect switchgrass fields in many states, especially southern states such as Georgia, Tennessee and Arkansas (Zale *et al*., 2008; Hirsch *et al*., 2010). As a result, switchgrass biomass yield is being severely impacted.

Life cycle of the genus *Puccinia*

The life cycle of *P. emaculata* is not completely understood. But similar to other *Puccinia* species, *P. emaculata* is expected to require two host species to complete its life cycle. Switchgrass is the primary (uredinial-telial) host, but the alternate (aecial) host plant has not been determined. Asexual reproduction happens in the primary host, and there is no genotype change other than by mutation or mitotic recombination. Sexual reproduction happens in the alternate host. *P. emaculata* has been reported to infect spurge (*Euphorbia*) as the alternate host (Demers et al., 2017), but attempts to inoculate *Euphorbia corollata* with *P. emaculata* were not successful.

P. emaculata is assumed to have five spore stages (Figure 1.4). In other *Puccinia*, stage I is the uredinial stage. In summer, aeciospores geminate on the primary host. Fungal hyphae form uredinium from which urediniospores are produced and released. Urediniospores are dispersed by wind and germinate within a few hours after landing on the plant leaves. After infection and haustorial growth, uredinia are formed. The uredinia produce urediniospores that are released again. Urediniospores infect the primary host

recursively. At this stage, urediniospores have two nuclei within one cell. Stage II is the telial stage. In later summer and fall, seasonal change triggers the formation of telia that yield teliospores instead of urediniospores. Teliospores are two-celled. Each cell has two nuclei. When released from telia, the two nuclei in each teliospore fuse (karyogamy). Then, teliospores undergo a resting period to get through winter. Stage III is the basidial stage. In the following spring, teliospores geminate, promycelia are produced, cells undergo meiosis and nuclei are separated by the formation of a cross wall. The four-celled structure is called a basidium and basidiospores are produced from it. Basidiospores represent the beginning of the haploid phase. Stage IV is the pycnidial stage. Basidiospores fall on the alternate host leaves, geminate to form pycnia (also known as spermogonia). In a pycnidium: flexuous hyphae are at the top and pycniospores are at the bottom. Both flexuous hyphae and pycniospores are haploid with one nucleus. The nucleus can be either " + " or " - ". The pycniospores of one type fuse with the flexuous hyphae of the opposite type, but the two nuclei are in a common protoplasm without fusion. This pair of nuclei is called a dikaryon. Stage V is the aecidial stage. Dikaryotic mycelia form aecidia that become aeciospores. Aeciospores are dispersed by wind and geminate and infect when they fall on the primary host. This completes the life cycle. The rust disease itself is caused by urediniospores repeatedly infecting switchgrass plants.

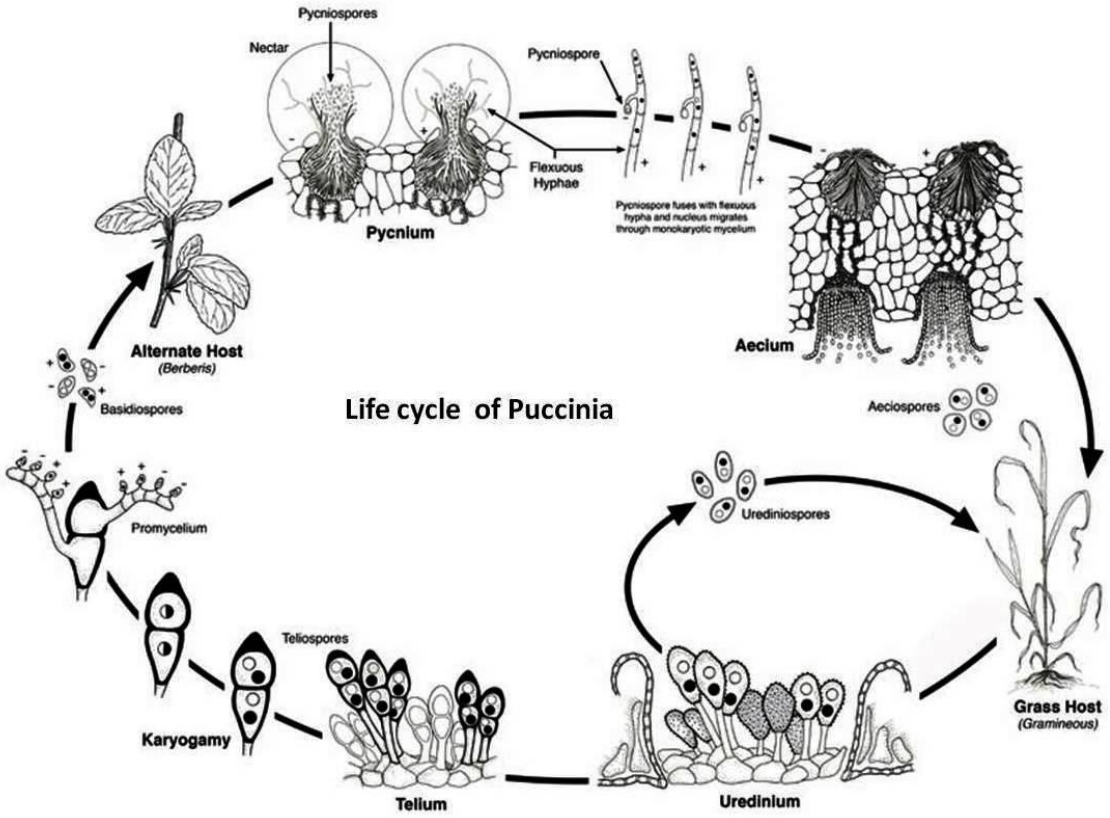


Figure 1.4 Life cycle of the genus *Puccinia*. The figure was downloaded from https://www.plantscience4u.com/2016/11/5-stages-in-life-cycle-of-puccinia.html#.XjDehC2UA6W

SECTION 1.4

ENDOPHYTES AND THEIR TRANSMISSION

Endophytes are microbes, including fungi and bacteria, that live inside a plant but do not cause disease (Wilson 1995). Endophytes can colonize diverse plant tissues, including

leaf and root. For example, *Rhizobia* colonize roots within a specially designed nodule where they can efficiently fix atmospheric nitrogen in legumes (Frank, 1889). Arbuscular mycorrhizal fungi (AMF) are root symbionts in most plants, and are noted for their roles in mineral uptake and drought tolerance (Gherbi et al., 2008). Many foliar endophytes have been reported in plants, including in grasses (Omacini et al., 2004; Novas et al., 2009; Gagne-Bourgue et al., 2013) and trees (Koukol et al., 2012; Rajamani et al., 2018). Endophytes spend all or part of their life cycle internal to plants. The compositions of endophytes are diverse and complex, ranging from mutualistic strains to latent pathogens.

Endophytes may or may not have any discerned effect on the plant. In some cases, the effects of endophytes on plants can be positive or negative (Hardoim et al., 2015). Endophytes can be classified into three categories, based on their effect on the plant. First, endophytes living inside a plant may have no apparent effect on the plant. Second, endophytes may have beneficial effects on the plant. For instance, some fungal endophytes produce antibiotic secondary metabolites that protect the plant from invasion by pathogens and/or damage by herbivores (Hartley & Gange, 2009; Ownley et al., 2010). Bacterial endophytes protect plants from biotic and abiotic stress. Some endophytes are also beneficial to nutrient uptake and growth of plant. Third, endophytes may be latent pathogens which can live in plants for a long time and only cause disease when plants are under stress. (Hardoim et al., 2015). For instance, *Botryosphaeriaceae* are endophytes and latent pathogens of woody plants (Slippers & Wingfield, 2007).

Endophytes can arrive in a plant via vertical transmission, horizontal transmission or both (Frank et al., 2017; Bright & Bulgheresi, 2010). Vertical transmission happens via

seed (through either or both gametes, or maternal tissue) such that endophytes are transmitted from parents to offspring. For example, *Epichloë* and *Neotyphodium* species, fungal endophytes of some grasses, are transmitted via host seeds (Schardl, 2002). Endophytes can be transmitted within the plant through xylem vessels. Endophytes can also be transmitted via the shoot apical meristem, including upon its terminal differentiation into reproductive organs (Frank et al., 2017). Horizontal transmission happens via the environment including through soil, air and insect vectors. For example, many ascomycetes or basidiomycetes endophytes can be transmitted from root to root (Rodriguez et al., 2009).

Transmission from soil to root is the major mode of horizontal transmission. Evidence that most endophytes are acquired from the soil environment is provided by the routine observation that vastly fewer root endophytes are found in plants growing in sterile conditions than in field soil (Frank et al., 2017). Endophytes can colonize the plant root through spermosphere ("the zone surrounding seeds where interactions between the soil, microbial communities and germinating seeds take place", defined by Schiltz et al., 2015) and rhizosphere (Schiltz et al., 2015; Mitter et al., 2017). Foliar endophytes can be transmitted by the wind or rain, and colonize inside leaves through epidermal openings such as stomates or wounds (Scott et al., 1996). Endophytes can also be transmitted via other vectors such as arthropods whose feeding on plant tissues provides wound access for microbes (Frank et al., 2017).

CHAPTER 2

GENOMIC ANALYSIS OF *HAFNIA PARALVEI* ISOLATED FROM PITCHER FLUIDS OF *SARRACENIA ROSEA*

Sarracenia rosea is a species of pitcher plant that is distributed along the Gulf Coastal Plain from southeast Mississippi to southwest Georgia. The pitcher of a pitcher plant contains a microecosystem that consists of diverse microbes and other small organisms. Ten bacterial isolates associated with archaea during plate selection were isolated from samples of pitcher fluids of *S. rosea* collected in Alabama. All ten were sequenced in Illumina MiSeq flow cells, and were assembled to the contig level. Six isolates were identified as *Hafnia paralvei* by analysis of average nucleotide identities (ANIs). This is the first time that *Hafnia* has been reported from a plant environment. Pan and core genome comparison between *H. paralvei* strains from pitcher and non-pitcher environments illustrated the distinct features of pitcher *H. paralvei*. A total of 276 gene clusters present in the core (that is, shared) genomes of all pitcher strains were found to not be present in any non-pitcher strain. These pitcher-specific genes included 210 gene clusters encoding hypothetical proteins. Of the 66 remaining pitcher-specific genes in *H. paralvei*, predicted functions were associated with 16 Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathways, including flagellar assembly, type VI secretion system, DNA replication, DNA mismatch repair and homologous recombination. All of these pathways were present in both pitcher and non-pitcher strains, but they use different orthologous gene family members to perform these functions, possibly due to environment-dependent selection on these traits. In addition, the diversity in the gene content within pitcher *H. paralvei* isolates was investigated by analysis of the "accessory" (i.e., variable) gene content of pitcher strains. This accessory gene content in pitcher *H. paralvei* consisted of 148 gene clusters, including 60 genes clusters annotated as encoding hypothetical proteins. Of the function-annotated remaining gene clusters in the accessory genome, activities such as nitrogen metabolism, were discovered. This study of *H. paralvei* strains from pitcher plants provides insights into intraspecies bacterial variation in a microbial community and between microbial communities.

Introduction

Sarracenia is a genus of pitcher plants that contains 13 species. *Sarracenia* inhabit sunny, aquatic, nutrient-poor places such as bogs where they use their pitchers to capture and digest insect prey. *Sarracenia rosea* has short, stout and decumbent pitchers with open hoods (Schnell, 2002), and is distributed in the Gulf Coastal Plain of southern Alabama, northwestern Florida, southwestern Georgia and southeastern Mississippi (NACZI et al., 1999). It was initially classified as *S. purpurea* var. *burkii* D.E. Schnell but was eventually separated from *S. purpurea* as a new species because of its distinct morphology (including pinker petals, larger flowers and shorter scapes) (NACZI et al.,1999; Ellison et al., 2004).

However, *S. purpurea* and *S. rosea* are more similar to each other than to other pitcher plants. They are classified into a single group that is distinct from other *Sarracenia* pitcher plants based on pollen structure and other morphological characters. (Oswald et al., 2011). Like *S. purpurea*, *S. rosea* collects rainwater because of its erect and open hoods (Robinson, 1908; NACZI et al., 1999). *S. rosea* occurs in few locations and its habitat is shrinking. Gulf Coast bogs covered approximately 3000 km^2 prior to European settlement. However, ~97% of this original area has been lost, and may be incapable of fully recovering because of human activities (eg. bog drainage for agriculture), climate change, rising sea levels and the increased frequency of intense hurricanes (Folkerts, 1982; Abbott & Battaglia, 2015).

Pitchers are special organs that are formed by modified leaves. They work as pitfalls to small prey organisms ranging from bacteria to vertebrates for the sake of acquiring nutrients (Adlassnig et al., 2010). Prey are digested at the bottom of pitchers by the pitcher community and by pitcher fluids. Fluid solvents can be secreted by pitcher plants or collected from rainfall. *S. rosea* are perennial plants. Pitchers senesce in the fall, but new pitchers form in the following spring. Pitchers are closed and empty when they first develop in the spring. Later, the pitchers' hoods open to collect water from rainfall (Arber, 1941; Hepburn et al., 1927; Peterson et al., 2008; Plummer & Jackson, 1963). Microbial communities are constructed from various inocula (wind, rain splash, bog flooding, prey arrival, etc.) The community is sustained by continual re-inoculations by these same processes and by the availability of nutrients from the digesting prey.

The peculiar characteristics of pitcher plants, including their carnivorous action, pitcher morphology and unusual growth habitats, have attracted the interest of researchers

for more than a century. The microecosystems of pitcher plants are unique, including dramatic variation within a single species at the same location at different seasons (Chaluvadi & Bennetzen, unpub. res.). The environment in the pitcher creates a food web where microbes and the plant host can be considered as mutualists. Pitcher plants are excellent model systems for the study of interactions among microbes and invertebrate inquilines. The microecosystem has served as a model to study food web dynamics, community genetics, trophic interactions, and population structure (Ellison et al., 2003). So far, more and more research is focused on the composition and internal interactions of pitcher microorganisms (Siragusa et al., 2007; Koopman et al., 2010; Butler et al., 2008; Peterson et al., 2008; Paisie et al., 2014; Young et al., 2018), but these studies are far from complete. Current data on the microbiology of pitcher fluids are spare and fragmentary. The ecology, taxonomy and interaction of the microbes in *Sarracenia* pitcher plants remain unclear. Variety and abundance of microbes and their metabolic pathways, especially catabolic pathways, are expected to be important to pitcher plants. The microbial community in the pitcher of *S. rosea* has never been described in any publication. Composition of the microbial community, diversity of inquilines and their roles in prey degradation remain largely unknown.

To study the microbial communities in pitchers, samples of pitcher fluids of *S. rosea* from the Splinter Hill Bog in Alabama were collected, and the Bennetzen lab isolated multiple microbes by growth on a medium that was expected to enhance the recovery of archaebacteria. Although no archaea were isolated, genomes of ten separate bacterial cultures from this isolation process were chosen and sequenced to get some idea of the

bacterial content of *S. rosea* pitchers. Genome assembly and annotation of all ten indicated that six were *Hafnia*. The genus *Hafnia* belongs to the *Enterobacteriaceae*, a family of facultative anaerobes that includes *Salmonella, Klebsiella* and *Escherichia coli*. *Hafnia* are often found in food products (Ridell and Korkeala,1997; Lindberg et al., 1998), in the gastrointestinal tracts of animals, and in various water sources (Janda and Abbott, 2006). *Hafnia alvei* was the first *Hafnia* species discovered. A subset of *H. alvei* was reclassified as *H. paralvei* in 2010 (Huys et al., 2010). *H. alvei* has been implicated in animal infections at many locations. In 1990, it was found to be associated with haemorrhagic septicaemia in rainbow trout (Gelev et al., 1990). *H. alvei* was reported to be associated with septicemia of commercial laying hens in Spain in 1997 (Real et al., 1997). In 2004, an outbreak of *H. alvei* infection in 12-week-old pullets was reported in Italy (Proietti et al., 2003). *Hafnia* have been isolated from various clinical materials and found to be related to human diseases such as bacteremia (Conte et al., 1996), respiratory tract infections, and extraintestinal infections (Huys et al., 2010). However, our understanding of the genus *Hafnia* remains incomplete. Although genomic information on *H. paralvei* has recently (since 2018) increased, most *H. paralvei* genome assemblies are incomplete. There is no analysis of gene content diversity among *H. paralvei* strains.

These six turned out to be *Hafnia paralvei* strains by taxonomic identification using genomic data. Although *H. paralvei* strains have been found from several sources (mostly mammal and fish related), it is the first time that any *Hafnia* strains were found in a plant environment. It has been found by pan-and-core genome analysis that intraspecific gene contents of many species, such as *Pseudomonas syringae* and *Streptomyces rimosus* differs

depending on hosts (Karasov et al., 2017; Park & Andam, 2019). Although the genomic analysis was undertaken at the genus level between *H. alvei* and *H. paralvei* (Yin et al., 2019), the gene content differences among *H. paralvei* has not been studied. Therefore, how gene contents of *H. paralvei* of pitcher environments differs from those of other environments were of strong interest. Exploring the unique features and gene diversities of *H. paralvei* from pitcher environment not only helps understand bacterial adaptations in the pitchers but also adds knowledge to the study of pitcher microbial communities.

Methods

Sample collection and isolation

Samples of the pitcher fluids of *S. rosea* were collected from several locations of the Splinter Hill Bog, Alabama (in March 2012). The collected fluids were pooled. Strains were isolated from pitcher fluids by plating. The fluids were placed on the plates of media used for the culture of the archaea *Pyrococcus furiosus* (Adams et al., 2001) in the hope that this would enrich for archaea, which are abundant in pitcher fluids (Chaluvadi and Bennetzen, unpub. res.). Single colonies were serially streaked onto the same media three times, and then individual colonies were picked into nutrient broth for growth and eventual storage at -80°C in 25% glycerol.

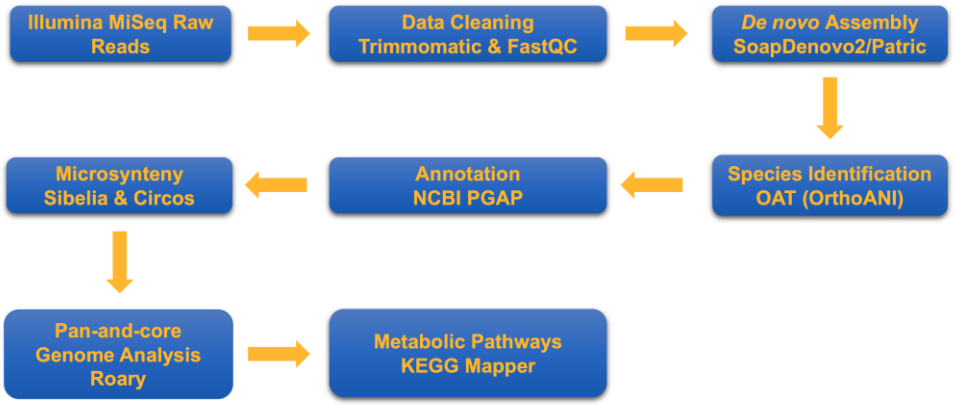
DNA extraction and whole genome sequencing

DNAs were extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research Cat No. D6005). Paired-end libraries were prepared using Nextera XT DNA

Library Preparation Kit V2 (Illumina Cat No. FC-131-1002). The sequencing platform used was Illumina MiSeq (PE250).

Overall pipeline of computational analyses

First, raw reads were collected from Illumina MiSeq platform. Second, raw reads were cleaned with Trimmonmatic to remove Illumina adaptor sequences and low-quality reads. FastQC were used to estimate the reads quality. Third, reads were put into *de novo* genome assemblers (SoapDenovo2 and Patric) for de *novo* genome assembly. Forth, species were identified by calculating average nucleotide identities between genomes with OAT. Fifth, genome assemblies were submitted and annotated with NCBI PGAP. Sixth, microsyntenic structures between genome assemblies and a complete reference genome were explored with Sibelia and visualized with Circos. Seventh, pan-and-core genomes were calculated with Roary. Then differences in gene contents were explored manually. Last, KEGG pathways associated with differential presence genes were discovered with KEGG Mapper. Software were chosen comprehensively based on the performance, accuracy, availability and accessibility. The workflow is shown below and details such as parameter settings were described in the following texts.



Genome assembly and annotation

Illumina raw reads were first quality filtered with Trimmomatic (version 0.36) (Bolger et al., 2014). FastQC (version 0.11.4) was employed to examine the quality of the trimmed reads (Andrews, 2010). SoapDenovo2 (version r240, parameter -K 83) was used to assemble the reads into contigs (Luo et al., 2012). Patric (Wattam et al., 2013; Wattam et al., 2016) was used to assemble the data set after I observed that this data set could not be assembled very well using SoapDenovo2. Contigs with low coverage (<20 x) or short lengths (<400 bp) were discarded. Then, genome assemblies were submitted to GenBank and annotated by the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2013). Genome coverage (also known as "coverage" or "depth") was estimated by the formula LN/G, where L was the read length, N was the number of reads and G was the genome length. Default parameters were used for all software tools unless otherwise noted.

Analysis of 16S rRNA genes

Reads associated with 16S rRNA genes were filtered by alignments with the Silva reference files downloaded from the Mothur database, an open source package that is

designed for analysis of 16S rRNA gene sequences (https://www.mothur.org/wiki/Silva_reference_files). Then, filtered reads were analyzed with the 16S Biodiversity tool in Geneious (http://www.geneious.com, Kearse et al., 2012). Default parameters were used unless otherwise specified.

Calculation of average nucleotide identities (ANIs)

Average nucleotide identity (ANI) is a nucleotide-level measurement of similarity between two genomes. The OrthoANI (Average Nucleotide Identity by Orthology) algorithm was used to calculate ANI values (Lee et al., 2016). ANI criteria to identify bacteria species were developed previously by large-scale comparison of bacterial genomes (Goris et al., 2007; Rodriguez-R et al., 2014). By this determination, it was decided that ANI values between two genomes of the same species are greater than or equal to 95%. ANI values between two genomes of different species is between 75% and 95%. ANI values below 75% are not trustworthy, so average amino acid identities should be calculated instead. Each of the assembled genomes from pitcher plants were compared to two published genomes of *H. alvei* and two published genomes of *Hafnia paralvei*. The two genomes of *H. alvei* are *H. alvei* FB1 (CP009706.1) and *H. alvei* HUMV-5920 (CP015379.1). The two genomes of *H. paralvei* are *H. paralvei* ATCC 29927 (GCA_001655005.1) and *H. paralvei* FDAARGOS_158 (CP014031.2). Default parameters in OrthoANI were employed.

Microsynteny analysis and circular genome visualization

Genomic microsynteny blocks were computed with Sibelia (Synteny Block ExpLoration tool). Parameters used were "-s -fine -m 400". Each genome assembly was compared to a complete reference genome. A version of Damien Richard's perl script (https://github.com/DamienFr/GC_content_in_sliding_window/blob/master/README.md) was revised and used to calculate GC content and GC skew. Gene content information of protein coding sequences (CDSs) and RNA genes were extracted from GFF files of genome annotation. Circos (http://circos.ca) was used to visualize genome comparison results. Input files for Circos were constructed and formatted with AWK in Linux command lines.

Analysis of pathogenicity, virulence and antibiotic resistance genes

PathogenFinder 1.1 (Cosentino et al., 2013) from the Center for Genomic Epidemiology (CGE) was used to predict bacterial pathogenicity of sequenced genomes. Genome sequences were inputted into the PathogenFinder server. Probability of a strain being a human pathogen was predicted by comparing genes of the input genome with genes that frequently show up in human pathogenic bacteria and in innocuous bacteria.

Bacterial virulence factors were predicted with VirulenceFinder 2.0 (Joensen et al., 2014). Virulence factors in strains of *Enterococcus, Staphylococcus aureus, Escherichia coli and Listeria* were used as the searching database. Strains of these genera/species are known to be human pathogens. Genome assemblies were inputted into the VirulenceFinder server. The minimum percentage of a nucleotides sequence that overlapped a virulence

gene that was considered informative was 60%. The minimum percentage of nucleotides that were identical between a virulence gene in the database and the corresponding sequence in a genome was set at 90% to be considered important.

Antimicrobial resistance genes were predicted with ResFinder 3.2 (Zankari et al., 2012). Antibiotic resistance genes for tetracycline, colistin, phenicol, *β*-lactam antibiotics, sulphonamide, oxazolidinone, fosfomycin, macrolide antibiotics, glycopeptide antibiotics, trimethoprim, quinolone, fusidicacid, nitroimidazole, aminoglycoside antibiotics, and rifampicin were screened for. The minimum percentage of a nucleotides sequence that was considered important if it overlapped a virulence gene was 60%. The minimum percentage of nucleotides that were identical between a resistant gene in the database and the corresponding sequence in a genome was set at 90% to be considered important.

Comparative analysis of gene contents and KEGG pathways

Genomic gene contents were analyzed with Roary (version 3.12.0) (Page et al., 2015). GFF files created with Prokka (version 1.13) were used as input (Seemann, 2014). Nucleotide sequences of coding regions were extracted and converted into predicted amino acid sequences. Predicted proteins were clustered with CD-HIT (Fu et al., 2012), resulting in a set of protein clusters. The protein cluster sets were compared to each other using BLASTP (cutoff 95%) and re-clustered with MCL (Altschul et al., 1990; Enright et al., 2002). Differences in gene cluster contents between two groups of strains (pitcher derived and non-pitcher derived) were computed using the "*query_pan_genome*" command under Roary. A single representative nucleotide sequence from each of the clusters was extracted. Nucleotide sequences were submitted to KAAS (KEGG Automatic Annotation Server) for

functional annotation (Moriya et al., 2007). Specifically, BLAST (BLASTX and TBLASTN) programs and the BBH (bi-directional best hit) method were used for KEGG Orthology (KO) assignments. Then, the Reconstruct Pathway tool of KEGG Mapper (Kanehisa & Sato, 2019) was used to map KO-assigned genes to KEGG pathways.

Results

Whole genome shotgun sequencing and data preprocessing

Genomic DNAs were extracted from ten pitcher bacterial isolates that were initially selected for their interaction with pitcher archaea (which was, unfortunately, not sustained during the serial culturing process). Paired-end libraries were prepared. Then, the samples were loaded on Illumina MiSeq flow cells for whole genome shotgun sequencing. In total, 19,967,488, 17,470,830, 14,819,740, 16,926,784, 14,943,144, 14,784,154, 15,543,750, 18,279,064, 17,972,284 and 15,359,394 raw reads were respectively generated for strains C1 to C10 (Table 2.1).

A good quality of sequence data should be guaranteed before carrying out any further analyses, because data with poor quality affects output devastatingly. The FastQC analysis suggested by Illumina is designed to inspect the quality of sequence data coming from high throughput sequencing pipelines. All raw reads were inputted into FastQC and the per base sequence quality was calculated (Figure 2.1, top). The per base sequence quality scores at the beginning of reads are good. However, the per base sequence quality scores at the tail of reads were lower than 30, indicating the necessity to preprocess raw reads to improve genome assembly quality. Raw reads were preprocessed in Trimmomatic for quality

filtering and the removal of sequence adapters. The number of reads that survived preprocessing removal are shown in Table 2.1. The quality of these reads was inspected by FastQC again. Per base quality scores of reads were improved after preprocessing (Figure 2.1, bottom). The preprocessed reads with quality scores above 30 were used for all subsequent analyses.

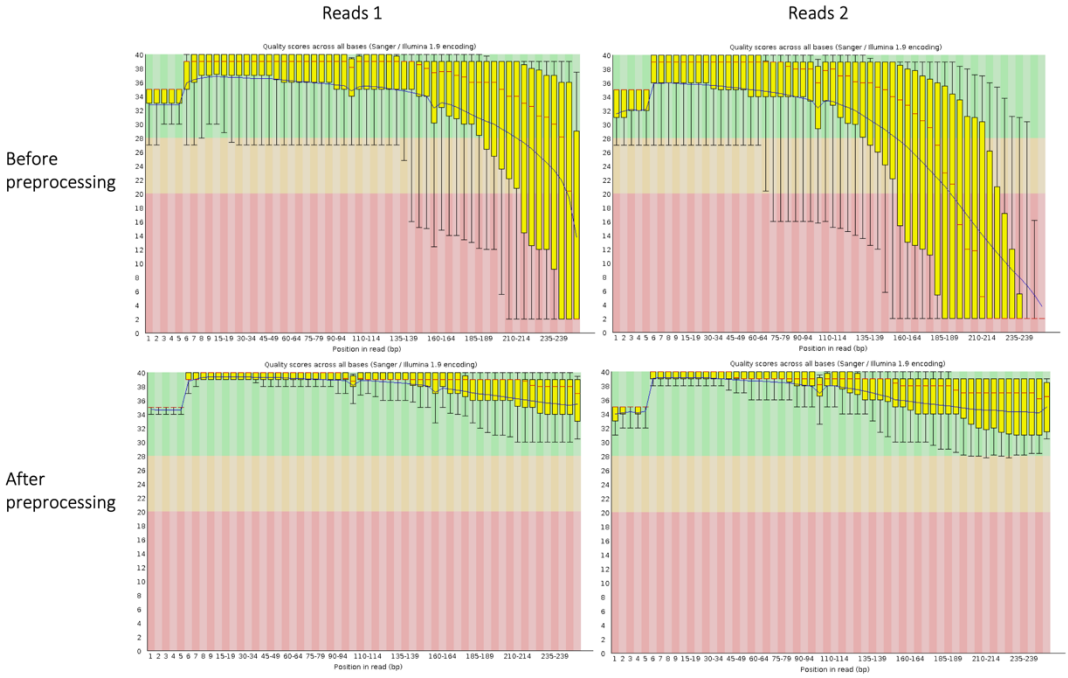


Figure 2.1 Per base sequence quality of raw reads (top) and preprocessed reads (bottom) of strain C2. Per base quality scores above 30 (green area) indicates good quality of Illumina sequence data. Per base sequence quality of raw reads and preprocessed reads of other genomes were similar to those for strain C2.

Table 2.1 Description of sequence data for the microbial samples

Microbial sample	Label	Number of Raw Reads	Number of Preprocessed Reads
1F7	C1	19,967,488	15,812,416
2B2-1	C2	17,470,830	15,809,772
2B2-2	C3	14,819,740	12,044,372
2D6-7	C4	16,926,784	9,629,704
2E1-8	C5	14,943,144	13,209,292
2E7	C6	14,784,154	9,779,180
2D1	C7	15,543,750	11,287,116
2D1	C8	18,279,064	15,816,968
1F7	C9	17,972,284	15,527,984
2E10	C10	15,359,394	12,913,488

Genome assembly

Reads that survived preprocessing were *de novo* assembled without a known reference genome. Two (C6 and C7) of these genome assemblies are described in the Appendices. The size of C1 assembly was ~10 Mb, which is unlikely to be an assembly from a single colony. About half of the C4 genome assembly was half *Clostridium* genome, and the other part is unknown sequences. Genome assemblies of C1 and C4 were not furthered studied. The initial assembly of C10 was ~9 Mb. The assembly was split into two parts based on the per base coverage. The small part of which the coverage of was <100X turned out to be a partial *Hafnia* genome. The coverage of the other part was >200X, and this part was kept as the final assembly of C10. Here, I focused on six genome assemblies (C2, C3, C5, C8, C9 and C10) whose features were very similar to each other (Table 2.2). The genome

sizes were all ~4.7 Mb. The GC contents were ~48%. N_{50}, the minimum contig length needed to cover half of a genome assembly, ranged from 69,659 bp to 430,846 bp. L_{50}, which is the number of contigs needed to cover half of a genome assembly, ranged from 5 to 23. Numbers of assembled contigs in the six genome assemblies ranged from 32 to 160.

Table 2.2 Features of genome assemblies.

Index	C2	C3	C5	C8	C9	C10
Size	4,722,064 bp	4,709,191 bp	4,694,262 bp	4,695,249 bp	4,696,175 bp	4,710,920 bp
GC Content	48%	48%	48.1%	48.1%	48.1%	48.1%
N_{50}	95,400 bp	69,659 bp	81,051 bp	80,989 bp	81,575 bp	430,846 bp
L_{50}	19	23	20	19	19	5
# of Contigs	159	160	120	120	120	32

Identification of Species

The genomes assembled were first identified at the genus level by 16S rRNA gene sequence analysis. Preprocessed reads associated with 16S rRNA genes were filtered by alignments with the Silva reference files of 16S rRNA genes. Then, the filtered reads were analyzed with the 16S Biodiversity tool in Geneious. The biodiversity results indicate that six of the strains isolated from pitcher plants belong to the genus *Hafnia*. I also analyzed the genomes of two of the four other strains that were sequenced. One turned out to be a *Seratia maracescens* ecotype and another was an apparently novel *Enterobacter* species, and these sequences plus gene content analyses were published (Zhang et al., 2020a; Zhang

et al., 2020b). The other two bacterial genomes were not further studied (Chaluvadi and Bennetzen, unpub. res.).

The genus *Hafnia* contains two species: *Hafnia alvei* and *Hafnia paralvei*. In order to figure out to which species of *Hafnia* these strains belong, average nucleotide identities (ANIs) of genomic sequences were calculated (Rodriguez-R and Konstantinidis, 2014). Typically, the ANI values between genomes equal to or greater than 95% indicates that they belong to the same species (Goris et al., 2007). Each of the genome assemblies was compared to two available genomes of *H. alvei* and two available genomes of *H. paralvei*. The two genomes of *H. alvei* were *H. alvei* FB1 (CP009706.1) and *H. alvei* HUMV-5920 (CP015379.1). *H. alvei* FB1 was isolated from a fish paste meatball and is a common bacterium in foods consumed by Malaysian and Chinese populations (Tan et al., 2014). The genome of *H. alvei* FB1 is the representative genome for the species chosen by the National Center for Biotechnology Information (NCBI). *H. alvei* HUMV-5920 was obtained from a urine sample from an adult patient, and its genome sequence is complete (Lázaro-Díez et al., 2016). The two genomes of *H. paralvei* were *H. paralvei* ATCC 29927 and *H. paralvei* FDAARGOS_158. *H. paralvei* ATCC 29927 (CDC 4510-73, Biosafety Level: 1) was isolated from a human clinical specimen and its genome is the representative genome designated by NCBI. *H. paralvei* FDAARGOS_158 was isolated from a hospital stool sample, and its genome sequence assembly is complete. ANI values of genome assemblies compared to *H. alvei* and *H. paralvei* strains are shown in Figure 2.2. I found that ANI values between each pitcher associated strain and either *H. alvei* strain were ~83%, suggesting that pitcher associated strains and *H. alvei* strains did not belong to the same

species. In contrast, the ANI values comparing the assemblies to *H. paralvei* strains were more than 99% percent, indicating that pitcher associated strains belong to the species *H. paralvei*.

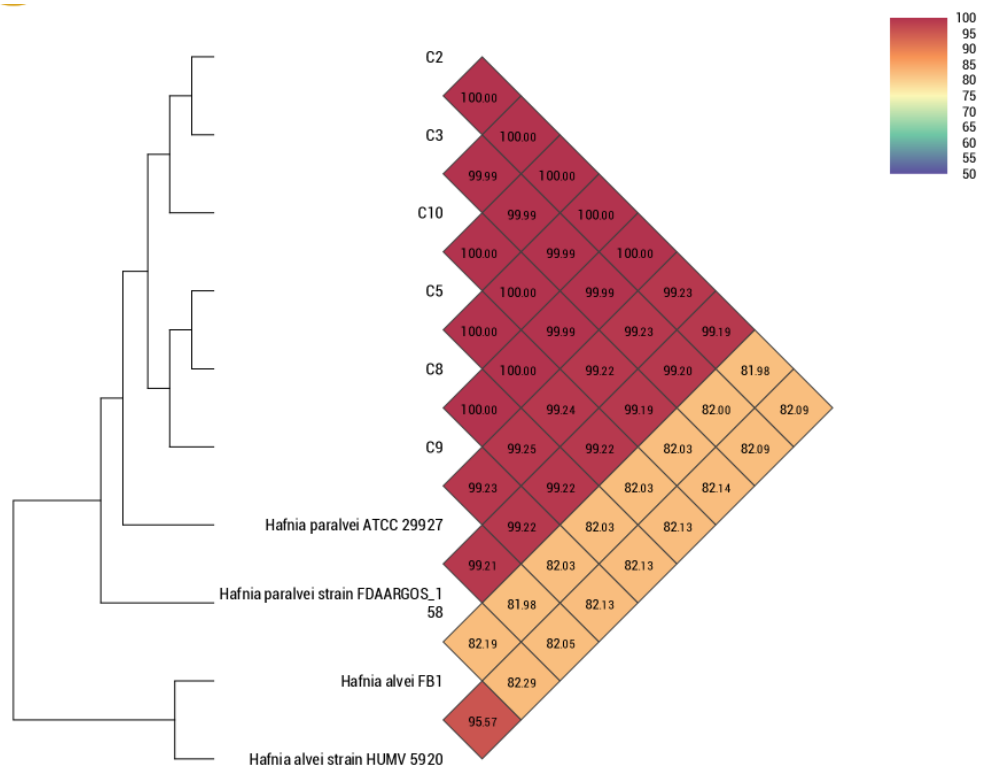


Figure 2.2 Comparisons of average nucleotide identities (ANIs) among *Hafnia* strains. Red block: ANIs are greater than or equal to 95%; yellow block: ANIs are less than 95%.

Genome annotation

Genome assemblies were submitted to GenBank and annotated by the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2013). Features of the genome annotations are shown in Table 2.3. Annotation features of the six genome assemblies were

similar but not all the same. In general, about 4200 protein-encoding genes and about 85 pseudogenes were predicted in each genome assembly. In addition, more than 80 RNA-encoding genes (rRNA genes, tRNA genes and ncRNA genes) were predicted in each genome assembly. More rRNA genes and tRNA genes were found in C10 than in other genome assemblies. In contrast, less protein-encoding genes were found in C10 than in the other genome assemblies.

Table 2.3 Annotation of genome assemblies

	C2	C3	C5	C8	C9	C10
Pseudogenes	85	87	85	84	88	86
Protein genes	4265	4256	4223	4218	4218	4199
RNA genes	89	89	81	84	86	126
rRNA genes	8	8	4	4	5	34
tRNA genes	74	74	70	73	74	85
ncRNA genes	7	7	7	7	7	7

Pairwise comparisons between genome assemblies and a complete reference genome

Each genome assembly was compared to the complete reference genome from *H. paralvei* (strain *FDAARGOS_158*). The comparisons are depicted by the circular visualization of Circos (Figures 2.3 to 2.8). The outermost circle represents the length of the reference genome (green) and the length of contigs of my genome assemblies (blue), ordered by their overall collinearity with the reference genome. Contigs of my genome assemblies shown in the circular figures are separated by gaps. The second to fifth outermost circles depict CDS genes on the positive strand, CDS genes on the negative strand, RNA genes on the positive strand and RNA genes on the negative strand. The sixth

and seventh outermost circles represent GC content and GC skew, which were calculated based on a sliding window of 400 bp.

Microsynteny blocks are non-overlapping highly conserved segments of a genome, and thus reflect similarities and variations between the structure of two genomes. The innermost circles in Figures 2.3-2.8 represent microsynteny blocks (based on a sliding window of 400 bp) between a genome assembly and the reference genome. Green and orange blocks represent the direction of microsynteny blocks on the positive and negative strands, respectively. Black blocks represent regions that were not syntenic. Microsynteny blocks between two genomes were linked with rainbow ribbons. Most parts of each genome assembly from pitcher plant strains were syntenic/colinear to the reference genome.

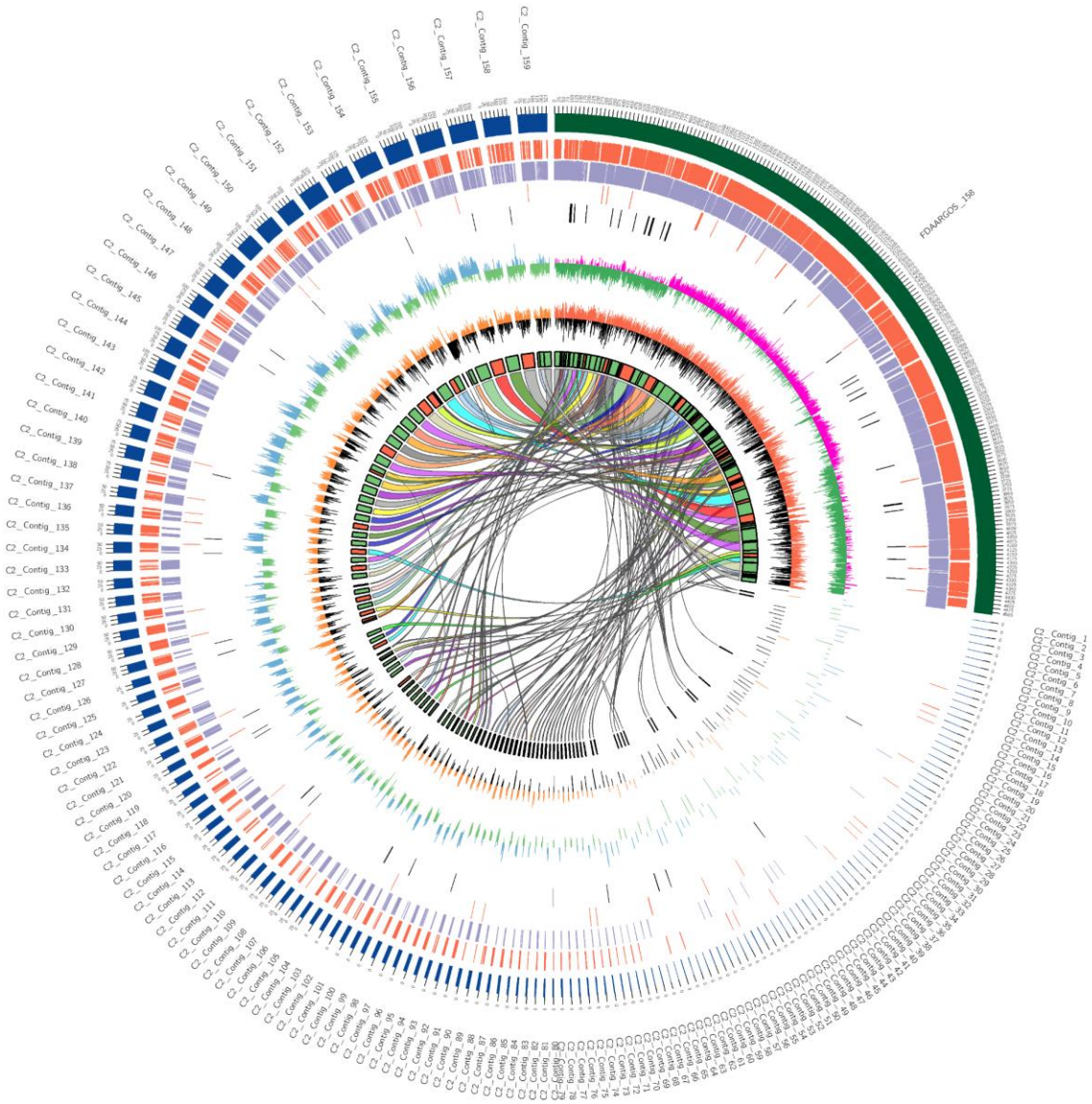


Figure 2.3 Genome comparison between *H. paralvei* strain C2 and *H. paralvei* strain *FDAARGOS_158*. Features in the figure were described in detail in the above 2 paragraphs.

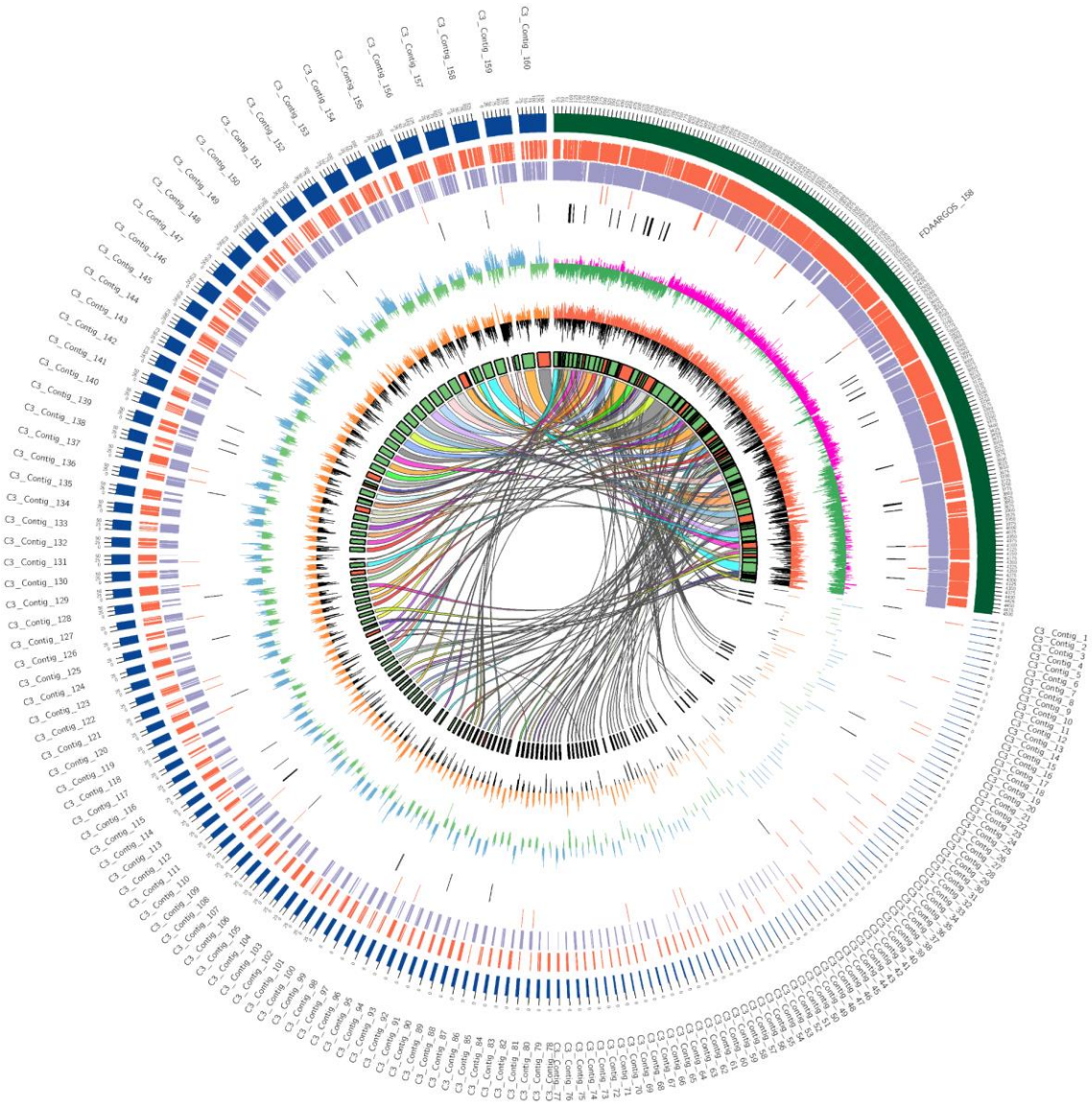


Figure 2.4 Genome comparison between *H. paralvei* strain C3 and *H. paralvei* strain *FDAARGOS_158.* Features in the figure were described in detail in the above 2 paragraphs.

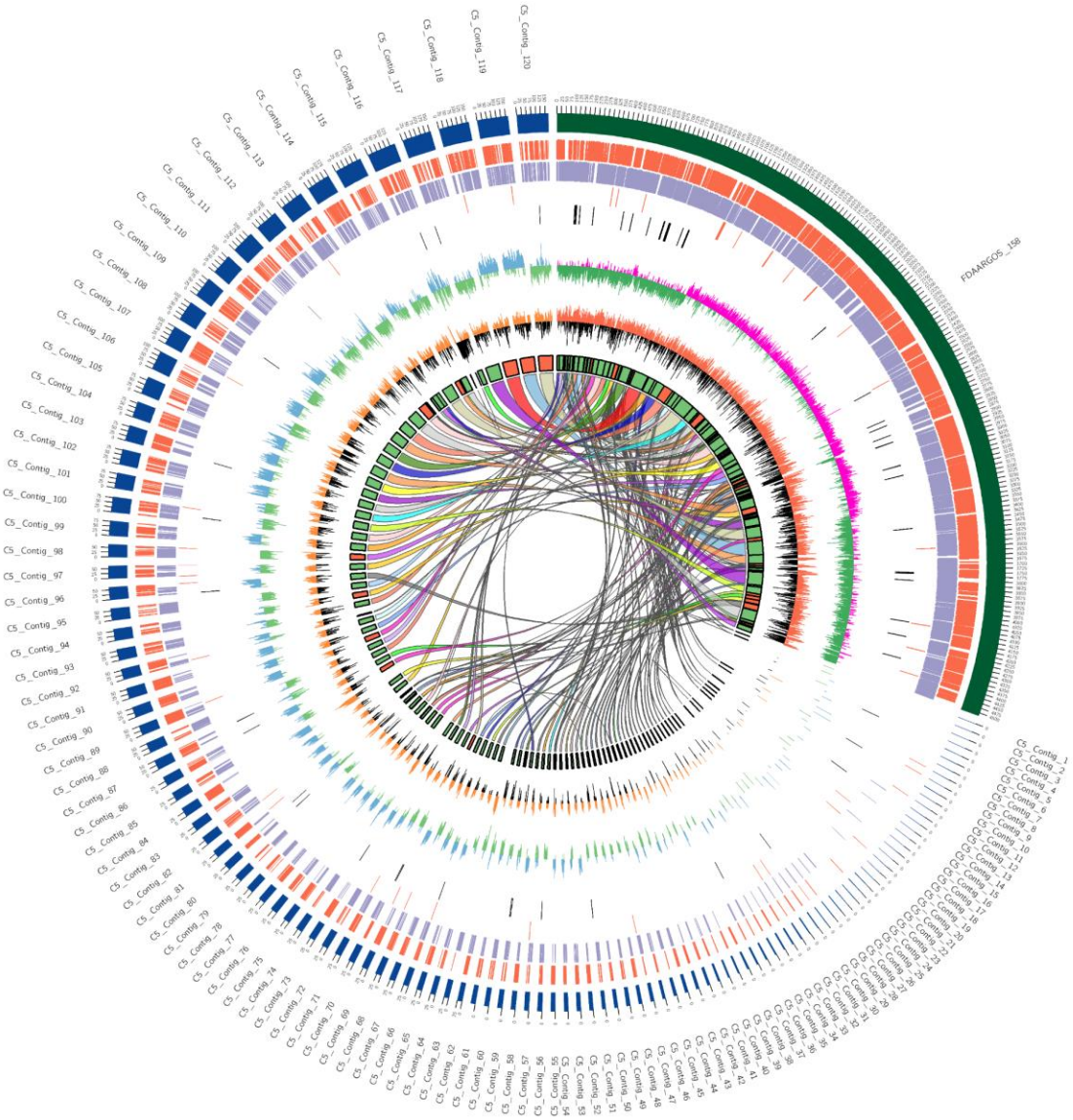


Figure 2.5 Genome comparison between *H. paralvei* strain C5 and *H. paralvei* strain *FDAARGOS_158.* Features in the figure were described in detail in the above 2 paragraphs.

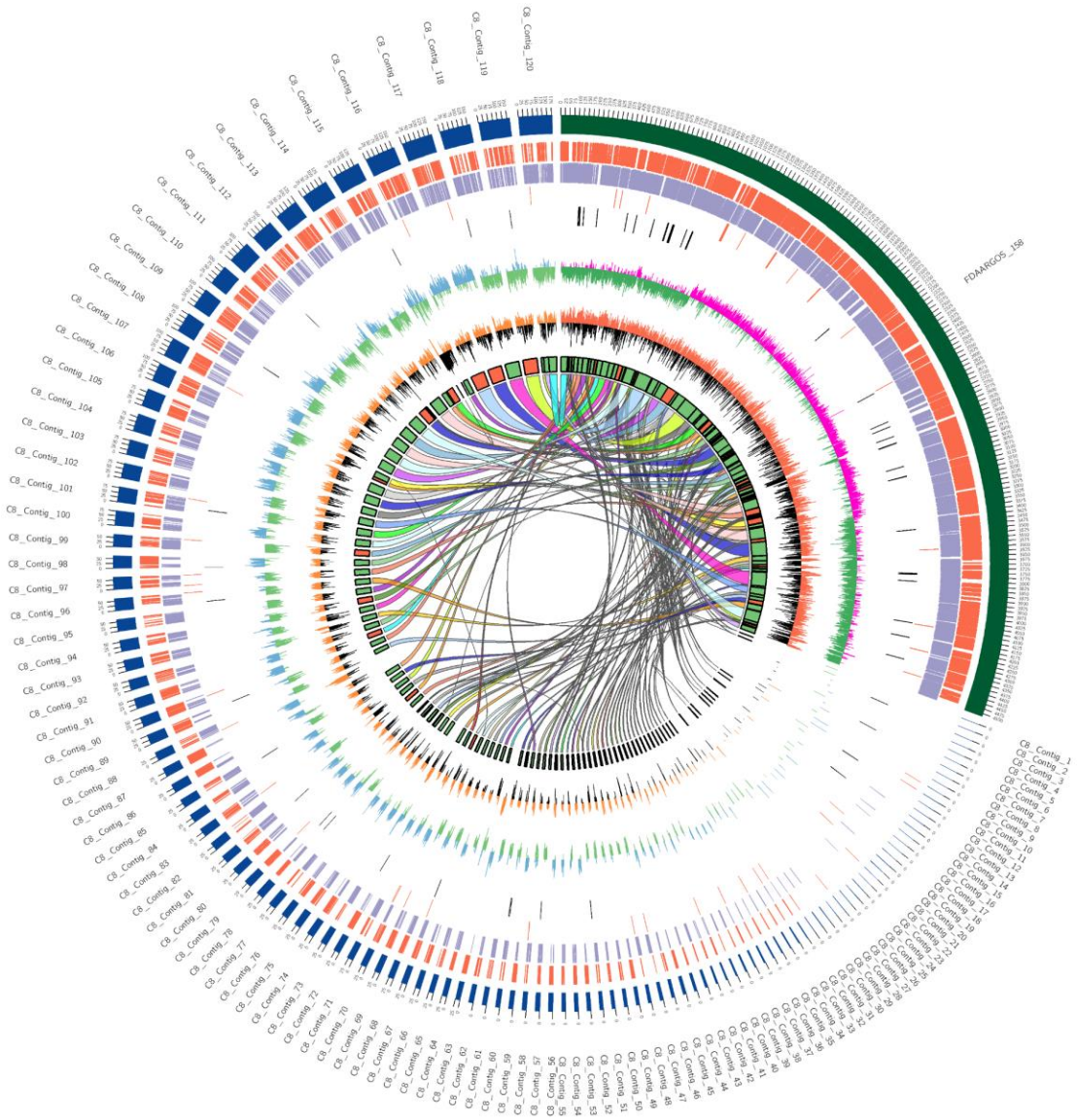


Figure 2.6 Genome comparison between *H. paralvei* strain C8 and *H. paralvei* strain *FDAARGOS_158.* Features in the figure were described in detail in the above 2 paragraphs.

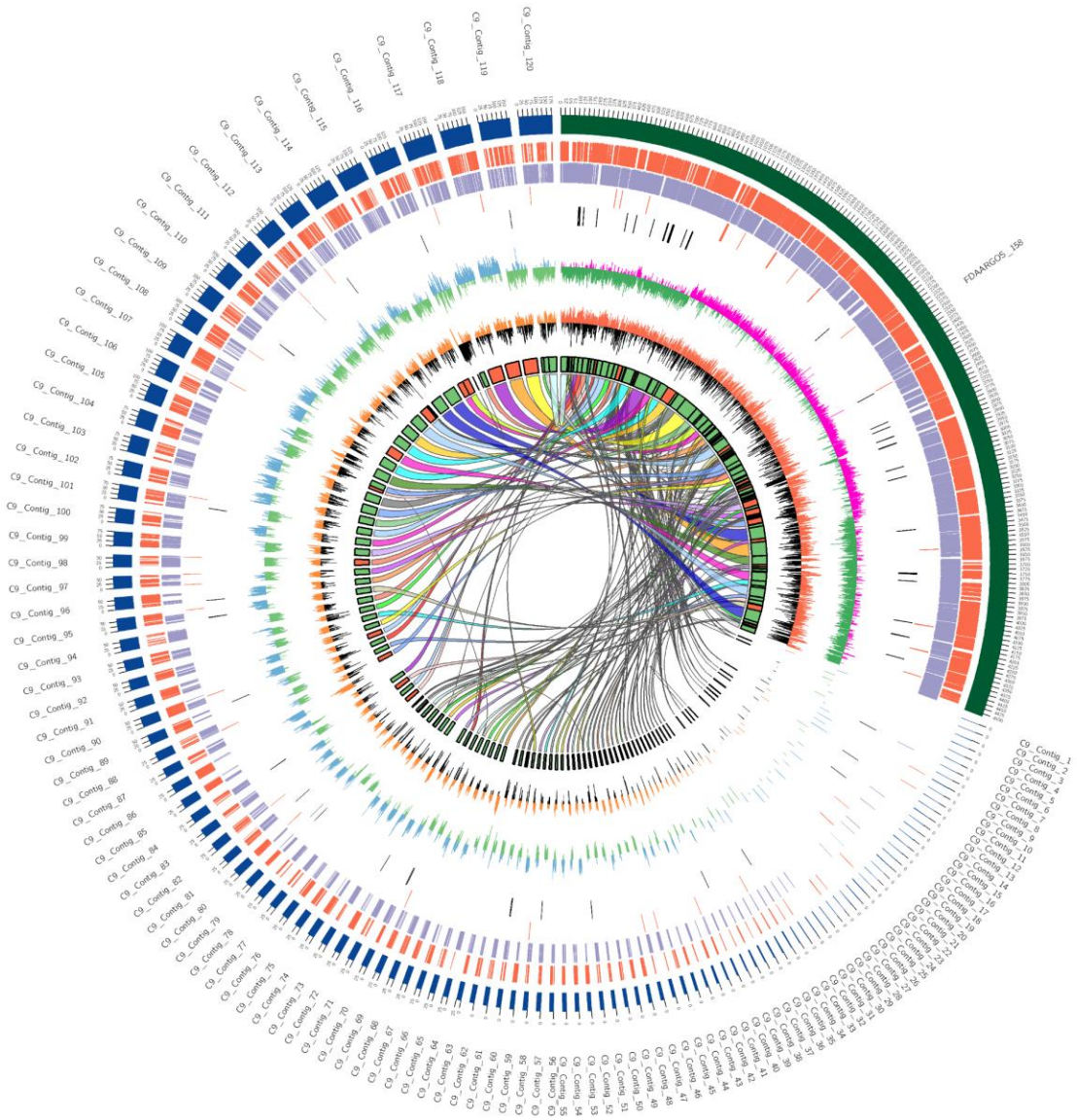


Figure 2.7 Genome comparison between *H. paralvei* strain C9 and *H. paralvei* strain *FDAARGOS_158.* Features in the figure were described in detail in the above 2 paragraphs.

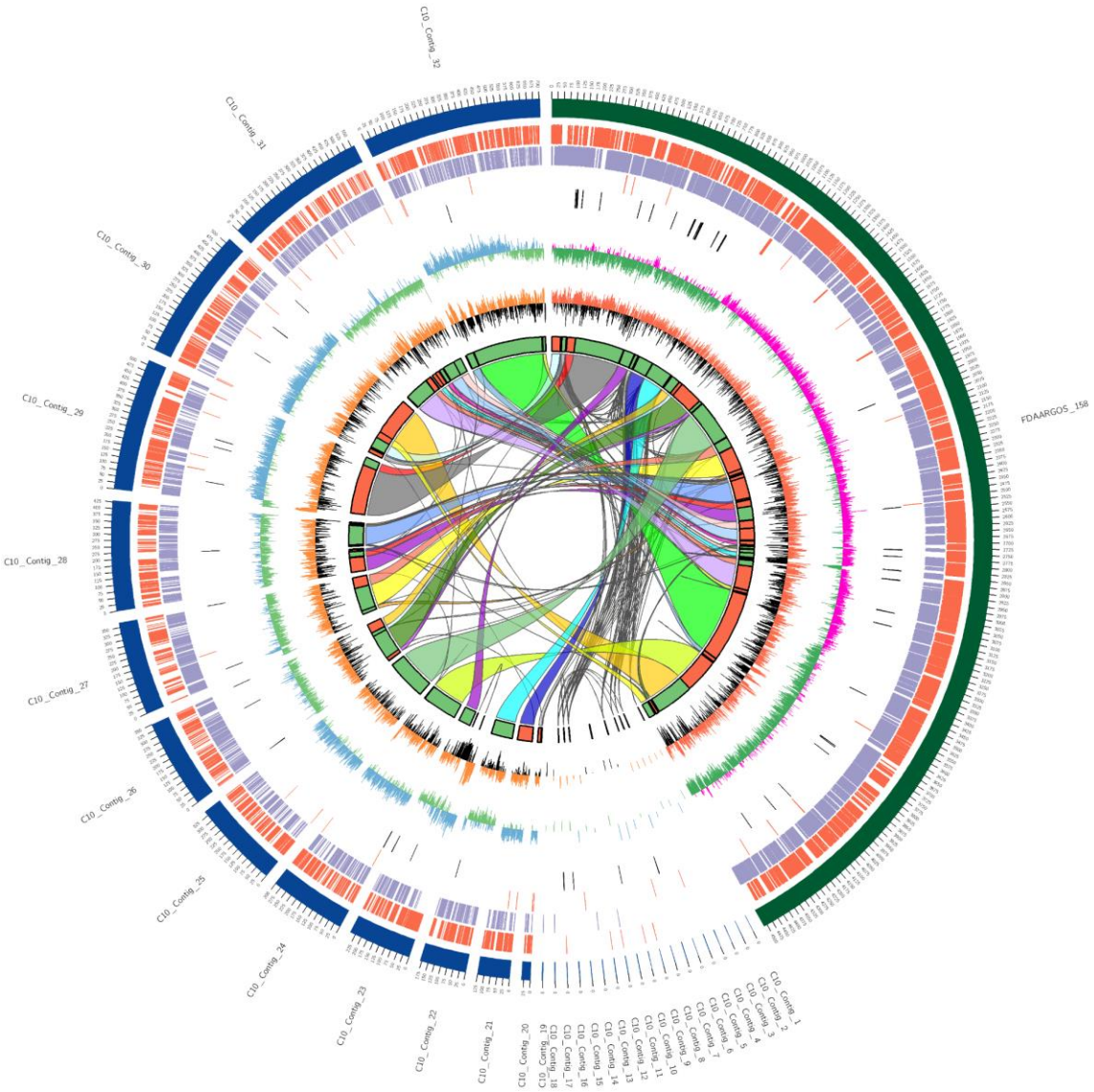


Figure 2.8 Genome comparison between *H. paralvei* strain C10 and *H. paralvei* strain *FDAARGOS_158.* Features in the figure were described in detail in the above 2 paragraphs.

Distribution of sequenced H. paralvei *strains*

In total, 22 genome sequences of *H. paralvei* are available in GenBank, including the 6 genome sequences that I generated and are presented above. A summary of available *H. paralvei* strain properties are shown in Table 2.4. *H. paralvei* strains were found from several geographic locations: Ireland, USA, Canada and Poland. Strains were isolated from different sources, two strains from food (meat), four strains from hospitals, three strains from fish, six strains from plants (strains identified in this research), and seven strains from the Polish Collection of Microorganisms. *Hafnia* are considered potential human pathogens. Four *H. paralvei* strains shown in Table 2.4 were found in human hosts. Particularly, *H. paralvei* strain ATCC 29927 has been listed as a strain to be treated at biosafety level 1. Genome sequences of strains except FDAARGOS_158 remain incomplete. The GC contents of *H. paravlei* genomes are around 48%. The genome sizes range from 4.69 Mb to 5.00 Mb.

Table 2.4 Summary of available genomic features of *H. paralvei* strains

Category	Strain	Isolation source	Geographic location	Genome completeness	Genome size (Mb)	GC content (%)
Food	UBA11299	Cheese	France	Incomplete	4.52	48.1
	CITHA-6	Raw bovine milk	Cork, Ireland	Incomplete	4.94	48.1
	GTA-HAF03	Beef trim broth	Ontario, Canada	Incomplete	5.00	47.9
Clinical material	ATCC 29927	Human	Unknown	Incomplete	4.83	48.1
	FDAARGOS_158	Stool from Children's National Hospital	Washington, D.C, USA	Complete	4.51	48.2
	FDAARGOS_230	Endotracheal aspirate	Washington, D.C, USA	Incomplete	4.91	48
	1133_RAQU	Human	Washington, USA	Incomplete	4.75	48
	PCM_1188	Polish Collection of Microorganisms	Poland	Incomplete	4.81	48.1

Lab culture	PCM_1192	Polish Collection of Microorganisms	Poland	Incomplete	4.85	47.9
	PCM_1194	Polish Collection of Microorganisms	Poland	Incomplete	4.81	48.2
	PCM_1198	Polish Collection of Microorganisms	Poland	Incomplete	4.84	48
	PCM_1211	Polish Collection of Microorganisms	Poland	Incomplete	4.94	47.7
	PCM_1218	Polish Collection of Microorganisms	Poland	Incomplete	4.7	48
	PCM_1223	Polish Collection of Microorganisms	Poland	Incomplete	4.8	48.1
Fish	UBA1836	Epidermal mucus of *Anguilla anguilla* fish	Ebro delta, Spain	Incomplete	4.52	48.2
	UBA4198	Epidermal mucus of *Anguilla anguilla* fish	Ebro delta, Spain	Incomplete	4.57	48
Plant	C2	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.72	48
	C3	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.71	48
	C5	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.69	48.1
	C8	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.7	48.1
	C9	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.7	48.1
	C10	Pitcher fluids of *S. rosea* pitcher plants	Splinter Hill Bog, Alabama, USA	Incomplete	4.71	48.1

a: biosafety level 1
b: uncultivated genomes

Analysis of virulence, pathogenicity and antibiotic resistance genes

Hafnia strains have been found to be associated with animal infections and are potential human pathogens. Several *H. paralvei* strains have been found in human hosts. Therefore, the virulence, pathogenicity and antibiotic resistance genes of *H. paralvei* strains were sought. VirulenceFinder 2.0 (Joensen et al., 2014) was used to predict bacterial virulence factors in *H. paralvei* strains. Virulence factors of *Enterococcus, S. aureus, E. coli*, and *Listeria* were used as database. Strains in these genera and species are human pathogens. However, no significant hit was found when matching genes of all 22 *H. paralvei* strains to the database of virulence factors. The results suggest that *H. paralvei*

strains are not virulent in humans by themselves. Probabilities of strains being human pathogens were estimated by CGE PathogenFinder 1.1 (Cosentino et al., 2013). The probabilities of *H. paralvei* strains being human pathogens ranged from 0.479 to 0.772 (Table 2.5). *H. paralvei* strains from pitcher plants had a lower chance of being human pathogens than most other *H. paralvei* strains (*p*-value <0.05).

Bacterial antibiotic resistances can play a key role in the survival of pathogens in various habitats. Antibiotic resistance genes that act against tetracycline, colistin, phenicol, *β*-lactam antibiotics, sulphonamide, oxazolidinone, fosfomycin, macrolide antibiotics, glycopeptide antibiotics, trimethoprim, quinolone, fusidicacid, nitroimidazole, aminoglycoside antibiotics, and rifampicin were screened across *H. paralvei* genome sequences with ResFinder (Zankari et al., 2012). The results are shown in Table 2.5. *β*-lactam resistance genes were found in all *H. paralvei* strains. The majority of *H. paralvei* strains contain *blaACC-1a* genes, including all six strains from pitcher fluids. Of these, four strains contain *blaACC-1b* genes and two strains contains *blaACC-5* genes. The abundance of microbes that produce *β*-lactam antibiotics (presumably, including in the pitcher microecosystem) to enhance competition with other microbes probably explains why these resistances are so common.

Table 2.5 Probabilities of human pathogenicity and predicted antibiotic resistance genes in *H. paralvei* strains

Strain	Probability	Resistant gene
CITHA-6	0.655	*blaACC-1a*
GTA-HAF03	0.681	*blaACC-1a*
ATCC 29927	0.674	*blaACC-1b*
FDAARGOS_158	0.634	*blaACC-1a*
FDAARGOS_230	0.701	*blaACC-1a*
1133_RAQU	0.631	*blaACC-1a*
PCM_1188	0.668	*blaACC-5*
PCM_1192	0.661	*blaACC-5*
PCM_1194	0.722	*blaACC-1a*
PCM_1198	0.619	*blaACC-1a*
PCM_1211	0.671	*blaACC-1a*
PCM_1218	0.642	*blaACC-1a*
PCM_1223	0.646	*blaACC-1a*
UBA1836	0.687	*blaACC-1b*
UBA4198	0.615	*blaACC-1b*
UBA11299	0.586	*blaACC-1b*
C2	0.521	*blaACC-1a*
C3	0.479	*blaACC-1a*
C5	0.63	*blaACC-1a*
C8	0.638	*blaACC-1a*
C9	0.63	*blaACC-1a*
C10	0.63	*blaACC-1a*

Pan-and-core genome analyses of H. paralvei *strains*

Gene contents of all available *H. paralvei* strains were investigated by pan and core genome analyses to explore the genetic diversity of *H. paralvei*. Genes of all available *H. paralvei* genome sequences were clustered into orthologous groups. The pan-and-core gene frequency is shown in Figure 2.9. In total, 10,446 clusters of orthologs were found

among all strains. Of these, 3044 clusters were found in at least 21 strains. A total of 1849 clusters were found only in 3 to 20 strains, while 5553 clusters (more than half of the total clusters) were found in less than 3 strains. These results indicate that *H. paralvei* has a large pan genome but a relatively small core genome. In other words, the gene content of *H. paralvei* species was found to be diverse. Pan genome gene number increased with the number of genomes analyzed ("genome number") (Figure 2.9 B), while the core genome gene number decreased with the number of genomes analyzed (Figure 2.9 C), as expected.

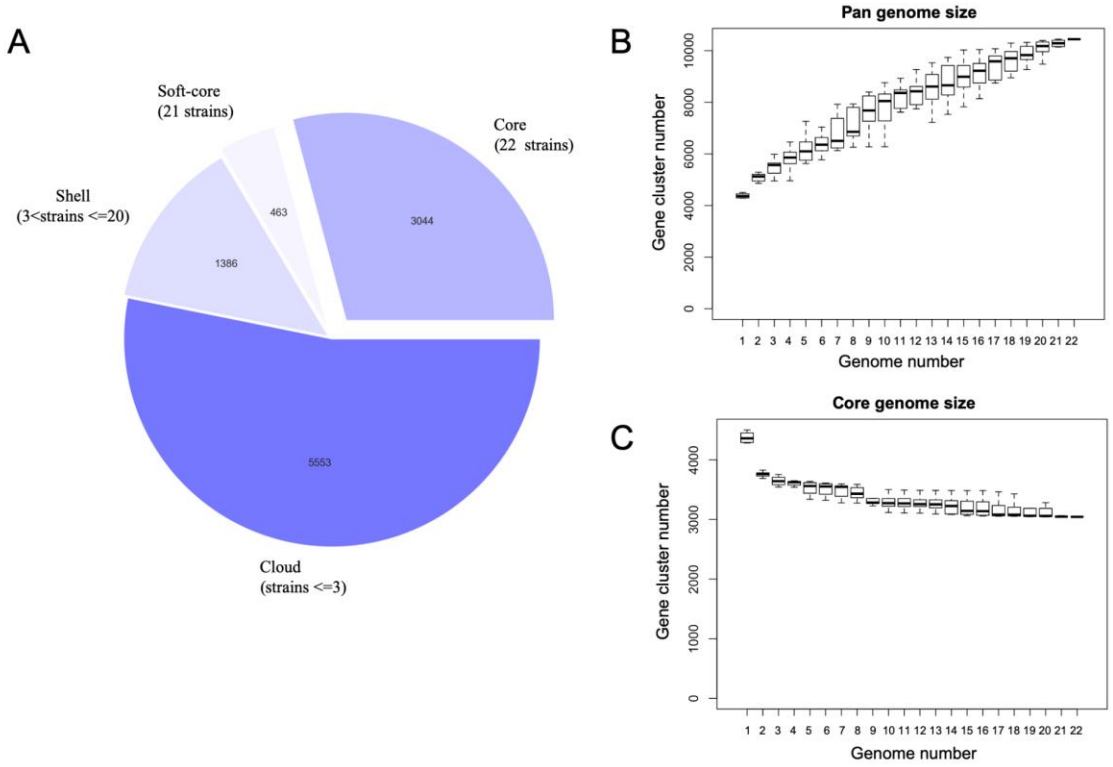


Figure 2.9 Pan-and-core genome analysis of *H. paralvei* strains. A): Pie chart of distributions of gene clusters; B) Side-by-side boxplot of pan genome size vs the number of additional genomes; C) Side-by-side boxplot of core genome size vs the number of additional genomes.

Among all *H. paralvei* strains whose genome sequences are available, only the six genome sequences generated in this project are associated with plants. All other strains are associated with either hospitals or meats. A gene presence-absence matrix of *H. paralvei* strains is shown in Figure 2.10. Gene contents among *H. paralvei* strains are quite variable. The core genome of all *H. paralvei* strains (represented by the solid blue region on the left) occupies less than half of the total gene clusters. The accessory genomes of *H. paralvei* strains were highly diverse. Relatively speaking, however, pitcher *H. paralvei* are highly similar to each other and strongly distinguished from non-pitcher strains in their gene contents. Gene contents of pitcher strains were the most conserved among the total set of strains that I analyzed.

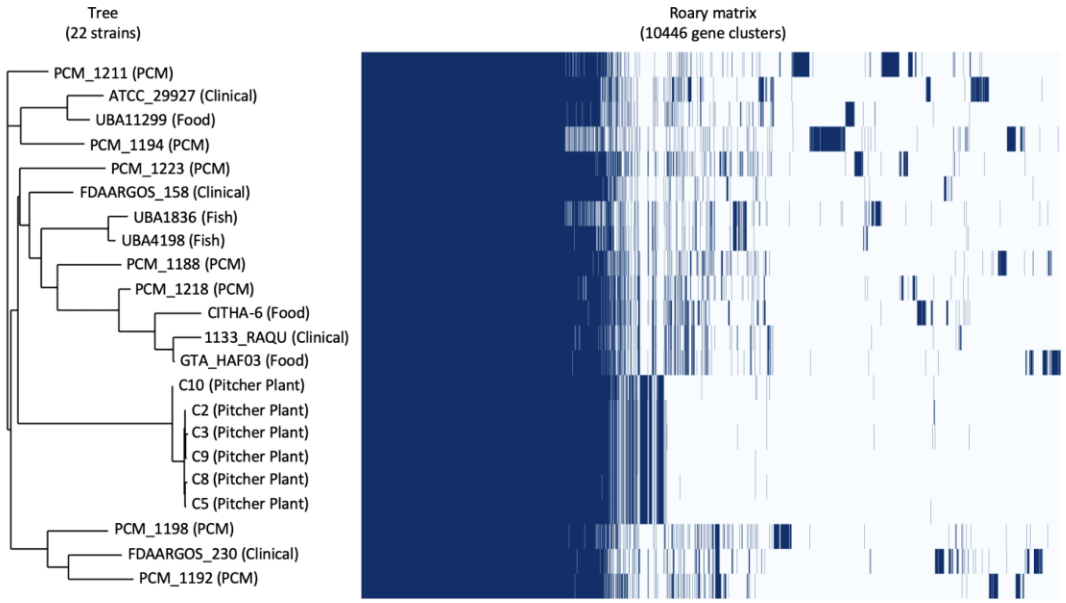


Figure 2.10 Comparison of gene content among 22 available genomes of *H. paralvei*. The presence of a gene cluster is shown in blue. The absence of a gene cluster is shown in white. Isolation sources are shown in parentheses.

Pan-and-core genome comparison between pitcher strains and non-pitcher strains

The pan and core genomes of pitcher strains and non-pitcher strains were compared to explore the genetic diversity of pitcher strains *versus* non-pitcher strains. (Table 2.6). *H. paralvei* strains were divided into two groups. The six strains from pitcher plants were considered as one group and the other 16 strains were considered as another group. The pan, core and accessory genomes of each group were obtained. The pan, core and accessory genomes of the group of pitcher strains consists of 4376, 4228 and 148 gene clusters, respectively. The pan, core and accessory genomes of the group of non-pitcher strains consists of 10,107, 3054 and 7053 gene clusters, respectively. The pan genome of pitcher strains was smaller than non-pitcher strains, the core genome of pitcher strains was bigger than non-pitcher strains, and the accessory genome of pitcher strains was much smaller than non-pitcher strains.

Table 2.6 Number of gene clusters in pan, core and accessory genomes of pitcher strains and non-pitcher strains

	Pitcher strains	Non-pitcher strains
Pan	4376	10107
Core	4228	3054
Accessory	148	7053

Genes in core genomes of pitcher strains and non-pitcher strains were compared first. A Venn diagram depicting the comparison results is shown in Figure 2.11 (Left). The light

blue set represents the core genome of pitcher strains (4228 gene clusters). The other set represents the core genome of non-pitcher strains (3051 gene clusters). A total of 3044 gene clusters were found to be present in both groups. 1184 gene clusters were present in the core genome of pitcher strains but absent in the core genome of non-pitcher strains. Among these 1184 gene clusters, 908 gene clusters were found in the accessory genomes of non-pitcher strains and 276 gene clusters were not present in any non-pitcher strain. Ten gene clusters were absent in the core genome of pitcher strains but present in the core genomes of non-pitcher strains. All 10 of these gene clusters were found in the accessory genome of pitcher strains. In short, core genomes of non-pitcher strains were included in the core genomes of pitcher strains except 10 gene clusters.

Pan genomes of the two groups were compared as well and the results are shown in Figure 2.11 (right). The light blue set represents the pan genome of pitcher strains that contained 4376 gene clusters. The other set represents the pan genome of non-pitcher strains that contained 10,107 gene clusters. Of these, 4037 gene clusters were shared by pitcher strains and non-pitcher strains. A total of 339 gene clusters were present in the pan genome of pitcher strains but absent in the pan genome of non-pitcher strains. Among these 339 gene clusters, 276 gene clusters were found to be present in the core genome of pitcher strains and 63 gene clusters were found to be present in the accessory genome of pitcher strains. There are 6070 gene clusters that are absent from the pan genome of pitcher strains but were found in the pan genome of non-pitcher strains. All of these 6070 gene clusters were found to be present in the accessary genome of non-pitcher strains. In summary, the

pan genome of pitcher strains was included in the pan genome of non-pitcher strains except 339 gene clusters.

The function of genes differentially present in pitcher strains and non-pitcher strains were of more interest. Many differentially present genes were found to be genes that are predicted to encode hypothetical proteins. In the core genome comparison, 13% (398/3044) of genes in the shared core genomes of pitcher and non-pitcher strains were genes of hypothetical proteins; 48% (568/1184) of genes that were present in the core genome of pitcher strains but were absent in the core genome of non-pitcher strains were annotated as hypothetical protein-encoding genes; and 20% (2/10) of the predicted genes that were absent in the core genome of pitcher strains but were present in the core genome of non-pitcher strains were hypothetical proteins genes. The proportions of hypothetical proteins genes in the unshared core genomes, especially the unshared core genome of pitcher strains, were much higher than those in the shared genomes. This difference in proportions of hypothetical proteins genes illustrated the mystery of gene functions within the pitcher environment.

In the pan genome comparison, 19% (773/4037) of genes in the shared core genomes of pitcher and non-pitcher strains were genes of hypothetical proteins. Similar with the comparison of core genomes, the proportions of hypothetical proteins genes in the unshared pan genomes was much higher, 75% (253/339). A total of ~72% (4341/6070) of genes that were not found in the pan genome of pitcher strains but were identified in the pan genome of non-pitcher strains were hypothetical proteins genes. The large number of hypothetical protein genes in the unshared pan/core genomes might play an important role in survival

and adaption of *H. paralvei* strains in the microecosystems of pitchers and other environments.

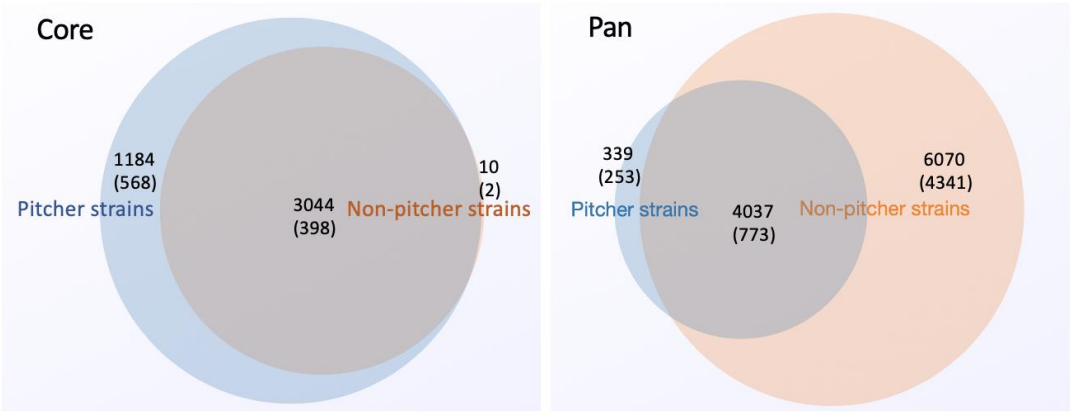


Figure 2.11 Pan-and-core genome comparisons between pitcher strains and non-pitcher strains. Sets of pitcher strains are shown in light blue. Sets of non-pitcher strains are shown in light orange. Shared sets are shown in grey. Number of hypothetical protein genes are shown in parenthesis.

The 339 genes found only in the pan genome of pitcher strains were observed to be involved to 34 pathways (Table 2.7). Eighteen of these 34 pathways were metabolic pathways, including one pathway of energy metabolism, eight pathways of amino acid metabolism, and three pathways of biosynthesis of other secondary metabolites. Three of the 34 pathways were DNA replication and repair pathways. Two of the 34 pathways were environmental information processing pathways: one pathway of membrane transport and one pathway of signal transduction. Five of 34 pathways were cellular processes pathways:

one pathway of cell growth and death, two pathways of cellular community of prokaryotes, and two pathways of cell motility. Three of 34 pathways are organismal systems pathways: one pathway related to bacterial immunity, one aging pathway and one pathway of environmental adaptation. Three of 34 pathways are bacterial infectious disease pathways related to human diseases.

A total of 6070 genes that were absent in the pan genome of pitcher strains were found to be associated with 150 pathways (Table 2.8). First, 96 of 150 pathways were metabolism pathways: nine pathways of global and overview maps (a special KEGG metabolic pathway maps that depicts global and overall pictures of metabolism), 15 pathways of carbohydrate metabolism, six pathways of energy metabolism, seven pathways of lipid metabolism, two pathways of nucleotide metabolism, 20 pathways of amino acid metabolism, five pathways of glycan biosynthesis and metabolism, eight pathways of metabolism of cofactors and vitamins, seven pathways of metabolism of terpenoids and polyketides, five pathways of biosynthesis of other secondary metabolites and 12 pathways of xenobiotic biodegradation and metabolism. Second, 12 of the 150 pathways were genetic information processing pathways: one pathway of transcription, two pathways of translation, four pathways of protein folding, sorting and degradation and five pathways of replication and repair. Third, six of 150 pathways were environmental information processing pathways: three pathways of membrane transport and three pathways of signal transduction. Fourth, 10 of 150 pathways were cellular processes pathways: two pathways of transport and catabolism, two pathways of cell growth and death, four pathways of prokaryotic cellular community, and two pathways of cell motility. Fifth, four of 150

pathways were organismal systems pathways: one pathway of the bacterial immune system, two pathways of aging and one pathway of environmental adaptation. Last, but not least, 23 of 150 pathways were related to human disease pathways: four pathways of cancer overview, one pathway of cancer specific types, one pathway of immune disease, three pathways of neurodegenerative disease, one pathway of cardiovascular disease, two pathways of endocrine and metabolic disease, eight pathways of bacterial infectious disease, one pathway of parasitic infectious disease and three pathways of antimicrobial drug resistance. Most of these pathways were related to the metabolism of various compounds, suggesting the differences in metabolism between pitcher-associated strains and other strains. It's noteworthy that genes that were missing in pitcher-associated genomes were annotated as genes related to human diseases including cancer, immune disease, neurodegenerative disease, cardiovascular disease, endocrine and metabolic disease. Some of these genes were mapped to pathways associated with drug resistance.

Table 2.7 KEGG pathways of genes present only in pan genomes of pitcher strains

Category	Subcategory	KEGG Pathway
Metabolism	Global and overview maps	Metabolic pathways
Metabolism	Global and overview maps	Biosynthesis of secondary metabolites
Metabolism	Global and overview maps	Microbial metabolism in diverse environments
Metabolism	Global and overview maps	Biosynthesis of antibiotics
Metabolism	Global and overview maps	2-Oxocarboxylic acid metabolism
Metabolism	Global and overview maps	Biosynthesis of amino acids
Metabolism	Energy metabolism	Nitrogen metabolism
Metabolism	Amino acid metabolism	Alanine, aspartate and glutamate metabolism
Metabolism	Amino acid metabolism	Cysteine and methionine metabolism

Metabolism	Amino acid metabolism	Arginine biosynthesis
Metabolism	Amino acid metabolism	Arginine and proline metabolism
Metabolism	Amino acid metabolism	Histidine metabolism
Metabolism	Amino acid metabolism	Tyrosine metabolism
Metabolism	Amino acid metabolism	Phenylalanine metabolism
Metabolism	Amino acid metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Isoquinoline alkaloid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Tropane, piperidine and pyridine alkaloid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Novobiocin biosynthesis
Genetic Information Processing	Replication and repair	DNA replication
Genetic Information Processing	Replication and repair	Mismatch repair
Genetic Information Processing	Replication and repair	Homologous recombination
Environmental Information Processing	Membrane transport	Bacterial secretion system
Environmental Information Processing	Signal transduction	Two-component system
Cellular Processes	Cell growth and death	Cell cycle - Caulobacter
Cellular Processes	Cellular community – prokaryotes	Biofilm formation - Vibrio cholerae
Cellular Processes	Cellular community – prokaryotes	Biofilm formation - Pseudomonas aeruginosa
Cellular Processes	Cell motility	Bacterial chemotaxis
Cellular Processes	Cell motility	Flagellar assembly
Organismal Systems	Immune system	NOD-like receptor signaling pathway
Organismal Systems	Aging	Longevity regulating pathway - multiple species
Organismal Systems	Environmental adaptation	Plant-pathogen interaction
Human Diseases	Infectious disease: bacterial	Salmonella infection
Human Diseases	Infectious disease: bacterial	Shigellosis
Human Diseases	Infectious disease: bacterial	Legionellosis

Table 2.8 KEGG pathways of genes absent in pan genomes of pitcher strains

Category	Subcategory	KEGG Pathway
Metabolism	Global and overview maps	Metabolic pathways
Metabolism	Global and overview maps	Biosynthesis of secondary metabolites
Metabolism	Global and overview maps	Microbial metabolism in diverse environments
Metabolism	Global and overview maps	Biosynthesis of antibiotics
Metabolism	Global and overview maps	Carbon metabolism
Metabolism	Global and overview maps	2-Oxocarboxylic acid metabolism
Metabolism	Global and overview maps	Fatty acid metabolism
Metabolism	Global and overview maps	Biosynthesis of amino acids
Metabolism	Global and overview maps	Degradation of aromatic compounds
Metabolism	Carbohydrate metabolism	Glycolysis / Gluconeogenesis
Metabolism	Carbohydrate metabolism	Citrate cycle (TCA cycle)
Metabolism	Carbohydrate metabolism	Pentose phosphate pathway
Metabolism	Carbohydrate metabolism	Pentose and glucuronate interconversions
Metabolism	Carbohydrate metabolism	Fructose and mannose metabolism
Metabolism	Carbohydrate metabolism	Galactose metabolism
Metabolism	Carbohydrate metabolism	Ascorbate and aldarate metabolism
Metabolism	Carbohydrate metabolism	Starch and sucrose metabolism
Metabolism	Carbohydrate metabolism	Amino sugar and nucleotide sugar metabolism
Metabolism	Carbohydrate metabolism	Pyruvate metabolism
Metabolism	Carbohydrate metabolism	Glyoxylate and dicarboxylate metabolism
Metabolism	Carbohydrate metabolism	Propanoate metabolism
Metabolism	Carbohydrate metabolism	Butanoate metabolism
Metabolism	Carbohydrate metabolism	C5-Branched dibasic acid metabolism
Metabolism	Carbohydrate metabolism	Inositol phosphate metabolism

Metabolism	Energy metabolism	Oxidative phosphorylation
Metabolism	Energy metabolism	Carbon fixation in photosynthetic organisms
Metabolism	Energy metabolism	Carbon fixation pathways in prokaryotes
Metabolism	Energy metabolism	Methane metabolism
Metabolism	Energy metabolism	Nitrogen metabolism
Metabolism	Energy metabolism	Sulfur metabolism
Metabolism	Lipid metabolism	Fatty acid biosynthesis
Metabolism	Lipid metabolism	Fatty acid degradation
Metabolism	Lipid metabolism	Steroid biosynthesis
Metabolism	Lipid metabolism	Steroid hormone biosynthesis
Metabolism	Lipid metabolism	Glycerophospholipid metabolism
Metabolism	Lipid metabolism	Ether lipid metabolism
Metabolism	Lipid metabolism	Sphingolipid metabolism
Metabolism	Nucleotide metabolism	Purine metabolism
Metabolism	Nucleotide metabolism	Pyrimidine metabolism
Metabolism	Amino acid metabolism	Alanine, aspartate and glutamate metabolism
Metabolism	Amino acid metabolism	Glycine, serine and threonine metabolism
Metabolism	Amino acid metabolism	Cysteine and methionine metabolism
Metabolism	Amino acid metabolism	Valine, leucine and isoleucine degradation
Metabolism	Amino acid metabolism	Valine, leucine and isoleucine biosynthesis
Metabolism	Amino acid metabolism	Lysine biosynthesis
Metabolism	Amino acid metabolism	Lysine degradation
Metabolism	Amino acid metabolism	Arginine biosynthesis
Metabolism	Amino acid metabolism	Arginine and proline metabolism
Metabolism	Amino acid metabolism	Histidine metabolism
Metabolism	Amino acid metabolism	Tyrosine metabolism
Metabolism	Amino acid metabolism	Phenylalanine metabolism
Metabolism	Amino acid metabolism	Tryptophan metabolism
Metabolism	Amino acid metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis
Metabolism	Metabolism of other amino acids	beta-Alanine metabolism

Metabolism	Metabolism of other amino acids	Taurine and hypotaurine metabolism
Metabolism	Metabolism of other amino acids	Selenocompound metabolism
Metabolism	Metabolism of other amino acids	Cyanoamino acid metabolism
Metabolism	Metabolism of other amino acids	D-Arginine and D-ornithine metabolism
Metabolism	Metabolism of other amino acids	Glutathione metabolism
Metabolism	Glycan biosynthesis and metabolism	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate
Metabolism	Glycan biosynthesis and metabolism	Glycosaminoglycan biosynthesis - heparan sulfate / heparin
Metabolism	Glycan biosynthesis and metabolism	Lipopolysaccharide biosynthesis
Metabolism	Glycan biosynthesis and metabolism	Peptidoglycan biosynthesis
Metabolism	Glycan biosynthesis and metabolism	Other glycan degradation
Metabolism	Metabolism of cofactors and vitamins	Thiamine metabolism
Metabolism	Metabolism of cofactors and vitamins	Nicotinate and nicotinamide metabolism
Metabolism	Metabolism of cofactors and vitamins	Pantothenate and CoA biosynthesis
Metabolism	Metabolism of cofactors and vitamins	Biotin metabolism
Metabolism	Metabolism of cofactors and vitamins	Lipoic acid metabolism
Metabolism	Metabolism of cofactors and vitamins	Folate biosynthesis
Metabolism	Metabolism of cofactors and vitamins	Porphyrin and chlorophyll metabolism
Metabolism	Metabolism of cofactors and vitamins	Ubiquinone and other terpenoid-quinone biosynthesis
Metabolism	Metabolism of terpenoids and polyketides	Terpenoid backbone biosynthesis

Metabolism	Metabolism of terpenoids and polyketides	Sesquiterpenoid and triterpenoid biosynthesis
Metabolism	Metabolism of terpenoids and polyketides	Limonene and pinene degradation
Metabolism	Metabolism of terpenoids and polyketides	Geraniol degradation
Metabolism	Metabolism of terpenoids and polyketides	Biosynthesis of ansamycins
Metabolism	Metabolism of terpenoids and polyketides	Polyketide sugar unit biosynthesis
Metabolism	Metabolism of terpenoids and polyketides	Biosynthesis of siderophore group nonribosomal peptides
Metabolism	Biosynthesis of other secondary metabolites	Tropane, piperidine and pyridine alkaloid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Monobactam biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Streptomycin biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Acarbose and validamycin biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Prodigiosin biosynthesis
Metabolism	Xenobiotics biodegradation and metabolism	Benzoate degradation
Metabolism	Xenobiotics biodegradation and metabolism	Aminobenzoate degradation
Metabolism	Xenobiotics biodegradation and metabolism	Fluorobenzoate degradation
Metabolism	Xenobiotics biodegradation and metabolism	Chloroalkane and chloroalkene degradation
Metabolism	Xenobiotics biodegradation and metabolism	Nitrotoluene degradation
Metabolism	Xenobiotics biodegradation and metabolism	Styrene degradation

Metabolism	Xenobiotics biodegradation and metabolism	Atrazine degradation
Metabolism	Xenobiotics biodegradation and metabolism	Caprolactam degradation
Metabolism	Xenobiotics biodegradation and metabolism	Naphthalene degradation
Metabolism	Xenobiotics biodegradation and metabolism	Metabolism of xenobiotics by cytochrome P450
Metabolism	Xenobiotics biodegradation and metabolism	Drug metabolism - cytochrome P450
Metabolism	Xenobiotics biodegradation and metabolism	Drug metabolism - other enzymes
Genetic Information Processing	Transcription	RNA polymerase
Genetic Information Processing	Translation	Ribosome
Genetic Information Processing	Translation	Aminoacyl-tRNA biosynthesis
Genetic Information Processing	Folding, sorting and degradation	Protein export
Genetic Information Processing	Folding, sorting and degradation	Protein processing in endoplasmic reticulum
Genetic Information Processing	Folding, sorting and degradation	Sulfur relay system
Genetic Information Processing	Folding, sorting and degradation	RNA degradation
Genetic Information Processing	Replication and repair	DNA replication
Genetic Information Processing	Replication and repair	Base excision repair
Genetic Information Processing	Replication and repair	Nucleotide excision repair
Genetic Information Processing	Replication and repair	Mismatch repair

Genetic Information Processing	Replication and repair	Homologous recombination
Environmental Information Processing	Membrane transport	ABC transporters
Environmental Information Processing	Membrane transport	Phosphotransferase system (PTS)
Environmental Information Processing	Membrane transport	Bacterial secretion system
Environmental Information Processing	Signal transduction	Two-component system
Environmental Information Processing	Signal transduction	MAPK signaling pathway - plant
Environmental Information Processing	Signal transduction	HIF-1 signaling pathway
Cellular Processes	Transport and catabolism	Lysosome
Cellular Processes	Transport and catabolism	Peroxisome
Cellular Processes	Cell growth and death	Cell cycle - Caulobacter
Cellular Processes	Cell growth and death	Necroptosis
Cellular Processes	Cellular community - prokaryotes	Quorum sensing
Cellular Processes	Cellular community - prokaryotes	Biofilm formation - Vibrio cholerae
Cellular Processes	Cellular community - prokaryotes	Biofilm formation - Pseudomonas aeruginosa
Cellular Processes	Cellular community - prokaryotes	Biofilm formation - Escherichia coli
Cellular Processes	Cell motility	Bacterial chemotaxis
Cellular Processes	Cell motility	Flagellar assembly
Organismal Systems	Immune system	NOD-like receptor signaling pathway

Organismal Systems	Aging	Longevity regulating pathway - multiple species
Organismal Systems	Environmental adaptation	Plant-pathogen interaction
Human Diseases	Cancer: overview	Pathways in cancer
Human Diseases	Cancer: overview	MicroRNAs in cancer
Human Diseases	Cancer: overview	Chemical carcinogenesis
Human Diseases	Cancer: specific types	Hepatocellular carcinoma
Human Diseases	Immune disease	Primary immunodeficiency
Human Diseases	Neurodegenerative disease	Amyotrophic lateral sclerosis (ALS)
Human Diseases	Neurodegenerative disease	Huntington disease
Human Diseases	Neurodegenerative disease	Prion diseases
Human Diseases	Cardiovascular disease	Fluid shear stress and atherosclerosis
Human Diseases	Endocrine and metabolic disease	Type I diabetes mellitus
Human Diseases	Endocrine and metabolic disease	Insulin resistance
Human Diseases	Infectious disease: bacterial	Epithelial cell signaling in Helicobacter pylori infection
Human Diseases	Infectious disease: bacterial	Salmonella infection
Human Diseases	Infectious disease: bacterial	Shigellosis
Human Diseases	Infectious disease: bacterial	Yersinia infection
Human Diseases	Infectious disease: bacterial	Pertussis
Human Diseases	Infectious disease: bacterial	Legionellosis
Human Diseases	Infectious disease: bacterial	Tuberculosis
Human Diseases	Infectious disease: bacterial	Bacterial invasion of epithelial cells
Human Diseases	Infectious disease: parasitic	Amoebiasis
Human Diseases	Drug resistance: antimicrobial	beta-Lactam resistance
Human Diseases	Drug resistance: antimicrobial	Cationic antimicrobial peptide (CAMP) resistance
Human Diseases	Drug resistance: antineoplastic	Platinum drug resistance

Analysis of genes present only in the core genome of pitcher strains

It's worth mentioning that the predicted functions of 276 gene clusters that are present only in the core genome of pitcher strains (absent in all non-pitcher strains) are particularly interesting. Functions of the core genome genes in pitcher strains suggested the unique and indispensable features of strains inhabiting the pitcher environment. Functions of these 276 gene clusters were analyzed. Of these, 210 genes were found to encode hypothetical proteins. The roles of the remaining 66 gene clusters were analyzed through KEGG. Sixteen KEGG pathways were found in total (Table 2.9): 3 pathways of DNA replication and repair, three pathways of bacterial infectious disease, two pathways of cellular community, one pathway of membrane transport, one signal transduction pathway, one pathway of cell motility, one immune system pathway, one aging pathway and one pathway of environmental adaption.

Table 2.9 Summary of pathways of 66 genes only present in the core genome of pitcher strains

Category	Subcategory	KEGG Pathway
Global and Overview Maps	Metabolism	Metabolic pathways
Glycan Biosynthesis and Metabolism	Metabolism	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Genetic Information Processing	Replication and repair	DNA replication
Genetic Information Processing	Replication and repair	Mismatch repair
Genetic Information Processing	Replication and repair	Homologous recombination

Environmental Information Processing	Membrane transport	Bacterial secretion system
Environmental Information Processing	Signal transduction	Two-component system
Cellular Processes	Cell growth and death	Cell cycle - Caulobacter
Cellular Processes	Cellular community – prokaryotes	Biofilm formation - Vibrio cholerae
Cellular Processes	Cellular community – prokaryotes	Biofilm formation - Pseudomonas aeruginosa
Cellular Processes	Cell motility	Flagellar assembly
Organismal Systems	Immune system	NOD-like receptor signaling pathway
Organismal Systems	Aging	Longevity regulating pathway - multiple species
Organismal Systems	Environmental adaptation	Plant-pathogen interaction
Human Diseases	Infectious disease: bacterial	Salmonella infection
Human Diseases	Infectious disease: bacterial	Shigellosis
Human Diseases	Infectious disease: bacterial	Legionellosis

Several pathways, including flagellar assembly, the type VI secretion system, DNA replication, DNA mismatch repair and homologous recombination were found to show different features between pitcher strains and non-pitcher strains. These pathways could all be impacted by environment factors. Differences of these pathways could well explain the differences between strains from pitcher and non-pitcher environments.

Bacterial flagella are more than mobility organelles. They are involved in bacterial adhesion and invasion of host cells (Haiko & Westerlund-Wikström, 2013). Flagella also play a vital role in bacterial virulence and pathogenicity (Ramos et al., 2004). Flagellar assembly differs among environments. FliC, the flagellin protein, is an essential component

of bacterial flagella. FliC has a conserved section of 22 amino acids at the N-terminus (flg22) that triggers plant defense (Garcia & Hirt, 2014). Pathways of flagellar assembly are shown in Figure 2.12. FliC proteins in pitcher strains were found to be different from those in non-pitcher strains (Figure 2.12, bottom). FliC in pitcher strains were highly conserved in this group, but they were very different from those in non-pitcher strains. Eight types of FliC protein sequences of non-pitcher strains were found in total, present in from 1 to 4 strains of the accessory genome of non-pitcher strains. Multiple amino acid sequence alignments of FliC proteins of *H. paralvei* strains indicated that the N-terminus and C-terminus of flagellin proteins were conserved, but the sequences in the middle were not conserved.

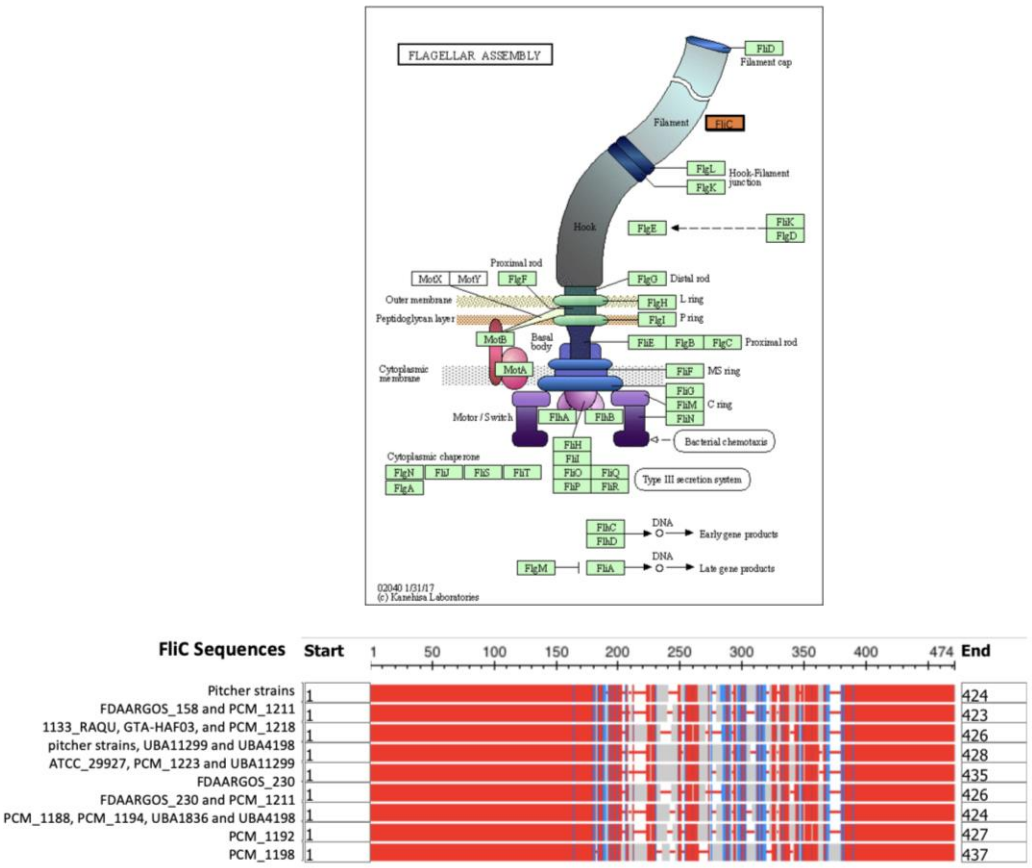


Figure 2.12 Flagellar assembly pathway and multiple sequence alignment of FliC proteins. Top: flagellar assembly pathway map downloaded from KEGG database. Proteins discovered in *H. paralvei* strains are shown in colored blocks. Specifically, proteins that are different between pitcher strains and non-pitcher strains are shown in orange blocks. Bottom: multiple sequence alignment of FliC proteins among *H. paralvei*. Matched region is shown in red, substitution is shown in blue, and insertion/deletion is shown in grey.

The type VI secretion system (T6SS) is one of the bacterial secretion systems that delivers effectors. T6SS genes are present in gram-negative bacteria that exist in polymicrobial environments, and play a key role in bacterial virulence, infection and competition in the environment (Cianfanelli et al., 2016). T6SS contributes to direct cell-to-cell signaling and interaction with eukaryotic hosts. Effectors delivered by T6SS are diverse, depending on the target cells in the environment (Russell et al., 2014). VgrG and Hcp proteins are essential components of T6SS. Multiple types of VgrG and Hcp proteins are found in bacteria. VgrG and Hcp proteins are also found outside of the main T6SS cluster working as effector proteins. VgrG and Hcp proteins are involved in antibacterial action, interactions with cells, colonization and virulence (Zong et al., 2019). The structure of T6SS is depicted in Figure 2.13 (left). Pitcher strains were found to have three types of VgrG protein sequences among which two types were uniquely present in the core genome of pitcher strains. VgrG proteins in pitcher strains were conserved. VgrG protein sequences were not conserved in non-pitcher strains, and 20 types were found. Nineteen of 20 types of VgrG protein sequences were present only in the accessory genome of non-pitcher strains. One of 20 types were present both in the core genome of pitcher strains and four non-pitcher strains. Besides, pitcher strains had four types of Hcp proteins (two types were found in the core genome of non-pitcher strains and two types were found in the accessory genome of pitcher strains). Five types of Hcp proteins were present in the accessory genome of non-pitcher strains. Multiple amino acid sequences alignment of Hcp and VgrG proteins of *H. paralvei* strains (Figure 2.13) indicated the sequence diversity of the two proteins in *H. paralvei*.

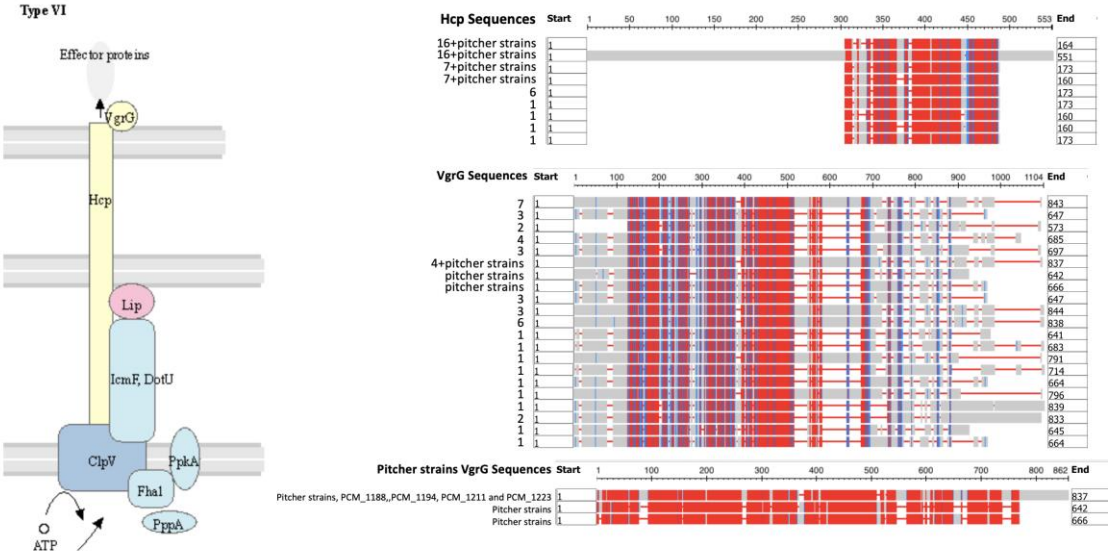


Figure 2.13 The type VI secretion system pathway and multiple sequence alignment of Hcp and VgrG proteins. Left: type VI secretion system pathway map downloaded from the KEGG database. Right: multiple sequence alignment of Hcp (right top) and VgrG proteins (right middle) among *H. paralvei*, and VgrG proteins within pitcher strains (right bottom). The number of non-pitcher strains that contain a protein with this highest homology is shown to the left of the sequence in an alignment. Matched region is shown in red, substitution is shown in blue, and insertion/deletion is shown in grey.

Environmental factors such as oxidizing chemicals can cause DNA damage and can affect DNA replication and repair capacity. During DNA replication, DnaB (a DNA helicase) opens the DNA replication fork and starts DNA synthesis in cooperation with the single-strand DNA-binding protein SSB (Stillman, 1994). SSB also participates in two

other pathways: 1) DNA mismatch repair, which repairs DNA mismatches that occur during DNA replication to prevent DNA mutation (Li, 2008); and 2) homologous recombination, which plays a vital role in repairing double-strand breaks in DNA (Li & Heyer, 2008). DNA replication pathways are shown in Figure 2.14 (top). In total, six types of DnaB genes and nine types of SSB protein genes were found among all *H. paralvei* strains. A specific pair of DnaB and SSB genes were found only in pitcher strains. Multiple amino acid sequence alignment of DnaB and SSB proteins of *H. paralvei* are shown in Figure 2.14 (bottom). DnaB is mostly conserved among *H. paralvei* strains but has some highly variable regions. SSB was found to be conserved at the N-terminal but is diverse towards the C-terminus.

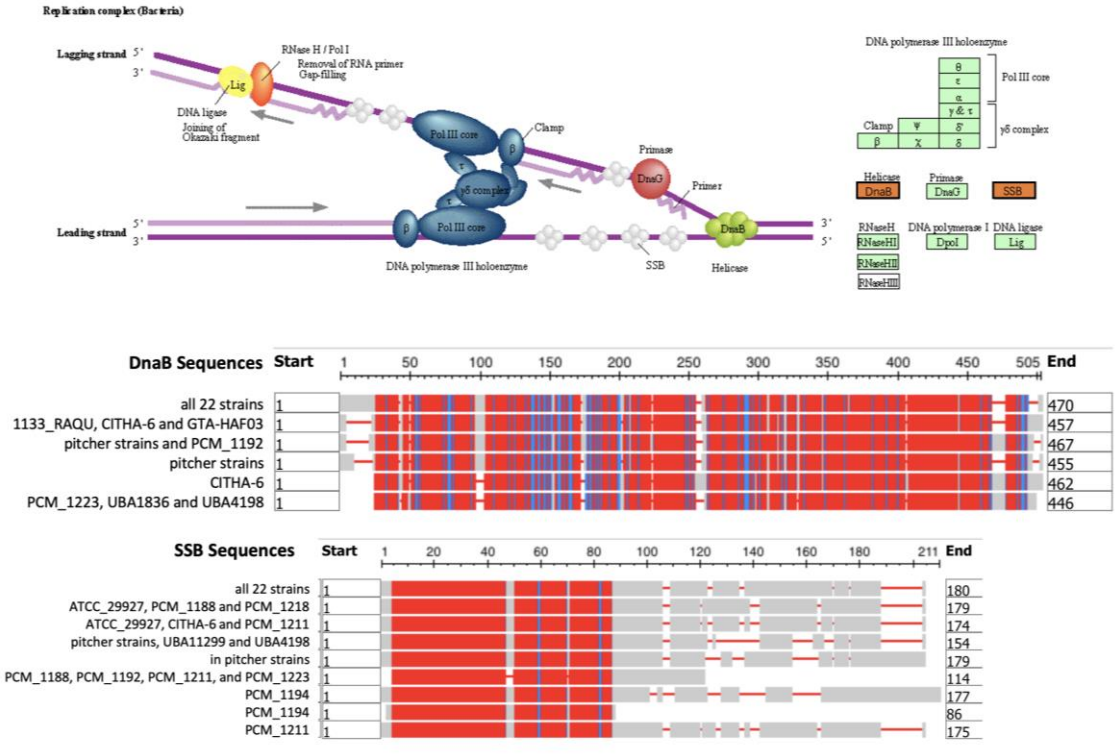


Figure 2.14 DNA replication pathway and multiple sequence alignment of DnaB and SSB proteins. Top: DNA replication pathway map downloaded from the KEGG database. Proteins discovered in *H. paralvei* strains are shown in colored blocks. Specifically, proteins that are different between pitcher strains and non-pitcher strains are shown in orange blocks. Middle: multiple sequence alignment of DnaB among *H. paralvei*. Bottom: multiple sequence alignment of SSB proteins among *H. paralvei*. Matched region is shown in red, substitution is shown in blue, and insertion/deletion is shown in grey.

Analysis of the accessory genome of pitcher strains

The above results illustrated that the predicted gene contents of the *H. paralvei* from pitcher plants were similar but not exactly the same. Pan, core and accessory genomes of pitcher strains that are described in Table 2.6 were used to explore gene content diversity. Genes present only in some of the pitcher strains demonstrate the diversities of *H. paralvei* strains in pitcher plant fluids, suggesting differences in *H. paralvei* strain functions/roles in the pitcher microecosystems.

Genes present only in some of the strains (accessory genes) were analyzed. Among 148 gene clusters of the accessory genome of pitcher strains, 60 gene clusters were found to encode hypothetical proteins, hence yielding no functional insights. KEGG pathways with function-associated accessory genes are shown in Table 2.10. In total, 12 KEGG pathways (nine for metabolism, two for environmental information processing and one

organismal system) were found. These metabolism pathways were: four for global and overview maps, two for energy metabolism, one for nucleotide metabolism and two for metabolism of cofactors and vitamins. In addition, one pathway of membrane transport, one pathway of signal transduction and one pathway of environmental adaptation were found to be encoded in the accessory genome.

Table 2.10 KEGG pathways of accessary genomes of pitcher strains

Category	Subcategory	KEGG Pathway
Metabolism	Global and overview maps	Metabolic pathways (11)
Metabolism	Global and overview maps	Biosynthesis of secondary metabolites (4)
Metabolism	Global and overview maps	Microbial metabolism in diverse environments (3)
Metabolism	Global and overview maps	Biosynthesis of antibiotics (3)
Metabolism	Energy metabolism	Nitrogen metabolism (3)
Metabolism	Energy metabolism	Sulfur metabolism (2)
Metabolism	Nucleotide metabolism	Purine metabolism (3)
Metabolism	Metabolism of cofactors and vitamins	One carbon pool by folate (1)
Metabolism	Metabolism of cofactors and vitamins	Ubiquinone and other terpenoid-quinone biosynthesis (1)
Environmental Information Processing	Membrane transport	Phosphotransferase system (PTS) (2)
Environmental Information Processing	Signal transduction	Two-component system (2)
Organismal Systems	Environmental adaptation	Plant-pathogen interaction (1)

Genes and pathways involved with the pitcher-specific environment such as nitrogen metabolism, nutrient condition and plant defense were of high interest because of the importance of the pitcher for the mobilization of nutrients from the insect prey to the plant host. Some of these pathways and genes are analyzed below. The 148 gene clusters were divided into 3 groups based on their presence and absence in non-pitcher strains. Ten gene clusters were present in the non-pitcher core genome, 75 gene clusters were present in the

non-pitcher accessory genome, and 63 gene clusters were not present in non-pitcher strains. Involved pathways and genes are described below.

Bacterial purine synthesis starts with converting phosphoribosyl-pyrophosphate (PRPP) to inosine monophosphate (IMP) by multiple steps (Figure 2.15). The third step in this pathway converts 5′-phosphoribosyl-glycinamide (GAR) to formyl-phosphoribosyl-glycinamide (fGAR), and is essential for the production of purines (Zhang et al., 2008). If this third step is blocked, GAR accumulates and downstream products are depleted (Nygaard & Smith, 1993). Two phosphoribosylglycinamide formyltransferase, PurN and PurT, are able to function in this third step. PurN exists in both eukaryotes and prokaryotes, but PurT has been found only in prokaryotes (Sampei et al., 2013; Zhang et al., 2008). PurN and PurT are not isoenzymes and use different formyl donors. PurN takes the formyl group from formyl tetrahydrofolate (fTHF) but PurT obtains the formyl group from formate (Nygaard & Smith, 1993). Formate may be enriched in the pitcher of *S. rosea* because ants are well known sources of formic acid. A predicted PurN gene was found in all pitcher strains, but PurT was found only in strains C2, C5, C8, C9 and C10. Strain C3 is thus predicted to not be able to use formate as formal donor to synthesize fGAR from GAR.

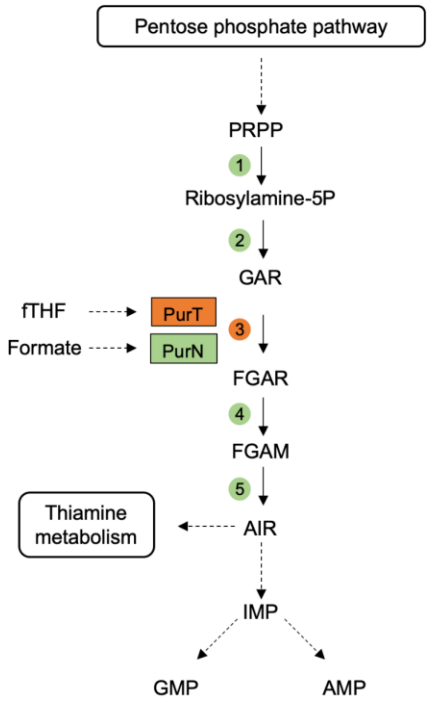


Figure 2.15 Pathway of part of purine metabolism. Proteins that were different among pitcher strains are shown in orange blocks.

A phosphotransferase system (PTS) is a major mechanism for bacteria to uptake and transport carbohydrates (Kotrba et al., 2001). BglF is the *β*-glucoside-specific EIIBCA component that transform *β*-glucoside to phospho-*β*-glucoside in PTS. BglF genes were found in strains C2, C3, C8, C9 and C10 (not in strain C5). An absence of BglF should affect the uptake of *β*-glucosides by strain C5.

Aerobic bacteria can produce ATP by respiration. Aerobic respiration usually requires oxygen as the terminal electron acceptor. Some facultative anaerobes like *H. paralvei* are able to perform anaerobic respiration using sulfur-related compounds such as DMSO as the reduced acceptor at the last step of the respiratory electron transport chain, rather than

oxygen (Zannoni, 1995). DMSO reductase has only been discovered in bacteria and archaea (kappler & schäfer, 2014). Two types of anaerobic dimethyl sulfoxide reductase (DmsB) were found among pitcher strains. A short DmsB protein existed in all pitcher strains. A long DmsB protein was found in strains C2, C5, C8 and C10 but was not found in strains C3 and C9. Pairwise alignment of DmsB amino acid sequences illustrated the variability in DmsB sequences (Figure 2.16). Different DmsB genes among pitcher *H. paralvei* strains suggest diversity in anaerobic respiration and energy generation within the pitcher environment.

Figure 2.16 Multiple amino acid sequence alignment of DmsB proteins within pitcher *H. paralvei* strains. Matched region is shown in red, substitution is shown in blue, and insertion/deletion is shown in grey.

The production of bacterial guanosine 5′-triphosphate 3′-diphosphate (pppGpp) increases under some types of stress such as amino acid shortages, resulting in the inhibition of RNA synthesis and translation. pppGpp can be synthesized by multiple enzymes, including the GTP pyrophosphokinase YjbM that converts GTP to pppGpp (Condon et al., 1995). All pitcher strains except strain C9 were predicted to contain YjbM. Strain C9 is thus speculated to be more vulnerable to amino acid shortages than other *H.*

paralvei strains in pitcher plants, although it seems likely that the pitcher environment usually contains a high level of amino acids because of prey decomposition.

Nitrogen metabolism (Figure 2.17) plays an important role in the pitcher microecosystem. A hydroxylamine reductase (hcp) that participate in nitrogen metabolism was found to be diverse among pitcher strains. Hcp not only converts nitrite to nitrate but also converts hydroxylamine to ammonia. One type of hcp was found in strain C2 but not in other pitcher strains or non-pitcher strains. The C2-specific hcp protein suggests different properties of nitrogen metabolism in strain C2.

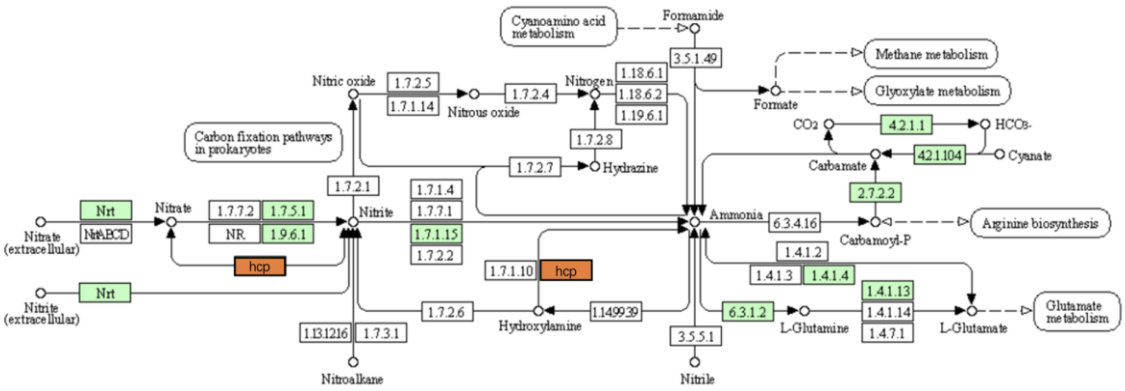


Figure 2.17 Pathway map of nitrogen metabolism downloaded from the KEGG database. Proteins that were discovered in pitcher *H. paralvei* strains are shown in color-filled blocks. Specifically, proteins that were different among pitcher strains are shown in orange blocks.

The gene *tufB*, which encodes the elongation factor Tu 2 protein, affects protein biosynthesis, cell growth and bacterial response to scarce nutrient condition. The protein encoded by *tufB* also serves as a microbe associated molecular pattern (MAMP) that can be recognized by plant extracellular pattern recognition receptors (PRRs) to trigger a PTI response. A unique TufB protein was found in strain C10, which suggests a unique characteristic of strain C10 in adaption to the pitcher environment.

Discussion and Conclusions

Pitchers create a unique environment for microbes, inquilines and visitors. These visitors include female mosquitoes that lay their eggs in the pitcher fluids, thereby generating the mosquito larvae that are often the top predator in the pitcher (Bradshaw, 1983). Various interactions occur in the microecosystems of these pitchers that consist of a variety of microorganisms: archaea, bacteria, fungi, protozoa and their viruses. Although six bacterial isolates from pitcher fluids of *S. rosea* were identified as *H. paralvei* by genomic sequences, other bacteria were identified at the same time. For example, two genomic sequencing data sets of identified bacteria were assembled and published (Appendices A and B). One strain was identified as *Serratia marcescens* (Zhang et al., 2020a) and the other sequence-analyzed strain was identified as a new species in the genus *Enterobacter* (Zhang et al., 2020b). Full 16S amplicon analysis of pitcher bacteria demonstrated that *Hafnia* were very rare in the pitchers (Chaluvadi and Bennetzen, unpub. res.), indicating that the culture conditions employed were highly preferential for *H. paralvei* growth. Moreover, the preferential selection of colonies with archaea-positive

PCR results from these plates suggests that *H. paralvei* and at least some archaea species are physically associated in pitchers. The inability to retain the archaea during subsequent culturing steps after the initial isolation was disappointing, but not surprising, because the most abundant types of archaea in pitcher plants have not ever been successfully cultured.

For any culturing strategy, sampling methods (e.g., culture media type) will partly determine which microbes are present (Prakash et al., 2013). Pitcher microecosystems from the same pitcher plant species have been shown to be similar in any given season (Chaluvadi and Bennetzen, unpub. res.), but each pitcher is a single unique microecosystem. Sampling dates have dramatic effects on the detected microbial composition inside pitchers. The *H. paralvei* strains identified in this study were isolated from pitcher fluids of *S. rosea* in spring. Sampling location (Alabama versus Florida) was found to have less effect on the composition of microbial communities than sampling time points and methods (Chaluvadi and Bennetzen, unpub. res.). In addition, the collected fluids from pitchers were pooled, and strains were isolated from the pooled fluids. Therefore, the strains may come from multiple pitchers or just a single pitcher. A discovery of strains from multiple pitcher indicates the extensive presence of strains and possesses more biological meaning. Separated fluids from individual pitchers, other than pooled fluids, can be used to test the extensive presence of *H. paralvei*.

Although sequencing of 16S rRNA gene amplicons has been widely used in metagenomic studies to determine taxonomic information of the bacteria in various environments, whole genome shotgun sequencing methods were applied in this research, but focused on only a few isolated microbes. On one hand, the taxonomic resolution

achieved by whole genome shotgun sequencing is higher than amplicon sequencing. The taxonomic resolution that amplicon sequencing achieved is usually at the genus level. In contrast, taxonomic information can be determined to the species level by whole genome shotgun sequencing of a purified genome. In addition, whole genome shotgun sequencing of this type covers the entire genome, so information about genome variability and potential microbial function can be acquired. Analysis of gene contents and metabolic pathways present in an environment is also possible with genomic data. The genome assemblies that I achieved were incomplete, only resolving at the level of contiguous sequences (contigs) of a few kb in most cases. This is partly because the Illumina sequencing strategy, chosen for its low cost, yields short raw reads. With the development of more advanced genome sequence technologies, the so-called Third Generation Sequencing, such techniques as PacBio or nanopore are able to provide a complete bacterial genome as one big contig (Lee et al., 2016). Third Generation Sequencing should be considered to sequence single colonies isolated from pitcher fluids for a complete genome assembly in future studies.

The presence or absence of genes in pitcher *H. paralvei* strains compared to other *H. paralvei* strains was determined by comparative genomic analysis in this study. These different gene compositions suggest the unique characteristics of *H. paralvei* lifestyles in the pitcher environment. Genes present only in pitcher strains, especially the genes encoding hypothetical proteins, are newly discovered genes that can enrich our understanding of bacterial genome diversity and inform our understanding of the gene pool of *H. paralvei*. Identified hypothetical protein genes that exist only in the group of pitcher

strains provide guidance for future studies of pitcher microbe functional diversity. Understanding the function of hypothetical proteins should help explain bacterial adaption to the polymicrobial environment of the pitcher. Genes unique to pitcher strains were found to be associated pathways of flagellar assembly, the type VI secretion system, DNA replication, DNA mismatch repair and homologous recombination. Differences in these pathways can help explain the functional differences between *H. paralvei* strains from pitcher and non-pitcher environments, but experiments were not performed to test the many hypotheses that such comparisons uncover. However, in this study, the comparative analysis of gene functions was undertaken at the gene sequence level only. Transcript levels and protein levels of gene products cannot be demonstrated by genome sequence data. Incorporation of transcriptomic and metabolic data will be helpful to understand the differences in gene functions and pathways among environments and microbial strains. Moreover, the presence or absence of genes associated with KEGG pathways were confirmed by scanning the preprocessed reads because of the incompleteness of genome assemblies. The presence or absence of genes can be further verified by PCR.

The diversity of gene content between different pitcher *H. paralvei* strains was demonstrated by analysis of the accessory genomes of pitcher isolates. The gene diversities imply different roles of *H. paralvei* in the pitcher. Genes involved in such processes as nitrogen metabolism and anaerobic respiration were found to be diverse between pitcher *H. paralvei* strains, a very interesting result considering the expected differences in these processes within the different environments in a single pitcher (e.g., more aerobic and less nutrient rich at the surface of the pitcher fluids, but more anaerobic and nutrient-rich at the

bottom of the pitcher because of prey precipitation and decomposition). For instance, C3 and C9 may play different roles from other isolates because they have enhanced abilities of reducing DMSO which is a terminal electron acceptor in anaerobic respiration. It's possible that C3 and C9 are present at the bottom of the pitcher (or at least deeper than others) where more anaerobic respiration is required.

In conclusion, six bacterial isolates from the pitcher fluids of *S. rosea* collected in Alabama were identified as *H. paralvei* by analysis of 16S rRNA genes and average nucleotide identities (ANI). The genomes of these six isolates were sequenced by Illumina MiSeq technology and were used to generate six partial genome assemblies. Each genome assembly was compared to a complete reference genome of *H. paralvei*. Both pan and core genomes in *H. paralvei* strains from pitcher and non-pitcher environments were compared. Distinct features of pitcher *H. paralvei* strains were found by exploring gene differences that could be associated with specific biological pathways in the KEGG classification system. A total of 276 gene clusters that were present only in the core genome of pitcher strains were discovered, including 210 gene clusters of hypothetical protein genes. Predicted genes unique to pitcher strains were found to be associated with 16 KEGG pathways. Pathways of flagellar assembly, the type VI secretion system, DNA replication, DNA mismatch repair and homologous recombination were found to show different features between pitcher strains and nonpitcher strains, primarily at the level of different orthologous genes present in different *H. paralvei* strains. In addition, the diversity of gene content among different pitcher *H. paralvei* was demonstrated. A total a 148 gene clusters were found to be present in some but not all pitcher isolates. These genes are involved in

such processes as nitrogen metabolism, purine synthesis, and anaerobic respiration. This study provides insights for future studies of microbial communities within pitchers, and information regarding how particular pitcher bacteria may preferentially associate with some archaea. Overall, the first discovery and description of *H. paralvei* associated with plants extends the possible roles of *H. paralvei* in the biosphere.

CHAPTER 3

GENOMIC ANALYSIS OF MICROBES ASSOCIATED WITH *PUCCINIA EMACULATA* SPORES

Switchgrass is a bioenergy crop that can be cultivated on marginal lands. *Puccinia emaculata* is the predominant fungal pathogen of switchgrass in the US. Rust disease on switchgrass leaves caused by *P. emaculata* severely affects biomass production. *P. emaculata* urediniospores infect host leaves via penetration with an appressorium structure. I discovered that numerous microbial species were associated with fungal spores. However, the microbial communities associated with *P. emaculata* spores have not been studied previously. This project hypothesized that bacteria associated with urediniospores of *P. emaculata* can be transferred into host plant leaves via urediniospores infection. To test this hypothesis, I inoculated fresh urediniospores of *P. emaculata* on the leaves of switchgrass seedlings that grew in sterile condition and had been treated to remove all endogenous microbes. Uninoculated plants were used as controls. Colony assay plate screening of homogenized leaf tissue indicated many more colonies on inoculated compared to uninoculated plants. The microbial communities inside inoculated and uninoculated leaves were analyzed by shotgun sequencing. Bacteria from different genera

were found to be transferred into plant leaves during the infection of *P. emaculata* urediniospores, but bacteria abundantly associate the spores were not transferred.

Introduction

Panicum virgatum, commonly known as switchgrass, is a C4 perennial plant that can be cultivated on marginal lands. It was selected as a "model" herbaceous bioenergy crop by the Department of Energy (DOE) in 1991 (Wright *et al*., 1992). *Puccinia emaculata* is the predominant fungal pathogen of switchgrass in the US. The life cycle of *P. emaculata* has not been fully described. Switchgrass is known to be the primary host for asexual reproduction. For other studied *Puccinia*, an alternate host species of plant is required to complete the sexual life cycle, but no alternate host has yet been verified. It has been proposed that some species of spurge (*Euphorbia*) may be the alternate host (Demers et al., 2017), but this has not been confirmed by successful infection in the lab. Rust disease is caused by urediniospores infecting switchgrass plants repeatedly, after each generation (one urediniospore generation requiring ~ 1.5 weeks) during the summer. Urediniospores of *P. emaculata* have been reported to infect switchgrass fields in many states, especially Georgia, Tennessee and Arkansas (Zale *et al*., 2008; Hirsch *et al*., 2010).

Urediniospores are dikaryotic, containing two haplontic nuclei that can be highly polymorphic (Gill *et al*., 2015). Infection of *P. emaculata* on host plants begins by the germination of urediniospores on the surface of plant leaves under a favorable environment such as high humidity. Germ tubes are formed when urediniospores germinate. Appressoria

form when germ tubes contact stomates (Szabo and Bushnell, 2001). The appressorium forms a penetration peg to punch a hole into the leaf through a stomate. The fungus then penetrates a plant cell wall, and degrades its own cell wall, so that there is extensive connection (but no penetration) of the two organisms' plasma membranes. Then, other infection structures such as substomatal vesicles, infection hyphae and haustorial mother cells develop inside the leaf (Hoch *et al*., 1987). The haustorium, a special organ, is formed in the space between plant cells. Haustoria play a key role in the uptake of nutrients from plants, suppression of host defense responses, and manipulation of host metabolism (Voegele and Mendgen, 2003; Voegele *et al*., 2001).

Fungal spore-associated microbes are not widely discovered. Microbes have been reported to be associated with fungal spores only in a few studies. For instance, a variety of bacterial communities have been found on spores of several arbuscular mycorrhizal fungi that undertake symbiotic interactions with most plant species (Agnolucci et al., 2015). Bacteria associated with spores affect colonization of arbuscular mycorrhizal fungi and plant growth in potato (Bharadwaj et al., 2008). I observed, in my lab's previous genomic sequence analysis of *P. emaculata*, that diverse bacteria and fungi were associated with urediniospores from generation to generation (Zhang, Chaluvadi, Smith and Bennetzen, unpub. res.). The microbial communities associated with spores of *P. emaculata* have never been studied. The roles of microbes associated with urediniospores remains unknown. None of bacteria has been found to penetrate and transfer into plant leaf tissues. Whether the associated microbes can be transferred into plant tissues during the urediniospore infection stage, and how associated microbes affect urediniospores infection

and plants are not clear. Various interactions can occur between fungi and other adjacent microbes. These interactions can be positive, negative or neutral (Boddy et al., 2016). Here, we are interested in the composition of spore-associated microbial communities surrounding and inside urediniospores of *P. emaculata*. It is hypothesized that microbes associated with urediniospores of *P. emaculata* can be transferred into host plant leaves via urediniospores infection . This hypothesis was tested by comparing the microbial communities of switchgrass leaves inoculated and uninoculated with *P. emaculata* spores.

Methods

Materials

Switchgrass lowland variety Alamo was used in this project. *P. emaculata* urediniospores were provided by Dr. S. Smith's lab in the Department of Plant Pathology at the University of Georgia (UGA). The *P. emaculata* spores utilized were isolated at UGA's Iron Horse Farm, and had been single-pustule purified. Urediniospores were harvested from Alamo plants grown on potting mix in a growth chamber or greenhouse, and were stored at -80 °C before use.

Sterile seeds and plant preparation

Alamo seed were placed into 10 mL centrifuge tubes. Eight mL of 10% sodium hypochlorite and 2 droplets of 0.01% Tween-20 were added into each tube. Tubes were put on a shaker and incubated gently at 150 rpm for 20 minutes. Supernatants were removed and seeds were washed three times with sterile water. Tubes were stored at 4 °C overnight. Sand was poured into magenta boxes (the depth of sand was about 2.5 cm). Then, the

magenta boxes with sand were autoclaved at 121°C for 30 minutes. Twenty mL of 0.5x MS media were poured into each magenta box. Then, 10 sterile seed were placed on the surface of the sand. Magenta boxes were incubated at 28°C with a 16/8-hour (light/dark) shift.

Determination of appropriate biocide concentration

Plant Preservative Mixture (PPM) (Plant Cell Technology, Inc) is a biocide that has been used to prevent or reduce microbial contamination in plant tissue culture. PPM was used in an attempt to achieve sterile seedlings in this study. Concentrations of 0%, 0.1%, 0.2% and 0.3% (w/v) of PPM were added into 0.5x MS media. Twenty mL of 0.5x MS media were poured into the magenta boxes with 1-inch of sand. Seed were placed on the surface of the sand. The magenta boxes were incubated at 28°C. The growth of plants was monitored every week. Four weeks later, plant leaf samples were collected into 2mL Synergy homogenization tubes. Two beads and 500 µL of PBS buffer were added into each tube. Then, the tubes were securely fastened between adapter plates and placed in a tissuelyzer (Qiagen). The leaf samples were homogenized at high speed (20 shakes/minute) for 3 minutes. The tubes were centrifuged at 12,000 rpm for 1 minute. Then, 50 µL of supernatant from each tube was spread on a plate of 0.5 X Difco™ Nutrient Broth media. Plates were incubated at room temperature for 24 hours. Colonies on plates were counted. Three repetitions were used for each PPM concentrations.

Spore rehydration

Tubes containing urediniospores stored in a -80°C freezer were placed in a water bath at 55°C for 5-10 minutes for a quick heat shock. Then, spores were placed on top of a weighing paper. The weighing paper was placed in a weighing boat, and another weighing boat was placed on top to cover the spores. The two weighing boats were sealed with tape. Some holes were made in the top weighing boat to let in the moisture to hydrate the spores. The boat was floated in a water bath and incubated at room temperature for ~18 hours.

Collection of fresh P. emaculata *urediniospores*

Fresh urediniospores were collected by infecting 4-week old Alamo seedlings with rehydrated urediniospores. Five mg of rehydrated urediniospores were mixed with 0.01g talc powder in a 1.5 mL Eppendorf tube. A small paint brush was used to pick up some mixture powder that was used to brush the leaves. A total of 2-3 leaves per magenta box were brushed. The plant boxes were covered with black bags, and distilled water was sprayed inside the bags to generate more moisture. Boxes were placed in the dark and left at room temperature for at least 18 hours. An inverted magenta box was placed on top of the box holding the seedlings to provide enough vertical space for seedling growth. The two magenta boxes were sealed with parafilm. Boxes were placed in the incubator at 28°C, 16/8-hour (light/dark) shift.

Urediniospores were ready to be collected about two weeks after infection. A spore collector (Pretorius et al., 2019) was autoclaved, cooled down and dried out before use. A vacuum hose was attached to the spore collector and the vacuum ran through the spore

collector for a few seconds to help clean it. An empty gelatin capsule was attached to the bottom of the spore collector. The spore collector was used to vacuum the leaves where rust urediniospores were erupting. The collected urediniospores in the gelatin capsule were transferred into a 1.5 mL Eppendorf tube, sealed with parafilm and stored at 4 °C for immediate use or -80 °C for long-term storage.

Seedling treatment for inoculation with fresh P. emaculata *urediniospores*

Alamo seedlings were grown with 0.1% PPM in sterile magenta boxes with sand and 0.5x MS media, as described above, for about four weeks. Then, four seedlings were transplanted to new sterile magenta boxes with no PPM. Two weeks later, the plants were inoculated with fresh urediniospores. Specifically, 10 mg of fresh spores were mixed with 0.02g talc powder in a 1.5 mL Eppendorf tube. A small paint brush was used to dip some mixture powder and brush 2-3 leaves per magenta box. The boxes were covered with black bags, and distilled water were sprayed inside the bags to generate more humidity. The plants were left in the dark and room temperature for at least 18 hours. Another magenta box was placed on top of the box holding the seedlings. The two magenta boxes were sealed with parafilm. Plants were placed in an incubator at 28°C, with a 16/8-hour (light/dark) regimen. Plants that were not inoculated with urediniospores were used as a control. Plants were checked daily for infection syndrome (rust colonies). Entire leaves were collected when infection syndromes first began to appear. Uninoculated leaves were harvested at the same time as the inoculated leaves were harvested. Leaves were washed with 10 mg/mL lysozyme and 10 U/ML zymolyase (shaken at 150 rpm for 30 min at room temperature), rinsed with sterile water (3 times) and stored at 4°C until use.

Extraction of metagenomic DNAs from plant leaf samples

The method used for metagenomic DNA isolation from plant leaf samples was modified from the Synergy protocol (OPS Diagnostics). Two beads and 0.1 g of plant tissue were added to each 2 mL Synergy homogenization tube. Tubes were frozen in liquid nitrogen as soon as leaves were placed in the tubes. Then, the tubes were placed in the tissuelyzer (Quiagen), and the leaf samples were homogenized as described previously. Then, 500 μL of homogenization buffer preheated at 65°C and 5 μL RNase A (100 mg/mL) were added into these 2 mL Synergy homogenization tubes. These tubes were vortexed for 5 – 10 seconds in order to fully resuspend tissues. Next, tubes were placed in a 65°C water bath for 15 min (mixed occasionally), and then were left on a bench top at room temperature for about 10 minutes to cool down. These tubes were spun in a microcentrifuge (14,000 rpm for 5 min). The clear supernatant in each tube was transferred into a clean 1.5 mL eppendorf tube. One volume of isopropanol was added into each of these tubes. Tubes were vortexed, and then were put into a -20 °C freezer for 15 minutes. The solution inside each tube was transferred into a spin column placed in a collection tube. The column was centrifuged at 800 x g for 1 minutes to pass through the wash solution. A total of 700 μL wash buffer (75% ethanol) was added into the column. Then, the column was spun in a microcentrifuge at 10,000 rpm for 1 minute. After discarding the flow thorough, the empty column in the collection tube was spin again, this time at 10,000 rpm for 2 minutes. DNAs were eluted by placing the column in a clean collection tube and by adding 100 μL of molecular biology grade water or TE buffer. The column was centrifuged at 10,000 rpm for 1 minute. The DNAs were stored at -20°C for future use.

Plating microbes from inside plant leaves

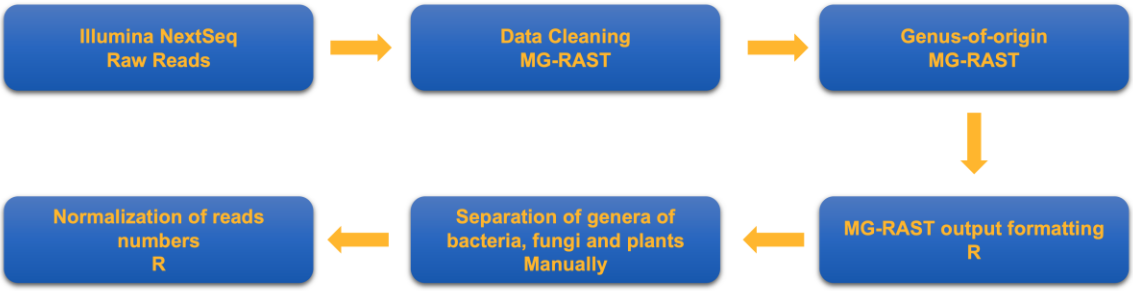
Two beads and 500 μL of PBS buffer were added into each Synergy homogenization tube with entire leaf samples. Then the tubes were securely fastened between adapter plates and placed in the tissuelyzer (Qiagen). The leaf samples were homogenized as described above. The tubes were centrifuged at 12,000 rpm for 1 minute to precipitate cell debris. Then, 50 μL of supernatant from each tube was spread on a plate of 0.5x Difco^{TM} Nutrient Broth media. The solutions were spread evenly with glass beads, and the plates placed at room temperature for 24 hours. Three repetitions were used.

Extraction of metagenomic DNAs from P. emaculata *urediniospores*

Metagenomic DNAs of urediniospores with or without lysozyme treatment were extracted. For lysozyme treatment, 5 mg urediniospores, 1 mL PBS buffer and 50 μL of 10 mg/mL lysozyme were added into 2 mL Synergy homogenization tubes. Then, the tubes were placed on a shaker and incubated at 150 rpm for 30 minutes. Then, the urediniospores were washed with sterile water three times. The method used for metagenomic DNA isolation from urediniospores was as described above.

Whole genome shotgun sequencing and data analysis

DNAs were sequenced at the Georgia Genomics and Bioinformatics Core (GGBC) using an Illumina NextSeq (PE150) sequencer. The sequencing libraries were prepared with the Nextera DNA Flex library preparation protocol that is used at GGBC. The length of paired end reads was 150 bp. The data analysis pipeline is shown below. Details are described in the following text.



MG-RAST were chosen for metagenomic analysis because of its easy accessibility, good performance of genera predictions and low computational cost (Meyer et al., 2008). The sequencing data were first analyzed for genus-of-origin using MG-RAST. : 1) raw reads in FASTQ format were preprocessed with SolexaQA (Cox et al., 2010) for removal of low-quality regions; 2) reads that matched with the human genome were filtered out; 3) the remaining reads were used for gene-calling. Protein encoding genes were predicted using FragGeneScan (Rho et al., 2010), and RNA genes were detected by searching the SILVA, Greengenes and RDP databases (90% identities); 4) predicted protein and RNA genes were clustered (97% identity) using CD-HIT (Fu et al., 2012), and the longest reads were selected as the cluster representatives; 5) taxonomies of reads were identified using the BLAT algorithm (Kent, 2002) and abundance profiles were generated by the best hit (90%). The taxonomy profiles from MG-RAST were merged and formatted in R. Then genera of bacteria, fungi and others were separated manually. Reads of genera were normalized by per 1 million *Puccnia* reads for metagenomic analysis of urediniospores, and by per 1 million plant reads for metagenomic analysis of switchgrass leaves with and without uredioniospore inoculation in R.

Results

Collection of fresh urediniospores

Fresh *P. emaculata* urediniospores were collected by infecting switchgrass plants. The so-called "rust infection syndrome" phenotype of switchgrass is shown in Figure 3.1. Obvious rust colonies were found on switchgrass leaves (Figure 3.1, left), indicating the successful infection of urediniospores. About 10 mg of fresh *P. emaculata* spores were collected into the gelatin capsule attached to the bottom of the spore collector (Figure 3.1, right).

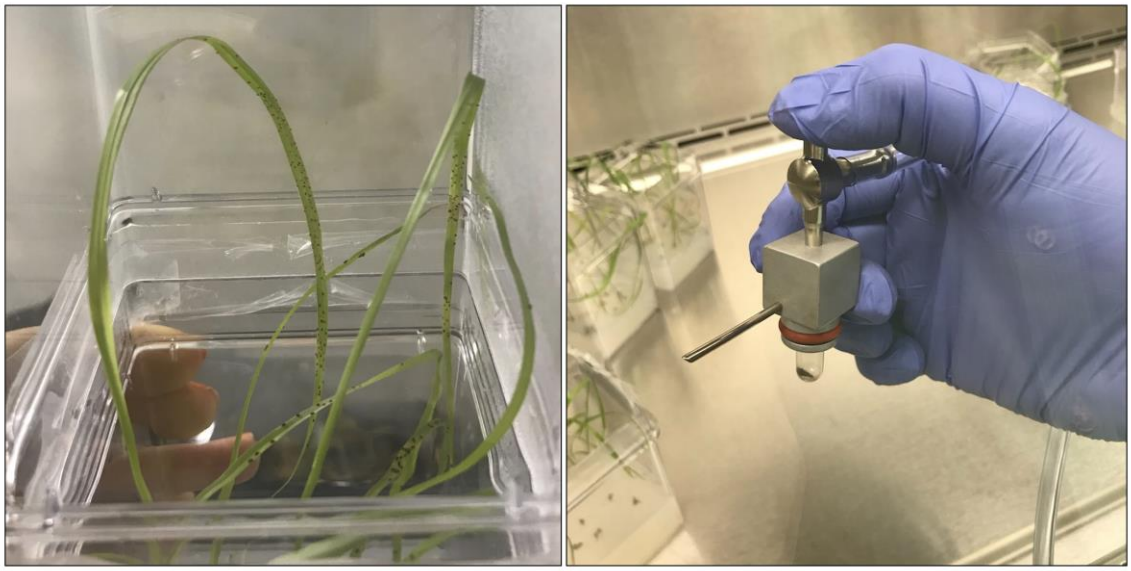


Figure 3.1 Successfully rust-infected switchgrass plants (left) and collection of fresh *P. emaculata* urediniospores with a spore collector (right). Urediniospores were collected into the gelatin capsule attached to the bottom of the spore collector**.**

Analysis of microbes associated with urediniospores

I was interested in the microbes that were associated with urediniospores. Metagenomes associated with urediniospores were sequenced and analyzed. Bacterial genera abundantly associated with urediniospores were listed in Table 3.1 (no lysozyme). The number of *Puccinia* reads was used as a positive control to determine the overall composition of bacteria relative to the *P. emaculata* spore host. The number of reads of each genus were shown as per 1 million *Puccinia* reads. *Stenotrophomonas* were found to be the most abundance genus associated with urediniospores. Other genera, including *Pantoea*, *Acidovorax*, *Agrobacterium* and *Chryseobacterium*, were found to be abundant as well. I was also interested in external versus internal microbes. Uredioniospores were washed with lysozyme so that bacteria on the surface were removed (or reduced). Lysozyme (N-acetylmuramide glycanhydrolase) attacks glycosidic bonds, thereby lysing many bacteria, especially Gram-negative bacteria. The bacteria associated with lysozyme-treated urediniospores were shown in Table 3.1 (lysozyme). *Acidovorax* were found to be the most abundant genera. Other genera, including *Xanthomonas*, *Conexibacter* and *Agrobacterium*, were found to be abundant as well. Amazingly, there were many more reads from several of the bacterial genera than there were from *P. emaculata*. For *Acidovorax*, for instance, there were more than 30 times more reads than for *Puccinia*, even though the *Puccinia* genome is about 20-fold larger (~130 Mb versus ~6 Mb). Fungal genera were also found from the metagenomic analysis, but they are less abundant than *Puccinia* and bacterial genera. Fungal genera in the top 50 most abundant microbial genera

are shown in Table 3.2. Fungal reads (other than *Puccinia*) were >40X fewer than bacterial reads.

Table 3.1 Bacterial compositions by read numbers of urediniospores with and without lysozyme treatment

Genera	Gram	No Lysozyme	Lysozyme
Puccinia	NA	1,000,000	1,000,000
Stenotrophomonas	Gram-negative	14,695,866	828,914
Pantoea	Gram-negative	5,725,178	146,895
Acidovorax	Gram-negative	3,423,416	8,885,193
Agrobacterium	Gram-negative	2,285,666	1,032,937
Chryseobacterium	Gram-negative	1,427,143	979,110
Xanthomonas	Gram-negative	1,018,468	2,293,665
Rhizobium	Gram-negative	404,947	541,949
Pseudomonas	Gram-negative	290,918	78,018
Conexibacter	Gram-positive	272,619	1,662,085
Methylobacterium	Gram-negative	208,065	650,539
Sinorhizobium	Gram-negative	172,992	106,795
Enterobacter	Gram-negative	147,747	3,620
Streptomyces	Gram-positive	131,481	412,410
Mycobacterium	Gram-positive	120,298	363,169
Escherichia	Gram-negative	117,418	5,647
Aeromicrobium	Gram-positive	93,697	45,168
Salmonella	Gram-negative	91,664	2,288
Vibrio	Gram-negative	79,634	7,433
Burkholderia	Gram-negative	79,295	128,843
Clavibacter	Gram-positive	78,787	184,311
Variovorax	Gram-negative	76,245	167,949
Erwinia	Gram-negative	75,229	2,857
Bacillus	Gram-positive	63,707	14,683
Klebsiella	Gram-negative	62,691	4,527
Mucilaginibacter	Gram-negative	56,422	159,406
Arthrobacter	Gram-positive	54,558	109,141

Yersinia	Gram-negative	54,049	5,020
Leifsonia	Gram-positive	51,508	105,250
Mesorhizobium	Gram-negative	50,322	44,145

* Top 30 abundant bacterial genera (no lysozyme) are listed

Table 3.2 Fungal compositions by read numbers of urediniospores with and without lysozyme treatment

Genera	No Lysozyme	Lysozyme
Puccinia	1,000,000	1,000,000
Metarhizium	309,726	1,158
Penicillium	209,590	13,621
Aspergillus	146,052	8,582
Nectria	70,146	2,404
Ustilago	53,033	1,186,985

Determination of an appropriate biocide concentration

Plants carry a variety of microorganisms, including endophytes, that are seed transmitted. Even treatment of seed with external sterilizing agents does not remove all these transmitted microbes (Khalaf et al., 2016). Plant Preservative Mixture (PPM) is a biocide that has been used to minimize contaminating microbial growth in plant tissue cultures. An appropriate concentration of PPM is essential to achieve sterile plants. Low concentrations of PPM may not be able to remove microbes, while high concentrations of PPM could be harmful to plants and/or inhibit subsequent microbial infection. Alamo seedlings were grown in magenta boxes to determine an appropriate biocide concentration,

using 0%, 0.1%, 0.2% or 0.3% of PPM concentrations. The phenotypes of plants were observed for four weeks after planting (Figure 3.2). The growth patterns of switchgrass seedlings under 0.1% PPM were not obviously worse than those under 0% of PPM. However, switchgrass seedlings grown under 0.2% and 0.3% PPM were much shorter than those under 0% of PPM.

Leaf samples of seedling growing under each PPM concentration were collected. Samples were homogenized and supernatants were plated on non-selective microbial media. The plating results are shown in Figure 3.2. Many colonies were found on the plate of switchgrass growing without PPM. No colony was observed on the plates of switchgrass seedlings growing with PPM at 0.1%, 0.2% or 0.3%. Given the results from observations of plant growth and plating of homogenized plant tissues, we concluded that 0.1% PPM should be an appropriate concentration to use for removing endogenous microbes transmitted on switchgrass seeds.

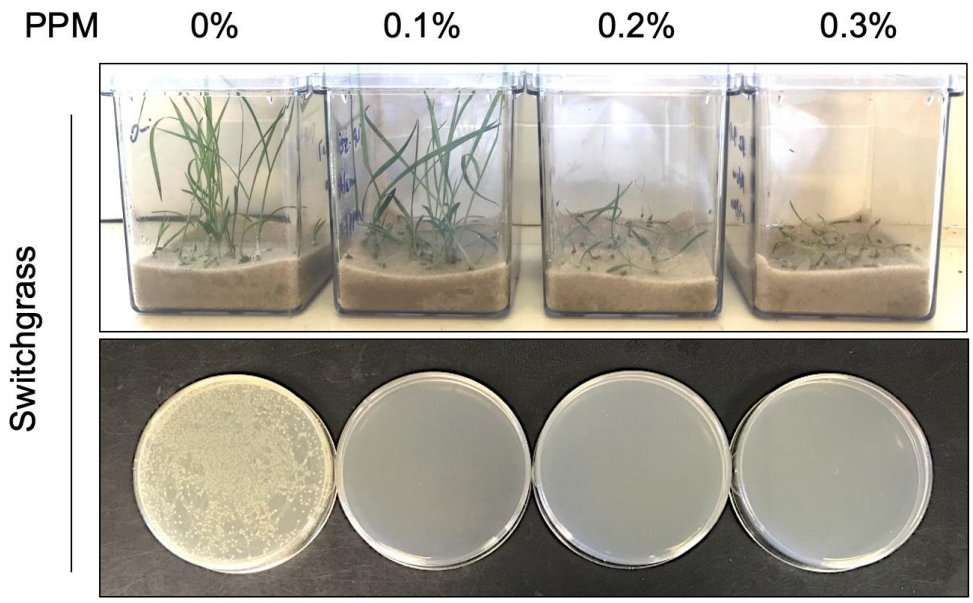


Figure 3.2 Plant growth and tissue extract plating for bacterial analysis under different concentrations of PPM

Comparison of microbial communities associated with inoculated and uninoculated plants

Alamo seedlings were inoculated with *P. emaculata* spores (or not inoculated, as a control) as described in Methods. Inoculated plants were monitored daily. Leaf samples were collected when "rust infection syndrome" started to appear. Specifically, whole switchgrass leaves were collected about 6 days after inoculation. Controls were harvested at the same time. Leaves were washed in lysozyme and zymolyase to sterilize the surface. Washed leaf tissues were homogenized and then plated on non-selective microbial media. Plating results are shown in Figure 3.3. Few colonies were found from the switchgrass leaves that were not inoculated with urediniospores. A variety of colonies were found on

the inoculated switchgrass leaves. It should be noted that these plates will not allow the growth of *P. emaculata,* which is an obligate biotroph. The plating results illustrated that diverse microbes were brought to switchgrass leaves by urediniospores during the infection stage.

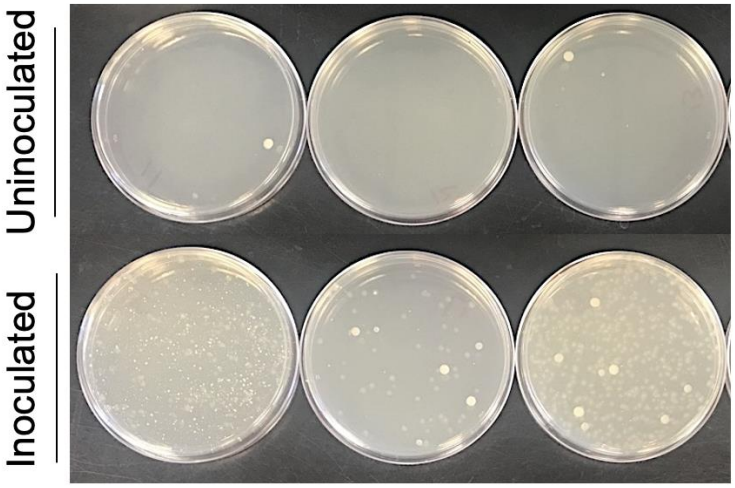


Figure 3.3 Plating of microbes from suspensions derived from leaves that were inoculated or uninoculated with *P. emaculata.* It should be noted that *P. emaculata*, an obligate biotroph, cannot grow on these plates.

Metagenomes of inoculated and uninoculated switchgrass leaves were shotgun sequenced, and annotated. Plant and microbial compositions of uninoculated and inoculated switchgrass leaves are shown in Table 3.3 and Table 3.4. Most of the reads were found to come from plant DNA, which thereby serves as a control to provide a consistent amount of tissue analyzed in each sample. The number of reads of each genus are shown as per 1 million plant reads. *Puccinia* was found only in inoculated leaves and was abundant.

More genera were found inside inoculated leaves than uninoculated leaves. Some bacterial genera (*Bifidobacterium*, *Prochlorococcus* and *Vibrio*), were found to be abundant regardless of inoculation. *Pseudomonas* were abundant in inoculated leaves only. Several bacterial genera, including *Methylobacterium*, *Xanthomonas*, and *Agrobacterium* were not found in uninoculated leaves but were found in inoculated leaves. These genera, *Vibrio* and *Pseudomonas*, were found to be associated with urediniospores as well (Table 3.1). Besides *Puccinia*, three fungal genera (*Phakopsora*, *Filobasidiella* and *Melampsora*) were found in inoculated leaves (Table 3.4).

Table 3.3 Bacterial compositions by read numbers of uninoculated (UI) and inoculated (I) switchgrass leaves *

	Gram	UI-1	UI-2	UI-3	I-1	I-2	I-3	UI-Average	UI-Sd	I-Average	I-Sd
Plant	NA	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	0	1,000,0000	
Puccinia	NA	0	0	0	18,534	15,156	37,891	0	0	23,860	12267
Bifidobacterium	Gram-positive	3410	3072	4374	5253	2733	4489	3619	676	4158	1292
Prochlorococcus	Gram-negative	462	798	1072	442	511	1213	777	306	722	426
Vibrio	Gram-negative	259	403	498	343	409	437	387	120	396	48
Pseudomonas	Gram-negative	0	0	20	842	130	0	7	0	324	537
Methylobacterium	Gram-negative	0	0	0	1128	600	54	0	12	594	453
Conexibacter	Gram-positive	0	0	0	71	486	25	0	0	194	254
Xanthomonas	Gram-negative	0	0	0	14	269	0	0	0	95	152
Agrobacterium	Gram-negative	0	0	0	0	96	25	0	0	40	50
Others	NA	364	428	327	871	2091	506	373	51	1156	830

* The number of reads of genera were normalized by per 1 million plant reads.

Table 3.4 Fungal compositions by read numbers of uninoculated (UI) and inoculated (I) switchgrass leaves *

	UI-1	UI-2	UI-3	I-1	I-2	I-3	UI-Average	UI-Sd	I_Average	I-Sd
Plant	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	0	1,000,000	0
Puccinia	0	0	0	18,534	15,156	37,891	0	0	23,860	12,267
Phakopsora	0	0	0	314	541	1,321	0	0	725	528
Filobasidiella	0	0	5	29	23	103	2	3	52	45
Melampsora	0	0	0	7	18	49	0	0	25	22

* The number of reads of genera were normalized by per 1 million plant reads.

Discussion and Conclusions

P. emaculata urediniospores were found to be associated with a very abundant supply of different microbes. The lysozyme results indicate that many bacteria are located on the surface of the spore where they are susceptible to lysozyme treatment. *Pantoea and Stenotrophomonas* numbers are dramatically (>15X) reduced by lysozyme treatment. Less abundant microbes (e.g., *Agrobacterium*) are also diminished by lysozyme treatment. Many, but not the majority, of microbes are apparently inside the urediniospores. How microbes associate with *P. emaculata* spores is an interesting question. They are most likely acquired from the endophytes of the host leaf they have infected. Some microbes appear to be packaged inside the urediniospores, and thus are most likely to be transmitted because they are protected from the external environment. However, many other microbes appear to increase in abundance after lysozyme treatment, a result that seems unlikely to be a biological phenomenon. Theoretically, the read number of a microbe in the lysozyme treatment should be no greater than that in the untreated samples when using *Puccinia* read

numbers as the control. This is consistent with the read number of many microbes. However, there are some microbes whose read number were higher in lysozyme treatment, which is inconsistent with the theoretical expectation. One possible reason is that the *Puccinia* DNA extraction was incomplete, resulting in the inaccuracy of using the number of *Puccinia* reads as a positive control. Moreover, the spores used for controls and lysozyme treatments were not treated the same because I did not soak the control spores in the same solution without lysozyme. Another possible explanation is contamination in the experimental process, but this is less likely.

PPM at a concentration of 0.1% was capable of minimizing seed-transmitted microbes without visibly hindering plant growth. Then the microbial communities of switchgrass seedlings that had been minimized for seed-transmitted microbial content by 0.1% PPM were investigated for their microbial composition with and without *P. emaculata* infection. Although most microbes were removed by PPM, several genera, including *Bifidobacterium* and *Prochlorococcus*, were still found in plants without *P. emaculata* infection. These genera were found in plants with *P. emaculata* infection as well. The abundance of these genera in inoculated plants were similar to that in uninoculated plants, suggesting that *P. emaculata* infection has little effects on the abundance of these genera.

Compared to uninoculated samples, the *P. emaculata*-inoculated seedlings contained many microbes, as assayed both by plating and shotgun DNA sequencing. The shotgun sequence analysis of microbes inside inoculated leaves discovered that the bacterial genera, *Methylobacterium*, *Xanthomonas*, and *Agrobacterium* were transmitted via infections of *Puccinia* spores, but not such genera as *Pantoea* and *Stenotrophomonas*, which were the

two most abundant bacterial genera found to be associated with *P. emaculata* spores. Some strains of *Pantoea* and *Stenotrophomonas* are biocontrol and bioremediation bacteria (Walterson et al., 2015; Ryan et al., 2009) that facilitate plant growth and health. It is possible that *Pantoea* and *Stenotrophomonas* are not inside the studied rust spores, and thus not transferred by *P. emaculata*, because they may hinder fungal invasion. A few fungal genera were found to be associated with *P. emaculata* urediniospores as well. However, none of these genera were transferred into switchgrass leaves. It's likely that these fungi are accompanying urediniospores. Although reads of three fungal genera were found from metagenomes of inoculated switchgrass leaves, they are not abundant (less than 0.1% of the total reads). Future studies can investigate how other fungi associate urediniospores from generation to generation and how those fungal genera in inoculated switchgrass leaves get into leaves.

Bacteria were found to be transferred into plant leaves during urediniospores infection on the plant leaves via penetration. This is a major discovery because none of bacteria are found to be able to get through the epidermis of an intact plant leaf. Moreover, movement of these transferred microbes after arriving inside plant leaves is a particularly interesting question that should be investigated in future studies. Diverse interactions between bacteria and fungus have been reported (Frey-Klettet al., 2011). It is possible that transferred microbes play an important role in uredioniospores infection, perhaps by helping to suppress the plant immune response. However, these microbes may have negative effects on rust infection as well, perhaps by competing with *P. emaculata* for nutrients. If one can fully sterilize but retain viability in a *P. emaculata* spore, the effects of a microbe on

urediniospore function could be tested by monitoring infection syndromes after inoculating the microbe on urediniospores.

In conclusion, this chapter has several major discoveries. Many microbial genera were found to be associated with urediniospores. Most of them are bacterial genera. More microbes were found inside inoculated switchgrass leaves than uninoculated leaves. Several urediniospores-associated bacteria (*Methylobacterium*, *Xanthomonas*, and *Agrobacterium*), were transferred into switchgrass leaves during *P. emaculata* urediniospores infection, probably vectored by urediniospores.

REFERENCES

Abbott, M. J., & Battaglia, L. L. (2015). Purple pitcher plant (*Sarracenia rosea*) dieback and partial community disassembly following experimental storm surge in a coastal pitcher plant bog. *PloS one, 10*(4), e0125475.

Adamec, L. (1997). Mineral nutrition of carnivorous plants: a review. *The Botanical Review*, *63*(3), 273-299.

Adams, M. W., Holden, J. F., Menon, A. L., Schut, G. J., Grunden, A. M., Hou, C., ... & Pan, G. (2001). Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus. Journal of Bacteriology*, *183*(2), 716-724.

Adlassnig, W., Peroutka, M., & Lendl, T. (2010). Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Annals of botany*, 107(2), 181-194.

Agnolucci, M., Battini, F., Cristani, C., & Giovannetti, M. (2015). Diverse bacterial communities are recruited on spores of different arbuscular mycorrhizal fungal isolates. *Biology and Fertility of Soils*, *51*(3), 379-389.

Allen, E. A., Hazen, B. E., Hoch, H. C., Kwon, Y., Leinhos, G. M. E., Staples, R. C., ... & Terhune, B. T. (1991). Appressorium formation in response to topographical signals by 27 rust species. *Phytopathology*, *81*(3), 323-331.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*, Babraham Institute, Cambridge, United Kingdom.

Arber, A. (1941). On the morphology of the pitcher-leaves in *Heliamphora*, *Sarracenia*, *Darlingtonia*, *Cephalotus*, and *Nepenthes*. *Annals of Botany*, *5*(20), 563-578.

Atwater, D. Z., Butler, J. L., & Ellison, A. M. (2006). Spatial distribution and impacts of moth herbivory on northern pitcher plants. *Northeastern Naturalist*, *13*(1), 43-57.

Batalin, F. A. (1880). Ueber die Function der Epidermis in den Schläuchen von Sarracenia und Darlingtonia. *Acta Horti Botanici Petropolitani*, *7*, 343-360.

Bauer, U., Willmes, C., & Federle, W. (2009). Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous *Nepenthes rafflesiana* plants. *Annals of Botany*, *103*(8), 1219-1226.

Bay, J. C. (1893). Nutrition of insectivorous plants. *Botanical Gazette*, *18*(3), 105-106.

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., & Moscou, M. J. (2014). Nonhost resistance to rust pathogens–a continuation of continua. *Frontiers in plant science*, *5*, 664.

Bharadwaj, D. P., Lundquist, P. O., & Alström, S. (2008). Arbuscular mycorrhizal fungal spore-associated bacteria affect mycorrhizal colonization, plant growth and potato pathogens. *Soil Biology and Biochemistry*, *40*(10), 2494-2501.

Bigeard, J., Colcombet, J., & Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Molecular plant*, *8*(4), 521-539.

Boddy, L. (2016). Interactions between fungi and other microbes. *The Fungi* (pp. 337-360). Academic Press.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Boller, T., & Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology*, *60*, 379-406.

Bradshaw, W. E. (1983). Interaction between the mosquito *Wyeomyia smithii,* the midge *Metriocnemus knabi*, and their carnivorous host *Sarracenia purpurea*. *Phytotelmata: terrestrial plants as hosts for aquatic insect communities*, 161-189.

Bradshaw, W. E., & Creelman, R. A. (1984). Mutualism between the carnivorous purple pitcher plant and its inhabitants. *American Midland Naturalist*, 294-304.

Bright, M., & Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology*, *8*(3), 218-230.

Butler, J. L., & Ellison, A. M. (2007). Nitrogen cycling dynamics in the carnivorous northern pitcher plant, *Sarracenia purpurea*. *Functional Ecology*, *21*(5), 835-843.

Butler, J. L., Atwater, D. Z., & Ellison, A. M. (2005). Red-spotted newts: an unusual nutrient source for northern pitcher plants. *Northeastern Naturalist*, *12*(1), 1-10.

Butler, J. L., Gotelli, N. J., & Ellison, A. M. (2008). Linking the brown and green: nutrient transformation and fate in the *Sarracenia* microecosystem. *Ecology*, *89*(4), 898-904.

Butler, J. L., Gotelli, N. J., & Ellison, A. M. (2008). Linking the brown and green: nutrient transformation and fate in the *Sarracenia* microecosystem. *Ecology*, *89*(4), 898-904.

Cianfanelli, F. R., Monlezun, L., & Coulthurst, S. J. (2016). Aim, load, fire: the type VI secretion system, a bacterial nanoweapon. *Trends in microbiology*, *24*(1), 51-62.

Condon, C., Squires, C., & Squires, C. L. (1995). Control of rRNA transcription in *Escherichia coli. Microbiol. Mol. Biol. Rev.*, *59*(4), 623-645.

Conte, M., Castagnola, E., Venzano, P., Tasso, L., & Giacchino, R. (1996). Bacteremia caused by *Hafnia alvei* in a human immunodeficiency virus-infected child. *The Pediatric infectious disease journal*, *15*(2), 182-183.

Cosentino, S., Larsen, M. V., Aarestrup, F. M., & Lund, O. (2013). PathogenFinder-distinguishing friend from foe using bacterial whole genome sequence data. *PloS one*, *8*(10), e77302.

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics*, *11*(1), 485.

Dangl, J. L., & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *nature*, *411*(6839), 826-833.

Deising, H., Jungblut, P. R., & Mendgen, K. (1991). Differentiation-related proteins of the broad bean rust fungus *Uromyces viciae-fabae*, as revealed by high resolution two-dimensional polyacrylamide gel electrophoresis. *Archives of Microbiology*, *155*(2), 191-198.

Demers, J. E., Liu, M., Hambleton, S., & Castlebury, L. A. (2017). Rust fungi on *Panicum*. *Mycologia*, *109*(1), 1-17.

Deslandes, L., & Rivas, S. (2012). Catch me if you can: bacterial effectors and plant targets. *Trends in plant science*, *17*(11), 644-655.

Eichenlaub, R., & Gartemann, K. H. (2011). The Clavibacter michiganensis subspecies: molecular investigation of gram-positive bacterial plant pathogens. *Annual review of phytopathology*, *49*, 445-464.

Ellis, J. G., Dodds, P. N., & Lawrence, G. J. (2007). Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annu. Rev. Phytopathol.*, *45*, 289-306.

Ellison, A. M., Buckley, H. L., Miller, T. E., & Gotelli, N. J. (2004). Morphological variation in *Sarracenia purpurea* (*Sarraceniaceae*): geographic, environmental, and taxonomic correlates. *American Journal of Botany*, *91*(11), 1930-1935.

Ellison, A. M., Gotelli, N. J., Brewer, J. S., Cochran-Stafira, D. L., Kneitel, J. M., Miller, T. E., ... & Zamora, R. (2003). The evolutionary ecology of carnivorous plants.

Elsik, C. G., Mackey, A. J., Reese, J. T., Milshina, N. V., Roos, D. S., & Weinstock, G. M. (2007). Creating a honey bee consensus gene set. *Genome biology*, *8*(1), R13.

Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, *2*(4), 953-971.

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology*, *300*(4), 1005-1016.

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, *30*(7), 1575-1584.

Estill, J. C., & Bennetzen, J. L. (2009). The DAWGPAWS pipeline for the annotation of genes and transposable elements in plant genomes. *Plant methods*, *5*(1), 8.

Fatima, U., & Senthil-Kumar, M. (2015). Plant and pathogen nutrient acquisition strategies. *Frontiers in plant science*, *6*, 750.

Flor, H. H. (1942). Inheritance of pathogenicity in Melampsora lini. *Phytopathology*, *32*, 653-669.

Folkerts, G. W. (1982). The Gulf Coast pitcher plant bogs [Mississippi, Alabama]. *American Scientist*.

Frank BA. 1889. Ueber die Pilzsymbiose der Leguminosen. Ber Dtsch

Frank, A. C., Saldierna Guzmán, J. P., & Shay, J. E. (2017). Transmission of bacterial endophytes. *Microorganisms*, *5*(4), 70.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152.

Gagne-Bourgue, F., Aliferis, K. A., Seguin, P., Rani, M., Samson, R., & Jabaji, S. (2013). Isolation and characterization of indigenous endophytic bacteria associated with leaves of switchgrass (*Panicum virgatum L.*) cultivars. *Journal of applied microbiology*, *114*(3), 836-853.

Gallie, D. R., & Chang, S. C. (1997). Signal transduction in the carnivorous plant *Sarracenia purpurea* (Regulation of secretory hydrolase expression during development and in response to resources). *Plant Physiology*, *115*(4), 1461-1471.

Garcia, A. V., & Hirt, H. (2014). *Salmonella enterica* induces and subverts the plant immune system. *Frontiers in microbiology*, *5*, 141.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.

Gelev, I., Gelev, E., Steigerwalt, A. G., Carter, G. P., & Brenner, D. J. (1990). Identification of the bacterium associated with haemorrhagic septicaemia in rainbow trout as *Hafnia alvei*. *Research in microbiology*, *141*(5), 573-576.

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., ... & Parniske, M. (2008). SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, *rhizobia*, and *Frankiabacteria*. *Proceedings of the National Academy of Sciences*, *105*(12), 4928-4932.

Gill, U. S., Uppalapati, S. R., Nakashima, J., & Mysore, K. S. (2015). Characterization of Brachypodium distachyon as a nonhost model against switchgrass rust pathogen *Puccinia emaculata*. *BMC plant biology*, *15*(1), 113.

Giraldo, M. C., & Valent, B. (2013). Filamentous plant pathogen effectors in action. *Nature Reviews Microbiology*, *11*(11), 800.

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.*, *43*, 205-227.

Glenn, A., & Bodri, M. S. (2012). Fungal endophyte diversity in *Sarracenia*. *PloS one*, *7*(3).

Göhre, V., & Robatzek, S. (2008). Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.*, *46*, 189-215.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International journal of systematic and evolutionary microbiology*, *57*(1), 81-91.

Gotelli, N. J., & Ellison, A. M. (2006). Food-web models predict species abundances in response to habitat change. *PLoS biology*, *4*(10), e324.

Grigoriev, I., Martinez, D., & Salamov, A. (2006). Fungal genomic annotation. In Aurora, D., Berka, R., Singh, G. (Vol. Eds.), *Applied mycology and biotechnology*: Vol. 6. Bioinformatics (pp. 123–142). Amsterdam: Elsevier.

Gustafson, D. M., Boe, A., & Jin, Y. (2003). Genetic Variation for Infection in Switchgrass. *Crop science*, *43*(3), 755-759.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., ... & Salzberg, S. L. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic acids research*, *31*(19), 5654-5666.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., ... & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology*, *9*(1), R7.

Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A., & Wortman, J. R. (2011). Approaches to fungal genome annotation. *Mycology*, *2*(3), 118-141.

Haiko, J., & Westerlund-Wikström, B. (2013). The role of the bacterial flagellum in adhesion and virulence. *Biology*, *2*(4), 1242-1267.

Hardoim, P. R., Van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., ... & Sessitsch, A. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.*, *79*(3), 293-320.

Hartley, S. E., & Gange, A. C. (2009). Impacts of plant symbiotic fungi on insect herbivores: mutualism in a multitrophic context. *Annual review of entomology*, *54*, 323-342.

Hatano, N., & Hamada, T. (2008). Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *The Journal of Proteome Research*, *7*(2), 809-816.

Hepburn, J. S., & Jones, F. M. (1927). The enzymes of the pitcher liquor of *Sarraceniaceae*. *Trans Wagner Free Inst Sci*, *11*, 49-68.

Hepburn, J. S., Saint John, E. Q., & Jones, F. M. (1927). The biochemistry of the American pitcher plants. *Transactions of the Wagner Free Institute of Science of Philadelphia*, *11*, 1-95.

Higashi, S., Nakashima, A., Ozaki, H., Abe, M., & Uchiumi, T. (1993). Analysis of feeding mechanism in a pitcher of *Nepenthes hybrida* [*Nepenthaceae*]. *Journal of Plant Research (Japan)*.

Hirsch, R. L., TeBeest, D. O., Bluhm, B. H., & West, C. P. (2010). First report of rust caused by *Puccinia emaculata* on switchgrass in Arkansas. *Plant Disease*, *94*(3), 381-381.

Hoch, H. C., Staples, R. C., Whitehead, B., Comeau, J., & Wolf, E. D. (1987). Signaling for growth orientation and cell differentiation by surface topography in *Uromyces*. *Science*, *235*(4796), 1659-1662.

Hölscher, D., Dhakshinamoorthy, S., Alexandrov, T., Becker, M., Bretschneider, T., Buerkert, A., ... & Heklau, H. (2014). Phenalenone-type phytoalexins mediate resistance of banana plants (*Musa spp*.) to the burrowing nematode *Radopholus similis*. *Proceedings of the National Academy of Sciences*, *111*(1), 105-110.

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research*, *35*(suppl_2), W585-W587.

Howe, K. L., Chothia, T., & Durbin, R. (2002). GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome research*, *12*(9), 1418-1427.

Huang, X., Adams, M. D., Zhou, H., & Kerlavage, A. R. (1997). A tool for analyzing and annotating genomic sequences. *Genomics*, *46*(1), 37-45.

Huys, G., Cnockaert, M., Abbott, S. L., Janda, J. M., & Vandamme, P. (2010). *Hafnia paralvei* sp. *nov*., formerly known as *Hafnia alvei* hybridization group 2. *International journal of systematic and evolutionary microbiology*, *60*(8), 1725-1728.

Janda, J. M., & Abbott, S. L. (2006). The genus *Hafnia*: from soup to nuts. *Clinical microbiology reviews*, *19*(1), 12-28.

Joel, D. M. (1986). Glandular structures in carnivorous plants: their role in mutual and unilateral exploitation of insects. *Insects and the plant surface. Edward Arnold, London*, 219-234.

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., & Aarestrup, F. M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *Journal of clinical microbiology*, *52*(5), 1501-1510.

Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *nature*, *444*(7117), 323-329.

Joye, K. (1989). Carnivorous plants.

Juniper, B. E., Robins, R. J., & Joel, D. M. (1989). The carnivorous plants. (Academic Press Limited: London).

Jürgens, A., El-Sayed, A. M., & Suckling, D. M. (2009). Do carnivorous plants use volatiles for attracting prey insects? *Functional Ecology*, *23*(5), 875-887.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, *338*(5), 1027-1036.

Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual review of phytopathology*, *44*.

Kanehisa, M., & Sato, Y. (2019). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science*.

Kappler, U., & Schäfer, H. (2014). Transformations of dimethylsulfide. *The metal-driven biogeochemistry of gaseous compounds in the environment* (pp. 279-313). Springer, Dordrecht.

Karagatzides, J. D., & Ellison, A. M. (2009). Construction costs, payback times, and the leaf economics of carnivorous plants. *American Journal of Botany*, *96*(9), 1612-1619.

Karasov, T. L., Barrett, L., Hershberg, R., & Bergelson, J. (2017). Similar levels of gene content variation observed for Pseudomonas syringae populations extracted from single and multiple host species. *PloS one*, *12*(9).

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., & Voegele, R. T. (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Molecular Plant-Microbe Interactions*, *18*(11), 1130-1139.

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome research*, *12*(4), 656-664.

Khalaf, E. M., & Raizada, M. N. (2016). Taxonomic and functional diversity of cultured seed associated microbes of the cucurbit family. *BMC microbiology*, *16*(1), 131.

Kobayashi, Y., Kobayashi, I., Funaki, Y., Fujimoto, S., Takemoto, T., & Kunoh, H. (1997). Dynamic reorganization of microfilaments and microtubules is necessary for the expression of non-host resistance in barley coleoptile cells. *The Plant Journal*, *11*(3), 525-537.

Koopman, M. M., Fuselier, D. M., Hird, S., & Carstens, B. C. (2010). The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. *Appl. Environ. Microbiol.*, *76*(6), 1851-1860.

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 adaptive k-mer weighting and repeat separation. *Genome research*, *27*(5), 722-736.

Korf, I. (2004). Gene finding in novel genomes. *BMC bioinformatics*, *5*(1), 59.

Kotrba, P., Inui, M., & Yukawa, H. (2001). Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *Journal of Bioscience and Bioengineering*, *92*(6), 502-517.

Koukol, O., Kolařík, M., Kolářová, Z., & Baldrian, P. (2012). Diversity of foliar endophytes in wind-fallen *Picea abies* trees. *Fungal Diversity*, *54*(1), 69-77.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, *305*(3), 567-580.

Kuo, A., Bushnell, B., & Grigoriev, I. V. (2014). Fungal genomics: sequencing and annotation. In *Advances in Botanical Research* (Vol. 70, pp. 1-52). Academic Press.

Laluk, K., & Mengiste, T. (2010). Necrotroph attacks on plants: wanton destruction or covert extortion?. *The Arabidopsis Book*, e0136.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.

Lawrence, G. J., Dodds, P. N., & Ellis, J. G. (2010). TECHNICAL ADVANCE: transformation of the flax rust fungus, Melampsora lini: selection via silencing of an avirulence gene. *The Plant Journal*, *61*(2), 364-369.

Lázaro-Díez, M., Redondo-Salvo, S., Arboleya-Agudo, A., Ocejo-Vinyals, J. G., Chapartegui-González, I., Ocampo-Sosa, A. A., ... & Ramos-Vivas, J. (2016). Whole-genome sequence of *Hafnia alvei* HUMV-5920, a human isolate. Genome Announc., 4(3), e00556-16.

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., ... & Schatz, M. (2016). Third-generation sequencing and the future of genomics. BioRxiv, 048603.

Lee, I., Kim, Y. O., Park, S. C., & Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International journal of systematic and evolutionary microbiology*, *66*(2), 1100-1103.

Lei, J., Finlayson, S. A., Salzman, R. A., Shan, L., & Zhu-Salzman, K. (2014). Botrytis-induced kinase1 modulates Arabidopsis resistance to green peach aphids via phytoalexin deficient4. *Plant Physiology*, *165*(4), 1657-1670.

Lewis, D. H. (1973). Concepts in fungal nutrition and the origin of biotrophy. *Biological Reviews*, *48*(2), 261-277.

Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell research*, *18*(1), 85-98.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.

Li, H., & Wysoker, A. (2009). Durbin R; 1000 Genome project data processing subgroup. The sequence alignment/Map format and SAMtools.

Li, X., & Heyer, W. D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell research*, *18*(1), 99-113.

Lindberg, A. M., Ljungh, Å., Ahrne, S., Löfdahl, S., & Molin, G. (1998). *Enterobacteriaceae* found in high numbers in fish, minced meat and pasteurised milk or cream and the presence of toxin encoding genes. *International journal of food microbiology*, *39*(1-2), 11-17.

Lindquist, J. A. (1975). Bacteriological and ecological observations on the northern pitcher plant, *Sarracenia purpurea* (Doctoral dissertation).

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., ... & Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual review of plant biology*, *66*, 513-545.

Luciano, C. S., & Newell, S. J. (2017). Effects of prey, pitcher age, and microbes on acid phosphatase activity in fluid from pitchers of *Sarracenia purpurea* (*Sarraceniaceae*). *PloS one*, *12*(7), e0181252.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., ... & Tang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*(1), 18.

Maier, W., Wingfield, B. D., Mennicken, M., & Wingfield, M. J. (2007). Polyphyly and two emerging lineages in the rust genera *Puccinia* and *Uromyces*. *Mycological research*, *111*(2), 176-185.

McLaughlin, S. B., & Kszos, L. A. (2005). Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*, *28*(6), 515-535.

Meir, P., Juniper, B. E., & Evans, D. E. (1991). Regulation of free calcium concentration in the pitchers of the carnivorous plant *Sarracenia purpurea*: a model for calcium in the higher plant apoplast? *Annals of Botany*, *68*(6), 557-561

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., ... & Wilkening, J. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, *9*(1), 386.

Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., ... & Taylor, J. M. (2018). De novo assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata f. sp. avenae*, the causal agent of oat crown rust. *mBio*, *9*(1), e01650-17.

Miller, T., Cassill, D., Johnson, C., Kindell, C., Lepis, J., McInnes, D., ... & Richard, B. (1994). Intraspecific and interspecific competition of *Wyeomyia smithii* (Coq.)(Culicidae) in pitcher plant communities. *American Midland Naturalist*, 136-145.

Mitter, B., Pfaffenbichler, N., Flavell, R., Compant, S., Antonielli, L., Petric, A., ... & Sessitsch, A. (2017). A new approach to modify plant microbiomes and traits by introducing beneficial bacteria at flowering into progeny seeds. *Frontiers in microbiology*, *8*, 11.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research*, *35*(suppl_2), W182-W185.

Mysore, K. S., & Ryu, C. M. (2004). Nonhost resistance: how much do we know? *Trends in plant science*, *9*(2), 97-104.

NACZI, R. F., SOPER, E. M., CASE JR, F. W., & CASE, R. B. (1999). *Sarracenia rosea* (*Sarraceniaceae*), a new species of pitcher plant from the southeastern United States. *SIDA, Contributions to Botany*, 1183-1206.

Nazareno, E. S., Li, F., Smith, M., Park, R. F., Kianian, S. F., & Figueroa, M. (2017). *Puccinia coronata f. sp. avenae*: a threat to global oat production. *Molecular plant pathology*.

Newell, S. J., & Nastase, A. J. (1998). Efficiency of insect capture by *Sarracenia purpurea* (*Sarraceniaceae*), the northern pitcher plant. *American Journal of Botany*, *85*(1), 88-91.

Nielsen, H., Brunak, S., & von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein engineering*, *12*(1), 3-9.

Novas, M. V., Iannone, L. J., Godeas, A. M., & Cabral, D. (2009). Positive association between mycorrhiza and foliar endophytes in *Poa bonariensis*, a native grass. *Mycological progress*, *8*(1), 75.

Nygaard, P., & Smith, J. M. (1993). Evidence for a novel glycinamide ribonucleotide transformylase in *Escherichia coli*. *Journal of bacteriology*, *175*(11), 3591-3597.

Oliver, R. P., & IPCHO, S. V. (2004). *Arabidopsis* pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. *Molecular Plant Pathology*, *5*(4), 347-352.

Oliver, R. P., & Ipcho, S. V. (2004). Arabidopsis pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. *Molecular Plant Pathology*, *5*(4), 347-352.

Omacini, M., J. Chaneton, E., M. Ghersa, C., & Otero, P. (2004). Do foliar endophytes affect grass litter decomposition? A microcosm approach using *Lolium multiflorum*. *Oikos*, *104*(3), 581-590.

Oswald, W. W., Doughty, E. D., Ne'eman, G., Ne'eman, R., & Ellison, A. M. (2011). Pollen morphology and its relationship to taxonomy of the genus *Sarracenia* (*Sarraceniaceae*). *Rhodora*, *113*(955), 235-251.

Ownley, B. H., Gwinn, K. D., & Vega, F. E. (2010). Endophytic fungal entomopathogens with activity against plant pathogens: ecology and evolution. *BioControl*, *55*(1), 113-128.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., ... & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691-3693.

Paisie, T. K., Miller, T. E., & Mason, O. U. (2014). Effects of a ciliate protozoa predator on microbial communities in pitcher plant (*Sarracenia purpurea*) leaves. *PloS one*, *9*(11).

Park, C. J., & Andam, C. P. (2019). Within-species genomic variation and variable patterns of recombination in the tetracycline producer Streptomyces rimosus. *Frontiers in Microbiology*, *10*, 552.

Peroutka, M., Adlassnig, W., Volgger, M., Lendl, T., Url, W. G., & Lichtscheidl, I. K. (2008). Utricularia: a vegetarian carnivorous plant?. *Plant Ecology*, *199*(2), 153-162.

Peterson, C. N., Day, S., Wolfe, B. E., Ellison, A. M., Kolter, R., & Pringle, A. (2008). A keystone predator controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem. *Environmental Microbiology*, *10*(9), 2257-2266.

Peterson, C. N., Day, S., Wolfe, B. E., Ellison, A. M., Kolter, R., & Pringle, A. (2008). A keystone predator controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem. *Environmental Microbiology*, *10*(9), 2257-2266.

Petre, B., Joly, D. L., & Duplessis, S. (2014). Effector proteins of rust fungi. *Frontiers in plant science*, *5*, 416.

Plummer, G. L., & Jackson, T. H. (1963). Bacterial activities within the sarcophagus of the insectivorous plant, *Sarracenia flava*. *American Midland Naturalist*, 462-469.

Plummer, G. L., & Kethley, J. B. (1964). Foliar absorption of amino acids, peptides, and other nutrients by the pitcher plant, *Sarracenia flava*. *Botanical Gazette*, *125*(4), 245-260.

Prakash, O., Shouche, Y., Jangid, K., & Kostka, J. E. (2013). Microbial cultivation and the role of microbial resource centers in the omics era. *Applied microbiology and biotechnology*, *97*(1), 51-62.

Prankevicius, A. B., & Cameron, D. M. (1991). Bacterial dinitrogen fixation in the leaf of the northern pitcher plant (Sarracenia purpurea). *Canadian Journal of Botany*, *69*(10), 2296-2298.

Pretorius, Z. A., Booysen, G., Boshoff, W., Joubert, J., & Els, J. (2019). Additive Manufacturing of Devices Used for Collection and Application of Cereal Rust Urediniospores. *Frontiers in plant science*, *10*, 639.

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, *21*(suppl_1), i351-i358.

Proietti, P. C., Passamonti, F., Franciosini, M. P., Rossi, E. D., & Asdrubali, G. (2003). *Hafnia alvei* infection in pullets in Italy. *Large Animals Review*, *9*(6), 77-78.

Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R., & Dodds, P. N. (2012). Challenges and progress towards understanding the role of effectors in plant–fungal interactions. *Current opinion in plant biology*, *15*(4), 477-482.

Rafiqi, M., Gan, P. H., Ravensdale, M., Lawrence, G. J., Ellis, J. G., Jones, D. A., ... & Dodds, P. N. (2010). Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen. *The Plant Cell*, *22*(6), 2017-2032.

Rajamani, T., Suryanarayanan, T. S., Murali, T. S., & Thirunavukkarasu, N. (2018). Distribution and diversity of foliar endophytic fungi in the mangroves of Andaman Islands, India. *Fungal ecology*, *36*, 109-116.

Ramos, H. C., Rumbo, M., & Sirard, J. C. (2004). Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in microbiology*, *12*(11), 509-517.

Ratsirarson, J., & Silander Jr, J. A. (1996). Structure and dynamics in *Nepenthes madagascariensis* pitcher plant micro-communities. *Biotropica*, 218-227.

Real, F., Fernández, A. C. O. S. T. A., Acosta, F., Acosta, B., Castro, P., Déniz, S., & Orós, J. (1997). Septicemia associated with *Hafnia alvei* in laying hens. *Avian diseases*, 741-747.

Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic acids research*, *38*(20), e191-e191.

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*, *13*(5), 278-289.

Ridell, J., & Korkeala, H. (1997). Minimum growth temperatures of *Hafnia alvei* and other *Enterobacteriaceae* isolated from refrigerated meat determined with a temperature gradient incubator. *International journal of food microbiology*, *35*(3), 287-292.

Robinson, W. J. (1908). A study of the digestive power of *Sarracenia purpurea*. *Torreya*, *8*(8), 181-194.

Rodriguez, R. J., White Jr, J. F., Arnold, A. E., & Redman, A. R. A. (2009). Fungal endophytes: diversity and functional roles. *New phytologist*, *182*(2), 314-330.

Rodriguez-R, L. M., & Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe*, *9*(3), 111-8.

Roller, M., Lucić, V., Nagy, I., Perica, T., & Vlahoviček, K. (2013). Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic acids research*, *41*(19), 8842-8852.

Russell, A. B., Peterson, S. B., & Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nature reviews microbiology*, *12*(2), 137-148.

Ryan, R. P., Monchy, S., Cardinale, M., Taghavi, S., Crossman, L., Avison, M. B., ... & Dow, J. M. (2009). The versatility and adaptation of bacteria from the genus Stenotrophomonas. *Nature Reviews Microbiology*, *7*(7), 514-525.

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio gene finding in Drosophila genomic DNA. *Genome research*, *10*(4), 516-522.

Salim, H. M., & Cavalcanti, A. R. (2008). Factors influencing codon usage bias in genomes. *Journal of the Brazilian Chemical Society*, *19*(2), 257-262.

Sampei, G. I., Kanagawa, M., Baba, S., Shimasaki, T., Taka, H., Mitsui, S., ... & Terao, K. (2013). Structures and reaction mechanisms of the two related enzymes, PurN and PurU. *The Journal of Biochemistry*, *154*(6), 569-579.

Saunders, D. G., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., & Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PloS one*, *7*(1), e29847.

Schardl, C. L. (2002). *Epichloë festucae* and related mutualistic symbionts of grasses. *Fungal Genetics and Biology*, *1*(35), 79.

Schnell, D. E. (2002). *Carnivorous plants of the United States and Canada* (No. Ed. 2). Timber Press.

Schiltz, S., Gaillard, I., Pawlicki-Jullian, N., Thiombiano, B., Mesnard, F., & Gontier, E. (2015). A review: what is the spermosphere and how can it be studied?. *Journal of applied microbiology*, *119*(6), 1467-1481.

Schoenwetter, T., Lyon, L. M., Hardesty, L. H., & Seddon, W. (2006). The use of *Nepenthes madagascariensis* in traditional medicine and healing among the people of Fort Dauphin, Madagascar. In *6th Conference of the International Carnivorous Plant Society. Frostburg*.

Schulze, W., Schulze, E. D., Pate, J. S., & Gillison, A. N. (1997). The nitrogen supply from soils and insects during growth of the pitcher plants *Nepenthes mirabilis*, *Cephalotus follicularis* and *Darlingtonia californica*. *Oecologia*, *112*(4), 464-471.

Schulze, W., Schulze, E. D., Schulze, I., & Oren, R. (2001). Quantification of insect nitrogen utilization by the venus fly trap *Dionaea muscipula* catching prey with highly variable isotope signatures. *Journal of Experimental Botany*, *52*(358), 1041-1049.

Schwessinger, B., Sperschneider, J., Cuddy, W., Miller, M., Garnica, D., Taylor, J., ... & Rathjen, J. (2017). A near complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis f. sp. tritici* reveals high inter-haplome diversity. *bioRxiv*, 192435.

Scott, R. I., Chard, J. M., Hocart, M. J., Lennard, J. H., & Graham, D. C. (1996). Penetration of potato tuber lenticels by bacteria in relation to biological control of blackleg disease. *Potato Research*, *39*(3), 333-344.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068-2069.

Siragusa, A. J., Swenson, J. E., & Casamatta, D. A. (2007). Culturable bacteria present in the fluid of the hooded-pitcher plant *Sarracenia minor* based on 16S rDNA gene sequence data. *Microbial ecology*, *54*(2), 324-331.

Slippers, B., & Wingfield, M. J. (2007). *Botryosphaeriaceae* as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal biology reviews*, *21*(2-3), 90-106.

Smith, J. G. (1893). Recent studies of carnivorous plants. *The American Naturalist*, *27*(317), 413-420.

Sonnhammer, E. L., Von Heijne, G., & Krogh, A. (1998, July). A hidden Markov model for predicting transmembrane helices in protein sequences. In *Ismb* (Vol. 6, pp. 175-182).

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., ... & Taylor, J. M. (2016). EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*, *210*(2), 743-761.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*(suppl_2), ii215-ii225.

Staples, R. C., & Hoch, H. C. (1987). Infection structures-form and function. *Exp. Myco*.

Staskawicz, B. J., Dahlbeck, D., & Keen, N. T. (1984). Cloned avirulence gene of Pseudomonas syringae pv. glycinea determines race-specific incompatibility on Glycine max (L.) Merr. *Proceedings of the National Academy of Sciences, 81*(19), 6024-6028.

Stillman, B. (1994). Smart machines at the DNA replication fork. *Cell*, *78*(5), 725-728.

Szabo, L. J., & Bushnell, W. R. (2001). Hidden robbers: the role of fungal haustoria in parasitism of plants. *Proceedings of the National Academy of Sciences*, *98*(14), 7654-7655.

Takemoto, D., Jones, D. A., & Hardham, A. R. (2003). GFP-tagging of cell components reveals the dynamics of subcellular re-organization in response to infection of *Arabidopsis* by oomycete pathogens. *The Plant Journal*, *33*(4), 775-792.

Tan, J. Y., Yin, W. F., & Chan, K. G. (2014). Gene clusters of *Hafnia alvei* strain FB1 important in survival and pathogenesis: a draft genome perspective. Gut pathogens, 6(1), 29.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Ciufo, S., & Li, W. (2013). Prokaryotic genome annotation pipeline. *The NCBI Handbook [Internet]. 2nd edition*. National Center for Biotechnology Information (US).

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., & Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome research*, *18*(12), 1979-1990.

Tian, M., Huitema, E., da Cunha, L., Torto-Alalibo, T., & Kamoun, S. (2004). A Kazal-like extracellular serine protease inhibitor from Phytophthora infestans targets the tomato pathogenesis-related protease P69B. *Journal of Biological Chemistry*, *279*(25), 26370-26377.

Torto, T. A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A., ... & Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. *Genome research*, *13*(7), 1675-1685.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105-1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, *28*(5), 511.

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36.

Trzcinski, M. K., Walde, S. J., & Taylor, P. D. (2005). Stability of pitcher-plant microfaunal populations depends on food web structure. *Oikos*, *110*(1), 146-154.

Tsuda, K., & Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current opinion in plant biology*, *13*(4), 459-465.

Upadhyaya, N. M., Mago, R., Staskawicz, B. J., Ayliffe, M. A., Ellis, J. G., & Dodds, P. N. (2014). A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Molecular Plant-Microbe Interactions*, *27*(3), 255-264.

Uppalapati, S. R., Serba, D. D., Ishiga, Y., Szabo, L. J., Mittal, S., Bhandari, H. S., ... & Saha, M. C. (2013). Characterization of the rust fungus, *Puccinia emaculata*, and evaluation of genetic variability for rust resistance in switchgrass populations. *Bioenergy research*, *6*(2), 458-468.

Van Der Biezen, E. A., & Jones, J. D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences*, *23*(12), 454-456.

Van der Hoorn, R. A., & Kamoun, S. (2008). From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell*, *20*(8), 2009-2017.

Van der Merwe, M., Ericson, L., Walker, J., Thrall, P. H., & Burdon, J. J. (2007). Evolutionary relationships among species of *Puccinia* and *Uromyces* (Pucciniaceae, Uredinales) inferred from partial protein coding gene phylogenies. *Mycological research*, *111*(2), 163-175.

Voegele, R. T., & Mendgen, K. (2003). Rust haustoria: nutrient uptake and beyond. *New Phytologist*, *159*(1), 93-100.

Voegele, R. T., Struck, C., Hahn, M., & Mendgen, K. (2001). The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proceedings of the National Academy of Sciences*, *98*(14), 8133-8138.

Walterson, A. M., & Stavrinides, J. (2015). Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiology Reviews*, *39*(6), 968-984.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., ... & Machi, D. (2013). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic acids research*, *42*(D1), D581-D591.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., ... & Gerdes, S. (2016). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic acids research*, *45*(D1), D535-D542.

Williams, D. D. (2006). *The biology of temporary waters*. Oxford University Press.

Wilson, D. (1995). Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 274-276.

Wright, L. L., Cushman, J. H., Ehrenshaft, A. R., McLaughlin, S. B., McNabb, W. A., Ranney, J. W., ... & Turhollow, A. F. (1992). *Biofuels Feedstock Development Program annual progress report for 1991* (No. ORNL--6742). Oak Ridge National Lab., TN (United States).

Wright, L., & Turhollow, A. (2010). Switchgrass selection as a “model” bioenergy crop: a history of the process. *Biomass and Bioenergy*, *34*(6), 851-868.

Wu, T. D., & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859-1875.

Yin, Z., Yuan, C., Du, Y., Yang, P., Qian, C., Wei, Y., ... & Liu, B. (2019). Comparative genomic analysis of the Hafnia genus reveals an explicit evolutionary relationship between the species alvei and paralvei and provides insights into pathogenicity. *BMC genomics*, *20*(1), 1-16.

Young, E. B., Sielicki, J., & Grothjan, J. J. (2018). Regulation of hydrolytic enzyme activity in aquatic microbial communities hosted by carnivorous pitcher plants. *Microbial ecology*, *76*(4), 885-898.

Zale, J., Freshour, L., Agarwal, S., Sorochan, J., Ownley, B. H., Gwinn, K. D., & Castlebury, L. A. (2008). First report of rust on switchgrass (*Panicum virgatum*) caused by *Puccinia emaculata* in Tennessee. *Plant Disease*, *92*(12), 1710-1710.

Zander, A. (2016). *Interactions of" Sarracenia purpurea" and its inquilines* (Doctoral dissertation, Université de Fribourg).

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., ... & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of antimicrobial chemotherapy*, *67*(11), 2640-2644.

Zannoni, D. (1995). Aerobic and anaerobic electron transport chains in anoxygenic phototrophic bacteria. In *Anoxygenic photosynthetic bacteria* (pp. 949-971). Springer, Dordrecht.

Zhang, S., Chaluvadi, S. R., & Bennetzen, J. L. (2020). Draft Genome Sequence of a *Serratia marcescens* Strain Isolated from the Pitcher Fluids of a *Sarracenia* Pitcher Plant. *Microbiology Resource Announcements*, *9*(2).

Zhang, S., Chaluvadi, S. R., & Bennetzen, J. L. (2020). Genome Sequence Analysis of *Enterobacter* sp. C6, Found in the Pitcher Fluids of *Sarracenia rosea*. *Microbiology Resource Announcements*, *9*(3).

Zhang, Y., Morar, M., & Ealick, S. E. (2008). Structural biology of the purine biosynthetic pathway. *Cellular and molecular life sciences*, *65*(23), 3699-3724.

Zhang, Y., Zalapa, J., Jakubowski, A. R., Price, D. L., Acharya, A., Wei, Y., ... & Casler, M. D. (2011). Natural hybrids and gene flow between upland and lowland switchgrass. *Crop Science*, *51*(6), 2626-2641.

Zong, B., Zhang, Y., Wang, X., Liu, M., Zhang, T., Zhu, Y., ... & Tan, C. (2019). Characterization of multiple type-VI secretion system (T6SS) VgrG proteins in the pathogenicity and antibacterial activity of porcine extra-intestinal pathogenic *Escherichia coli*. *Virulence*, *10*(1), 118-132.

APPENDIX A

DRAFT GENOME SEQUENCE OF A *SERRATIA MARCESCENS* STRAIN ISOLATED FROM THE PITCHER FLUIDS OF A *SARRACENIA* PITCHER PLANT[1]

---

[1]Zhang S, Chaluvadi SR, Bennetzen JL. 2020. Draft genome sequence of a *Serratia marcescens* strain isolated from the pitcher fluids of a *Sarracenia* pitcher plant. Microbiol Resour Announc 9:e01216-19. Reprinted here with permission of publisher.

ABSTRACT

The genome of a *Serratia marcescens* strain that was found in the pitcher fluids of a *Sarracenia rosea* pitcher plant was sequenced using the Illumina platform. A 5,543,750 bp genome assembly was obtained. A total of 6278 coding sequences are predicted from this assembly.

ANNOUNCEMENT

Serratia marcescens is a species of facultatively anaerobic bacteria that stains as Gram-negative rods (1). The genus *Serratia* is in the family Enterobacteriaceae (1). *S. marcescens* strains have been previously isolated from soil (2) and plant tissue (3). Some strains can promote enhanced plant growth and confer abiotic and/or biotic stress tolerance (3, 4, 5). *S. marcescens* has also been isolated from humans (6). Some *S. marcescens* strains cause various diseases, such as meningitis (7), urinary tract infections, and wound infections (8). *S. marcescens* has also been found in infections of other animals, including some insects (9).

The strain investigated in this study was isolated from the pitcher fluids of *Sarracenia rosea* pitcher plants from Splinter Hill Bog, Alabama (collected in March 2012). The strain was isolated from pitcher fluids by plating on media used for the culture of *Pyrococcus furiosus* (10), and single colonies were serially streaked onto the same media three times, and then an individual colony was picked into nutrient broth (11) for growth and eventual storage at -80°C in 25% glycerol. DNA was extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research Cat No. D6005). A paired-end library was

prepared using Nextera XT DNA Library Preparation Kit V2 (Illumina Cat No. FC-131-1002) with average insert size of ~400 bp for Illumina MiSeq sequencing. Out of 17,917,430 Illumina raw reads, we retained 5,643,558 reads after quality filtering with Trimmomatic (version 0.36) (12) . Then, FastQC (version 0.11.4) was employed to check the quality of the trimmed reads (13) . SoapDenovo2 (version r240, parameter -K 83) was used to assemble the reads into contigs (14). Contigs with lengths <400 bp were discarded. In total, 1306 contigs were selected to be ordered in Mauve (15) using *Serratia marcescens* strain FDAARGOS_65 (GenBank accession No. NZ_CP026050.1) as the reference genome. The draft genome was annotated using the online RAST server (16). Default parameters were used for all software tools unless otherwise noted.

The final genome assembly is 5,543,750 bp, with a 58.1% GC content. The genome coverage is ~300 X. The N50 is 22,377 bp and the L50 value is 80. There are a predicted 6278 coding sequences in total, consisting of 6189 protein coding sequences and 89 RNA genes. Only 34.2% of the protein coding sequence (2122 genes) for this genome could be grouped into the functional subsystems of the RAST SEED server (17). The subsystem category distribution is shown in Figure 1.

CGE ResFinder (18) was used to predict antibiotic resistance genes. The genome assembly was predicted to contain Beta-lactam resistance gene *blaACT-6* and aminoglycoside resistance gene *aac(6')-Ic*. CGE PathogenFinder (19) predicted that the strain has a 77.2% probability of being a human pathogen. The Average Nucleotide Identities (20) from comparing to the two most closely related bacteria (identified with BLASTn) in the NCBI database, *Serratia marcescens* strain AS1 (CP010584) and *Serratia*

marcescens strain UMH5 (CP018917), are both ~94.2%. The average amino acid identities are 96.1% and 96.0%, respectively. Using Patric (21, 22) for comparison with these two other *S. marcescens* strains (both from clinical samples), the pitcher *S. marcescens* has a unique fluorobenzoate degradation pathway and an increase in the number of genes involved in sulfur metabolism. This comparison also demonstrated that the pitcher strain lacks carbamate kinase (EC 2.7.2.2), but has two more glutamate ammonia ligase (EC 6.3.1.2) and 2 more glutamate synthase NADPH (EC 1.4.1.13) genes involved in nitrogen metabolism.

Data availability. This Whole Genome Shotgun project has been deposited in GenBank under accession no. NZ_QPFX01000000. The raw read accession number is SRX6867789. The version described in this paper is the first version. The BioProject number is PRJNA481376. The BioSample number is SAMN09666492.

ACKNOWLEDGMENTS

We wish to thank Willie L Rogers and Dr. Russell L Malmberg for providing pitcher fluids collected from the Splinter Hill Bog, Dr. Sanjeev Chandrayan for his help with anaerobic culturing, and Macrogen, Inc. for Illumina sequencing. This research was supported by the Giles Endowment to JLB.

REFERENCES

1. Hejazi A, Falkiner F. 1997. *Serratia marcescens*. *J Med Microbiol* 46:903-912.

2. Ashelford KE, Fry JC, Bailey MJ, Day MJ. 2002. Characterization of Serratia isolates from soil, ecological implications and transfer of *Serratia proteamaculans* subsp. quinovora Grimont et al. 1983 to Serratia quinivorans corrig., sp. nov. *Int J Syst Evol Microbiol* 52:2281-2289

3. Gyaneshwar P, James EK, Mathan N, Reddy PM, Reinhold-Hurek B, Ladha JK. 2001. Endophytic colonization of rice by a diazotrophic strain of *Serratia marcescens*. *J Bacteriol* 183:2634-2645

4. Matteoli FP, Passarelli-Araujo H, Reis RJA, da Rocha LO, de Souza EM, Aravind L, Olivares FL, Venancio TM. 2018. Genome sequencing and assessment of plant growth-promoting properties of a *Serratia marcescens* strain isolated from vermicompost. *BMC Genomics* 19:750-750

5. Singh RP, Jha PN. 2016. The multifarious PGPR *Serratia marcescens* CDP-13 augments induced systemic resistance and enhanced salinity tolerance of wheat (*Triticum aestivum* L.). *PloS one* 11:e0155026-e0155026

6. Laupland K, Parkins M, Gregson D, Church D, Ross T, Pitout J. 2008. Population-based laboratory surveillance for *Serratia* species isolates in a large Canadian health region. *Eur J Clin Microbiol Infect Dis* 27:89.

7. Rabinowitz K, Schiffrin R. 1952. A ward-contamination by *Serratia marcescens*. *Acta Medica Orientalia* 11:181-4.

8. LANCASTER LJ. 1962. Role of *Serratia* species in urinary tract infections. *Arch Intern Med (Chic)* 109:536-539.

9. Mahlen SD. 2011. *Serratia* infections: from military experiments to current practice. *Clin Microbiol Rev* 24:755-791.

10. Adams MW, Holden JF, Menon AL, Schut GJ, Grunden AM, Hou C, Hutchins AM, Jenney FE, Kim C, Ma K. 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* 183:716-724.

11. Greenberg AE, Trussell RR and Clesceri LS. (Eds.). 1985. Standard methods for the examination of water and wastewater, 16th ed., APHA, Washington, D.C.

12. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

13. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*, Babraham Institute, Cambridge, United Kingdom.

14. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18.

15. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394-1403.

16. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.

17. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691-5702.

18. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640-2644.

19. Cosentino S, Larsen MV, Aarestrup FM, Lund O. 2013. PathogenFinder-distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 8:e77302.

20. Rodriguez-R L, Konstantinidis K. 2014. Bypassing cultivation to identify bacterial species. *Microbe* 9: 111–118.

21. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R. 2013. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581-D591.

22. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL. 2016. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 45:D535-D542.

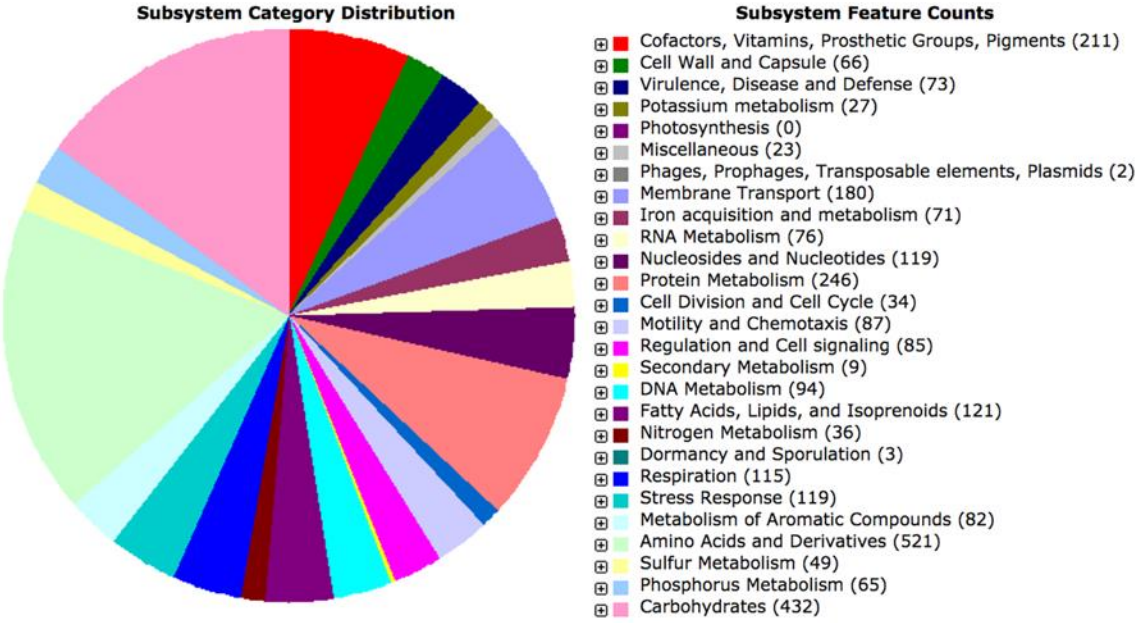


FIG 1 Subsystem distribution based on RAST SEED. The pie chart organizes presented subsystem by cellular process. “Subsystem Feature Counts” indicates the number of protein encoding genes (in parentheses) that are predicted to be involved in that cellular process.

APPENDIX B

GENOME SEQUENCE ANALYSIS OF *ENTEROBACTER* SP. C6 FOUND IN THE PITCHER FLUIDS OF *SARRACENIA ROSEA*[2]

---

[2]Zhang S, Chaluvadi SR, Bennetzen JL. 2020. Genome sequence analysis of *Enterobacter* sp. C6, found in the pitcher fluids of *Sarracenia rosea*. Microbiol Resour Announc 9:e01214-19. Reprinted here with permission of publisher.

ABSTRACT

An unclassified *Enterobacter* strain was isolated from the pitcher fluids of a *Sarracenia rosea* pitcher plant growing at Splinter Hill Bog in Alabama, USA. Its genome was sequenced using the Illumina platform. A genome assembly of 4,673,815 bp was obtained. 4646 protein-encoding sequences and 72 RNA genes are predicted from this assembly.

ANNOUNCEMENT

The genus *Enterobacter* was first proposed in 1960 (1). *Enterobacter* belongs to the Enterobacteriaceae, a group of gram-negative, facultatively anaerobic, motile and straight rod bacteria (2). *Enterobacter* are widespread in aquatic environments, and are also present in the animal intestine (3). *Enterobacter* species are sometimes found as pathogens that cause nosocomial infections, such as urinary tract infections and bacteremia (4).

The strain used in this study was isolated from the fluids of *S. rosea* pitcher plants from Splinter Hill Bog in March of 2012. Pitcher fluids were plated on media used for the culture of *Pyrococcus furiosus* (5), and streaked from single colonies onto the same media. An individual colony was then picked into nutrient broth for growth and eventual storage at -80°C in 25% glycerol. DNA was extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research Cat No. D6005). A paired-end library was prepared using Nextera XT DNA Library Preparation Kit V2 (Illumina Cat No FC-131-1002) with average insert size of ~400 bp for Illumina MiSeq sequencing. Out of 14,784,154 Illumina raw reads, we retained 4,889,590 reads after quality filtering with Trimmomatic (version 0.36)

(6). FastQC (version 0.11.4) was employed to evaluate the quality of the cleaned reads (7). SoapDenovo2 (version r240, parameter -K 83) was used to assemble the reads into contigs (8). In total, 252 contigs (all >400 bp), with sequence coverage of more than 20X were achieved. Both the NCBI Prokaryotic Genome Annotation Pipeline (9) and the online RAST server (RASTtk Toolkit) were used for annotation (10). Annotation by the NCBI Prokaryotic Genome Annotation Pipeline was provided with the the GenBank submission. Default parameters were used for all software tools unless otherwise noted.

The final assembly has 4,673,815 bp. The GC content is ~55.9%. The N50 and L50 are 41,959 bp and 33, respectively. The sequence depth of the genome assembly is ~260 X. The genome contains a predicted 4718 genes, consisting of 4646 protein-encoding sequences and 72 RNA genes. About 38% of the protein-encoding sequence (1764 genes) for this genome were grouped in the functional subsystems of the RAST SEED server (10). Beta-lactam resistance gene *blaACT-6* and fosfomycin resistance gene *fosA* were predicted by the CGE ResFinder (11). The strain is predicted to have a 75.5% probability of being a human pathogen according to CGE PathogenFinder (12). The average nucleotide identities (ANI) when compared to the two most closely related genomes, discovered by an NCBI BLASTn analysis using the nucleotide collection (nt/nr) database, are *Enterobacter cloacae* strain R11 (NZ_CP019839) and *Enterobacter cloacae* subsp. *cloacae* ATCC 13047 (CP001918). The nucleotide identities of these two strains, when compared to *E.* sp. C6, are 93.7% and 87.9%, respectively (13). Several recent studies have demonstrated that a >95% ANI value is indicative of the same species (14). Because both ANI values are

<95% in comparison betweens *E.* sp. C6 and either *E. cloacae* strain, this suggests that *E.* sp. C6 is a new species of *Enterobacter*.

Data availability. This Whole Genome Shotgun project has been deposited in GenBank under accession no. QPFW00000000.1. The raw read accession number is SRX6867760. The version described in this paper is the first version. The BioProject number is PRJNA476218, and the BioSample number is SAMN09428499.

ACKNOWLEDGMENTS

We wish to thank Willie L Rogers and Dr. Russell L Malmberg for providing pitcher fluids collected from the Splinter Hill Bog, Dr. Sanjeev Chandrayan for his help with anaerobic culturing and Macrogen, Inc. for Illumina sequencing. This research was supported by the Giles Endowment to JLB.

REFERENCES

1. Hormaeche E, Edwards P. 1960. A proposed genus *Enterobacter*. *Int Bull Bacteriol Nomencl Taxon* 10:71-4.

2. Tan W-S, Yunos N, Tan P-W, Mohamad N, Adrian T-G-S, Yin W-F, Chan K-G. 2014. Freshwater-borne bacteria isolated from a Malaysian rainforest waterfall exhibiting quorum sensing properties. *Sensors* 14:10527-10537.

3. Cabral JP. 2010. Water microbiology. Bacterial pathogens and water. *Int J Environ Res Public Health* 7:3657-3703.

4. Grimont F, Grimont PA. 2006. The Genus *Enterobacter*. The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass:197-214.

5. Adams MW, Holden JF, Menon AL, Schut GJ, Grunden AM, Hou C, Hutchins AM, Jenney FE, Kim C, Ma K. 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* 183:716-724.

6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

8. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18.

9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucl Acids Res* 44:6614-6624.

10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics* 9:75.

11. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640-2644.

12. Cosentino S, Larsen MV, Aarestrup FM, Lund O. 2013. PathogenFinder-distinguishing friend from foe using bacterial whole genome sequence data. *PloS one* 8:e77302.

13. Rodriguez-R L, Konstantinidis K. 2014. Bypassing cultivation to identify bacterial species. *Microbe* 9: 111–118.

14. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9:5114.