ABSTRACT

LIEBELT, DONNA JEAN. Analysis of Temporal and Tissue Specific *Humulus lupulus* Isoforms by Long-Read Sequencing (Under the direction of Dr. Colleen Doherty).

The phytochemical composition in plants is dependent on environmental factors, both directly and through daily and seasonal rhythms. We attribute the value of many consumer products and pharmaceuticals to plant phytochemicals. As influential environmental factors continue to change, including the global rise in temperatures, and the feedback this has on extreme weather events, we must understand the role environmental factors have in signaling regulatory pathways within specialty crops. This understanding has the potential to aid decisions in agricultural management and plant breeding.

Humulus lupulus (hops) can be a model perennial specialty crop for studying the effects of the environment on metabolites as breeders consider composition carefully, and they are grown in diverse climates and regions. Hops, while most known for their role in beer production, are also a rich source of nutraceuticals, including antioxidants, antimicrobials, and essential oils. Though their metabolite profiles are extensive, there are restrictions in the tools available to further examine biochemical pathways due to a lack of detailed annotations. One way to address this challenge is to use long-read sequencing such as single-molecule real-time (SMRT) sequencing.

SMRT sequencing, a third-generation sequencing technology developed by Pacific Biosciences optimized to capture entire transcripts. Like any new technology of its kind, the bioinformatic analysis and compatible software for isoform processing are continuously evolving. PacBio's SMRT Analysis software and compatible packages for post-processing are enormously valuable in analyzing Iso-Seq and isoform data; however, they can be complicated and challenging to understand and troubleshoot for biological scientists without a history in coding and bioinformatics. With the advancement of sequencing applications as well as the rise in implementation of interdisciplinary teams, inexperienced researchers can either expect to have to learn some level of bioinformatics to analyze the data they generate or pay to outsource the analysis. Basic processes to analyze the data have been established. However, due to the diversity of available software, operating systems, and hardware, it is difficult for a non-bioinformatician to understand the end-to-end workflow. Herein we provide a beginner-friendly end-to-end workflow for Iso-Seq and isoform analysis. The workflow commentary can be used to both understand the coding required as well as the biological relevance of these steps, bridging the divide between them.

With this workflow applied, we analyzed long-read isoforms in *Humulus lupulus* Teamaker variety leaves, roots, and shoots sampled at dawn and dusk. We produced 373,555 full-length reads, which accounted for only ~40% of the total reads detected when mapped to the current Teamaker genome assembly. Over 5,000 isoforms are time-of-day or tissue-specific, and nearly 4,000 unique isoforms that

are dependent on both tissue type and time of day. Future directions include analyzing cone tissues, as well as environmental factors such as temperature.

While temperatures are steadily increasing, night temperatures are warming faster than day temperatures. Daily temperatures cycles are known to entrain endogenous rhythms in plants. The decreasing difference between day and night temperatures poses an interesting question of how plant composition will be affected by a change in this signal. As proof of principle, we examine the effects of three-degree warmer nights on the metabolite composition of broccoli and radish sprouts. Further exploration could reveal plant varieties and environmental conditions that maximize the desired phytochemical profiles to benefit human health. Additionally, we could gain insights into the impact of changing environments on the regulation of phytochemicals, which could be applied to other crop systems targeted for the production of pharmaceuticals and other consumer products.

Analysis of Temporal and Tissue Specific *Humulus lupulus* Isoforms by Long-Read Sequencing

by
Donna Jean Liebelt

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Biochemistry

Raleigh, North Carolina

2020

APPROVED BY:

_____	_____
Colleen Doherty	Flora Meilleur
Chair of Advisory Committee	

_____
Michael Goshe

BIOGRAPHY

Donna was born in 1995 in central New Jersey. In 2013 she graduated from Allegan High School in western Michigan and by 2017 completed her bachelor's degree in Biochemistry at Michigan State University with a minor in Biotechnology. Her undergraduate experience working both in a biochemistry research laboratory and at the university greenhouses led Donna to continue her education in a marriage of the two; plant biochemistry. She went on to higher education in Biochemistry, joining the College of Agriculture and Life Sciences Graduate School at North Carolina State University, where she completed her masters under the direction of Dr. Colleen Doherty. Donna also has a passion for art and adventure; she enjoys concerts and dancing as well as hiking and camping, and occasionally shredding on the ukulele. She finds inspiration in music and nature.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

TIMING IS EVERYTHING: THE IMPACT OF TEMPORAL RHYTHMS ON PHYTOCHEMICALS

1.1 Abstract

Daily and seasonal fluctuations in temperature, light, humidity, and precipitation directly influence the rhythmic regulation of plant molecular activity. These environmental inputs signal a complex network of interconnecting and overlapping elements that not only affect each other but also the plant's endogenous circadian clock. The resulting regulatory network tightly coordinates a dynamic phytochemical composition to the plant's immediate needs. Temporal control of metabolism ensures phytochemicals are in tune with the demands of the environment, including available resources. Consequently, plant phytochemical composition varies both daily and seasonally (Figure1.1). Therefore, experimental results and conclusions can be directly affected by the time in which the plant sample is analyzed. It is critical that we understand temporal variations in phytochemical composition to reveal underlying regulatory connections that can later be used in the improvement of phytochemical products. This chapter will review the mechanisms underlying temporal variations in phytochemicals with examples across phytochemical classes, and the implications of climate change on plant quality and composition.

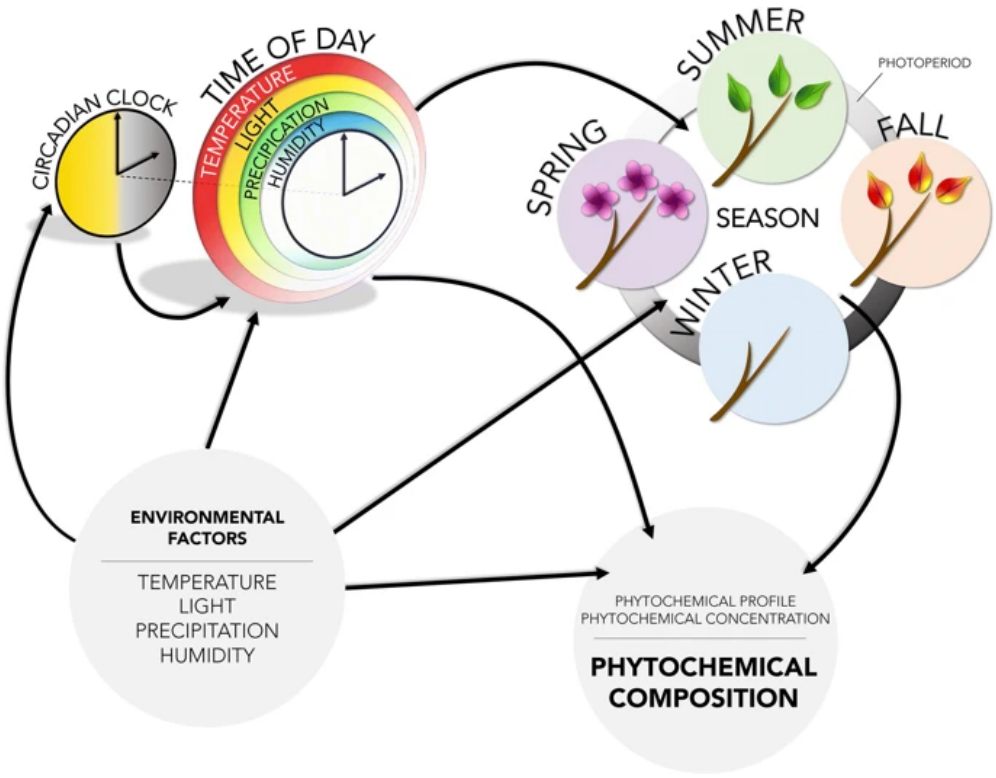


Figure 1.1 ***Temporal Regulation of Phytochemical Composition.*** Plant phytochemical composition is controlled by a complex amalgamation of environmental signals and endogenous circadian rhythms. Environmental signals can vary dramatically throughout the day and year. Daily cycles in environmental factors entrain the endogenous circadian clock which imposes additional regulation on the abundance and activity of many phytochemicals. This concerted effort to tune phytochemicals to the demands of the environment results in a temporally dynamic metabolome. As a result, misleading or inconsistent estimations of the composition and potency of phytochemical extractions can occur with a failure to consider the time of day and year of the extraction. Integrating temporal factors will improve our understanding of the underlying regulatory connections and ultimately improve the quality of phytochemical products. (Liebelt et. al 2019)

1.2 Introduction

The composition, relative concentration, and potency of bioactive molecules determine the perceived value of a plant as a source of phytochemicals. All of these factors can vary as highly plastic plant biochemistry responds dynamically to the environment. Research on specialty crops producing phytochemicals has shown that value-added phenotypes from aesthetics to flavor and texture vary with environmental changes. These variations can impact shelf life, food safety, and health benefits which in

turn drive value and consumer perception of quality (Ahmed and Stepp 2016). These variations can have a financial impact on producers and retailers, as well as impact food waste as devalued or unaesthetic products are discarded (Buzby et al. 2014). For example, consumers can taste the difference between tea leaves that are grown in the sun versus shade (Ahmed et al. 2010) or from the monsoon harvest season versus the dry spring harvest (Ahmed et al. 2014). Understanding environmental factors that alter the phytochemical composition of specialty crops can improve breeding and agricultural production.

Although variation in primary metabolism at different times of day or seasons of the year have been well-established (Farré and Weise 2012), little is known on the temporal regulation of specialized metabolites. Nutraceutical efficacy including phytochemical toxicity and potency can be altered by temporal changes to metabolite concentrations. Moreover, understanding the temporal variation in specialized metabolites can improve selection and breeding strategies, increasing the consistency of phytochemical production and research. This chapter will focus on the impact of recurring daily and seasonal cycles on phytochemical biochemistry through a general overview of the mechanisms that can drive daily and seasonal variation in phytochemicals with examples of compounds in various phytochemical classes that show daily and seasonal variation. A final discussion will review how current trends are changing these rhythmic environmental patterns and the impact this may have on phytochemical research and production.

Several plant species have shown either circadian, diel, or photoperiod-driven rhythmic expression of genes involved in specialized metabolite regulatory and biosynthetic pathways (Alabadí et al. 2002; Filichkin et al. 2011; Pavarini et al. 2012; Gyllenstrand et al. 2014; Soni et al. 2015; Fenske and Imaizumi 2016; Zeng et al. 2017; Greenham et al. 2017; Koda et al. 2017; Weiss et al. 2018; Yin et al. 2018; Yoshida and Oyama-okubo 2018). The correlation between changes in transcript and metabolite levels will depend on the environmental conditions and the specific pathway of interest.

1.3 Daily cycles drive variation in phytochemical production and processing

Plants coordinate their molecular activities by time to operate at optimum efficiency. Resource availability, primarily light and water, provide timing cues for plant growth, stomatal opening, photosynthesis, metabolism, nutrient, and water uptake (Nozue and Maloof 2006; Harmer 2010). Four factors (light intensity, light quality, temperature, and humidity) have a direct impact on nutrient uptake and photosynthesis which impacts the abundance and distribution of carbon and nitrogen in the plant within a 24 hour period (Rufty et al. 1989; Bläsing et al. 2005; Ruts et al. 2012), thus modulating the availability of the building blocks necessary to generate specialized metabolites. Diel environmental factors also directly influence the pathways that regulate specialized metabolite composition. For example, rhythmic fluctuations in temperature, humidity, light intensity, and quality also modulate the

plant's circadian clock so that it is in sync with the local environment (McClung 2006, 2008; Harmer 2010; Mwimba et al. 2018). Production and processing of specialized metabolites are regulated by a plant's endogenous circadian clock (Bläsing et al. 2005). Altogether, the complex pathways generated by these daily cycles help 'gate' physiological and molecular responses to abiotic and biotic stresses so that the potential for response is maximal when the stress stimulus is most likely to occur.

1.4 Regulation of Phytochemicals by Changes in Environment

For many phytochemicals, accumulation levels are directly affected by changes in light, temperature, and water availability. Variations in these environmental factors throughout the day and year can result in their altered phytochemical compositions depending on the time of day or year the plant is extracted.

1.4.1 Light quality

Earth's atmosphere can affect light intensity depending on the position of the sun as it moves through the sky both daily and seasonally. Additionally, inconsistent atmospheric filtering across wavelengths, as well as the angle of the sun, impact light spectrum availability to plants. Rayleigh scattering strongly affects blue light and other shorter wavelengths, resulting in a lower abundance of these in winter and at dawn and dusk, leading to comparatively increased amounts of the longer, red wavelengths. Red to far-red ratios impact developmental transitions such as flowering. In Arabidopsis, light quality sensitivity is achieved with phytochromes by a mechanism of switching between activity states (Ulijasz and Vierstra 2011; Ushijima et al. 2017; Sethe et al. 2017). These phytochromes impact photosynthesis and primary metabolism (Yang et al. 2016a, b; Kreslavski et al. 2018). Phytochrome loss increases sugars and amino acids, a change that is similar to abiotic stress responses. Specifically, the accumulation of proline and raffinose (Yang et al. 2016a, b). Phytochrome mutants also induce transcriptional response consistent with abiotic stress, suggesting a connection between light quality sensing and abiotic stress pathways.

Several phytochemical pathways are sensitive to changes in light quality. Both red and blue light enhance total antioxidant activity in the Chinese medicinal herb, *Rehmannia glutinosa* (Manivannan et al. 2015). Total phenol content and flavonoid levels were responsive to specific wavelengths. Blue light resulted in higher phenol content, while red-light favored an increase in flavonoid levels. Red light induction of flavonoids was also reported in *Pisum sativum* (Bottomley et al. 1966). However, in other species, flavonoid levels increase with blue light. In the medicinal plant, *Cyclocarya paliurus* total flavonoid levels were the most responsive to blue light (Liu et al. 2018). The levels of the flavonoids quercetin and 3-malonylglucoside and the phenylpropanoid chicoric acid increase under prolonged

enhanced blue light in red leaf lettuce and basil (Taulavuori et al. 2016). In *Betula pendula* (silver birch) seedlings, UV-B supplementation also increased levels of quercetins, kaempferols, and chlorogenic acids (Tegelberg et al. 2004). Total essential oils increased under blue and red light in *Mentha piperita, M. spicata*, and *M. longifolia* compared to white light or sunlight (Sabzalian et al. 2014). However, like flavonoids, the effect of light quality on essential oils tends to be species-specific. In the Sabzalian et al. (2014) study, the three *Mentha* species showed significantly higher levels of essential oils under red light compared to blue light. Red light also increased menthol levels, the primary essential oil compound in *Mentha arvensis,* higher than blue light (Nishioka et al. 2008). In *Mentha piperita,* menthol biosynthesis was reduced by 25% when supplemented with blue light and resulted in a significant decrease in essential oil and total phenol content (Maffei and Scannerini 1999). A recent review by Dou et al. (2017) provides other examples of light quality impacts on phytochemicals in herb species. The seasonal changes in light quality are the greatest farthest from the equator, suggesting that there will be an interaction between seasonal variability and latitude.

1.4.2 Temperature

The variation in temperature throughout the year, particularly in temperate climates, can be quite drastic. Perhaps less obvious is that the day to night temperature differential can also be significant, even in tropical climates. This variation in temperature across the time of day and season of the year can also result in a corresponding variability in phytochemical levels. Total glucosinolate and ascorbic acid levels increased in sprouts germinated at 30°C compared to those at 20°C or 10°C in broccoli (*Brassica oleracea*) and rocket sprouts (*Eruca sativa*) (Ragusa et al. 2017). Increasing temperature enhanced the total polyphenol content of broccoli sprouts but reduced the total polyphenol amount in rocket sprouts. Increasing temperature (20°C, 25°C, and 30°C) also resulted in significantly higher phenolic acid and flavonoid production in wheat (*Triticum spp*), irrespective of genotype (Shamloo et al. 2017). Campesterol levels decreased across genotypes, while other phytosterols in wheat were not significantly affected by the temperature change. St. John's wort (*Hypericum perforatum)* harvested after 15 days of elevated temperatures (30°C or 35°C) show increased concentration of anthraquinones, hypericin and pseudohypericin, and the phloroglucinol, hyperforin (Zobayed et al. 2005). These are considered the active compounds in *H. perforatum* (Bauer et al. 2001) and their sensitivity to temperature may result in different potencies for harvested plant material.

In *Zea mays* (Christie et al. 1994), Arabidopsis (Leyva et al. 1995; Rowan et al. 2009), *Nicotiana tabacum* (Huang et al. 2012) and apple (*Malus spp*) (Ubi et al. 2006) anthocyanin levels are increased by lower temperatures and reduced in higher temperatures. Changes in phytochemical levels in response to temperature are dynamic processes and evaluating the response to temperature at a single time point may

not provide a clear picture of the full intersection between the changing temperature and the biochemical response. For example, in plum fruit (Pr*unus salicina),* the initial response to high temperature is an increase in anthocyanin, however, after nine days at high temperatures, the total anthocyanin dropped below control levels (Niu et al. 2017). Monitoring the activity of a hydrogen peroxide-based degradation mechanism showed that the accumulation of anthocyanin is a balance between synthesis and degradation. The reactive oxygen-based degradation suggests other environmental factors may modulate the temperature-induced degradation of anthocyanin in plum fruit.

1.4.3 Water availability

For most plants, the availability of water varies daily and seasonally. Diel oscillations in water potential are driven by active photosynthesis and associated stomatal transpiration. At dawn, the humidity drops and simultaneously stomata open, resulting in increased water losses from aerial plant tissue (Klepper 1968). The water potential continues to drop throughout the day as the open stomata facilitate CO2 acquisition. In most species, water potential is the lowest in the late afternoon. Once the stomata close, the rate of water uptake exceeds the rate of water loss, and leaves and fruits regain their water potential. In some plant species, the increased water potential at night fuels the hydraulic redistribution of moisture in the soil (Richards and Caldwell 1987; Caldwell et al. 1998). This daily rhythm persists even with reduced water levels. Low water potential in the root environment, through the loss of soil moisture throughout the daytime period, can impact nutrient uptake and the transport of metabolites or the precursors needed to generate phytochemicals (Plaut and Reinhold 1965; Greenway et al. 1969).

Changes in water availability throughout the year have a higher amplitude than daily changes. In most locations, precipitation and water availability have a strong seasonal component with a wet and dry season (Snyder and Tartowski 2006; Taxak et al. 2014). Water availability affects the accumulation of several phytochemicals. In Coffea arabica leaves, irrigation reduced the levels of the quinone pheophytin, a photosynthetic pigment (Scheel et al. 2016). Tea (*Camellia sinensis*) harvested during the dry season had higher concentrations of desirable methylxanthines and polyphenols than in the monsoon season (Ahmed et al. 2014). Drought results in phenolic accumulation in *Amaranthus tricolor* (Sarker and Oba 2018) and *Zea mays* (Hura et al. 2008). However, phenolic content shows variability in response to water availability across plant species. In peanuts (*Arachis hypogaea L.*), phenolic content increased in leaves and stems in response to drought stress but decreased in the seeds (Aninbon et al. 2016). Using a rain-out shelter Cheruiyot et al. (2007) observed that tea cultivars had lower phenolic levels in drought conditions. There is a similar decrease in phenolic content in cotton leaves (Shah et al. 2011; Shallan et al. 2012), and under progressive water deficit, both Shiraz and Cabernet Sauvignon grape cultivars (*Vitis vinifera*) showed a decrease in flavonols and non-flavonoid phenolic compounds (Hochberg et al. 2013). In the

Shiraz cultivar, quercetin-3-O-galactoside and rutin increased under water deficit. Phenolic compounds decreased in drought-sensitive tomato *(Solanum lycopersicum L.)* cultivars, but increased in response to drought in Zarina, a tolerant cultivar (Sánchez-Rodríguez et al. 2011). The tolerant Zarina cultivar also showed an increase in rutin and flavonoid levels. In other plant species, drought induces the accumulation of flavonoids (Hernández et al. 2004; Yang et al. 2007; Ma et al. 2014; Shojaie et al. 2016). Under mild drought stress, the diel accumulation patterns of non-structural carbohydrates showed stress-induced changes in Brassica rapa plants, suggesting an interaction between diel regulation and abiotic stress responses (Greenham et al. 2017).

Under both mild and moderate water stress, the components of essential oils, geraniol, and citral increased in two species of lemongrass, *Cymbopogon nardus* and *Cymbopogon pendulus* (Singh-Sangwan et al. 1994). In medicinal plants, the concentration of desired phytochemicals can be increased by growth in water limiting conditions (Jaafar et al. 2012; Selmar and Kleinwächter 2013). For example, water-stress induced accumulation of hyperforin in St. John's wort (*Hypericum perforatum*) (Zobayed et al. 2007) and ajmalicine in *Catharanthus roseus* (Jaleel et al. 2008).

Like changes in the pattern of temperature cycles, climate change is affecting the timing of precipitation events, with a seasonal component to even extreme precipitation events (Unal et al. 2012; Pal et al. 2013; Keggenhoff et al. 2014; Ganguli and Ganguly 2016; Gitau 2016; Rahimpour et al. 2016; Tye et al. 2016; Mallakpour and Villarini 2017; Roque-Malo and Kumar 2017). In the US Ohio/Missouri River valleys, the start of the dry season is altered by 2–3 weeks while the wet season in east New York arrives about three weeks earlier than a century ago (Pal et al. 2013). These disruptions to the seasonal patterns of water availability may alter the phytochemical composition of plants even if harvested at the same time of year.

Additional examples of the effects of light, temperature, water availability and other environmental factors on the accumulation of specialized metabolites are reviewed by Ramakrishna and Ravishankar (2011), Ncube et al. (2012), Verma and Shukla (2015), Borges et al. (2017), Yang et al. (2018).

1.5 Circadian Regulation of Phytochemicals

The plant's endogenous circadian clock is crucial to the coordination and control of molecular activities in alignment with daily and seasonal variability of resources (Green et al. 2002; Dodd et al. 2005; Yerushalmi et al. 2011; Bendix et al. 2015). Evidence of diel variation in metabolism and environmental responses is mainly derived from transcriptional evidence that shows rhythmic diel or circadian-driven patterns of expression (Alabadí et al. 2002; Bläsing et al. 2005; Michael et al. 2008; McClung 2014). The relative accessibility of transcriptomics facilitates the evaluation of daily rhythmic

expression patterns across various plant species. While they are perhaps less well-studied, protein and metabolite levels also show daily rhythmic patterns of accumulation (Gibon et al. 2006; Graf et al. 2010; Hwang et al. 2011; Flis et al. 2016). Both primary and specialized metabolites in *Arabidopsis thaliana* such as maltose, fructose, GABA, and isoleucine show daily oscillation in their abundance (Stitt et al. 2007; Espinoza et al. 2010; Augustijn et al. 2016). Upon entrainment, the accumulation of some specialized metabolites shows a rhythmic pattern in the absence of light, temperature, or humidity cues implicating regulation by the circadian clock (Gibon et al. 2006), the disruption of which can significantly alter metabolism (Fukushima et al. 2009).

1.5.1 Gating Response to Environmental Signals

Biotic stressors from bacterial and fungal pathogens to herbivores have rhythmic daily activity (Hevia et al. 2015). Abiotic stressors can also have rhythmicity such as peaks in heat stress both at certain times of day and year. Plants respond to these stresses with temporal coordination, or 'gating.' This gating secures maximal molecular response at the time that coincides with the highest probability of the stress occurring. Circadian regulation occurs in part through gating the response to phytohormones such as auxin, abscisic acid, jasmonic acid, and salicylic acid (Covington et al. 2008; Goodspeed et al. 2012). At dawn, *Arabidopsis thaliana* shows decreased susceptibility to the pathogens *Pseudomonas syringae* pv. tomato DC3000 (Bhardwaj et al. 2011) and *Botrytis cinerea* (Ingle et al. 2015). Evidence of circadian regulation of basal defense mechanisms (Wang et al. 2011) suggests that this observation is not unique to these pathogens. Symbiotic relationships with beneficial species such as pollinators and parasitoids rely on coordination with the insects' daily rhythms. The primary mediators of insect attraction, volatile organic compounds (VOCs), show variation in emissions throughout the day. Fig (Ficus racemosa) volatiles, for example, can be distinguished by their diel variation in temporal patterns (Borges et al. 2013). Fig plants offer a distinctive volatile profile to match the biological activity of parasitoid and pollinator populations.

Many of the metabolites induced by generalist herbivores peak in accumulation in target tissues at specific times of the day. For example, *Nicotiana attenuata* has a diel rhythm of jasmonic acid (JA) and 12-oxo-phytodienoic acid (OPDA) accumulation in roots, but not in leaves (Kim et al. 2011). Shoot-oscillating metabolites mostly peak during the day such as citric acid, tyrosine, phenylalanine, and lyciumoside. In contrast, the metabolites that oscillate in roots, such as disaccharides and JA mostly peak at dusk or night. Lyciumoside which is a precursor to diterpene glycosides, peaks at dusk and is more prevalent in sink leaves than in source leaves. Circadian-gated herbivory response alters the interaction between the plant and the herbivore. *Trichoplusia ni*, a cabbage looper insect, shows rhythmic feeding behavior that peaks in the afternoon (Goodspeed et al. 2012). *Arabidopsis* plants can defend against the *T.*

ni herbivory through concerted timing of jasmonate induction and reduced salicylates. However, disruption of the plant circadian clock or switching the phase of *T. ni* feeding reduces these justifications. Like jasmonates and salicylates, many phytochemicals have a role in plant defenses. Thus, this gating of defense responses combined with the effects of daily rhythms in primary metabolism, environmental factors and circadian regulation of specialized metabolites likely drive significant variation in phytochemical levels throughout the day.

Abiotic stress responses also show time-dependence, such as temperature (Fowler et al. 2005; Dong et al. 2011; Grenivich et al. 2019), UVB (Takeuchi et al. 2014; Horak and Farré 2015) and reactive oxygen species (Lai et al. 2012). The master regulators of low-temperature responses, CBF transcription factors, show the highest response to a 4°C cold shock in the morning and a reduced induction in response to the same stress at night. The CBFs control, in part, the rewiring of metabolism that occurs under low temperature (Cook et al. 2004). In Arabidopsis, low temperatures (4°C) provoke metabolic rewiring, including induction of phenylpropanoid levels (Kaplan et al. 2007). Constitutive expression of CBF3 recapitulates 79% of the metabolic changes observed under low temperature in control conditions (Cook et al. 2004). Circadian gating of CBF3 cold response suggests the induction of the low-temperature-responsive metabolites regulated by CBF3 is also likely gated.

1.6 Seasonal changes

Perhaps more drastic than daily changes, environmental factors vary in seasonally recurring patterns, including changes in precipitation, temperature, and daylength. Biotic and abiotic stresses also show seasonal variation. Also, developmentally regulated changes in phytohormones auxin, gibberellin, ethylene, and jasmonate alter the sensitivity of the plant to different stimuli depending on the time of year (Davies 2010; Footitt et al. 2011; Singh and Mas 2018).

1.6.1 Photoperiod

Both north and south of the equator there is an asymmetry to the amount of light or day length throughout the year. The amplitude of the difference between the long days of summer and the shorter days of winter increases with the distance from the equator. Seasonal changes in photoperiod are concomitant with seasonal changes in light intensity and temperature. Therefore, when isolating the factor of photoperiod, chamber-based experiments comparing different day lengths are often performed. One challenge with these experiments is that the chamber-experiments are often performed with square waves where the changes in light intensity and light quality which also vary across the day and season are ignored. Ignoring these subtle changes can result in missing important mechanisms that control for the integration of signals (Liao et al. 2017). Most plant species are sensitive to changes in photoperiod and

respond by adjusting growth and developmental stages. For example, silver birch, *Betula pendula,* integrates both temperature and photoperiod cues to control the timing of bud burst (Linkosalo and Lechowicz 2006). The chilling requirement is met long before the modeled initiation date, suggesting a second cue is integrated to control the timing of bud burst. Linkosalo and Lechowicz compared the effects of a reduced red to far-red ratio throughout the day or only at dawn and dusk and demonstrated that the diurnal timing of the change in light quality is an important factor in this developmental transition. The integration of both light quality timing and temperature changes may protect some species from frost damage due to early budding associated with warm early-spring weather. These developmental changes can impact phytochemical composition and distribution in the plant (Degu et al. 2014; El Senousy et al. 2014). For example, in cherry radish (*Raphanus sativus L. var. '*Changfeng') polyphenol and antioxidant content increased in the root, but not in the shoot when grown in a longer day period (Guo et al. 2019). Photoperiod changes impact the flavonoid, phenolic, amino acid, anthocyanin, alkaloid, and glucosinolate levels (Bernarth and Tetenyi 1979; Xu et al. 2011; Steindal et al. 2015). Phytochemical responses to photoperiod have been reviewed in Verma and Shukla (2015). In addition to these general impacts on phytochemicals, for many compounds, specific response to photoperiod have been described.

1.7 Polyphenols and phenolic compounds

Polyphenols are generally high molecular weight molecules derived from the shikimate, phenylpropanoid, flavonoid, anthocyanin, lignin (Mouradov and Spangenberg 2014) and polyketide pathways in plants. A large number of phenolic hydroxyl groups provide these compounds with unique metabolic properties and compounds in this class have a range of functions *in planta* including UV protectants, defense compounds, signaling molecules, antimicrobials, and antioxidants. Polyphenols have been confirmed as the source of biological activity in many medicinal plants, particularly those used to prevent and ameliorate metabolic diseases (Cvejić et al. 2017; Chen et al. 2018).

1.7.1 Circadian, Diel, and Seasonal Effects on Total Polyphenols

The regulatory and protective functions of phenols in oxidative stress-induced signaling may explain the observed rhythmic accumulation in many plant species. In Arabidopsis, the response to reactive oxygen is circadian regulated (Lai et al. 2012). Consistent with this observation, antioxidant activity was found to be circadian regulated in other Brassicaceae (Soengas et al. 2018). However, the temporal signature of antioxidant activity was not consistent across diverse species. Coordinating antioxidant activity with the recurring timing of local stresses, other phenological activity, or breeding pressures could drive species-specific changes in the waveform of expression that could be locally adaptive. Soengas et al. (2018) observed circadian driven rhythmicity in the pattern of total phenolic

compound accumulation in broccoli, cabbage, Chinese cabbage, and turnip greens. Although all species accumulated the highest total phenolic content in the period surrounding dawn, the specific phasing of the peak in accumulation varied between species. Broccoli and cabbage accumulation was the highest in the dark period before dawn while Chinese cabbage and turnip greens peaked in accumulation at dawn. Specific compounds also showed variability in their waveform of selection discussed below. Selection for desired traits or the need to meet the pressures of local environments may result in observed altered temporal profiles between species. We provide examples below where the time of day and season of year variation in accumulation has been observed for polyphenolic compounds.

1.7.2 Flavonoids

Flavonoids function as pigments, signaling molecules, UV-protectants, and antimicrobials. Daily rhythms of flavonoids have been reported in the tropical tree, *Anacardium excelsum* and the fern, *Cryptogramma crispa* (Veit et al. 1996). In grape berries, (*Vitis vinifera*) (Reshef and Agam 2019) the diel pattern of expression for some flavonoids was sensitive to the orientation of the berry in relationship to the sun. The flavonoid quercetin showed a rhythmic accumulation in all three orientations, however other flavonoids, such as cyan-3glu, were only arrhythmic in some positions but showed no change in signal across the day in others. The observable difference in flavonoid accumulation in response to the small change in solar irradiation due to the position of the berry relative to the large daily oscillatory light suggests that the sensitivity of the phytochemical response is attuned to account for this daily change in light intensity. Seasonal variation in anthocyanin levels can manifest if the plant is harvested at different times of the year. For strawberry fruit (*Fragaria x ananassa*), plants at the same ripening stage and harvested at different times of the year showed variation in anthocyanin accumulation levels (Pincemail et al. 2012; Ariza et al. 2015). Strawberries harvested earlier in the year had the lowest content of organic acids and antioxidant compounds, perhaps indicating a role for increasing day length in the accumulation of these phytochemicals (Ariza et al. 2015). Over all genotypes, strawberries in the mid to late harvest showed overall better health-related properties. There was an interaction between cultivar and the compound time of harvest suggesting that healthy harvesting could be achieved by staggering plantings by cultivar and should be a factor considered during breeding selection.

1.7.3 Phenylpropanoids, Phenolic acids, and Aldehydes

Both the time of day and time of season plants are harvested can impact the phenolic acid content for some plant species. Broccoli head (*Brassica oleracea* L.) harvested at the end of the day had higher phenolic content and antioxidant activity during storage than those harvested in the morning or at midday (Hasperué et al. 2011). The time of year plants are harvested can impact the nutritional quality of the

product as much as the genotypic variation. Four different cloudberry (*Rubus chamaemorus*) genotypes were examined throughout the year and seasonal variations were observed in gallic acid, ellagic acid, and anthocyanins (Hykkerud et al. 2018). Across all genotypes, cloudberry plants harvested early- and mid-season contained significantly higher ellagic acid levels than those harvested in late season (7.41, 7.03 and 6.35 mg/g dry weight, respectively). This pattern reflected the total phenol measurement, which was on average higher in early-season, decreasing in mid- and late-season (22.12, 20.97, 20.09 mg/g dry weight, respectively). Although the variation between genotype was more significant than the seasonal variation, ellagic acid was still significantly responsive to environmental variation. In an earlier study in *Rubus* species, raspberry fruits (*Rubus idaeus*), Mazur et al. (2014) compared ten genotypes of raspberry fruit for the effects of harvest season and genotype on phenolics, ellagic acid, anthocyanins, and ascorbic acid content. Only ascorbic acid content was consistent across harvest seasons. Phenolics, ellagic acid, and anthocyanins showed significant variation depending on the time of year the berries were harvested. The study concluded that the "quality of red raspberry fruits were significantly affected both by genotype and harvest season."

Total phenolic content and phenolic profile varied in globe artichoke (*Cynara cardunculus* L. var. scolymus (L.) Fiori) between plants harvested in the winter or spring (Lombardo et al. 2010). The phenolic content increased sixteen times in the later spring harvest. Several compounds, 1- and 3-*O*-caffeoylquinic acids, caffeic acid, narirutin, and naringenin 7-*O*-glucoside were only detected in the spring harvest, changing the overall polyphenol profile. The variation in total phenolic content of *Vochysia divergens* Poh., a pioneer tree species was greater across the season than between geographic areas in Brazil (Uriu et al. 2018). Harvest time also affected the total level of phenols and antioxidant activity in grape cultivars (*Vitis vinifera*) (Piazzolla et al. 2016). In this study, the harvest time of grapes was found to influence the distribution of volatiles including acetaldehyde, 2-butenal, Hexanal, ethyl acetate, ethanol, and D-Limonene resulting in an overall change in grape quality. In the Piazzolla et al. (2016) study, since the grapes were not at the same developmental stage, the effect observed is likely a combination of developmental and environmental effects.

1.7.4 Hydroxycinnamic acids

Leaf oils showed a time of day and time of year variation in *Lippia origanoides* Kunth (Ribeiro et al. 2014). Examining plants collected near a mine in Brazil, Ribeiro et al. identified many metabolites with significant variation throughout the year. The major specialized metabolites with a time of year change in concentration were (E)-methyl cinnamate, (E)-nerolidol, p-cymene, 1,8-cineole, carvacrol, a-pinene, (E)-caryophyllene and g-terpinene. Cinnamate levels also varied by time of year with no accumulation from March through June, followed by moderate accumulation in July, peaking in

concentration in August, and returning to moderate levels from September through February. Cinnamate also showed significant variation in concentration depending on the time of day samples were harvested. In kimchi (green Chinese cabbage (*Brassica rapa*) and red cabbage (*Brassica oleracea)*), fall and spring-sown cultivars were compared for two genotypes. The phenols caffeic acid, p-coumaric acid, ferulic acid, and sinapic acid and flavonols (quercetin and kaempferol) showed significant differences in levels between fall and spring in both red cabbage cultivars tested (Lee et al. 2018).

1.8 Naphthoquinones and Quinones

The medicinal plant, *Euclea undulata* produces epicatechin and α-amyrin-3O-β-(5-hydroxy) ferulic acid, both desired for their potential value in the treatment of diabetes. However, *E. undulata* also produces the naphthoquinone, 7-methyl-juglone, which is cytotoxic. Botha et al. (2018) evaluated the environmental and seasonal effects on the accumulation of these three compounds to determine if altering the harvest time could eliminate the presence of 7-methyl-juglone. Comparing plants harvested in the rainy (December) and dry (August) season, in three locations, they observed both geographic and seasonal effects on the metabolite profile. While the variation in metabolites between geographic locations was greater than the variation between seasons, at a given location, seasonal differences had a significant impact on the metabolic profile. In particular, the 7-methyl-juglone was not detectable in leaves in the rainy season in the summer rainfall region, while plants from the same region accumulated 7-methyl-juglone in the dry season. In this same region, the desired compounds epicatechin and α-amyrin-3O-β-(5-hydroxy) ferulic acid, also showed seasonal variation and the stems and leaves of *E. undulata Thunb. var. myrtina* contained both epicatechin and α-amyrin-3O-β-(5-hydroxy) ferulic acid and lacked 7-methyl-juglone only in the rainy season. Therefore, the authors concluded, “for the safe and effective use of *E. undulata* it would be best to collect leaf material during the dry season in the summer rainfall areas.” This example highlights the importance of considering not only the temporal profile of the phytochemicals of interest but also the temporal profile of any negative or even synergistic compounds (e.g., Figure 1.2).

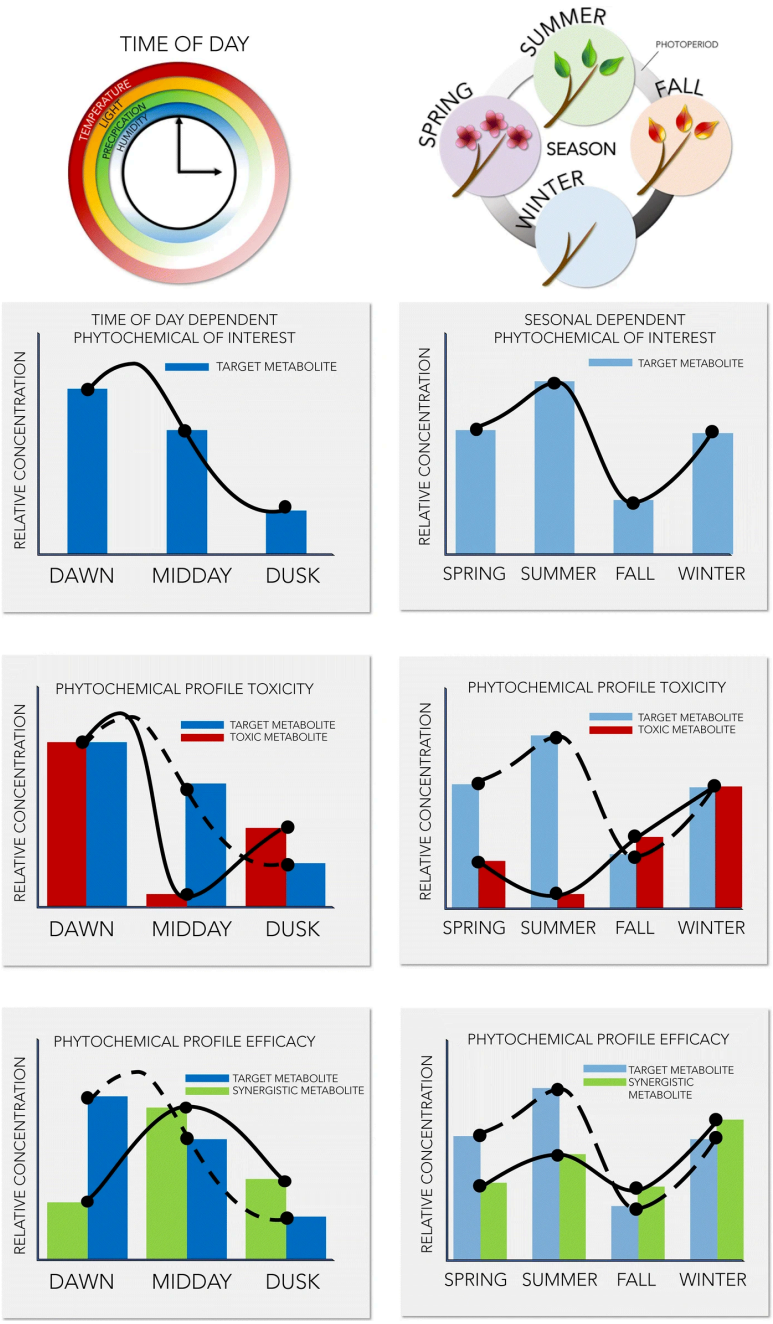


Figure 1.2 ***Time of day and season of the year impact the composition and abundance of phytochemical extracts.*** This figure from Liebelt et al. 2019 shows extract composition quality is a combination of the relative concentration of the phytochemical of interest and the antagonistic or synergistic effects of the remaining profile. Time of day and seasonal variations in the target metabolite, toxins, and synergistic metabolites can be taken into consideration to determine the optimal growing season and time of day to harvest extracts. In the simplified graphic above an extract with the patterns shown would be optimally harvested during the middle of the day to account for diel variations.

Harvesting in the summer would provide the maximum efficacy at the lowest toxicity across the yearly seasonal variation.

1.9 Glucosinolates

Glucosinolates are amino acid-derived phytochemicals found in the *Brassicaceae* family of plants and have a role in defense against herbivory (Mithöfer and Boland 2012; Singh and Mas 2018). Much of this amino-acid derived phytochemical's economic value is from its link to carcinogen detoxification properties of *Brassica* such as broccoli, cauliflower, cabbage, and Brussel sprouts when hydrolyzed to isothiocyanate upon consumption. These also contribute to taste and flavor profiles. Glucosinolates have been the object of numerous seasonal variation studies with sometimes conflicting findings. Studies mainly focus on the two relatively abundant types of glucosinolates: aliphatic glucosinolates derived from the amino acids' alanine, valine, leucine, isoleucine, and methionine; and indoles derived from tryptophan.

Seasonal factors of higher mean temperatures and longer photoperiods were found to correlate with the highest concentrations of glucosinolate across ten *Brassica oleracea* cultivars (Charron et al. 2005). One study reports a sweeter and more desired taste correlated to lower temperatures (Mølmann et al. 2015) while higher glucosinolate levels generally correspond to warmer temperatures. Consumers reportedly selected the sweeter tasting broccoli perhaps at the detriment of the nutraceutical value of the plant. Despite the correlation between generally higher levels of glucosinolate in high temperature, not all glucosinolates were affected similarly within each condition. These results support the findings in another report showing differential accumulation of glucosinolates in broccoli (*Brassica oleracea*) between spring and winter seasons (Vallejo and Toma 2003).

1.10 Thermal Climate Trends

Temperature is often the main focus of climate change discussions. While in general, it is accepted that the world is warming, what is less known is that this warming is not happening equally. There is a disparity in the effects of global warming impacts by both season and latitude (Figure 1.3). Colder months in northern latitudes, in particular, are heavily impacted by warming temperatures.

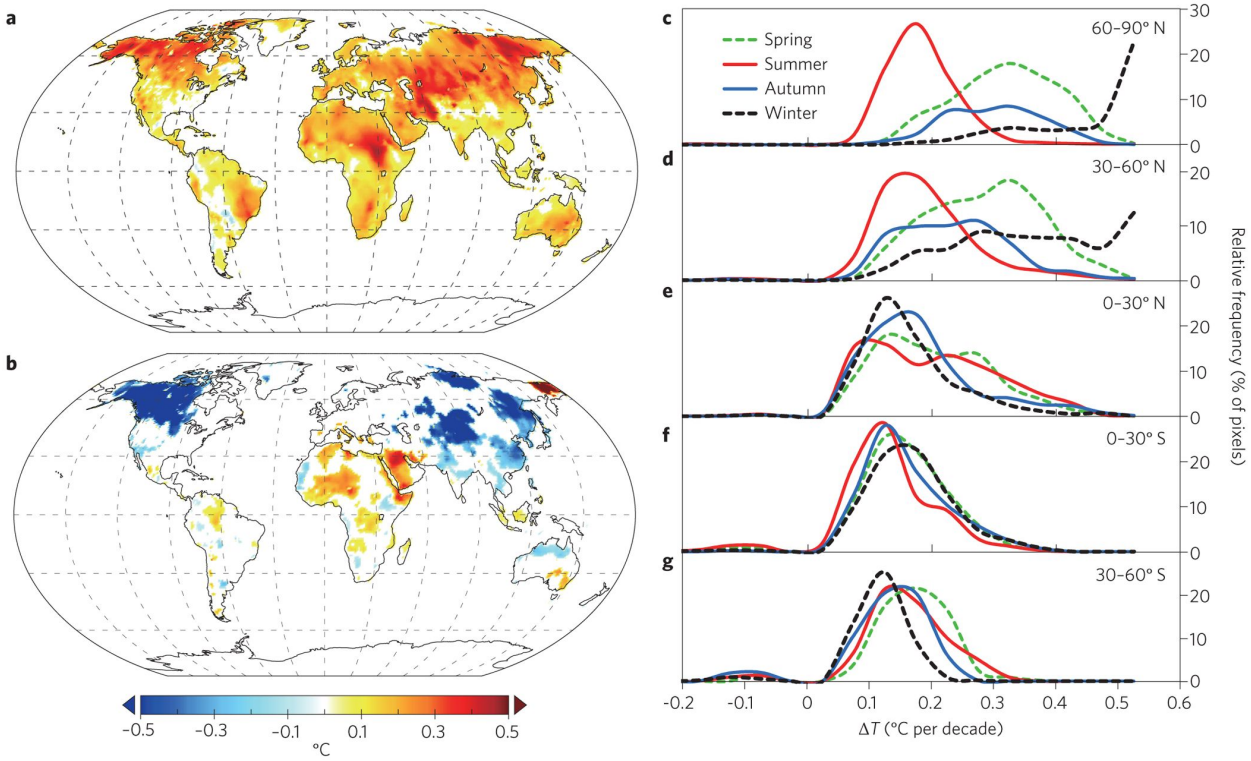


Figure 1.3 ***Seasonal Warming between 1948 and 2010.*** This figure from Xia et al. (2014) illustrates global trends in seasonal warming for the annual mean temperature (a), temperature range (b), as well as seasonal trends by latitude (c-g).

Not only are we experiencing record-breaking temperatures around the world, but also the change in temperature between day and night is decreasing. A 1.5-2.0 degree global rise in temperature has a larger impact on the coldest nights than on hottest days (Xia et al. 2014) Night temperatures are also increasing at a faster rate than day temperatures (Easterling et al. 1997, Vose et al. 2005) (Figure 1.4), leading us to wonder what phytochemical profile changes we should expect as a major component of the circadian clock is disappearing.

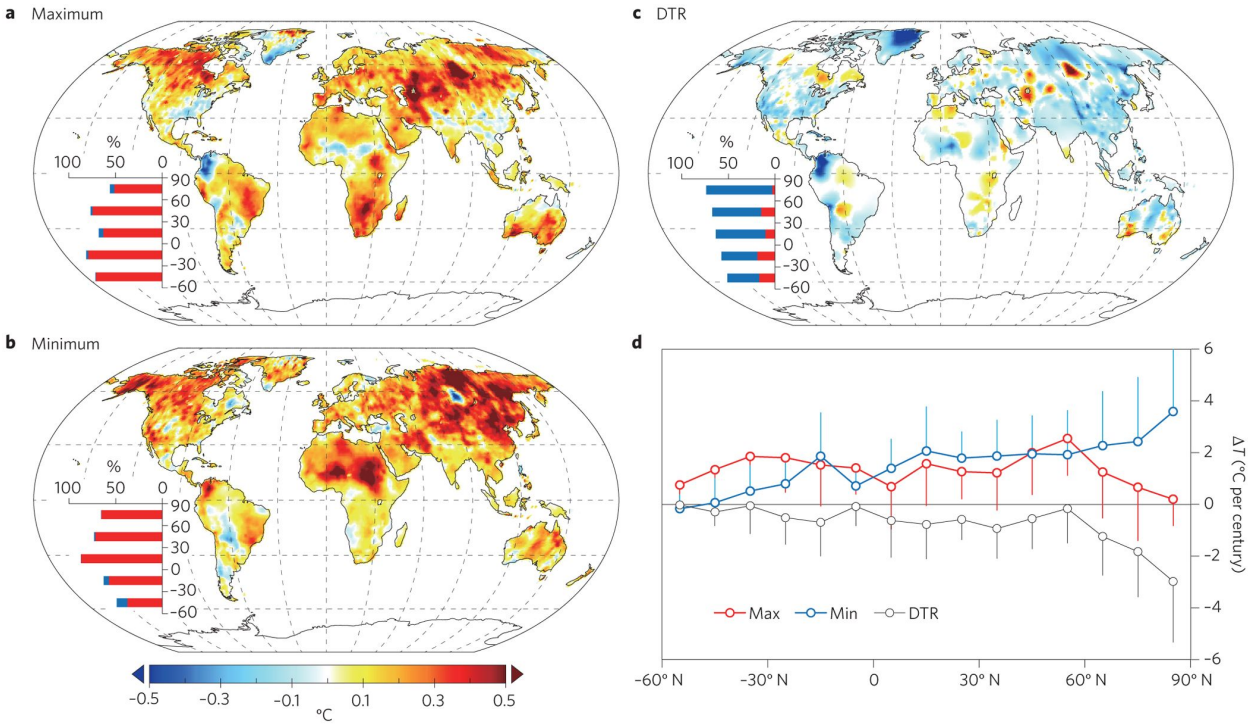


Figure 1.4 ***Diurnal Warming between 1948 and 2010.*** This graph shows global warming trends in °C per decade for mean maximum (a), minimum (b), diurnal temperature range (c), as well as temperature of a-c by latitude (d). (Xia et al. 2014)

Plant response to temperature changes at night are distinct from temperature perturbations during the day (Grinevich et al. 2019). The difference between day and night temperature impacts the overall growth of tomato (*Lycopersicum esculentum),* fruit quality, and the secretion of low molecular weight organic acids from the root which can impact nutrient uptake (Papadopoulos and Hao 2000; Yang et al. 2014, 2016). A mild increase in nighttime temperature appears to impact primary metabolism resulting in reduced yield and grain quality in rice and wheat (Ziska and Manalo 1996; Peng et al. 2004; Counce et al. 2005; Lobell and Ortiz-Monasterio 2007; Kanno et al. 2009; Mohammed and Tarpley 2009; Welch et al. 2010; Glaubitz et al. 2015; Laza et al. 2015; Shi et al. 2017; Impa et al. 2018). An enrichment for stress-responsive and reactive oxygen scavenging proteins was observed specifically under increased night temperatures in rice grains (Li et al. 2011). If these altered changes occur in other plant species, the impacts of changes in night temperature on specialized metabolites could be distinct from the impacts of changes in temperature during the day. Most studies to date investigate the response of specialized metabolites to increasing temperatures by employing higher daytime temperatures or a consistent increase in both day and night temperatures. Ibrahim et al. (2010) investigated the specific role of altered night temperatures on VOCs and found that increasing night temperatures increased the VOC emissions

detected during the day. Most VOCs examined showed a linear relationship between increasing night temperature and emission levels. However, in contrast to previous studies that examined daytime temperature changes, isoprene emission was not correlated with rising temperatures at night (Gouinguene and Turlings 2002; Tarvainen et al. 2005; Helmig et al. 2007; Sharkey et al. 2008; Copolovici et al. 2012). The VOC isoprene may provide an adaptive role in protecting plants from heat-induced damage (Sharkey et al. 2008). Therefore, understanding if isoprene emission is differentially responsive between night and day temperatures could impact our understanding of how the perception and response to stress vary across time. The asymmetric pattern of global warming (IPCC 2018 and Xia et al. 2014) suggests that it will be essential to understand the impacts of warming nights on specialized metabolite production.

1.11 Discussion

Environmental inputs play a vital role in phytochemical composition. Plant attunement to their environment can be seen on the scale of the positioning of a berry on a vine (Rashef et al. 2019) to seasonal variations in accumulation such as artemisinin in sweet wormwood (Ferreira et al. 2018). The value of specialty crops is attributed in part to their phytochemical composition. Transcriptomic and metabolomic databases can be bridged to structure a complex whole-genome metabolic network that can aid in candidate gene or pathway discovery. A lack of high-quality genomes or detailed annotations restricts the tools available to examine biochemical pathways further. Isoform analysis using long-read sequencing is one way to address this issue. Integrating interdisciplinary tools from genomics and biochemistry to bioinformatics aids in research efforts to understand the complexity of environmental influence on regulatory pathways in plants.

Temperatures are a central research topic globally in nearly every field as interdisciplinary teams of researchers and policymakers are attempting to come together to implement climate adaptation strategies to combat the consequences we are now experiencing. Global warming has brought several changes that directly affect agriculture from extreme weather anomalies (e.g., hurricane intensity and frequencies, length of drought and fire season) to the often-overlooked factor of warming nights. Minimized daily temperature fluctuations pose an interesting question of how plant composition will be affected by a change in this signal. Research on warming nights could be used to identify varieties and environmental conditions that maximize the desired phytochemical profiles to benefit human health. Additionally, when applied to specialty crop systems, we could gain insights into the environmental regulation of phytochemicals targeted for the production of pharmaceuticals and other consumer products.

CHAPTER 2

LONG READ SEQUENCING ISOFORM ANALYSIS: END TO END ISO-SEQ ANALYSIS WORKFLOW

2.1 Abstract

Single-molecule real-time sequencing, a third-generation sequencing technology developed by Pacific Biosciences optimized to capture entire transcripts, is revolutionizing isoform analysis. While the uses of these long-read sequences continue to grow, they are being used throughout the plant research community to provide insight into alternative splicing and improve the accuracy of genome annotations. Like any new technology of its kind, the bioinformatic analysis and compatible software for isoform processing is constantly evolving. PacBio's SMRT Analysis software and compatible packages for post-processing are enormously valuable in analyzing Iso-Seq and isoform data, however, they can be complicated and difficult to understand and troubleshoot for biological scientists without a history in coding and bioinformatics. With the advancement of sequencing applications as well as the rise in implementation of interdisciplinary teams, inexperienced researchers can either expect to have to learn some level of bioinformatics to analyze the data they generate or pay to outsource the analysis. Basic processes to analyze the data have been established, but due to the diversity of available software, operating systems, and hardware, it is difficult for a non-bioinformatician to understand the end-to-end workflow. Herein we provide a comprehensive end-to-end guide to isoform analysis of PacBio Iso-Seq data and provide compatible R source scripts for isoform data visualization.

2.2 Introduction

Iso-Seq is set apart from other sequencing technologies by its capability of sequencing full-length transcripts. Pacific Biosciences (PacBio) achieves these long-reads by closing linear double-stranded cDNA with 5' and 3' hairpin adapters to generate a circular molecule, referred to as a SMRTbell template, that will be loaded into the SMRTcell (Rhoads et al. 2015). Within each SMRTcell are 150,000 Zero Mode Waveguides (ZMW). Waveguides are a physical structure that manipulate a signal, to establish constraints on the area and frequency thresholds that can be detected. Using nanophotonic ZMW, PacBio has constrained the area where signal can be detected to a zeptoliter. Within this volume, a polymerase is fixed to the bottom of each ZMW and binds one SMRTbell molecule. Fluorescent tagged base-pairs come into range of the detector as the polymerase adds them to the replicating molecule and the fluorophore emission is detected at a wavelength unique to a specific base (Figure 2.1). One template is sequenced repeatedly as the polymerase continues around the circularized molecule with multiple

passes that have the coding and complementary strand separated by the adapters all on one long single stranded molecule or concatenated read (Figure 2.2). The fluorophores are detected in real time and recorded as a movie.subreads bam file, which stores the kinetics of the read in a way that can be used to determine other features of the transcript such as methylation state.

This comprehensive guide covers the general workflow from receiving a subreads.bam files from PacBio through Iso-Seq and post-clustering isoform analysis. Further computational analysis of the data will be able to draw complex comparisons between isoform states associated with specificity of tissue or time of day, expanding the applications of Iso-Seq from its original purpose of *de novo* genome assembly into answering potent questions regarding gene regulation.

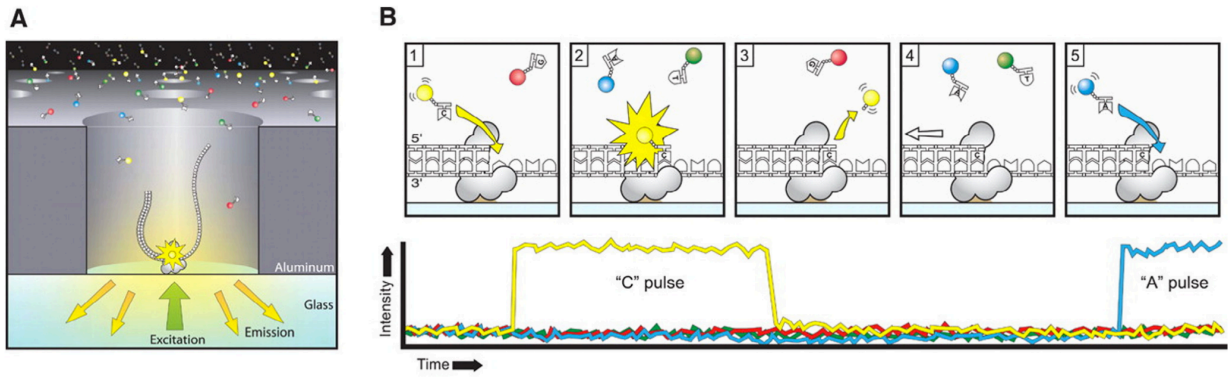


Figure 2.1 ***SMRT Sequencing*****.** This figure from Rhoads et al. (2015) depicts the Zero-Mode Waveguide (ZMW) containing the fixed polymerase in the zeptoliter detection space (A) and illustrates the pulse detection of fluorophores emitting a base-specific wavelength when in range of the detector as it is added to the nascent chain (B).

2.3 Results

2.3.1 Single v. Cluster Server

Due to the demands for memory and computational power required to process Iso-Seq data, all analysis was done through a remote server. Access to a server can be granted by most universities' IT or Data Management departments. Analysis can be performed on a standalone server or a cluster.
Remote servers are incredibly valuable in data analysis largely due to their core processing power. A core is often colloquially referred to as the "brain" of the computer and the more cores available the greater the number of simultaneous processes. The standalone server used to model the analysis is a RStudio virtual machine operating on Ubuntu 18.04. The software layer of the virtual machine was designed to use 16 of the cores provided by the real machine which has a total of 24 hyper-threaded cores. Hyper-threading

makes it appear like double the cores are present to double the number of simultaneous processes. The virtual machine therefore runs 16 out of 48 cores. While clusters can provide advantages in project distribution, load balancing and a higher processing power, standalone servers like the one described above are sufficient for Iso-Seq analysis.

2.3.2 Terminal Coding Environments

The minimum software requirements were examined in light of our goal to increase accessibility of Iso-Seq analysis workflow. The server's RStudio software layer provides both a console and Terminal workspace. All analysis was done in Terminal using Linux commands and visuals were performed in both Terminal and Console.

The SMRTLink SMRT Tools package provided by PacBio is developed to be used through a Bioconda platform, or as a standalone software for clustering and aligning reads to a reference genome. As a standalone, installing SMRT Tools loads the command scripts from the package without the graphic user interface. Not only is the user expected to understand and execute the commands in the proper sequence and correct syntax, but also to manually manage file paths and directories in Linux.

While the SMRT Tools platform is enough to process Iso-Seq raw subreads files up to clustering and mapping to a reference genome, the compatible post-Iso-Seq isoform analysis software requires Bioconda and bx-python. Docking SMRTLink SMRT Tools commands through the Bioconda platform not only provides a foundation for future analysis software but also processes faster circular consensus sequence calling per ZMW (https://github.com/PacificBiosciences/ccs).

2.3.4 Clustering - Iso-Seq Analysis

This stage of the analysis is generally referred to as Iso-Seq. In order to understand the clustering workflow, it is important to have a grasp of the contents of the raw files. In the case used to model this workflow, eight raw data files were received from PacBio after sequencing the samples collected and processed in Waddell (2017). Each file contains data of size separated cDNA sequences multiplexed from six samples representing three tissues at two times of day (Figure 2.2). Due to the nature of PacBio sequencing, each read has multiple passes concatenated into a long-read single molecule.

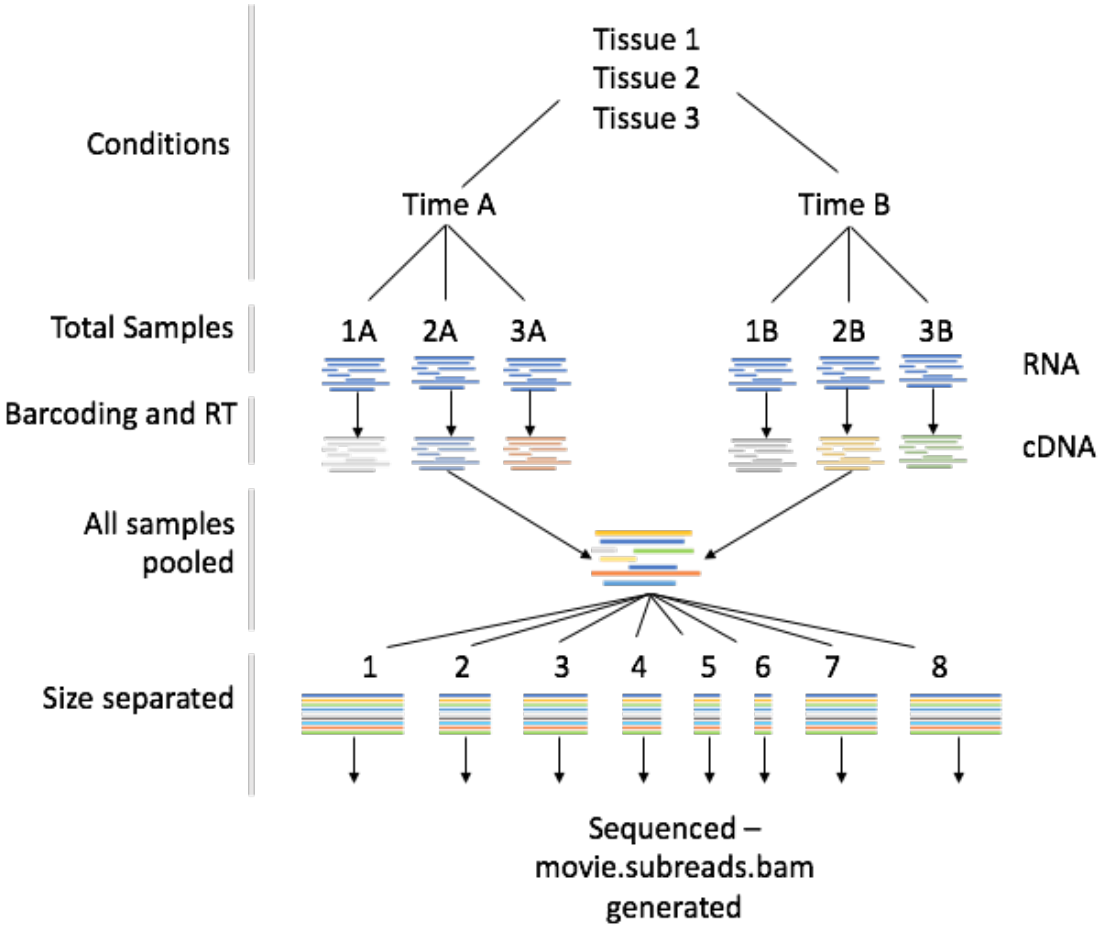


Figure 2.2 ***Sample Origin Schematics.*** Hops plants were sampled in Waddell (2017). In summary, three tissues at two times of day were sampled and RNA extracted. A universal 5' adaptor was added and with unique 3' primers they circularize the reverse transcribed cDNA. The cDNA is then pooled, and size separated into eight subsamples. Each of these samples were sequenced to generate the eight files used in data analysis.

2.3.4-I Circular Consensus Sequence Calling

The first step in processing the raw files is to take each multi-pass ZMW concatenated sequence and generate a single circular consensus sequence (CCS) (Figure 2.3A). This step in processing the raw data is the most time-consuming step of the workflow. In earlier versions the CCS script could take anywhere from days up to a week to process, though with the newer versions 4.0.0 and 4.2.0 this step can be done within several hours (https://github.com/PacificBiosciences/ccs). This step also benefits from using a server with more cores as you can open multiple Terminals and run each data file through CCS in parallel to reduce the total amount of time that this step consumes.

The CCS report shows the total number of recognized ZMWs as well as a count of those that generated a CCS or were filtered out. Of the reads that were filtered, 95-97% were due to lacking full passes (Table 2.1). In the latest versions of CCS script, setting a minimum predicted accuracy at 0.9 polishes reads so it would be redundant to polish again after clustering. Each file represents multiplexed data set so no comparative biological information can be gained from this step.

Table 2.1 ***CCS Percent Attributed Error Analysis***. Percent ZMW generating CCS and ZMW's filtered are normalized to the total input ZMW's recognized. Of the filtered ZMW's the majority were sorted due to lacking full passes. Minor errors include a draft generation error, heteroduplexes, and minimum coverage violation.

	1	2	3	4	5	6	7	8
ZMWs input (%)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
ZMWs generating CCS (%)	72.46	71.93	77.07	34.91	27.54	36.78	77.20	42.96
ZMWs filtered (A) (%)	27.54	28.07	22.93	65.09	72.46	63.22	22.80	57.04
Lacking full passes (% of A)	96.42	96.58	95.87	97.28	96.65	96.92	96.02	96.35
Heteroduplexes (% of A)	1.04	0.97	0.93	0.37	0.25	0.37	0.92	0.43
Min coverage violation (% of A)	0.94	0.94	1.17	0.53	0.47	0.56	1.12	0.64
Draft generation error (% of A)	1.57	1.46	1.98	1.75	2.58	2.08	1.86	2.52
CCS below minimum RQ (% of A)	0.03	0.05	0.05	0.06	0.05	0.06	0.07	0.06

2.3.4-II Classifying reads - Demultiplexing

Depending on the author and year of publication the next part of the workflow is identified as either classifying reads or demultiplexing. Demultiplexing runs the primers used to barcode the samples against the CCS reads generated in the previous step to linearize the sequence and separate based on sample identity (https://github.com/PacificBiosciences/barcoding). Sequences are identified and sorted based on the match between the universal primer and unique 3'primer. Although there are several script options available to complete this step, for the sake of consistency and a streamlined approach, we chose to use SMRT Tools 'lima' command.

The primer fasta file formatting may require troubleshooting. The format is highly dependent on the version of the script being used. In this case lima 1.10.0 (commit SL-release-8.0.0) is used therefore the requirements are that the universal 5' primer, or left adapter, can only be listed once and the 3' primers are the reverse complement of the sequence including the right adapter and the unique primer. In the model used for this workflow the right and left adapter are the same.

The output sequences are the full length reads for each sample. Some reads may still be concatemers, so refining is needed to remove repeating sequences (Figure 2.3B). Once refined the full length non-concatemer reads can be pooled by tissue and time of day to eliminate separation by size in preparation for clustering. At this point the files have been pooled with six representing a specific tissue at a time A or time B, three that represent each tissue at all time points, two that are all tissues at either time A or time B and one that pools all tissues at all times of day. This can be streamlined by first creating corresponding directories within an umbrella analysis directory. All files that apply were copied to each descriptive directory before merging into one flnc.fofn file for clustering.

2.3.4-III Clustering

This algorithm groups sequences based on their biological relevance so that isoforms of a single gene origin are clustered. Notable products of this step are the hq.fasta.gz and cluster_report.csv. The hq.fasta file is required for aligning reads to the genome and the cluster report is used when calculating the abundance of unique isoforms.

Once clustered, Iso-Seq is finished and the files are ready for alignment and isoform analysis using compatible open source software.

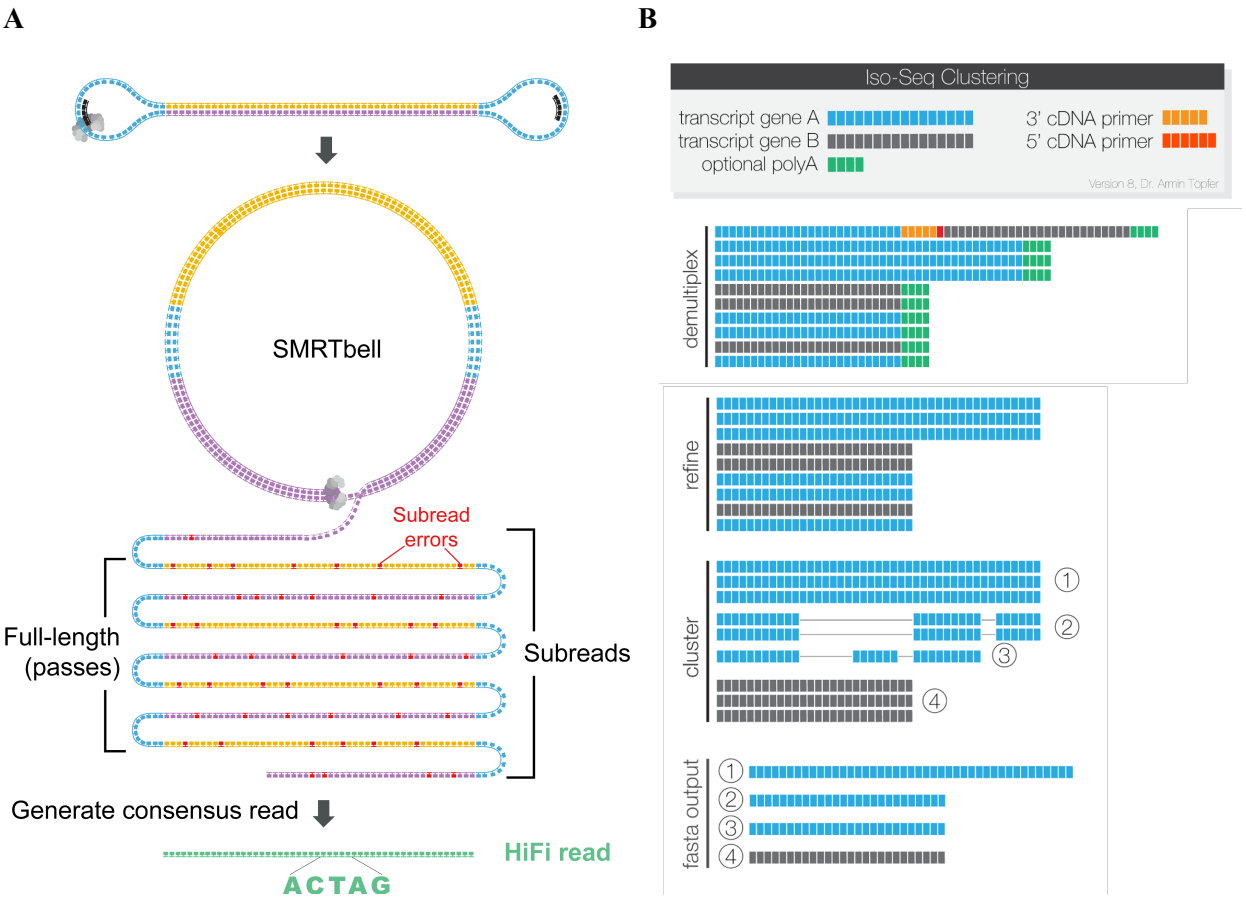


Figure 2.3 ***CCS generation and Iso-Seq Workflow.*** This figure adapted from Pacific Biosciences Iso-Seq guide and GitHub repositories, respectively, show A) SMRTbell concatenated subreads collapse into a consensus read and B) the midlevel workflow illustrating sequence manipulation at each stage of Iso-Seq analysis.

2.3.5 Mapping Reads

There are several methods for aligning clustered Iso-Seq reads, some even applicable without a reference genome. Splice aligners are constantly being updated, therefore when referencing a previous publication, it is increasingly important to understand the limitations of the aligner used at the time of publication and at the very least, what improvements and capabilities current aligners have. Current recommendations for "best practices" of aligning Iso-Seq data include minimap2, GMAP, deSALT, BLAT and STAR (https://github.com/Magdoll/cDNA_Cupcake/wiki/). We chose to compare minimap2 and GMAP in this workflow.

Minimap2 is analyzed in this workflow because it is included in the SMRT Tools command suite, so it does not require any added software. The command line is relatively easy and intuitive to use which is another advantage for someone who has no background in bioinformatics. The input reference genome

does not require preprocessing and even works in .gz zipped format. There is an option within this algorithm to index the genome first. In order to compare with Genome Mapping Alignment Program (GMAP) which requires indexing, we ran the alignment with both the original genome as well as the ".mmi" indexed genome. Indexing the genome with minimap2 reduces the alignment run time with only minor changes in abundance of unique reads (Figure2.4).

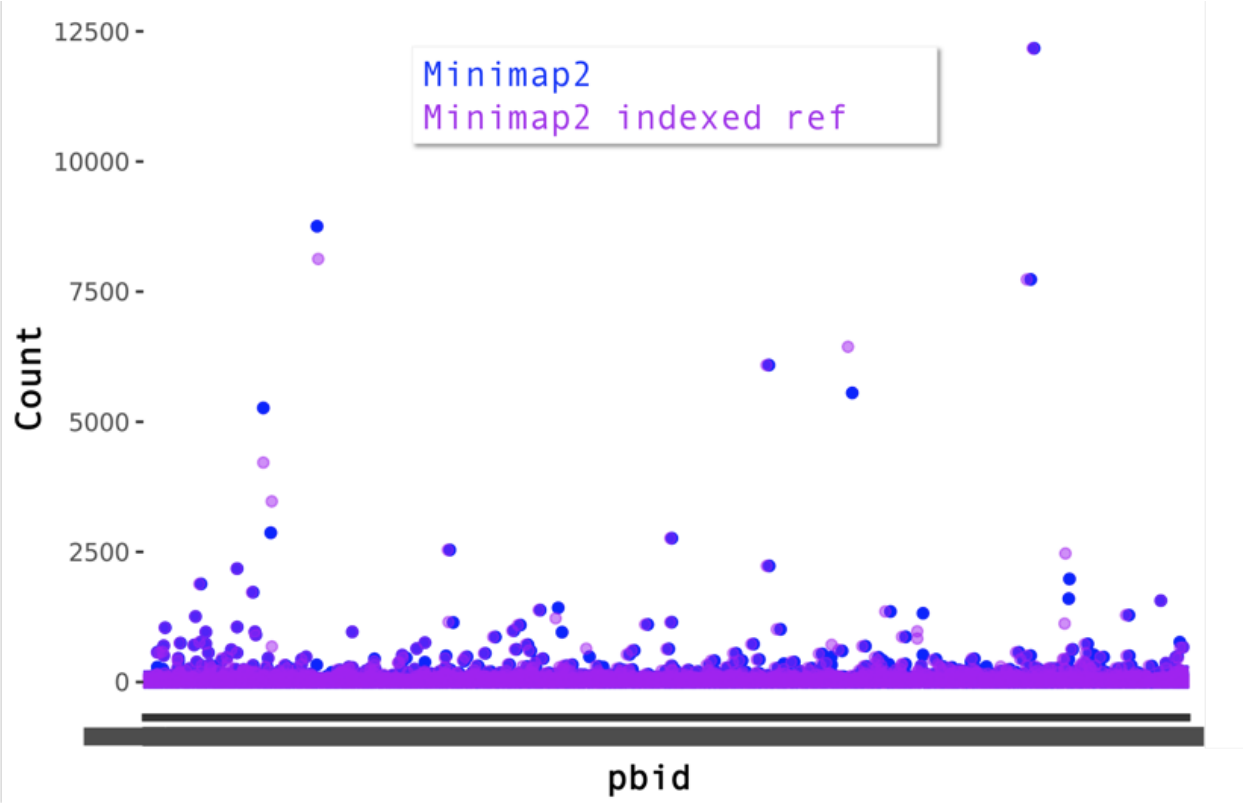


Figure 2.4 ***Count Abundance of Indexed Reference Comparison of Minimap2 Aligner.*** A comparison of full-length unique reads aligned to the raw reference genome (blue) and the minimap2 indexed reference genome (purple) show minor differences in count abundance calls.

GMAP was chosen to compare to minimap2 due to its prevalence in related research. GMAP has been used in high-profile plant isoform analysis papers (e.g. Abdel-Ghany et al. 2016 and Wang et al. 2016). Before the GMAP algorithm can be applied, the input genome requires pre-processing using the "GMAP_build" feature that indexes the genome to a format compatible with the alignment algorithm. Comparing the collapsed full-length count abundance grouped by pbid of minimap2 and GMAP showed that the majority of reads are similar with the exception of a few extreme reads (Figure2.5). While we chose minimap2 moving forward for streamlined workflow and greatly reduced alignment run time, the results from GMAP are also compatible with the following steps.

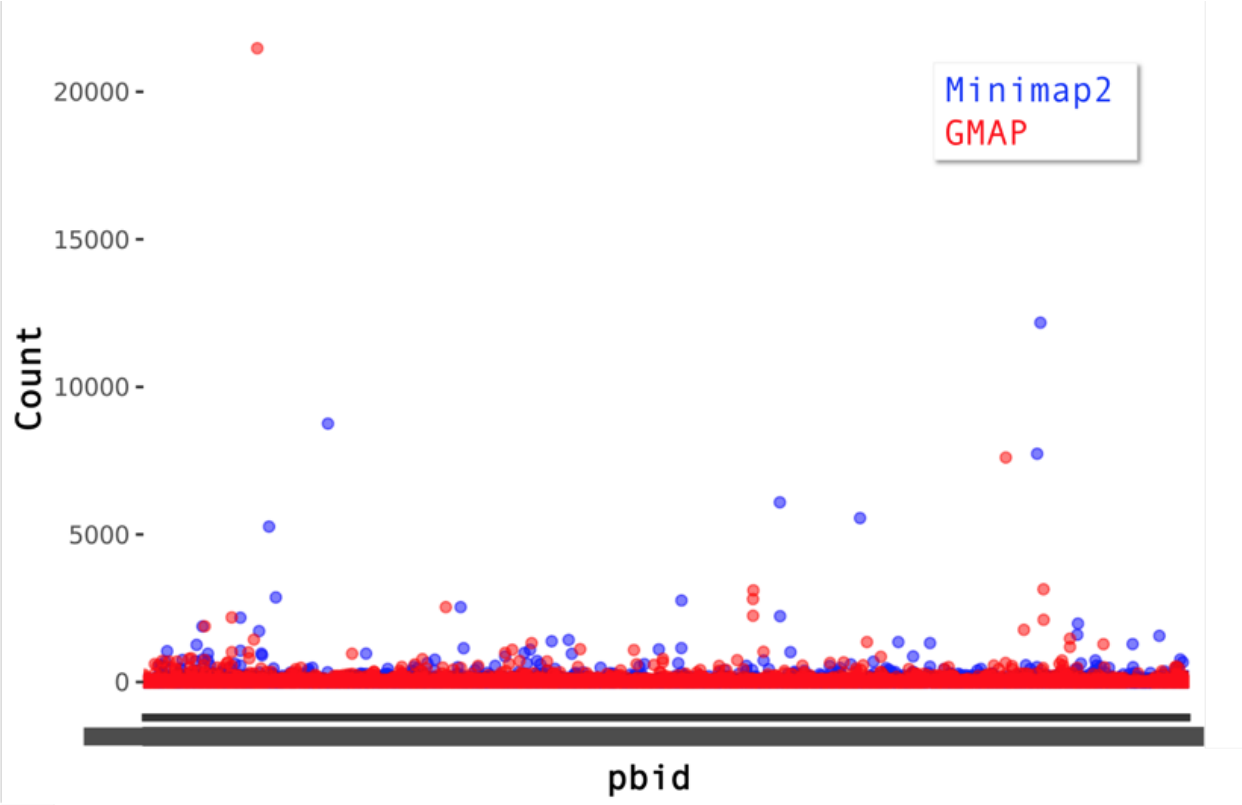


Figure 2.5 ***Count Abundance Comparison of Minimap2 vs. GMAP aligners***. Minimap2 (blue) and GMAP (red) have majority similar unique full length read count calls with the exception of count maximums.

2.3.6 Isoform Analysis

Cupcake ToFu scripts collapse redundant reads and calculate abundance of unique FL transcripts. The text files generated from getting the abundance of collapsed reads "abundance.txt" and "read_stat.txt" can be used to evaluate the percent of reads that were mapped to the genome, full length counts associated with each PacBio identifier (pbid), as well as identifying overlap in isoform composition between samples. Of the total isoforms detected, 60% were unmapped.

2.4 Materials and Methods

2.4.1 Hardware

Analysis was completed using a MacOS laptop with access to a remote server, in this case a virtual machine, running on 16 cores of 48 (24 hyperthreaded) cores available on the hardware. The MacBook used has a 1.1 GHz Intel Core m3 processor and 8 GB 1867 MHz LPDDR3 memory.

2.4.2 Software

2.4.2-I SMRT Tools Command Suite - Standalone

This protocol is used to install the SMRT Tools commands from SMRT Analysis suite without Bioconda. If Bioconda is installed first, then the package can be installed with conda run. The following assumes SMRT Tools will be added as a standalone software suite. Bioconda can be installed later, so the order isn't essential, however, once Bioconda is installed adding other packages becomes streamlined and is more intuitive than the following protocol.

SMRT Analysis software was obtained via Pacific Biosciences website and retrieved to a labeled folder in the MacOS desktop (https://www.pacb.com/support/software-downloads). Remote access to the server was established in Google Chrome, and a new folder labeled "pacbio" was added to the user's home directory. All commands were executed in Terminal running on 64-bit Linux. The file was transferred to the remote server using the secure copy protocol, "scp," command in a new Terminal on the MacBook. This command can be used whenever transferring files from a local computer to a server, or vice versa with minor changes to the command line. Here we use the path to the file on the local computer with the target file name at the end of the string and "~" in Linux represents the home directory.

```
MacBook:~ user$ scp ~/Path/to/local/file/smrtlink_8.0.0.80529.zip
user@remote.server.edu:~/pacbio
```

The following code is automatically generated in Terminal to log in to the remote server that will be accepting the file. The password must be entered carefully as the characters cannot be seen while they are being typed.

```
user@remote.server.edu's password:
```

Alternatively, the computer can be bypassed completely using the command "wget" in the Terminal window on the server. This command downloads the file directly from the provided link to the current directory, which can be navigated to in the command line using the "cd," call directory command:

```
user@remote.server.edu:~$ cd pacbio/
user@remote.server.edu:~/pacbio$ wget
https://downloads.pacbcloud.com/public/software/installers/smrtlink_8.0.0.80529.zip
```

2.4.2-II Installing SMRT Tools as a Standalone

Navigate to the directory that contains the software package and unzip the file.

```
$ cd pacbio
$ unzip smrtlink_8.0.0.80529.zip
```

Make the ".run" file an executable command using chmod.

```
$ chmod +x smrtlink_8.0.0.80529.run
```

SMRTLink installation instructions provide a --smrttools only option, the command line was retrieved from https://www.pacb.com/wp-content/uploads/SMRT-Link-Software-Installation-v8.0.pdf. Before this command line can be executed, "SMRT_ROOT" needs to be defined. SMRT_ROOT tells the system where to create the folder "smrtlink." It is important that this file does not exist at the time the command is executed since the installation software is designed to create the directory.

```
$ SMRT_ROOT=~/pacbio/smrtlink/
```

Next we run the SMRTLink --smrttools-only installation with a command modified from PacBio's instructions by adding "./" in front of the file name so that Linux recognizes it as executable.

```
$ ./smrtlink_8.0.0.80529.run --rootdir $SMRT_ROOT --smrttools-only
```

Once SMRT Tools are installed the "export PATH" command is used to tell the system where to find the package. This code can be executed at the start of every session or "~/.bashrc" can be added to the end of the file.

```
$ export PATH=~/pacbio/smrtlink/smrtcmds/bin:$PATH
```

2.4.2-III Bioconda and compatible Packages

Conda is a package management system that streamlines the installation of software independent of programming language and operating systems. Bioconda is a channel of conda designed specifically for bioinformatics (Grüning et al. 2018). The channel is added to Miniconda which was retrieved and installed per instructions in the "Getting Started" github provided by The Bioconda Team (https://bioconda.github.io/user/install.html).

Once the Bioconda channel is established, the following commands can be executed in the command line to install packages including SMRT Tools, if not already installed through the standalone method, and downstream analysis softwares. All installation commands can be found at https://anaconda.org/ and executed in Terminal. While installation commands can be found at multiple locations, we used the following to add Biopython and Bx-Python to the system (Table 2.2).

Table 2.2 ***Conda Run Commands.*** Conda executable command lines for installing SMRT Tools, GMAP, Biopython, and Bx-Python and their corresponding source.

Package	Conda Install Command	Origin
SMRT Tools	$ conda install -c hcc smrtlink-tools	https://anaconda.org/hcc/smrtlink-tools
GMAP	$ conda install -c bioconda gmap	https://anaconda.org/bioconda/gmap
Biopython	$ pip install biopython	https://biopython.org/wiki/Download
Bx-Python	$ conda install -c bcbio bx-python	https://anaconda.org/bcbio/bx-python

2.4.2-IV cDNA Cupcake

Cupcake is a compilation of supporting scripts downstream of Iso-Seq analysis. Within the Cupcake repository is ToFu, which can be used to collapse redundant reads and obtain the associated count information, or abundance, of unique reads. Installing cDNA-Cupcake was completed following a guide provided by Elizabeth Tseng with some minor modifications as follows (https://github.com/Magdoll/cDNA_Cupcake/wiki/Cupcake:-supporting-scripts-for-Iso-Seq-after-clustering-step#count).

First, activate the anaCogent environment and download the software package onto the server.

```
$ conda activate anaCogent
(anaCogent)$ git clone https://github.com/Magdoll/cDNA_Cupcake.git
```

Next, the python scripts need to be compiled to be able to use ToFu. This step assumes that the user has python 3.7.

```
(anaCogent)$ cd cDNA_Cupcake
(anaCogent)$ python setup.py build
(anaCogent)$ python setup.py install
```

Scripts location in the users bin can be confirmed using the “which” command. The output is a file path that displays where the system retrieves the target command. If the command is not found, use “export PATH” to define the directory path where the target script is located.

2.4.2-V Cloning GitHub Repositories

Several github repositories contain scripts that can be used to visualize reports and data generated throughout the Iso-Seq workflow. One of the easiest ways to access these scripts in RStudio is to clone the desired repository into the RStudio workspace through creating a new project. To navigate to git go to file, new project, version control, then git and paste the repository clone url obtained from the github page above. Check if the scripts are recognized in Terminal using “which” command; if not, then use “export PATH” to define the directory where the script is located. One possible step in troubleshooting the script is to use the “chmod +x” command to change mode and make the file accessible as an executable command. To run a R script in the Terminal command line use “Rscript --vanilla.”

2.4.3 Iso-Seq Analysis

Iso-seq Analysis was completed using SMRT Tools commands from PacBio’s SMRT Analysis software suite SMRT Link v8.0. Command lines were retrieved from the low-level-workflow in Pacific Biosciences Iso-Seq/isoseq-clustering.md GitHub repository based on Gordon et al. (2015). All versions were reported with --version option, for example:

```
$ ccs --version
```

If the system produces an error “command ‘ccs’ not found” then double check the “export PATH” command used after installation, or if using in a new Terminal window the export PATH command will need to be re-executed anyway.

Raw subreads.bam movie files include a prefix that begins with “m” followed by a string of unique numbers. In this example, the asterisk is a metacharacter recognized in Linux code as ‘any character’ at that position in the pattern. When generating one representative circular consensus sequence we set the minimum predicted accuracy, “--min-rq,” to 0.9 using ccs 4.0.0 (commit SL-release-8.0.0):

```
$ ccs m*.subreads.bam m*.ccs.bam --min-rq 0.9
```

The primer file format requirements are specific to the version of the lima command used. We used lima 1.10.0 (commit SL-release-8.0.0), each line in the primer fasta file must contain a unique sequence. Since a universal forward primer was used in barcoding this means there was simply one line that has the suffix “5p” to indicate a 5’ primer with the corresponding sequence, also referred to as the left adapter. The other six lines have the suffix “3p” and list the reverse complement of a concatenated sequence consisting of the universal adapter and unique tissue-time-of-day dependent primer. The prefix for the primers was annotated “time_tissue.” With a properly formatted primer.fasta the following command can be invoked to demultiplex the data. This command uses the options “--isoseq” and “--peek-guess” to activate specialized Iso-Seq scripts and to remove false positive barcode hits, respectively. Input m*.ccs.bam from the previous step to output full length reads, m*.fl.5p--tissue_3p.bam.

```
$ lima m*.ccs.bam barcoded_primers.fasta m*.fl.bam --isoseq --peek-guess
```

One of the outputs from lima is “lima.report.” This file can be passed through detail_report.R scripts obtained from https://github.com/PacificBiosciences/barcoding for quality control and troubleshooting. The scripts were cloned into RStudio workspace following the “Cloning GitHub Repositories” guide above.

Some full-length reads may still have concatemers. The next command refines the reads to full length non-concatemers (flnc) using isoseq3 3.2.2 (commit SL-release-8.0.0) for every tissue-time combination in all eight samples. Use the same primer set from the previous lima command.

```
$ isoseq3 refine m*.fl.5p--tissue_time_3p.bam primers.fasta  m*_tissue.flnc.bam
```

At this point all of the samples are still separated into the eight size-dependent folders. Next, samples need to be collected and merged based on tissue and time of day. One way to manage the files for this step is to make a directory for each sample type being analyzed. For example, two directories will be labeled with either time A or time B, three directories with tissue 1, 2, and 3 and six directories corresponding with tissue and time. Next, the "cp" command can be executed to transfer all corresponding flnc files to the directories above. The code below copies all source arguments to the target directory in the syntax: destination source. The command line assumes the user has navigated to the directory where the above directories are located and can therefore use "." in the path to the target directory; in this case the directories were created under ~/smrtanalysis/cluster/.

```
# collecting all flnc reads
$ cp -t ./all/ ~/ngs/*/m*.flnc.bam
# time dependent reads
$ cp -t ./timeA/ ~/ngs/*/m*timeA*.flnc.bam
# tissue dependent reads
$ cp -t ./tissue1/ ~/ngs/*/m*tissue1.flnc.bam
# time and tissue dependent
$ cp -t ./timeA_tissue1/ ~/ngs/*/m*timeA_tissue1.flnc.bam
```

All flnc files are copied and sorted based on their biological relevance and can now be merged using the "ls" command.

```
$ ls m*.flnc.bam > prefix.flnc.fofn
```

Once the files have been merged, they can be clustered into consensus sequences by invoking "isoseq3 cluster," an iterative cluster algorithm. The options "verbose" and "use-qvs" provide process details and use CCS quality values with a partial order alignment algorithm set to 100.

```
$ isoseq3 cluster prefix.flnc.fofn prefix.bam --verbose --use-qvs
```

2.4.4 Mapping to a Reference Genome

The reference genome can be downloaded into the workspace using the "wget" with the url in the Terminal command line. In this model, *Humulus lupulus* v. Teamaker is used and the full genome is retrieved from HopBase. Notice that the file is gzipped "gz" to save space. Both Minimap2 and GMAP can be fed the .gz file directly without the need to unzip first. The file retrieved will be referred to as "ref.fasta.gz" in the command lines provided in this section.

```
$ wget http://hopbase.org/content/Teamaker/hopbase.teamaker.v1.0.20150713.fasta.gz
```

2.4.2-I Minimap2

The minimap2 script is included in SMRT Tools so no further installation is required. This algorithm inputs the reference genome and hq.fasta from isoseq3 clustering to generate a SAM file that can later be collapsed into unique reads. The hq.fasta file generated in the clustering step contains the suffix ".gz" then it will need to be unzipped before it can be used. This can be done with the Linux "gunzip" command followed by the file name.

Indexing the reference genome with minimap2 is an option that can reduce the run time of the alignment algorithm. To index the reference genome in minimap, use the "-d" option to output the ".mmi" minimap-indexed reference genome from the fasta.gz raw reference genome.

```
#index the reference genome
$ minimap2 -d prefix.mmi ref.fasta.gz
#align hq.fasta reads to the indexed genome
$ minimap2 -ax splice -uf --secondary=no prefix.mmi hq.fasta > prefix.sam
```

To align reads without indexing the reference genome use the following command line.

```
$ minimap2 -ax splice -uf --secondary=no ref.fasta.gz prefix.hq.fasta.gz > prefix.sam
```

Whether the reference genome is indexed or not the same options can be used. The first "-ax splice" tells the minimap2 algorithm: "-a" to generate a CIGAR and output alignments in SAM format and "x splice" specifies that it is mapping long read spliced alignment so that the CIGAR denotes long

deletions with “N” to indicate an intron. CIGAR stands for Concise Idiosyncratic Gapped Alignment Report and uses a string of numbers and letters in shorthand to indicate position alignment of the query fasta file with the reference genome. To find canonical splicing sites, GT-AG, on the transcript strand use “-uf.” The “--secondary=no” option prevents the output of secondary alignments, a step that is required for the output SAM file to be compatible with the cDNA_Cupcake post-processing scripts.

2.4.4-II GMAP

Once GMAP is installed using the Conda Install Command, the first step is to index the reference genome with “gmap_build.” In the following command line “-D” tells the software what directory the reference genome is in, “-d” is the genome name, and “-g” indicates that the reference genome fasta file is gzipped.

```
$ gmap_build -D ~/path/to/ref/ -d ref -g ref.fasta.gz
```

The output of this command is a directory with the name indicated after “-d.” Next the clustered hq.fasta can be aligned to the indexed genome. Be sure to use “gunzip” to unzip the hq.fasta file if it contains a “.gz” suffix. Both “-D” and “-d” are the same as in the indexing step above. To generate a SAM file as the output format use “-f samse.” Setting “-n” as 0 gets both single and chimeric alignments. The user can set the number of worker threads designated to the alignment using “-t.” This number is determined by the cores available on the server. It is recommended to use as many as possible to reduce the time it takes to align reads, however, if the server is needed to run other processes or if it is shared among multiple users it is beneficial to leave a few cores available for other projects. The option “--cross-species” increases the algorithm's canonical splice search sensitivity. Max length for first or last intron default is set to 10000 and can be increased to the recommended 200000 with the “--max-intronlength-ends” option. Iso-Seq analysis outputs sequences in the sense direction so “-z sense_force” is used to indicate cDNA direction to the algorithm.

```
$ gmap -D ~/path/to/ref -d ref -f samse -n 0 -t 13 --cross-species --max-intronlength-ends 200000 -z sense_force cluster.hq.fasta >  prefix.sam
```

2.4.5 Isoform Analysis

After the Iso-Seq fasta files have been aligned to the reference genome the resulting SAM file can be processed for isoform analysis. Third-party Python and R scripts compatible with clustered and aligned

isoforms have been compiled by Elizabeth Tseng into one GitHub repository named cDNA_Cupcake. This repository also contains the scripts for ToFu which has been referenced in both maize and sorghum Iso-Seq studies (Abdel-Ghany et al. 2016, Wang et al. 2016). Instructions on the invocation and use of these scripts can be retrieved from https://github.com/Magdoll/cDNA_Cupcake/wiki.

2.4.6 Visualizing Isoform Data in RStudio

The following scripts can be used to visualize isoform data from "abundance.txt" and "read_stat.txt" files obtained from Isoform Analysis.

This source script can be executed in RStudio console to visualize the number of collapsed isoforms that were mapped to an annotation during alignment, referred to as unique reads, as compared to the unmapped reads across six samples. The user inputs information including the path to the directory that contains all sample "read_stat.txt" files. The Sample prefix prompt assumes the user will input "A_Tissue1" for the source to locate the file "A_Tissue1.collapsed.read_stat.txt" by parsing the user inputs together and importing "~/smrtanalysis/clustered/read_stat.txt/A_Tissue1.collapsed.read_stat.txt" as a dataframe that can then be analyzed and plotted. The data is appended to a table, plotted on a bar graph, and all outputs are saved to the sample directory. The source code can easily be customized to add or reduce the number of samples as the source code relies on a numbering template that is consistent throughout (see Appendix).

```
#load the source script to the console
source('~/path/to/6_stat.analysis.R')

#example user inputs
path to SAMPLE directory: ~/smrtanalysis/clustered/read_stat.txt
SAMPLE 1 prefix: A_Tissue1
SAMPLE 2 prefix: B_Tissue1
SAMPLE 3 prefix: A_Tissue2
SAMPLE 4 prefix: B_Tissue2
SAMPLE 5 prefix: A_Tissue3
SAMPLE 6 prefix: B_Tissue3

# example console output
[1] "A_Tissue1 Count Unmapped:  53198 Count Unique:  59633 Percent Unique reads:
52.8516099298951"
[1] "B_Tissue1 Count Unmapped:  68818 Count Unique:  61225 Percent Unique reads:
47.0805810385795"
[1] "A_Tissue2 Count Unmapped:  71895 Count Unique:  49486 Percent Unique reads:
40.769148384014"
[1] "B_Tissue2 Count Unmapped:  86015 Count Unique:  43027 Percent Unique reads:
33.3434075727283"
[1] "A_Tissue3 Count Unmapped:  71983 Count Unique:  62618 Percent Unique reads:
46.5211996939101"
[1] "B_Tissue3 Count Unmapped:  91411 Count Unique:  76145 Percent Unique reads:
45.444508104753"
[1] "Analysis complete. All files saved to ~/smrtanalysis/clustered/read_stat.txt"
```

The user can cross check output by matching “Unique” count numbers obtained from the source above to “Total Number of FL reads:” in the corresponding “abundance.txt” file.

2.5 Discussion

Iso-Seq analysis and post-cluster processing can be a daunting undertaking for beginners in bioinformatics. Not only does it require competency across multiple coding languages and familiarity with layering software requirements, but also proficient understanding of the algorithms used to select appropriate options, arguments, and syntax; the explanations of which are often overlooked in available workflows which often cater to developers and IT personnel.

Biologically relevant comparisons between samples can be obtained as early as the demultiplexing stage of Iso-Seq analysis. Though there are several long-read compatible alignment software available, we found that minimap2 is the most accessible for a streamlined end-to-end workflow. Managing package downloads and software installation through Conda, Miniconda, or Bioconda provides a considerable advantage for beginners with consistent syntax and compatibility.

Herein we provide an end-to-end workflow from processing raw data files obtained from PacBio SMRT sequencing through alignment to a reference genome and isoform count analysis. With the tools and tips provided, interdisciplinary and non-bioinformatician users can process Iso-Seq reads and develop skills that will increase accessibility to downstream analysis. We also provide two Rscript source codes that can be used in RStudio Console to output visuals of abundance.txt and read_stat.txt files obtained from collapsing aligned reads. The files obtained from this workflow can also be used in genome annotation as well as in-depth isoform classification.

CHAPTER 3

TEMPORAL AND TISSUE SPECIFIC HOPS ISOFORMS ANALYSIS USING LONG-READ SEQUENCING

3.1 Abstract

Isoforms are proteins of similar functionality with differing amino acid sequences. While some may arise from entirely separate loci in the genome, many originate from the same gene that has undergone alternative splicing. The latter illustrates biology's ability to amplify its complexity from a simple fragment of DNA to vast biochemical networks. The complexity of these networks aids plant survival despite their mobility constraints by orchestrating detection and response to changes in the environment.

Hop plants, while most known for their role in beer production, are a rich source of nutraceuticals, including antioxidants, antimicrobials, and essential oils. Though their metabolite profiles are extensive, there are restrictions in the tools available to further examine biochemical pathways due to a lack of a high-quality genome. One way to address this challenge is to use long-read sequencing. Here we analyze long-read isoforms through PacBio's Iso-Seq in *Humulus lupulus* Teamaker variety leaves, roots, and shoots sampled at dawn and dusk. We produced 373,555 full-length reads from three tissues at two times of day that mapped to the current Teamakers genome assembly, accounting for only ~40% of the total reads detected. Herein we found that over 5,000 isoforms are time of day or tissue specific, and nearly 4,000 unique isoforms that are dependent on both tissue type and time of day.

3.2 Introduction

Hops plants have extensive medicinal and cultural histories. Medieval European cures used the cones as a primary ingredient to treat various ailments from blood impurities to liver, spleen, and skin diseases. Ironically, hops are now known to cause acute dermatitis in 1 of every 30 handlers. The physical record of their cultivation dates back circa 760 CE, where germanic records indicate Charlamagne's father, Pepin le Bref, donated a hops garden to the Monastery of St. Denis (Edwardson et al. 1952). The origin of the scientific name *Humulus lupulus* is debated by etymologists still today. However, some argue it may be an homage to these germanic roots, *Humulus* being similar to the low german word for hops. Monasteries were well known for their brews, and within 300 years of Pepin le Bref's donation, hops overtook previous bittering agents used in brews and are now predominantly used in crafting beer. Initially valued for their flavoring profile, hops also contain antimicrobial phytochemicals that increase the shelf life of the brew, thus opening the gates to commercialization. Before the monastery

commercialization of breweries, women were the majority in this now male-dominated field. Often referred to as ale-wives, their pointed hats, large cauldrons, and reputation would later be associated with witches and witchcraft, a perception that has survived millennia. Another perception that prevailed upon commercialization of hops cultivation was the poor work ethic of hops farmers. Farmers at the time were known to fall asleep in the fields as they were collecting cones. Later, dried hops cones would become a recommended sleep aid; a property eventually attributed primarily to the metabolite lupulin. Sleeping on a hops-stuffed pillow is a remedy that even King George III is said to have used.

Despite its rich history, commercial and medicinal value, and complex composition, functional genomics and biochemical research on hops are just breaking the surface. The disparity in research progress is due in part to a lack of a high-quality, well-annotated genome. In the era of data mining, high-quality genome assemblies are an invaluable resource for proposing biochemical regulatory networks and specialized metabolite biosynthetic gene identification. Candidate biochemical reactions can be pulled from matched genome assemblies and metabolic databases to draft whole-genome metabolic models (reviewed in Rai et al. 2019). Another challenge in biochemical research of specialized metabolite-rich plants like hops is the high plasticity in metabolite composition. The phytochemical composition is dynamically affected by diel and seasonal variations in environmental inputs, including light cycles (Liebelt et al. 2019). On a transcriptional level, these variations can be connected in part back to alternative splicing and the generation of isoforms. Tissue specificity of isoforms involved in metabolite biosynthesis pathways has been identified, such as lignifying cell-type-specific isoform expression of 4-coumarate:CoA ligase in phenylpropanoid metabolism (Li et al. 2015). Isoforms also show differential expression throughout the day. Typically, cell composition can be broken down into the three layers: transcriptome, proteome, and metabolome. Typically, biochemical interactions such as targeted degradation explain fluctuations in cellular abundance of observed proteins or metabolites. A steady rise in research on alternative splicing has revealed several examples of isoform involvement in diurnal regulatory pathways. For example, the abundance of CONSTANS (CO), involved in triggering photoperiodic flowering, has recently shown self-modulation using two distinct isoforms in addition to canonical regulation by sequential E3 ubiquitin ligases (Gil et al. 2017).

A unified hops genome assembly database, HopBase, ushered this plant species into the age of functional genomics in 2017 (Hill et al. 2017). While recent progress has been notable, Teamaker genome assemblies, and more critically, gene annotations can still be improved. Long-read sequencing technologies, such as Pacific Biosciences (PacBio) small-molecule real-time sequencing, can be applied to improve even high-quality genome assemblies like *Zea mays* (Wang et al. 2016).

Herein we use PacBio SMRT Analysis Iso-Seq to analyze Teamaker isoforms in leaves, roots and shoots at dawn and dusk. The processed Iso-Seq data can be used further to improve the precision of the

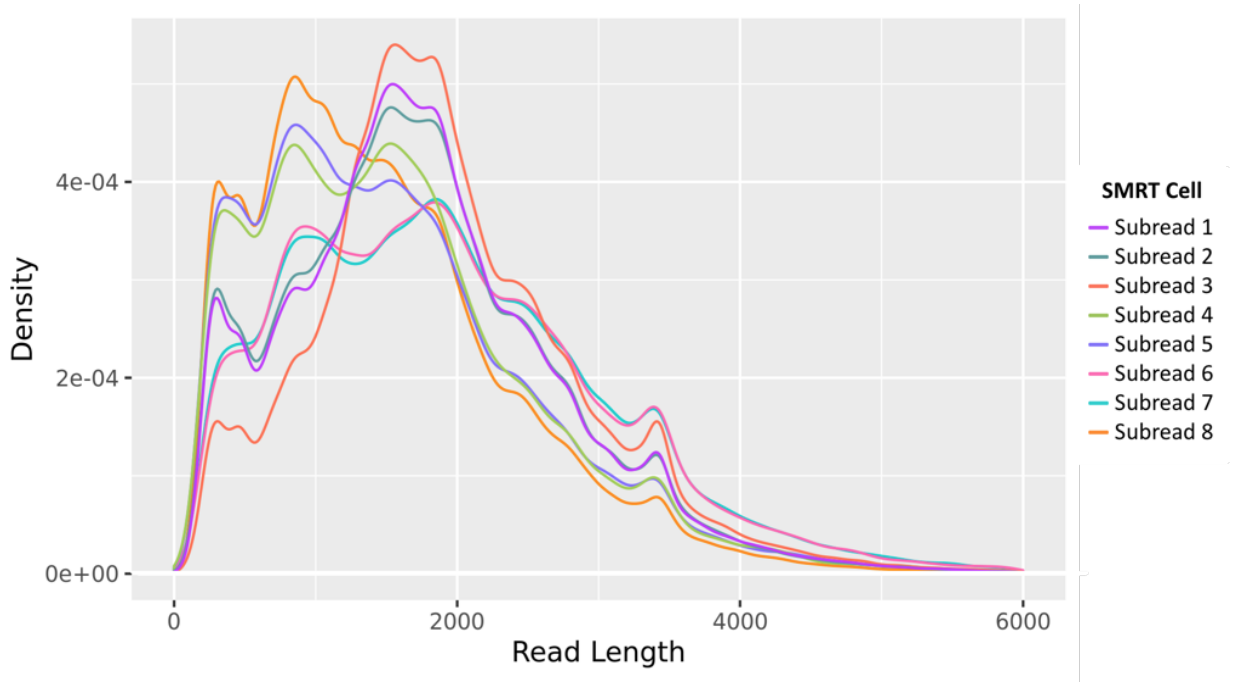
current Teamaker genome assembly, identify tissue, and time of day specific gene networks and has the potential to be used in drafting a whole-genome metabolic model for *Humulus lupulus*.

3.3 Results.

3.3.1 Iso-Seq bioinformatics pipeline

We analyzed sequenced subreads from eight libraries of barcoded cDNA obtained from reverse-transcribed high-quality RNA, extracted from three tissues of the *Humulus lupulus* Teamaker variety at two times of the day in Waddell (2018). We obtained a total of 3,456,715 reads from raw sequences. Of those, less than half generated circular consensus sequences (1,460,128 reads which equate 42%), and nearly all reads discarded were filtered due to a lack of full pass (1,931,961 reads which equate 97%). The majority of CCS full-length (FL) reads range between 100-4,000 bp (Fig. 3.1A). Demultiplexed reads showed similar distribution across all libraries and barcoded pairs (Fig. 3.1B). Barcode scores are an indicator of how well the pairs match the sequence by taking the average of both the forward and reverse primer match to the sequence (Armin Töpfer 2018), overall barcodes scored high across all libraries (Fig. 3.2). FL demultiplexed reads were merged by barcode identifiers. After removing concatemers and applying the Isoseq3 iterative clustering algorithm, we yielded a total of 924,361 reads with 108,271 high-quality (HQ) transcript sequences.

A.



B.

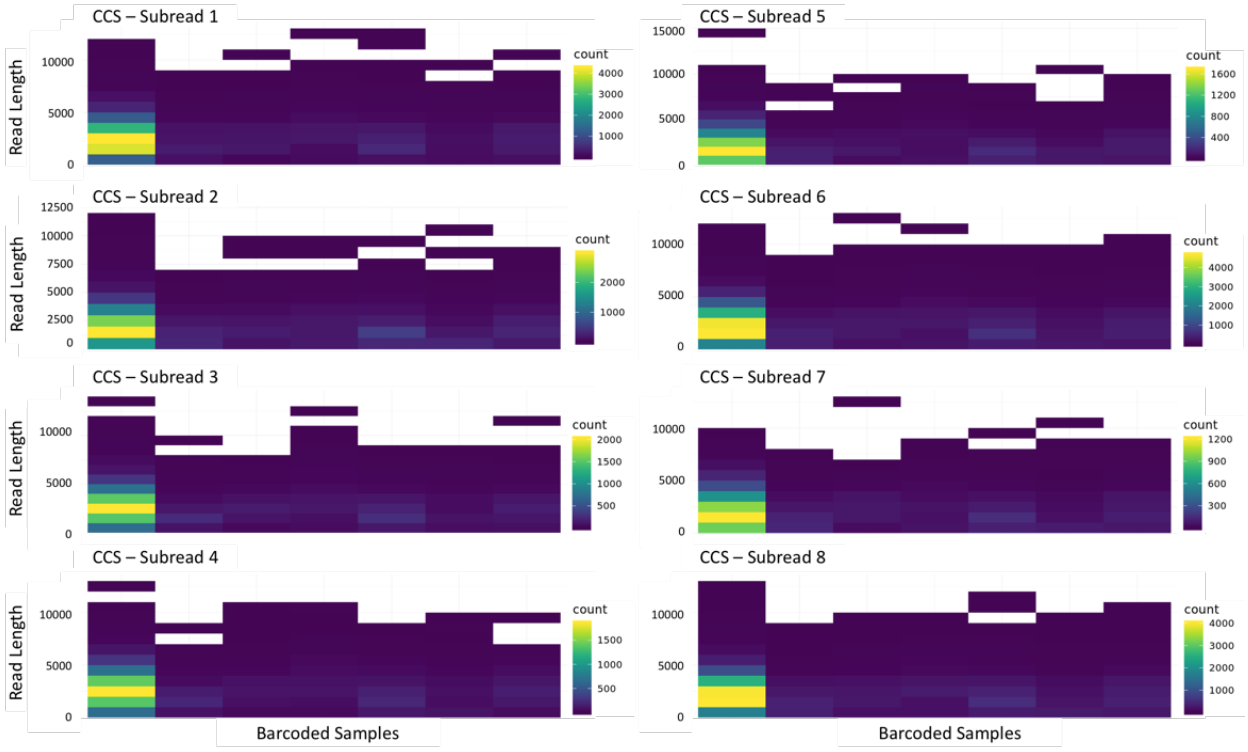


Figure 3.1 ***SMRT Cell Read Length Distribution.*** A) FL CCS library size range. B) Summarized count distribution by read length and barcoded pair. All subread samples consistently show the highest density of reads of less than 5000 base pairs.

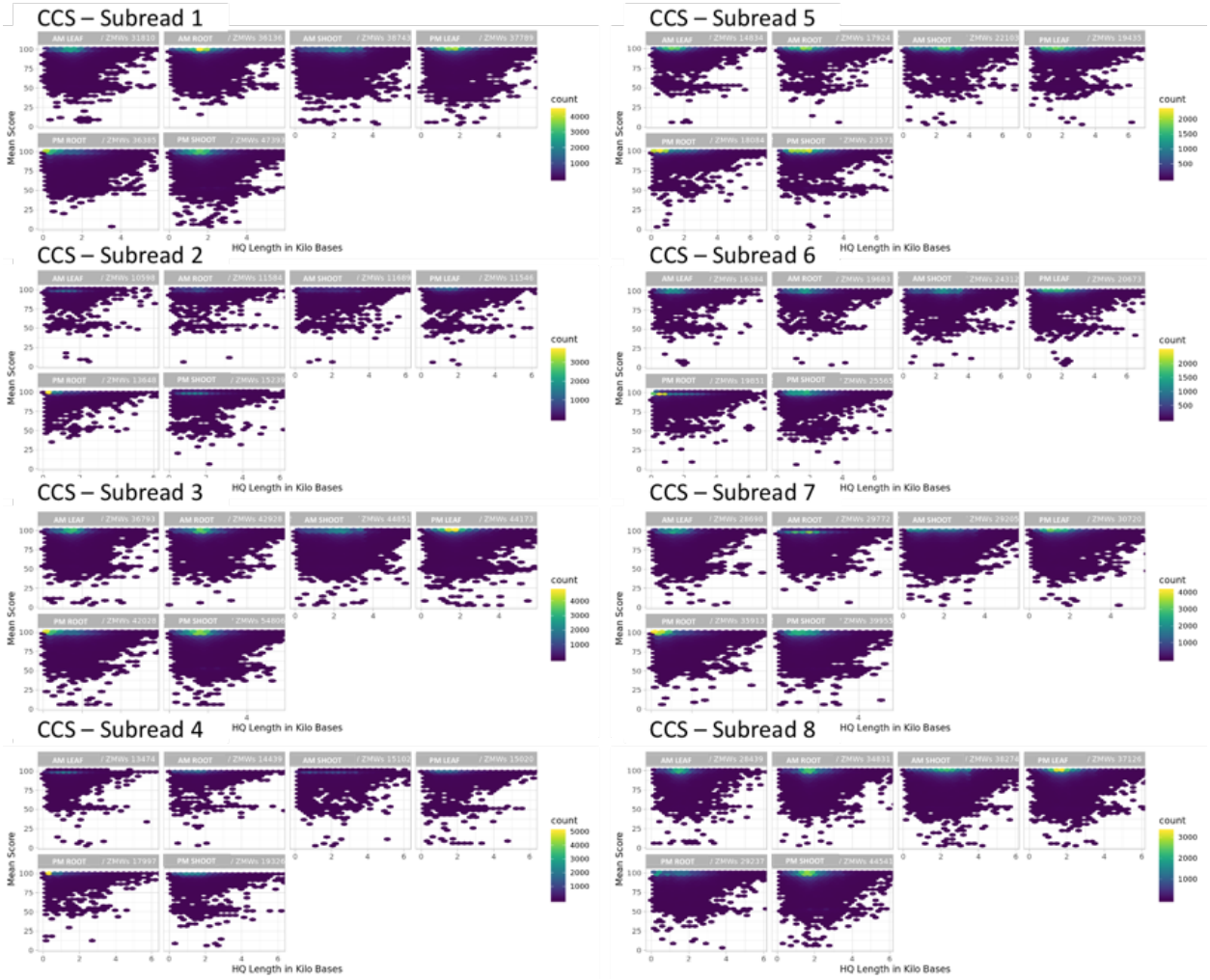


Figure 3.2 ***Count Density by Mean Barcode Score vs. HQ Length.*** Distribution of mean score of barcode pairs against HQ Length in kilo bases (Kb) by count across eight subread libraries.

3.3.2 Genome Alignment

HQ Transcripts were mapped to the Teamaker Genome using Minimap2 aligner included in the SMRT Tools software suite. Reads were then collapsed using cDNA Cupcake ToFu (https://github.com/Magdoll/cDNA_Cupcake). In total we yielded 14,801 mapped, non-redundant transcripts with a total of 373,555 FL reads. Samples collected at dusk generally yielded a higher number of isoform transcripts overall as well as in leaves and shoots, excepting dawn root samples (Table 3.1). Between 30-50% of the FL reads were mapped to the reference genome. Despite yielding a lower number of isoforms, dawn samples had a higher percent match than dusk, with leaves having the highest percent mapped and roots the lowest (Table 3.2). Dusk samples had consistently higher total FL counts by tissue

type, however, upon collapsing the counts associated with non-redundant isoforms dawn root samples retained more counts and surpassed dusk counts per unique read (Figure3.3).

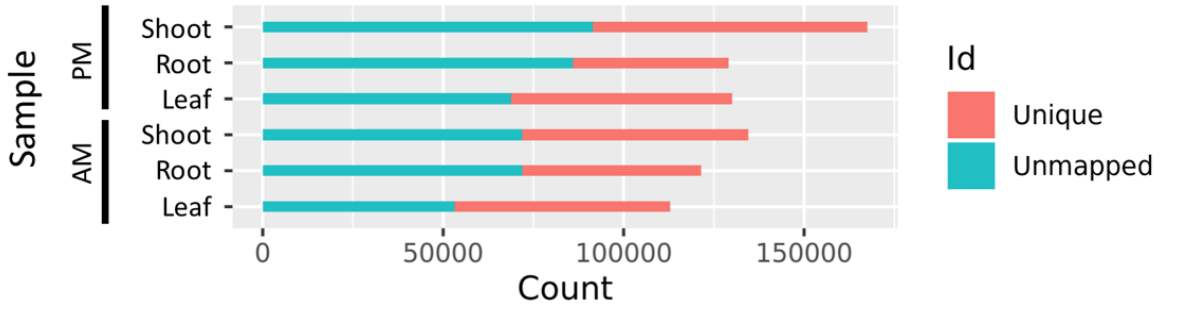
Table 3.1 ***HQ and Non-Redundant Isoform Sequences by Tissue and Time of Day.*** HQ transcripts represent the number of transcripts obtained after clustering, and unique transcripts are the number of non-redundant isoforms yielded upon collapsing mapped HQ transcripts.

	HQ Transcripts	Unique Transcripts
Time of day		
AM	57746	9337
PM	62338	9813
Tissue		
Leaf	37192	6986
Root	37844	6038
Shoot	47876	8173
Sample		
AM Leaf	17489	3963
AM Root	20985	4072
AM Shoot	22950	4516
PM Leaf	21730	4612
PM Root	18030	3166
PM Shoot	27813	5614

Table 3.2 ***Abundance of Unmapped and Unique Isoforms.*** Abundance of unmapped and unique isoform transcripts by tissue and time of day aligned to the Teamaker reference genome and percent counts retained after collapsing aligned transcripts.

	Count		
	Unmapped	Mapped	Percent Match
Time of Day			
AM	223388	189209	46
PM	273555	199688	42
Tissue			
Leaf	133096	129558	49
Root	169082	99915	37
Shoot	179350	148776	45
Sample			
AM Leaf	53198	59633	53
AM Root	71895	49486	41
AM Shoot	71983	62618	47
PM Leaf	68818	61225	47
PM Root	86015	43027	33
PM Shoot	91411	76145	45

A.



B.

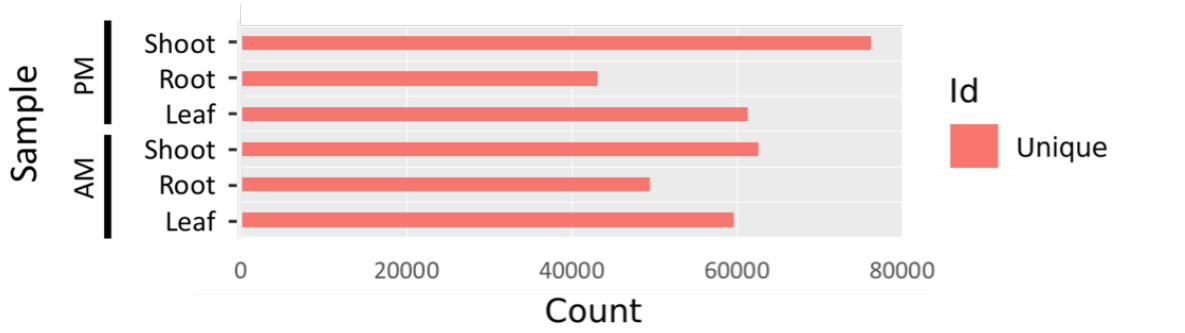
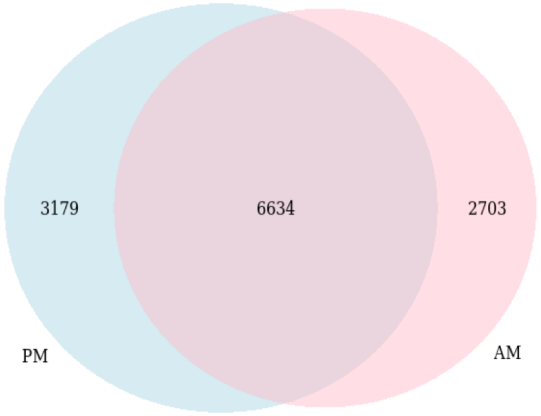


Figure 3.3 ***Abundance of Unique and Unmapped Transcripts by Tissue and Time of Day.*** A) Total abundance of unique and unmapped HQ transcripts aligned to the Teamaker genome. B) Abundance of non-redundant collapsed transcripts.

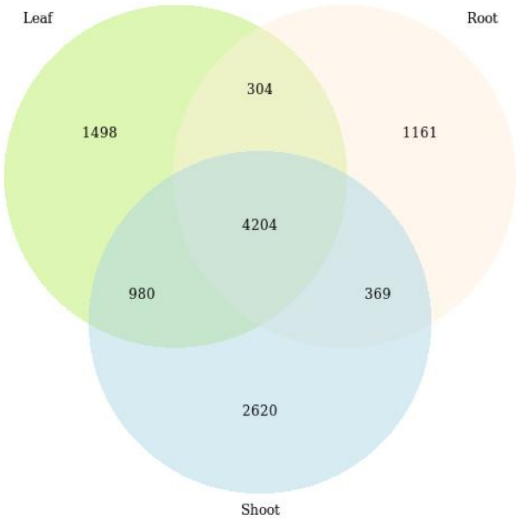
3.3.3 Tissue and Time of Day Isoforms

Non-redundant isoforms were compared across all samples to determine overlapping reads (Table 3.3). Nearly half of the transcripts were time of day specific (5,882 47%) with 3,179 transcripts unique to dusk and 2,703 unique to dawn (Figure3.4A). We found 5,279 transcripts that are tissue dependent with leaf and shoot isoforms having ~3x more overlap with each other than with root isoforms (Figure3.4B). We discovered that 3,917 transcripts were dependent upon both tissue and time of day, particularly dusk shoot samples which contained 1,537 unique isoforms (Figure3.4C). Interestingly, PM Shoots had twice as many common isoforms with PM Leaf than AM Shoot.

A.



B.



C.

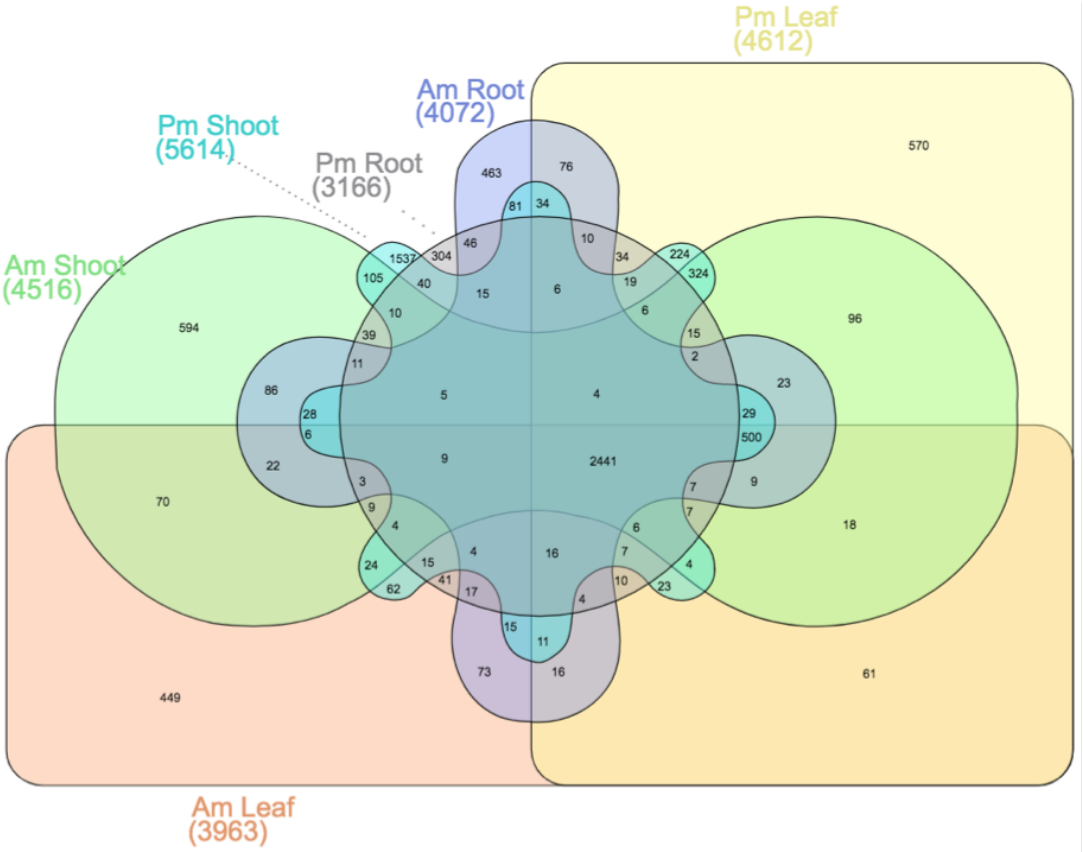


Figure 3.4 ***Overlap of Non-Redundant Isoforms by Tissue and Time of Day.*** A) Time of day dependent isoforms. B) Tissue dependent isoforms. C) Time of day and tissue dependent isoforms.

Table 3.3 ***Isoform Specificity.*** A comparison of the number of isoforms unique to the sample type versus those that are common across all types by time of day, tissue, and sample type.

	Isoforms	Shared
Time of day		
AM	2703	6634
PM	3179	
Tissue		
Leaf	1498	
Root	1161	4204
Shoot	2620	
Sample		
AM Leaf	449	
AM Root	463	
AM Shoot	594	2441
PM Leaf	570	
PM Root	304	
PM Shoot	1537	

3.4 Materials and Methods

3.4.1 Data Acquisition and Iso-Seq Analysis

All data analyzed herein were obtained from the RNA extracted by Eric Waddell from three-month old hops tissues grown in a 15-hour photoperiod just after dawn and dusk (Waddell 2018). Iso-seq Analysis was completed using SMRT Tools based on Gordon et al. (2015). Samples were processed and visuals presented as described in Chapter 2 above.

3.4.2 Barcoding Library

All samples were barcoded using a universal forward primer, or left adapter, (5'-AAGCAGTGGTATCAACGCAGAGTAC -3') and a sample specific reverse primer (Table 3.4).

Table 3.4 ***cDNA Barcodes***. 3'-cDNA tissue specific barcodes containing the universal right adapter concatenated with sample specific primer.

	3'-cDNA Barcodes
AM Leaf	GAGTGCTACTCTAGTAGTACTCTGCGTTGATACCACTGCTT
AM Root	GCATATAGTAGAGATCGTACTCTGCGTTGATACCACTGCTT
AM Shoot	ATGACGCATCGTCTGAGTACTCTGCGTTGATACCACTGCTT
PM Leaf	CATGTACTGATACACAGTACTCTGCGTTGATACCACTGCTT
PM Root	CAGCAGTATAGACTGTGTACTCTGCGTTGATACCACTGCTT
PM Shoot	GCAGAGTCATGTATAGGTACTCTGCGTTGATACCACTGCTT

3.4.3 Mapping PacBio Data

Clustered isoforms were aligned to the hops Teamaker genome assembly using Minimap2. Aligned reads were then collapsed and abundance was calculated using scripts from cDNA Cupcake ToFu (https://github.com/Magdoll/cDNA_Cupcake).

3.4.4 Tissue and Time of Day Dependent Isoforms Quantification

Sample transcript overlap was calculated by comparing assigned PBIDs. The visual models were created using VennDiagram (Chen 2018) and the R script generated in chapter 2 above (see Appendix) for comparisons by time of day and by tissue. Due to a five-sample limit in VennDiagram, InteraciVenn was used to visualize the six-sample comparison (Heberle et al. 2015).

3.5 Discussion

Functional genomics in hops is just scratching the surface. Numerous plant species have shown either circadian, diel, or photoperiod-driven rhythmic expression of genes involved in specialized metabolite regulatory and biosynthetic pathways (Alabadí et al. 2002; Filichkin et al. 2011; Pavarini et al. 2012; Gyllenstrand et al. 2014; Soni et al. 2015; Fenske and Imaizumi 2016; Zeng et al. 2017; Greenham et al. 2017; Koda et al. 2017; Weiss et al. 2018; Yin et al. 2018; Yoshida and Oyama-okubo 2018). With the introduction of next-generation long-read sequencing tools, we can now capture entire full-length transcripts to detect alternatively spliced isoforms accurately. With the availability of a high-quality genome and a metabolite database, isoform data can bridge the two to draft a genome-wide metabolic model.

In this study, we examine isoforms in three-month-old *Humulus lupulus* Teamaker leaves, roots, and shoots just after dawn and dusk in a 15-hour photoperiod. We found that non-redundant isoforms overall were influenced dramatically by both tissue-type and time of day. Using the current Teamaker genome assembly, PM shoot samples had the highest number of unique isoforms. While there is a

possibility that this is indicating some biologically significant regulation that is happening in the shoot at dusk, it is essential to note that the current state of the Teamaker genome is not considered high-quality. Thus, the majority of full-length isoform transcripts detected were discarded after alignment to the genome due to low mapping. The clustered transcripts generated from Iso-Seq analysis in this study can be used to improve the quality of the Teamaker genome and to build a more accurate annotation.

CHAPTER 4

PRELIMINARY ANALYSIS OF BROCCOLI AND RADISH MICROGREEN METABOLIC RESPONSE TO WARM NIGHTTIME TEMPERATURES

4. 1 Abstract.

Fluctuations of internal phytochemical composition optimize plants to their surroundings at any given time. Phytochemicals are not only essential to plant survival but also are a vital contributor to the world's nutraceutical supply. Consumer products from food to pharmacy rely on plant derivatives. The production and accumulation of many phytochemicals are tightly controlled based on environmental conditions; therefore, the yield and quality of these chemicals are at the mercy of a dynamic environment. As demand for plant derivatives continues to grow, producers face challenges of inconsistent products due to fluctuating environmental inputs.

The effects of temperature is a major topic in plant research. Many of these studies focus on extremes such as heat or cold stress. Plants, however, are sensitive to less extreme changes in temperature in their environment. Daily temperature cycles can even be used to entrain the circadian clock. There is a disparity in the way the world is warming; while temperatures are steadily increasing all around, nights are warming faster than day temperatures. Warming nights as a result of climate change pose an interesting question of how plant compositions may change in response to the reduction of the differences in daily temperature. As proof of principle, we examine the effects of three-degree warmer nights on metabolite composition of broccoli and radish sprouts. Further exploration could reveal plant varieties and environmental conditions that maximize the desired phytochemical profiles to benefit human health. Additionally, we will gain insights into the impact of changing environments on the regulation of phytochemicals, which could be applied to other crop systems targeted for the production of pharmaceuticals and other consumer products.

4.2 Introduction

The coordination between plants and the daily and seasonally recurring environmental changes has been optimized through millennia of evolution. However, these environmental patterns are changing. Nights are warming (Peng et al. 2004; Vose et al. 2004; Welch et al. 2010; Kumar et al. 2017), seasonal patterns are shifting (reviewed in Parmesan 2006), and patterns of precipitation are changing (Easterling et al. 2017; Unal et al. 2012; Pal et al. 2013; Keggenhoff et al. 2014; Ganguli and Ganguly 2016; Gitau 2016; Rahimpour et al. 2016; Tye et al. 2016; Mallakpour and Villarini 2017; Roque-Malo and Kumar 2017; González-Zeas et al. 2019). Management and harvesting practices refined to optimize

phytochemical production may no longer be suitable to the altered patterns of a changing environment. Yet it is challenging to predict how these changes will impact the phytochemical profile. The changing environmental patterns are complex and may have antagonistic effects on phytochemical profiles, making the combined impact challenging to predict. For example, an increase in CO_2 can lead to an increase in phenolics and a decrease in terpenoids in forest trees in the northern hemisphere while warmer weather leads to an increase in terpenoids and a decrease in phenolics. The combined changes overall are predicted to result in an increase in phenolics in foliage and decrease in woody tissue (Holopainen et al. 2018). How these changes in temperature and CO_2 combine with changes in precipitation patterns on phytochemicals has not yet been explored. Understanding how phytochemicals are affected by daily and seasonal cycles in current conditions and integration of the timing of these events will improve the accuracy of these predictions.

Changes in environmental cycles could also impact the plant's temporal balance with the environment. For example, new climate patterns could allow a herbivore to move into a different temporal niche (Belesky and Malinowski 2016; Porqueddu et al. 2016). If a plant produces defense compounds at a specific time of day or season coincident with the historical highest activity of the herbivores the plant's defenses may not be sufficient to protect the plant in the new conditions. As plants respond and adapt to changing environments, their phytochemical profiles will likely change and can ripple through an entire ecosystem starting with the changes in chemicals that deter herbivores or attract pollinators (Akula and Ravishankar 2011).

Variation in temperature daily and seasonally have been shown to influence phytochemical levels in Brassica. Sprouts germinated at 30°C show higher total glucosinolate and ascorbic acid than 20°C or 10°C germinated broccoli (*Brassica oleracea*) and rocket sprouts (*Eruca sativa*) (Ragusa et al. 2017). The total polyphenol content of broccoli sprouts increases in higher temperatures but decreases in rocket sprouts under the same conditions. Protective functions of phenols in oxidative stress-induced signaling may explain the observed rhythmic accumulation in many plant species. Coordinating antioxidant activity with the recurring timing of local stresses, other phenological activity, or breeding pressures could drive species-specific changes in the waveform of expression that could be locally adaptive. Soengas et al. (2018) observed circadian driven rhythmicity in the pattern of total phenolic compound accumulation in broccoli, cabbage, Chinese cabbage, and turnip greens. Although all species accumulated the highest total phenolic content in the period surrounding dawn, the peak phasing varied between species. Broccoli and cabbage accumulation were the highest in the dark period before dawn, while Chinese cabbage and turnip greens peaked in accumulation at dawn. Other studies, however, show broccoli head (*Brassica oleracea* L.) harvested at the end of the day had higher phenolic content and antioxidant activity during storage than those harvested in the morning or at midday (Hasperué et al. 2011).

4.3 Results

Seven-day old Broccoli sprouts grown under a three-degree higher night temperature had significantly higher Total Phenolic Content (TPC), an average of 62.1 mg GAE/100 g fresh weight (FW) compared to the 31.7 mg GAE/100 g FW of those grown under normal night conditions (Figure 4.1A). Radish sprouts, however, showed no significant differences in TPC between normal night (NNT) and warm night (WNT); 57.0 and 55.8 mg GAE/100 g FW, respectively. UV-HLPC at 280 nm was used to compare specialized metabolite composition across NNT and WNT samples. Both broccoli and radish samples show differences in peak intensities and retention times between NNT and WNT conditions. Broccoli sprouts show variation in peaks at retention times of 14, 14.5, 23, and 33 min (Figure 4.1B, D). Radishes show notable differences in peaks at 23.5, 27, 32, 32.5 and 33-35 min (Figure 4.1C); specifically, between 33 and 35 min WNT radishes have an additional peak compared to NNT.

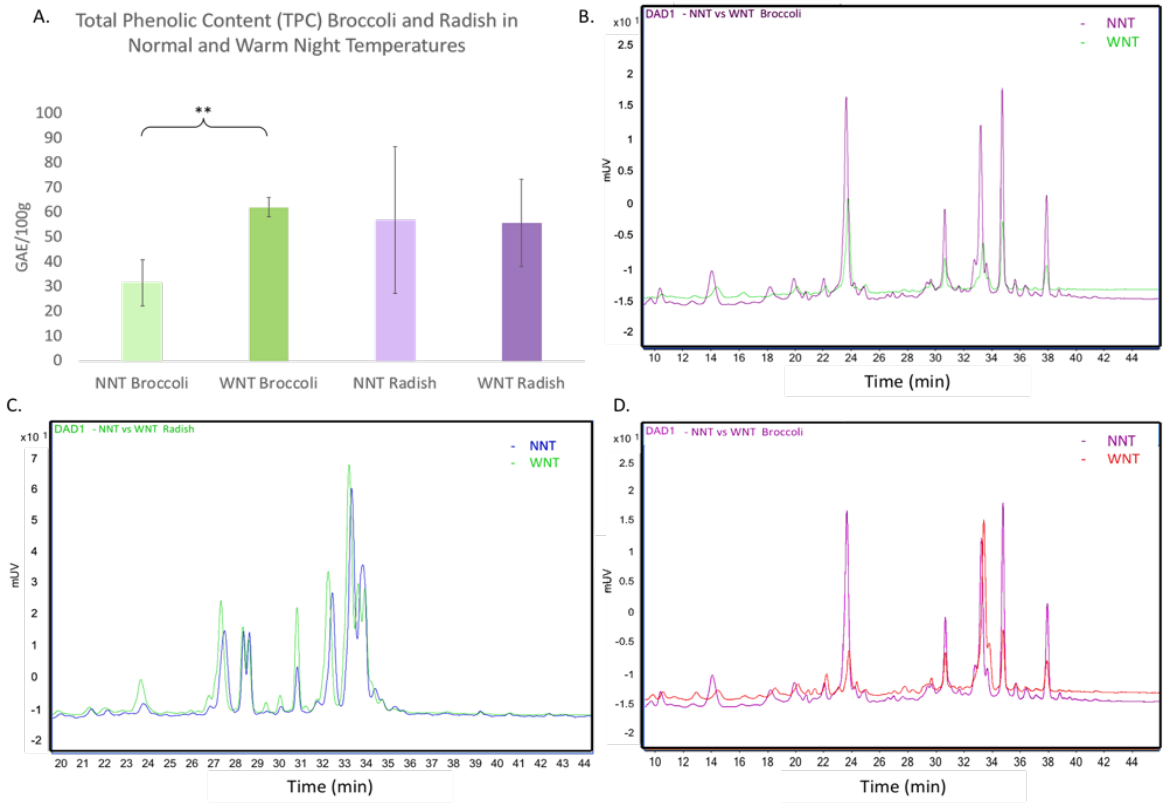


Figure 4.1 ***Comparison of Normal and Warm Night treated Broccoli and Radish Sprout Phytochemical Composition.*** (A) Total Phenol Content (TPC) in mg GAE per 100 g FW between NNT and WNT Broccoli de Cicco and Radish French Breakfast varieties. NNT vs WNT (B) Broccoli sprout and (C) Radish sprout UV-HPLC at 280 nm. (D) A second UV-HPLC chromatogram of NNT vs. WNT Broccoli shows some variability within WNT Broccoli at 23 min when compared to B. ** significance determined with students t-test ($p<0.001$).

4.4 Materials and Methods

4.4.1 Plant Material and Growth Conditions

Broccoli (De Cicco) and Radish (French Breakfast), both in the Brassicaceae family, were chosen for this experiment due to the value of their metabolite profiles with applications to microgreen cultivation as well as prior evidence of high day-temperature responsive specialized metabolite profiles. Each variety was planted in potting soil at ¼ in. depth in a randomized grid within a 48 -ell planting tray and grown in a Percival chamber under a 12-hour photoperiod. The normal night temperature (NNT) group was grown with cycling day and night temperatures of 22°C and 16°C whereas warm night temperature (WNT) group was three degrees warmer at night. Six biological replicates for each variety and experimental condition were harvested six hours after dawn (midday) on the 7th day after sowing and flash frozen in liquid nitrogen for processing (Figure 4.2).

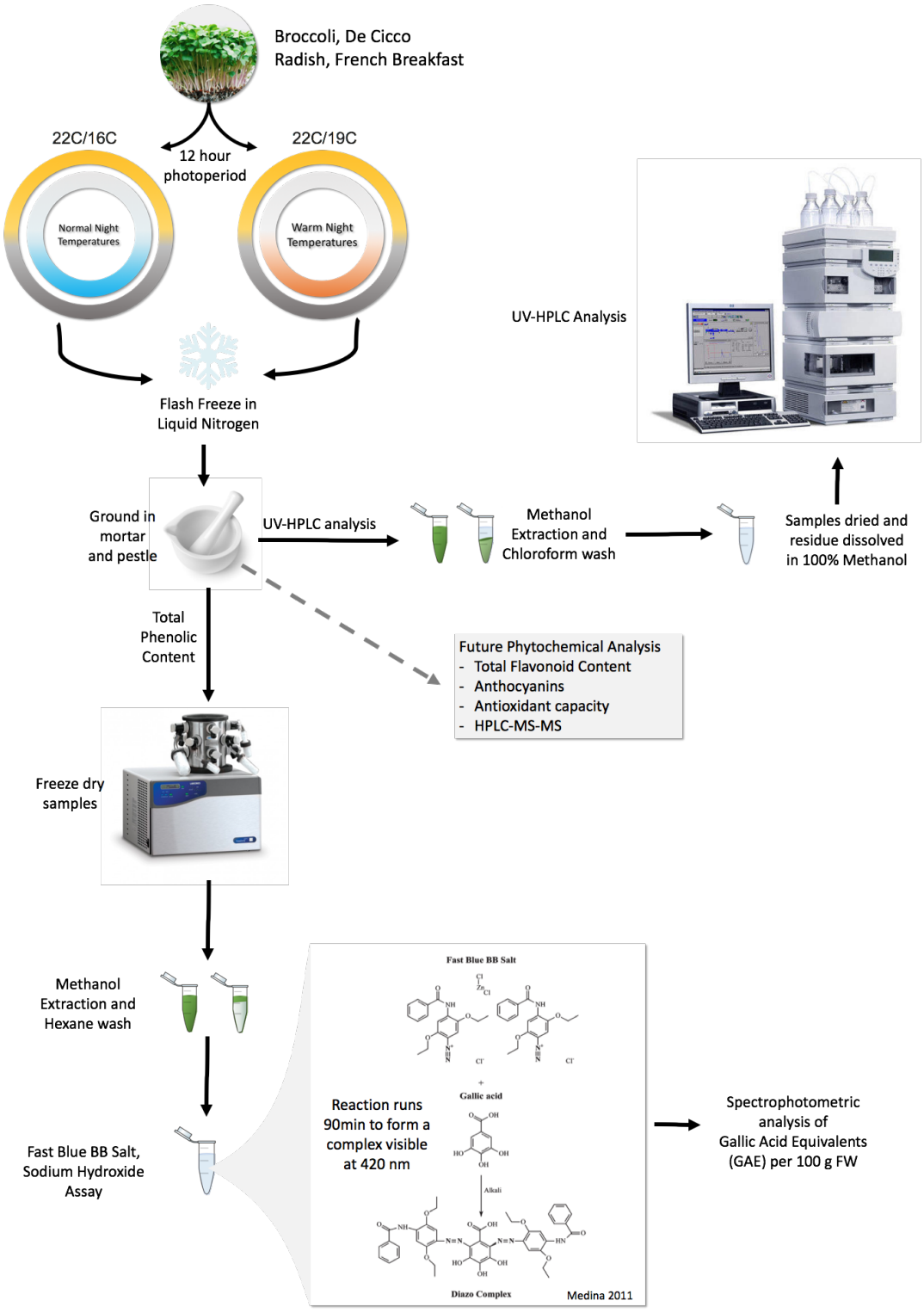


Figure 4.2 ***Plant growth and metabolite analysis workflow.*** Broccoli and Radish sprouts were sampled normal and warm-night conditions at midday and flash frozen in liquid nitrogen to be processed for UV-HPLC and total phenolic analysis.

4.4.2 Total Phenolic Content

Total Phenolic content is determined using a methanol-based extraction modified from Lester et al. 2013 and Xiao et al. 2019. First, lyophilized (freeze dried) powder is washed with hexane to reduce chlorophyll. Once metabolites are extracted, 0.1% Fast Blue BB salt and 5M Sodium Hydroxide are added and react with phenolic compounds to form a diazo dye complex (Medina 2011). This absorbance at 420nm is presented in Gallic Acid Equivalents (GAE) and normalized by weight.

4.4.2 UV-HPLC Analysis

UV-HPLC analysis was completed using an Eclipse XDB-C18 (5uM) column. Data was analyzed using Agilent MassHunter Qualitative analysis B.05.00 Software. Analysis was run by Yue Zhu with methods previously described in Yuzuak et al. (2018).

4.5 Discussion

While it is clear that three-degree warmer nights alter the phytochemical composition of both Broccoli de Cicco and Radish sprouts, as detectable by UV-HPLC, only Broccoli sprouts show significantly different TPC. This supports our hypothesis that total phytochemical concentrations are not enough to determine the effects of incremental increases in night temperatures. Understanding the variation in total concentrations, while still useful, may mask changes in composition.

The temporal variation of phytochemicals across time of day and time of year appears to occur in a diverse species of plants. Cognizance of this temporal control can improve efforts to enhance the production of select phytochemicals. When engineering or selecting plants for enhanced phytochemical activity it will be critical to consider that the timing of the target may vary throughout the day and year. During selection, testing candidates for breeding will require monitoring the desired phytochemical at consistent times as harvesting at different times can result in different results (Hasperué et al. 2011; Pincemail et al. 2012; Ariza et al. 2015). The plant's endogenous clock continues to cycle post-harvest (Goodspeed et al. 2013), and interactions between harvest time and storage length and conditions (Hasperué et al. 2011) will be important to consider during selection.

Temporal and environmental variation can be used to researchers' advantage as a tool to identify regulatory and biosynthetic pathways of desired phytochemicals. Using environmental perturbations to identify regulatory relationships between genes and metabolites has been a successful approach to elucidating the biosynthetic and regulatory components of specialized metabolites. Genes in glucosinolate metabolism were identified by pairing transcriptional and metabolic responses to sulfur depletion (Hirai et al. 2005, 2006). In a study by Li et al. (2018) that used Dimocarpus longan Lour. embryonic calli, to evaluate the impact of photoperiod, light intensity, and light quality on flavonoid content, they mapped

the variation across these environmental perturbations to identify regulatory components correlated (or anti-correlated) with the changes in flavonoid content, identifying candidate mechanisms for miRNA as regulators of flavonoid accumulation. This approach can be used to exploit variation that extends across time of day and year to identify biosynthetic components and regulatory pathways for a specific metabolite. The temporal resolution also provides enhanced separation in clustering approaches. Clustering volatiles across time has been used to identify their potential environmental sensitivity (Borges et al. 2013). Knowing the sensitivity of each group can be used to interrogate their relationships, their potential regulation mechanisms, and how each group may respond to changes in climate; making temporal variation a powerful tool for identifying regulatory relationships. Not only is the accuracy of identifying transcriptional regulators through gene networks improved by the addition of temporal information (Madar et al. 2009; Desai et al. 2017; Sanchez-Castillo et al. 2018; Yang et al. 2018), but also can be extended to identifying regulators of metabolite accumulation (Hannah et al. 2010; Pérez-Schindler et al. 2017).

Engineering phytochemical accumulation may also be facilitated by analyzing the daily and seasonal flux in the phytochemical profile. For example, efforts to engineer plants to enhance artemisinin production or other compounds for pharmaceutical or industrial use requires increasing the concentration, storage, deliverability, stability, purity, and performance of the desired phytochemical (Pulice et al. 2016, De Martinis et al. 2016). Each of these steps can be affected by environmental factors, and production of phytochemicals can be improved by considering environmental stresses (reviewed in Rai et al. 2011; Naeem et al. 2017; Aftab et al. 2017). As plants tightly control the timing of their metabolism and stress responses, working within the framework of this temporal regulation will aid efforts to enhance production. Strategies to enhance production through altering environmental stimuli (Fujiuchi et al. 2016; Huang et al. 2016) will likely benefit from considering the native rhythm of environmental stresses and the endogenous plant response to maximize the effectiveness of the treatments. Designing molecular and biochemical strategies that integrate the temporal variation in primary and specialized metabolism may improve the success in optimizing phytochemical production.

For some medicinal products, whole-plant extracts have added benefits over single-molecule products (Adwan and Mhanna 2009; Deharo and Ginsburg 2011; Herranz-López et al. 2018). Inconsistencies in whole-extract profiles can impact their efficacy and acceptance as a viable alternative to single products. Considering the temporal variation in both the primary bioactive product and synergistic molecules may reduce variation between producers, which could lead to improved acceptance. To ensure the efficacy of the product, producers will need to understand the best time to grow and harvest the crop to optimize the concentration of the target metabolite while also quantifying active metabolites in the overall profile with antagonistic or synergistic effects.

References

Abdel-Ghany SE, Hamilton M, Jacobi JL, et al (2016) A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun 7:1–11. doi: 10.1038/ncomms11706

Adwan G, Mhanna M (2009) Synergistic Effects of Plant Extracts and Antibiotics on Staphylococcus aureus Strains Isolated from Clinical Specimens

Aftab T, Naeem M, Khan MMA, et al (2017) Artemisia annua. CRC Press, Boca Raton : Taylor & Francis, 2018. | "A CRC title, part of the Taylor & Francis imprint."

Ahmed S, Stepp JR (2016) Beyond yields: Climate change effects on specialty crop quality and agroecological management. Elem Sci Anthr 4:92. doi: 10.12952/journal.elementa.000092

Ahmed S, Stepp JR, Orians C, et al (2014) Effects of extreme climate events on tea (*Camellia sinensis*) functional quality validate indigenous farmer knowledge and sensory preferences in Tropical China. PLoS One 9:. doi: 10.1371/journal.pone.0109126

Ahmed S, Unachukwu U, Stepp JR, et al (2010) Pu-erh tea tasting in Yunnan , China : Correlation of drinkers ' perceptions to phytochemistry. J Ethnopharmacol 132:176–185. doi: 10.1016/j.jep.2010.08.016

Akula R, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav 6:1720–1731. doi: 10.4161/psb.6.11.17613

Alabadí D, Yanovsky MJ, Más P, et al (2002) Critical Role for CCA1 and LHY in Maintaining Circadian Rhythmicity in Arabidopsis. Curr Biol 12:757–761. doi: 10.1016/S0960-9822(02)00815-1

Aninbon C, Jogloy S, Vorasoot N, et al (2016) Effect of end of season water deficit on phenolic compounds in peanut genotypes with different levels of resistance to drought. Food Chem 196:123–129. doi: 10.1016/j.foodchem.2015.09.022

Ariza MT, Martínez-Ferri E, Domínguez P, et al (2015) Effects of harvest time on functional compounds and fruit antioxidant capacity in ten strawberry cultivars. J Berry Res 5:71–80. doi: 10.3233/JBR-150090

Armin Töpfer (2018) Lima - Demultiplex Barcoded PacBio Samples. (c) Pacific Biosciences of California, Inc. https://github.com/PacificBiosciences/barcoding

Asghari G, Houshfar G, Mahmoudi Z, Asghari M (2014) Diurnal variation of essential of the oil components of pycnocycla spinosa decne. ex boiss. Jundishapur J Nat Pharm Prod 9:35–38. doi: 10.17795/jjnpp-12229

Augustijn D, Roy U, Van Schadewijk R, et al (2016) Metabolic profiling of intact Arabidopsis thaliana leaves during circadian cycle using 1H high resolution magic angle spinning NMR. PLoS One 11:1–17. doi: 10.1371/journal.pone.0163258

Bauer S, Störmer E, Graubaum HJ, Roots I (2001) Determination of hyperforin, hypericin, and pseudohypericin in human plasma using high-performance liquid chromatography analysis with fluorescence and ultraviolet detection. J Chromatogr B Biomed Sci Appl 765:29–35. doi: 10.1016/S0378-4347(01)00390-5

Belesky DP, Malinowski DP (2016) Grassland communities in the USA and expected trends associated with climate change. Acta Agrobot 69:1–25. doi: 10.5586/aa.1673

Bendix C, Marshall CM, Harmon FG (2015) Circadian Clock Genes Universally Control Key Agricultural Traits. Mol Plant 8:1135–1152. doi: 10.1016/j.molp.2015.03.003

Bernarth J, P T (1979) Effect of Environmental-Factors on Growth - Development and Alkaloid Production of Poppy (*Papaver-Somniferum* L) .1. Responses to Day-Length and Light-Intensity. Biochem und Physiol der Pflanz 174:468–478

Bhardwaj V, Meier S, Petersen LN, et al (2011) Defence Responses of Arabidopsis thaliana to Infection by Pseudomonas syringae Are Regulated by the Circadian Clock. PLoS One 6:e26968

Bläsing OE, Gibon Y, Günther M, et al (2005) Sugars and Circadian Regulation Make Major Contributions to the Global Regulation of Diurnal Gene Expression in Arabidopsis. Plant Cell Online 17:3257–3281

Borges CV, Minatel IO, Gomez-Gomez HA, Lima GPP (2017) Medicinal Plants: Influence of Environmental Factors on the Content of Secondary Metabolites. In: Ghorbanpour M, Varma A (eds) Medicinal Plants and Environmental Challenges. Springer International Publishing, Cham, pp 259–277

Borges RM, Bessière JM, Ranganathan Y (2013) Diel Variation in Fig Volatiles Across Syconium Development: Making Sense of Scents. J Chem Ecol 39:630–642. doi: 10.1007/s10886-013-0280-5

Botha LE, Prinsloo G, Deutschländer MS (2018) Variations in the accumulation of three secondary metabolites in *Euclea undulata* Thunb. var. myrtina as a function of seasonal changes. South African J Bot 117:34–40. doi: 10.1016/j.sajb.2018.04.016

Bottomley W, Smith H, Galston AW (1966) Flavonoid complexes in *Pisum sativum*—III. : The effect of light on the synthesis of kaempferol and quercetin complexes. Phytochemistry 5:117–123. doi: https://doi.org/10.1016/S0031-9422(00)85089-X

Buzby J, Farah-Wells H, Hyman J (2014) The Estimated Amount, Value, and Calories of Postharvest Food Losses at the Retail and Consumer Levels in the United States

Caldwell MM, Björn LO, Bornman JF, et al (1998) Effects of increased solar ultraviolet radiation on terrestrial ecosystems. J Photochem Photobiol B Biol 46:40–52. doi: 10.1016/S1011-1344(98)00184-5

Charron CS, Saxton AM, Sams CE (2005) Relationship of climate and genotype to seasonal variation in the glucosinolate – myrosinase system . I . Glucosinolate content in ten cultivars of *Brassica oleracea* grown in fall and spring seasons. 681:671–681. doi: 10.1002/jsfa.1880

Chen L, Cao H, Xiao J (2018) 2 - Polyphenols: Absorption, bioavailability, and metabolomics. In: Galanakis Recovery, and Applications CMBT-PP (ed). Woodhead Publishing, pp 45–67

Cheruiyot E k., Mumera L m., Ng'etich W k., et al (2007) Polyphenols as Potential Indicators for Drought Tolerance in Tea (C*amellia sinensis* L.). Biosci Biotechnol Biochem 71:2190–2197. doi: 10.1271/bbb.70156

Christie PJ, Alfenito MR, Walbot V (1994) Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: Enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. Planta 194:541–549. doi: 10.1007/BF00714468

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci 101:15243–15248. doi: 10.1073/pnas.0406069101

Copolovici L, Kännaste A, Pazouki L, Niinemets Ü (2012) Emissions of green leaf volatiles and terpenoids from *Solanum lycopersicum* are quantitatively related to the severity of cold and heat shock treatments. J Plant Physiol 169:664–672. doi: 10.1016/j.jplph.2011.12.019

Counce PA, Bryant RJ, Bergman CJ, et al (2005) Rice milling quality, grain dimensions, and starch branching as affected by high night temperatures. Cereal Chem 82:645–648. doi: 10.1094/CC-82-0645

Covington MF, Maloof JN, Straume M, et al (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biol 9:. doi: 10.1186/gb-2008-9-8-r130

Cvejić JH, Krstonošić MA, Bursác M, Miljić U (2017) Polyphenols. Elsevier Inc.

Davies PJ (2010) The Plant Hormones: Their Nature, Occurrence, and Functions. In: Davies PJ (ed) Plant Hormones: Biosynthesis, Signal Transduction, Action! Springer Netherlands, Dordrecht, pp 1–15

De Paoli HC, Tuskan GA, Yang X (2016) An innovative platform for quick and flexible joining of assorted DNA fragments. Sci Rep 6:1–14. doi: 10.1038/srep19278

Degu A, Hochberg U, Sikron N, et al (2014) Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. BMC Plant Biol 14:1–20. doi: 10.1186/s12870-014-0188-4

Deharo E, Ginsburg H (2011) Analysis of additivity and synergism in the anti-plasmodial effect of purified compounds from plant extracts. Malar J 10:1–5. doi: 10.1186/1475-2875-10-S1-S5

Desai JS, Sartor RC, Lawas LM, et al (2017) Improving Gene Regulatory Network Inference by Incorporating Rates of Transcriptional Changes. Sci Rep 7:1–12. doi: 10.1038/s41598-017-17143-1

Dodd AN, Toth R, Kevei E, et al (2005) Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. Science 309:630–633. doi: 10.1126/science.1115581

Dong MA, Farré EM, Thomashow MF (2011) CIRCADIAN CLOCK-ASSOCIATED 1 and LATE ELONGATED HYPOCOTYL regulate expression of the C-REPEAT BINDING FACTOR (CBF) pathway in Arabidopsis. Proc Natl Acad Sci 108:7241–7246

Dou H, Niu G, Gu M, Masabni J (2017) Effects of Light Quality on Growth and Phytonutrient Accumulation of Herbs under Controlled Environments. Horticulturae 3:36. doi: 10.3390/horticulturae3020036

Dudareva N, Andersson S, Orlova I, et al (2005) From The Cover: The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc Natl Acad Sci 102:933–938. doi: 10.1073/pnas.0407360102

Easterling DR, Kunkel KE, Arnold JR, et al (2017) Precipitation change in the United States. In: Wuebbles DJ, Fahey DW, Hibbard KA, et al. (eds) Climate Science Special Report: Fourth National Climate Assessment, Volume I. U.S. Global Change Research Program, Washington, DC, USA, pp 207–230. doi: 10.7930/J0H993CC

Edwardson JR, Barry S, Muggah EM, et al (1952) Hops : Their Botany , History , Production and Utilization. Econ Bot 6:804–811. doi: 10.1111/ijfs.13656

El Senousy AS, Farag MA, Al-Mahdy DA, Wessjohann LA (2014) Developmental changes in leaf phenolics composition from three artichoke cvs. (*Cynara scolymus*) as determined via UHPLC-MS and chemometrics. Phytochemistry 108:67–76. doi: 10.1016/j.phytochem.2014.09.004

Elfawal MA, Towler MJ, Reich NG, et al (2012) Dried Whole Plant *Artemisia annua* as an Antimalarial Therapy. PLoS One 7:e52746. doi: 10.1371/journal.pone.0052746

Espinoza C, Degenkolbe T, Caldana C, et al (2010) Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in arabidopsis. PLoS One 5:. doi: 10.1371/journal.pone.0014101

Fairbairn JW, Suwal PN (1961) The Alkaloids of Hemlock (*Conium maculatum* L.) Evidence for a Rapid Turnover of the Major Alkaloids. Phytochemistry 1:38–46

Farré EM, Weise SE (2012) The Interactions between the Circadian Clock and Primary Metabolism. Curr Opin Plant Biol 15:293–300. doi: 10.1016/j.pbi.2012.01.013

Fenske MP, Imaizumi T (2016) Circadian Rhythms in Floral Scent Emission. Front Plant Sci 7:1–6. doi: 10.3389/fpls.2016.00462

Ferreira JFS, Benedito VA, Sandhu D, et al (2018) Seasonal and Differential Sesquiterpene Accumulation in Artemisia annua Suggest Selection Based on Both Artemisinin and Dihydroartemisinic Acid may Increase Artemisinin in planta. Front Plant Sci 9:1–12. doi: 10.3389/fpls.2018.01096

Filichkin SA, Breton G, Priest HD, et al (2011) Global profiling of rice and poplar transcriptomes highlights key conserved Circadian-controlled pathways and cis-regulatory modules. PLoS One 6:. doi: 10.1371/journal.pone.0016907

Flis A, Sulpice R, Seaton DD, et al (2016) Photoperiod-dependent changes in the phase of core clock transcripts and global transcriptional outputs at dawn and dusk in Arabidopsis. Plant Cell Environ 39:1955–1981. doi: 10.1111/pce.12754

Footitt S, Douterelo-Soler I, Clay H, Finch-Savage WE (2011) Dormancy cycling in Arabidopsis seeds is controlled by seasonally distinct hormone-signaling pathways. 108:. doi: 10.1073/pnas.1116325108

Fowler SG, Cook D, Thomashow MF (2005) Low Temperature Induction of Arabidopsis CBF1, 2, and 3 Is Gated by the Circadian Clock. Plant Physiol 137:961–968. doi: 10.1104/pp.104.058354

Fujiuchi N, Matoba N, Matsuda R (2016) Environment Control to Improve Recombinant Protein Yields in Plants Based on Agrobacterium-Mediated Transient Gene Expression. Front Bioeng Biotechnol 4:1–6. doi: 10.3389/fbioe.2016.00023

Fukushima A, Kusano M, Nakamichi N, et al (2009) Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. Proc Natl Acad Sci 106:7251–7256. doi: 10.1073/pnas.0900952106

Ganguli P, Ganguly AR (2016) Space-time trends in U.S. meteorological droughts. J Hydrol Reg Stud 8:235–259. doi: 10.1016/j.ejrh.2016.09.004

Gibon Y, Usadel B, Blaesing OE, et al (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. Genome Biol 7:R76. doi: 10.1186/gb-2006-7-8-r76

Gil KE, Park MJ, Lee HJ, et al (2017) Alternative splicing provides a proactive mechanism for the diurnal CONSTANS dynamics in Arabidopsis photoperiodic flowering. Plant J 89:128–140. doi: 10.1111/tpj.13351

Gitau M (2016) Long-term seasonality of rainfall in the southwest Florida Gulf coastal zone. Clim Res 69:93–105. doi: 10.3354/cr01399

Glaubitz U, Erban A, Kopka J, et al (2015) High night temperature strongly impacts TCA cycle, amino acid and polyamine biosynthetic pathways in rice in a sensitivity-dependent manner. J Exp Bot 66:6385–6397. doi: 10.1093/jxb/erv352

González-Zeas D, Erazo B, Lloret P, et al (2019) Linking global climate change to local water availability: Limitations and prospects for a tropical mountain watershed. Sci Total Environ 650:2577–2586. doi: 10.1016/j.scitotenv.2018.09.309

Goodspeed D, Chehab EW, Min-Venditti A, et al (2012) Arabidopsis synchronizes jasmonate-mediated defense with insect circadian behavior. Proc Natl Acad Sci 109:4674–4677. doi: 10.1073/pnas.1116368109

Goodspeed D, Liu JD, Chehab EW, et al (2013) Postharvest Circadian Entrainment Enhances Crop Pest Resistance and Phytochemical Cycling. Curr Biol 23:1235–1241. doi: 10.1016/J.CUB.2013.05.034

Gordon, S. P. et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. PLoS ONE 10, e0132628 (2015).

Gouinguene SP, Turlings TCJ (2002) The Effects of Abiotic Factors on Induced Volatile Emissions in Corn Plants. Plant Physiol 129:1296–1307. doi: DOI: 10.1104/pp.001941

Graf A, Schlereth A, Stitt M, Smith AM (2010) Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. Proc Natl Acad Sci 107:9458–9463. doi: 10.1073/pnas.0914299107

Green RM, Tingay S, Wang Z-Y, Tobin EM (2002) Circadian Rhythms Confer a Higher Level of Fitness to Arabidopsis Plants. Plant Physiol 129:576–584. doi: 10.1104/pp.004374

Greenham K, Guadagno CR, Gehan MA, et al (2017) Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in *Brassica rapa*. 1–26

Greenway H, Hughes PG, Klepper B (1969) Effects of Water Deficit on Phosphorus Nutrition of Tomato Plants. Physiol Plant 22:199–207. doi: 10.1111/j.1399-3054.1969.tb07856.x

Grinevich DO, Desai JS, Stroup KP, et al (2019) Novel transcriptional responses to heat revealed by turning up the heat at night. Plant Mol Biol 1–19. doi: 10.1007/s11103-019-00873-3

Grüning, Björn, Ryan Dale, Andreas Sjödin, Brad A. Chapman, Jillian Rowe, Christopher H. Tomkins-Tinch, Renan Valieris, the Bioconda Team, and Johannes Köster. 2018. “Bioconda: Sustainable and Comprehensive Software Distribution for the Life Sciences”. Nature Methods, 2018 doi:10.1038/s41592-018-0046-7.

Guo R, Li W, Wang X, et al (2019) Effect of photoperiod on the formation of cherry radish root. Sci Hortic (Amsterdam) 244:193–199. doi: 10.1016/j.scienta.2018.09.044

Gyllenstrand N, Karlgren A, Clapham D, et al (2014) No time for spruce: Rapid dampening of circadian rhythms in *Picea abies* (L. Karst). Plant Cell Physiol 55:535–550. doi: 10.1093/pcp/pct199

Han X, Yu H, Yuan R, et al (2019) Arabidopsis Transcription Factor TCP5 Controls Plant Thermomorphogenesis by Positively Regulating PIF4 Activity. iScience 15:611–622. doi: 10.1016/j.isci.2019.04.005

Hannah MA, Caldana C, Steinhauser D, et al (2010) Combined Transcript and Metabolite Profiling of Arabidopsis Grown under Widely Variant Growth Conditions Facilitates the Identification of Novel Metabolite-Mediated Regulation of Gene Expression. Plant Physiol 152:2120–2129. doi: 10.1104/pp.109.147306

Harmer S (2010) Plant Biology in the Fourth Dimension. Plant Physiol 154:467–470. doi: 10.1104/pp.110.161448

Hasperué JH, Chaves AR, Martínez GA (2011) End of day harvest delays postharvest senescence of broccoli florets. Postharvest Biol Technol 59:64–70. doi: 10.1016/j.postharvbio.2010.08.005

Heberle, H.; Meirelles, G. V.; da Silva, F. R.; Telles, G. P.; Minghim, R. (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 16:169. doi: 10.1186/s12859-015-0611-3

Helmig D, Ortega J, Duhl T, et al (2007) Sesquiterpene emissions from pine trees - Identifications, emission rates and flux estimates for the contiguous United States. Environ Sci Technol 41:1545–1553. doi: 10.1021/es0618907

Hernández I, Alegre L, Munné-Bosch S (2004) Drought-induced changes in flavonoids and other low molecular weight antioxidants in Cistus clusii grown under Mediterranean field conditions. Tree Physiol 24:1303–1311

Herranz-López M, Losada-Echeberría M, Barrajón-Catalán E (2018) The Multitarget Activity of Natural Extracts on Cancer: Synergy and Xenohormesis. Medicines 6:6. doi: 10.3390/medicines6010006

Hevia MA, Canessa P, Müller-Esparza H, Larrondo LF (2015) A circadian oscillator in the fungus Botrytis cinerea regulates virulence when infecting Arabidopsis thaliana. Proc Natl Acad Sci 112:8744–8749. doi: 10.1073/pnas.1508432112

Hill, S.T., Sudarsanam,R., Henning,J., et al. (2017) HopBase: a unified resource for Humulus genomics. Database Vol. 2017: article ID bax009; doi:10.1093/database/bax00

Hirai MY, Klein M, Yano M, et al (2005) Elucidation of Gene-to-Gene and Metabolite-to-Gene Networks in Arabidopsis by Integration of Metabolomics Supplemental material : Elucidation of Gene-to-Gene and Metabolite-to-Gene Networks in. doi: 10.1074/jbc.M502332200

Hirai MY, Yano M, Goodenowe DB, et al (2006) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. doi: 10.1073/pnas.0403218101

Hochberg U, Degu A, Toubiana D, et al (2013) Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. BMC Plant Biol 13:. doi: 10.1186/1471-2229-13-184

Holopainen JK, Virjamo V, Ghimire RP, et al (2018) Climate Change Effects on Secondary Compounds of Forest Trees in the Northern Hemisphere. Front Plant Sci 9:1–10. doi: 10.3389/fpls.2018.01445

Horak E, Farré EM (2015) The regulation of UV-B responses by the circadian clock. Plant Signal Behav 10:1–4. doi: 10.1080/15592324.2014.1000164

Huang X, Yao J, Zhao Y, et al (2016) Efficient Rutin and Quercetin Biosynthesis through Flavonoids-Related Gene Expression in Fagopyrum tataricum Gaertn. Hairy Root Cultures with UV-B Irradiation. Front Plant Sci 7:1–11. doi: 10.3389/fpls.2016.00063

Huang ZA, Zhao T, Fan HJ, et al (2012) The Upregulation of NtAN2 Expression at Low Temperature is Required for Anthocyanin Accumulation in Juvenile Leaves of Lc-transgenic Tobacco (*Nicotiana tabacum* L.). J Genet Genomics 39:149–156. doi: 10.1016/j.jgg.2012.01.007

Hura T, Hura K, Grzesiak S (2008) Contents of Total Phenolics and Ferulic Acid, and PAL Activity during Water Potential Changes in Leaves of Maize Single-Cross Hybrids of Different Drought Tolerance. J Agron Crop Sci 194:104–112. doi: 10.1111/j.1439-037X.2008.00297.x

Hwang H, Cho M-H, Hahn B-S, et al (2011) Proteomic identification of rhythmic proteins in rice seedlings. Biochim Biophys Acta - Proteins Proteomics 1814:470–479

Hykkerud AL, Uleberg E, Hansen E, et al (2018) Seasonal and yearly variation of total polyphenols , total anthocyanins and ellagic acid in different clones of cloudberry (*Rubus chamaemorus* L .). J Appl Bot Food Qual 91:96–102. doi: 10.5073/JABFQ.2018.091.013

Ibrahim MA, Mäenpää M, Hassinen V, et al (2010) Elevation of night-time temperature increases terpenoid emissions from *Betula pendula* and *Populus tremula*. J Exp Bot 61:1583–1595. doi: 10.1093/jxb/erq034

Ierodiakonou K (2016) Theophrastus. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, Summer 201. Metaphysics Research Lab, Stanford University

Impa SM, Sunoj VSJ, Krassovskaya I, et al (2018) Carbon balance and source-sink metabolic changes in winter wheat exposed to high night-time temperature. Plant Cell Environ 1–14. doi: 10.1111/pce.13488

Ingle RA, Stoker C, Stone W, et al (2015) Jasmonate signalling drives time-of-day differences in susceptibility of Arabidopsis to the fungal pathogen *Botrytis cinerea.* Plant J 84:937–948

IPCC, 2018: Summary for Policymakers. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock,

S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.

Itenov K, Mølgaard P, Nyman U (1999) Diurnal fluctuations of the alkaloid concentration in latex of poppy *Papaver somniferum* is due to day-night fluctuations of the latex water content. Phytochemistry 52:1229–1234. doi: 10.1016/S0031-9422(99)00420-3

Jaafar HZE, Ibrahim MH, Fakri NFM (2012) Impact of soil field water capacity on secondary metabolites, phenylalanine ammonia-lyase (PAL), maliondialdehyde (MDA) and photosynthetic responses of Malaysian Kacip Fatimah (*Labisia pumila* Benth). Molecules 17:7305–7322. doi: 10.3390/molecules17067305

Jaleel CA, Sankar B, Murali P V, et al (2008) Water deficit stress effects on reactive oxygen metabolism in *Catharanthus roseus*; impacts on ajmalicine accumulation. Colloids Surfaces B Biointerfaces 62:105–111. doi: 10.1016/j.colsurfb.2007.09.026

Kanno K, Mae T, Makino A (2009) High night temperature stimulates photosynthesis, biomass production and growth during the vegetative stage of rice plants. Soil Sci Plant Nutr 55:124–131. doi: 10.1111/j.1747-0765.2008.00343.x

Kaplan F, Kopka J, Sung DY, et al (2007) Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. Plant J 50:967–981

Keggenhoff I, Elizbarashvili M, Amiri-Farahani A, King L (2014) Trends in daily temperature and precipitation extremes over Georgia, 1971-2010. Weather Clim Extrem 4:75–85. doi: 10.1016/j.wace.2014.05.001

Kim SG, Yon F, Gaquerel E, et al (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native Tobacco, *Nicotiana attenuata*. PLoS One 6:. doi: 10.1371/journal.pone.0026214

Klepper B (1968) Diurnal Pattern of Water Potential in Woody Plants. Plant Physiol 43:1931--. doi: 10.1104/pp.43.12.1931

Koda S, Onda Y, Matsui H, et al (2017) Diurnal Transcriptome and Gene Network Represented through Sparse Modeling in Brachypodium distachyon. Front Plant Sci 8:1–11. doi: 10.3389/fpls.2017.02055

Kumar KN, Molini A, Ouarda TBMJ, Rajeevan MN (2017) North Atlantic controls on wintertime warm extremes and aridification trends in the Middle East. Sci Rep 7:1–11. doi: 10.1038/s41598-017-12430-3

Lai AG, Doherty CJ, Mueller-Roeber B, et al (2012) CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. Proc Natl Acad Sci 109:17129–17134. doi: 10.1073/pnas.1209148109

Lavoir A V., Staudt M, Schnitzler JP, et al (2009) Drought reduced monoterpene emissions from the evergreen Mediterranean oak quercus ilex: Results from a throughfall displacement experiment. Biogeosciences 6:1167–1180. doi: 10.5194/bg-6-1167-2009

Laza MRC, Sakai H, Cheng W, et al (2015a) Differential response of rice plants to high night temperatures imposed at varying developmental phases. Agric For Meteorol 209–210:69–77. doi: 10.1016/j.agrformet.2015.04.029

Laza MRC, Sakai H, Cheng W, et al (2015b) Differential response of rice plants to high night temperatures imposed at varying developmental phases. Agric For Meteorol 209–210:69–77. doi: 10.1016/j.agrformet.2015.04.029

Lee H, Oh IN, Kim J, et al (2018) Phenolic compound profiles and their seasonal variations in new red-phenotype head-forming Chinese cabbages. LWT - Food Sci Technol 90:433–439. doi: 10.1016/j.lwt.2017.12.056

Lester, G. E., Makus, D. J., Hodges, D. M., & Jifon, J. L. (2013). Summer (subarctic) versus winter (subtropic) production affects spinach (Spinacia oleracea L.) leaf bionutrients: vitamins (C, E, Folate, K1, provitamin A), lutein, phenolics, and antioxidants. Journal of agricultural and food chemistry, 61(29), 7019-7027. doi: 10.1021/jf401461z

Leyva A, Jarillo JA, Salinas J, Martinez-Zapater JM (1995) Low Temperature Induces the Accumulation of Phenylalanine Ammonia-Lyase and Chalcone Synthase mRNAs of *Arabidopsis thaliana* in a Light-Dependent Manner. Plant Physiol 108:39–46. doi: 10.1104/pp.108.1.39

Li H, Chen Z, Hu M, et al (2011) Different effects of night versus day high temperature on rice quality and accumulation profiling of rice grain proteins during grain filling. Plant Cell Rep 30:1641–1659. doi: 10.1007/s00299-011-1074-2

Li H, Lin Y, Chen X, et al (2018) Effects of blue light on flavonoid accumulation linked to the expression of miR393, miR394 and miR395 in longan embryogenic calli. PLoS One 13:1–22. doi: 10.1371/journal.pone.0191444

Li Y, Kim JI, Pysh L, Chapple C (2015) Four isoforms of arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol 169:2409–2421. doi: 10.1104/pp.15.00838

Liao K-L, Jones RD, McCarter P, et al (2017) A Shadow Detector for Photosynthesis Efficiency. J Theor Biol 414:231–244. doi: doi:10.1016/j.jtbi.2016.11.027

Liebelt, D.J., Jordan, J.T. & Doherty, C.J. Only a matter of time: the impact of daily and seasonal rhythms on phytochemicals. Phytochem Rev 18, 1409–1433 (2019). doi: 10.1007/s11101-019-09617-z

Liu Y, Fang S, Yang W, et al (2018) Light quality affects flavonoid production and related gene expression in *Cyclocarya paliurus*. J Photochem Photobiol B Biol 179:66–73. doi: 10.1016/j.jphotobiol.2018.01.002

Lobell DB, Ortiz-Monasterio JI (2007) Impacts of day versus night temperatures on spring wheat yields: A comparison of empirical and CERES model predictions in three locations. Agron J 99:469–477. doi: 10.2134/agronj2006.0209

Lombardo S, Pandino G, Mauromicale G, et al (2010) Influence of genotype, harvest time and plant part on polyphenolic composition of globe artichoke [*Cynara cardunculus* L. var. scolymus (L.) Fiori]. Food Chem 119:1175–1181. doi: 10.1016/j.foodchem.2009.08.033

Lu S, Xu R, Jia J, et al (2002) Cloning and functional characterization of a β-pinene synthase from Artemisia annua that shows a circadian pattern of expression 1. Plant Physiol 130:477–486. doi: 10.1104/pp.006544.the

Ma D, Sun D, Wang C, et al (2014) Expression of flavonoid biosynthesis genes and accumulation of flavonoid in wheat leaves in response to drought stress. Plant Physiol Biochem 80:60–66. doi: 10.1016/j.plaphy.2014.03.024

Madar A, Greenfield A, Ostrer H, et al (2009) The inferelator 2.0: A scalable framework for reconstruction of dynamic regulatory network models. Proc 31st Annu Int Conf IEEE Eng Med Biol Soc Eng Futur Biomed EMBC 2009 5448–5451. doi: 10.1109/IEMBS.2009.5334018

Maffei M, Scannerini S (1999) Photomorphogenic and chemical responses to blue light in *Mentha piperia.* J Essent Oil Res 11:730–738. doi: 10.1080/10412905.1999.9712007

Mallakpour I, Villarini G (2017) Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. Theor Appl Climatol 130:345–363. doi: 10.1007/s00704-016-1881-z

Manivannan A, Soundararajan P, Halimah N, et al (2015) Blue LED light enhances growth, phytochemical contents, and antioxidant enzyme activities of *Rehmannia glutinosa* cultured in vitro. Hortic Environ Biotechnol 56:105–113. doi: 10.1007/s13580-015-0114-1

Mazur SP, Nes A, Wold AB, et al (2014) Quality and chemical composition of ten red raspberry (*Rubus idaeus* L.) genotypes during three harvest seasons. Food Chem 160:233–240. doi: 10.1016/j.foodchem.2014.02.174

McClung CR (2006) Plant Circadian Rhythms. Plant Cell 18:792–803. doi: 10.1105/tpc.106.040980

McClung CR (2008) Comes a time. Curr Opin Plant Biol 11:514–520. doi: 10.1016/j.pbi.2008.06.010

McClung CR (2014) Wheels within wheels: new transcriptional feedback loops in the Arabidopsis circadian clock. F1000Prime Rep 6:1–6. doi: 10.12703/P6-2

Medina, M. B. (2011). Simple and rapid method for the analysis of phenolic compounds in beverages and grains. Journal of Agricultural and Food Chemistry, 59(5), 1565-1571. doi: 10.1021/jf103711c

Michael TP, Mockler TC, Breton G, et al (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. PLoS Genet 4:. doi: 10.1371/journal.pgen.0040014

Mithöfer A, Boland W (2012) Plant Defense Against Herbivores: Chemical Aspects. Annu Rev Plant Biol 63:431–450. doi: 10.1146/annurev-arplant-042110-103854

Mohammed AR, Tarpley L (2009) High nighttime temperatures affect rice productivity through altered pollen germination and spikelet fertility. Agric For Meteorol 149:999–1008. doi: 10.1016/j.agrformet.2008.12.003

Mølmann JAB, Steindal ALH, Bengtsson GB, et al (2015) Effects of temperature and photoperiod on sensory quality and contents of glucosinolates , flavonols and vitamin C in broccoli florets. Food Chem 172:47–55. doi: 10.1016/j.foodchem.2014.09.015

Mouradov A, Spangenberg G (2014) Flavonoids: a metabolic network mediating plants adaptation to their real estate. Front Plant Sci 5:1–16. doi: 10.3389/fpls.2014.00620

Mwimba M, Karapetyan S, Liu L, et al (2018) Daily humidity oscillation regulates the circadian clock to influence plant physiology. Nat Commun 9:1–10. doi: 10.1038/s41467-018-06692-2

Naeem M, Aftab T, Masroor M, Khan A (2017) Strategies for Enhancing Artemisinin Production in *Artemisia annua* Under Changing Environment. In: Ghorbanpour M, Varma A (eds) Medicinal Plants and Environmental Challenges,. Springer International Publishing, pp 227–246

Ncube B, Finnie JF, Van Staden J (2012) Quality from the field: The impact of environmental factors as quality determinants in medicinal plants. South African J Bot 82:11–20. doi: 10.1016/j.sajb.2012.05.009

Nishioka N, Nishimura T, Ohyama K, et al (2008) Light Quality Affected Growth and Contents of Essential Oil Components of Japanese Mint Plants. Acta Horticulturae. 797:431-436

Niu J, Zhang G, Zhang W, et al (2017) Anthocyanin concentration depends on the counterbalance between its synthesis and degradation in plum fruit at high temperature. Sci Rep 7:1–16. doi: 10.1038/s41598-017-07896-0

Nozue K, Maloof JN (2006) Diurnal regulation of plant growth. Plant, Cell Environ 29:396–408. doi: 10.1111/j.1365-3040.2005.01489.x

Pal I, Anderson BT, Salvucci GD, Gianotti DJ (2013) Shifting seasonality and increasing frequency of precipitation in wet and dry seasons across the U.S. Geophys Res Lett 40:4030–4035. doi: 10.1002/grl.50760

Papadopoulos AP, Hao X (2000) Effects of day and night air temperature on growth, productivity and energy use of long English cucumber. Can J Plant Sci 80:143–150. doi: 10.4141/P99-021

Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate Change. Annu Rev Ecol Evol Syst 37:637–669. doi: 10.1146/annurev.ecolsys.37.091305.110100

Pavarini DP, Pavarini SP, Niehues M, Lopes NP (2012) Exogenous influences on plant secondary metabolite levels. Anim Feed Sci Technol 176:5–16. doi: 10.1016/j.anifeedsci.2012.07.002

Peng S, Huang J, Sheehy JE, et al (2004) Rice yields decline with higher night temperature from global warming. PNAS 101:9971–9975. doi: https://doi.org/10.1073/pnas.0403720101

Pérez-Schindler J, Kanhere A, Edwards L, et al (2017) Exercise and high-fat feeding remodel transcript-metabolite interactive networks in mouse skeletal muscle. Sci Rep 7:1–11. doi: 10.1038/s41598-017-14081-w

Piazzolla F, Pati S, Amodio ML, Colelli G (2016) Effect of harvest time on table grape quality during on-vine storage. J Sci Food Agric 96:131–139. doi: 10.1002/jsfa.7072

Pincemail J, Kevers C, Tabart J, et al (2012) Cultivars, Culture Conditions, and Harvest Time Influence Phenolic and Ascorbic Acid Contents and Antioxidant Capacity of Strawberry (*Fragaria x ananassa*). J Food Sci 77(2) C205-10. doi:10.1111/j.1750-3841.2011.02539.x

Plaut Z, Reinhold L (1965) Effect of water stress on 14C sucrose transport in bean plant. Aust J Biol Sci 18:1143–1155

Porqueddu C, Ates S, Louhaichi M, et al (2016) Grasslands in “Old World” and “New World” Mediterranean-climate zones: Past trends, current status and future research priorities. Grass Forage Sci 71:1–35. doi: 10.1111/gfs.12212

Pulice G, Pelaz S, Matías-Hernández L (2016) Molecular Farming in Artemisia annua, a Promising Approach to Improve Anti-malarial Drug Production. Front Plant Sci 7:329. doi: 10.3389/fpls.2016.00329

Ragusa L, Picchi V, Tribulato A, et al (2017) The effect of the germination temperature on the phytochemical content of broccoli and rocket sprouts. Int J Food Sci Nutr 68:411–420. doi: 10.1080/09637486.2016.1248907

Rahimpour V, Zeng Y, Mannaerts CM, Su Z (Bob) (2016) Attributing seasonal variation of daily extreme precipitation events across The Netherlands. Weather Clim Extrem 14:56–66. doi: 10.1016/j.wace.2016.11.003

Rai A, Yamazaki M, Saito K (2019) A new era in plant functional genomics. doi: 10.1016/j.coisb.2019.03.005

Rai R, Meena RP, Smita SS, et al (2011) UV-B and UV-C pre-treatments induce physiological changes and artemisinin biosynthesis in *Artemisia annua* L. - An antimalarial plant. J Photochem Photobiol B Biol 105:216–225. doi: 10.1016/j.jphotobiol.2011.09.004

Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav 6:1720–1731. doi: 10.4161/psb.6.11.17613

Rasoanaivo P, Wright CW, Willcox ML, Gilbert B (2011) Whole plant extracts versus single compounds for the treatment of malaria: synergy and positive interactions. Malar J 10:S4. doi: 10.1186/1475-2875-10-S1-S4

Reshef N, Fait A, Agam N Grape berry position affects the diurnal dynamics of its metabolic profile. Plant, Cell Environ. doi: 10.1111/pce.13522

Rhoads A, Au KF (2015) PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma. 13:278–289

Ribeiro AF, Andrade EHA, Salimena FRG, Maia JGS (2014) Circadian and seasonal study of the cinnamate chemotype from *Lippia origanoides* Kunth. Biochem Syst Ecol 55:249–259. doi: 10.1016/j.bse.2014.03.014

Richards JH, Caldwell MM (1987) Hydraulic lift: substantial nocturnal water transport between soil layers by Artemisia tridentata roots. Oecologia 73:486–489

Robinson T (1974) Metabolism and Function of Alkaloids in Plants. 184:430–435

Roque-Malo S, Kumar P (2017) Patterns of change in high frequency precipitation variability over North America. Sci Rep 7:10853. doi: 10.1038/s41598-017-10827-8

Rowan DD, Cao M, Lin-Wang K, et al (2009) Environmental regulation of leaf colour in red 35S:PAP1 Arabidopsis thaliana. New Phytol 182:102–115. doi: 10.1111/j.1469-8137.2008.02737.x

Rufty TW, MacKown CT, Volk RJ (1989) Effects of Altered Carbohydrate Availability on Whole-Plant Assimilation of 15NO3-. Plant Physiol 89:457–463. doi: 10.1104/pp.89.2.457

Ruts T, Matsubara S, Wiese-Klinkenberg A, Walter A (2012) Diel patterns of leaf and root growth: Endogenous rhythmicity or environmental response? J Exp Bot 63:3339–3351. doi: 10.1093/jxb/err334

Sabzalian MR, Heydarizadeh P, Zahedi M, et al (2014) High performance of vegetables, flowers, and medicinal plants in a red-blue LED incubator for indoor plant production. Agron Sustain Dev 34:879–886. doi: 10.1007/s13593-014-0209-6

Sanchez-Castillo M, Blanco D, Tienda-Luna IM, et al (2018) A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. Bioinformatics 34:964–970. doi: 10.1093/bioinformatics/btx605

Sánchez-Rodríguez E, Ruiz JM, Ferreres F, Moreno DA (2011) Phenolic Metabolism in Grafted versus Nongrafted Cherry Tomatoes under the Influence of Water Stress. J Agric Food Chem 59:8839–8846. doi: 10.1021/jf201754t

Sarker U, Oba S (2018) Drought stress enhances nutritional and bioactive compounds, phenolic acids and antioxidant capacity of Amaranthus leafy vegetable. BMC Plant Biol 18:. doi: 10.1186/s12870-018-1484-1

Scheel GL, Daiane E, Rakocevic M, et al (2016) Environmental stress evaluation of Coffea arabica L. leaves from spectrophotometric fingerprints by PCA and OSC-PLS-DA. Arab J Chem 7. doi: 10.1016/j.arabjc.2016.05.014

Selmar D, Kleinwächter M (2013) Influencing the product quality by deliberately applying drought stress during the cultivation of medicinal plants. Ind Crops Prod 42:558–566. doi: 10.1016/j.indcrop.2012.06.020

Shah AR, Khan TM, Sadaqat HA, Chatha AA (2011) Alterations in leaf pigments in cotton (*Gossypium hirsutum*) genotypes subjected to drought stress conditions. Int J Agric Biol 13:902–908

Shallan MA, Hassan HMM, Namich AAM, Ibrahim AA (2012) Effect of Sodium Niroprusside, Putrescine and Glycine Betaine on Alleviation of Drought Stress in Cotton Plant. J Agric Environ Sci 12:1252–1265. doi:10.5829/idosi.aejaes.2012.12.09.1902

Shamloo M, Babawale EA, Furtado A, et al (2017) Effects of genotype and temperature on accumulation of plant secondary metabolites in Canadian and Australian wheat grown under controlled environments. Sci Rep 7:1–13. doi: 10.1038/s41598-017-09681-5

Sharam GJ, Turkington R (2005) Diurnal cycle of sparteine production in *Lupinus arcticus*. Can J Bot 83:. doi: 10.1139/b05-104

Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: Why and how. Ann Bot 101:5–18. doi: 10.1093/aob/mcm240

Shi W, Yin X, Struik PC, et al (2017) High day- and night-time temperatures affect grain growth dynamics in contrasting rice genotypes. J Exp Bot 68:5233–5245. doi: 10.1093/jxb/erx344

Shojaie B, Mostajeran A, Ghannadian M (2016) Flavonoid dynamic responses to different drought conditions: Amount, type, and localization of flavonols in roots and shoots of Arabidopsis thaliana L. Turkish J Biol 40:612–622. doi: 10.3906/biy-1505-2

Singh-Sangwan N, Farooqi AHA, Sangwan RS (1994) Effect of Drought Stress on Growth and Essential Oil Metabolism in Lemongrasses. New Phytol 128:173–179

Singh M, Mas P (2018) A Functional Connection between the Circadian Clock and Hormonal Timing in Arabidopsis. Genes (Basel) 9:567. doi: 10.3390/genes9120567

Snyder KA, Tartowski SL (2006) Multi-scale temporal variation in water availability: Implications for vegetation dynamics in arid and semi-arid ecosystems. J Arid Environ 65:219–234. doi: 10.1016/j.jaridenv.2005.06.023

Soengas P, Cartea ME, Velasco P, Francisco M (2018) Endogenous Circadian Rhythms in Polyphenolic Composition Induce Changes in Antioxidant Properties in Brassica Cultivars. J Agric Food Chem 66:5984–5991. doi: 10.1021/acs.jafc.8b01732

Soni U, Brar S, Gauttam VK (2015) Effect of Seasonal Variation on Secondary Metabolites of Medicinal Plants. Int J Pharm Sci Res 6:3654–3662. doi: 10.13040/IJPSR.0975-8232.6(9).3654-62

Steindal ALH, Rdven R, Hansen E, Mlmann J (2015) Effects of photoperiod, growth temperature and cold acclimatisation on glucosinolates, sugars and fatty acids in kale. Food Chem 174:44–51. doi: 10.1016/j.foodchem.2014.10.129

Stitt M, Gibon Y, Lunn JE, Piques M (2007) Multilevel genomics analysis of carbon signalling during low carbon availability: Coordinating the supply and utilisation of carbon in a fluctuating environment. Funct Plant Biol 34:526–549. doi: 10.1071/FP06249

Takeuchi T, Newton L, Burkhardt A, et al (2014) Light and the circadian clock mediate time-specific changes in sensitivity to UV-B stress under light/dark cycles. J Exp Bot 65:6003–6012. doi: 10.1093/jxb/eru339

Tarvainen V, Hakola H, Hellén H, et al (2005) Temperature and light dependence of the VOC emissions of Scots pine. Atmos Chem Phys 5:989–998. doi: 10.5194/acp-5-989-2005

Taulavuori K, Hyöky V, Oksanen J, et al (2016) Species-specific differences in synthesis of flavonoids and phenolic acids under increasing periods of enhanced blue light. Environ Exp Bot 121:145–150. doi: 10.1016/j.envexpbot.2015.04.002

Taxak AK, Murumkar AR, Arya DS (2014) Long term spatial and temporal rainfall trends and homogeneity analysis in Wainganga basin, Central India. Weather Clim Extrem 4:50–61. doi: 10.1016/j.wace.2014.04.005

Tegelberg R, Julkunen-Tiitto R, Aphalo PJ (2004) Red:Far-red light ratio and UV-B radiation: Their effects on leaf phenolics and growth of silver birch seedlings. Plant, Cell Environ 27:1005–1013. doi: 10.1111/j.1365-3040.2004.01205.x

Towler E, Pal I (2018) Characterizing and understanding the variability of streamflow drought indicators within the USA AU - Pournasiri Poshtiri, Maryam. Hydrol Sci J 63:1791–1803. doi: 10.1080/02626667.2018.1534240

Tye MR, Blenkinsop S, Fowler HJ, et al (2016) Simulating multimodal seasonality in extreme daily precipitation occurrence. J Hydrol 537:117–129. doi: 10.1016/j.jhydrol.2016.03.038

Ubi BE, Honda C, Bessho H, et al (2006) Expression analysis of anthocyanin biosynthetic genes in apple skin: Effect of UV-B and temperature. Plant Sci 170:571–578. doi: 10.1016/j.plantsci.2005.10.009

Unal YS, Deniz A, Toros H, Incecik S (2012) Temporal and spatial patterns of precipitation variability for annual, wet, and dry seasons in Turkey. Int J Climatol 32:392–405. doi: 10.1002/joc.2274

Uriu DM, Godoy BS de A, Battirola LD, et al (2018) Temporal Variation Of The Total Phenolic Compounds Concentration In Vochysia Divergens Pohl. (Vochysiaceae) Leaves In The Brazilian PantanaL. Rev Árvore 41:. doi: 10.1590/1806-90882017000300016

Vallejo F, Toma FA (2003) Total and individual glucosinolate contents in inflorescences of eight broccoli cultivars grown under various climatic and fertilisation. J. Sci. Food Agric., 83: 307-313. doi.org/10.1002/jsfa.1320

Veit M, Bilger W, Mühlbauer T, et al (1996) Diurnal changes in flavonoids. J Plant Physiol 148:478–482. doi: 10.1016/S0176-1617(96)80282-3

Verma N, Shukla S (2015) Impact of various factors responsible for fluctuation in plant secondary metabolites. J Appl Res Med Aromat Planta 2:105–113. doi:10.1016/j.jarmap.2015.09.002

Vose RS, Easterling DR, Gleason B (2004) Maximum and minimum temperature trends for the Globe: An update through 2004. Geophys Res Lett 32:. doi:org/10.1029/2005GL024379

Waddell EM (2018) Framing Stress in the Context of Time of Day for Establishing Metabolomic and Transcriptomic Networks to Advance the Study of Specialized Metabolites in Humulus lupulus.

Wang GY, Shi JL, Ng G, et al (2011) Circadian clock-regulated phosphate transporter PHT4;1 Plays an important role in Arabidopsis defense. Mol Plant 4:516–526. doi: 10.1093/mp/ssr016

Weiss J, Terry MI, Martos-Fuentes M, et al (2018) Diel pattern of circadian clock and storage protein gene expression in leaves and during seed filling in cowpea (*Vigna unguiculata*). BMC Plant Biol 18:1–20. doi: 10.1186/s12870-018-1244-2

Welch JR, Vincent JR, Auffhammer M, et al (2010) Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. PNAS 107:14562–14567. doi: https://doi.org/10.1073/pnas.1001222107

Wink M, Witte L (1984) Turnover and transport of quinolizidine alkaloids. Diurnal fluctuations of lupanine in the phloem sap, leaves and fruits of *Lupinus albus* L. Planta 161:519–524. doi: 10.1016/j.lfs.2007.10.023

World Health Organization (2015) Guidelines for the Treatment of Malaria. Third Edition. World Health Organization

Xia J, Chen J, Piao S, et al (2014) Terrestrial carbon cycle affected by non-uniform climate warming. Nat Geosci 7:173–180. doi: 10.1038/ngeo2093

Xiao, Z., Rausch, S. R., Luo, Y., Sun, J., Yu, L., Wang, Q., ... & Stommel, J. R. (2019). Microgreens of Brassicaceae: Genetic diversity of phytochemical concentrations and antioxidant capacity. LWT, 101, 731-737. doi: 10.1016/j.lwt.2018.10.076

Xu C, Zhang Y, Zhu L, et al (2011) Influence of growing season on phenolic compounds and antioxidant properties of grape berries from vines grown in subtropical climate. J Agric Food Chem 59:1078–1086. doi: 10.1021/jf104157z

Yang L, Wen KS, Ruan X, et al (2018) Response of plant secondary metabolites to environmental factors. Molecules 23:1–26. doi: 10.3390/molecules23040762

Yang D, Seaton D, Krahmer et al (2016) Photoreceptor effects on plant biomass, resource allocation, and metabolic state. PNAS 113:767-762. doi: 10.1073/pnas.1601309113

Yang Y, He F, Yu L, et al (2007) Influence of Drought on Oxidative Stress and Flavonoid Production in Cell Suspension Culture of Glycyrrhiza in flata Batal. Yunnan Zhiwu Yanjiu 410–416. doi: https://doi.org/10.1515/znc-2007-5-615

Yang Z, Li Y, Li P, et al (2016) Effect of difference between day and night temperature on tomato (Lycopersicon esculentum Mill.) root activity and low molecular weight organic acid secretion. Soil Sci Plant Nutr 62:423–431. doi: 10.1080/00380768.2016.1224449

Yang Z, Wang X, Peng X, et al (2014) Effect of difference between day and night temperature on nutrients and dry mass partitioning of tomato in climate chamber

Yerushalmi S, Yakir E, Green RM (2011) Circadian clocks and adaptation in Arabidopsis. Mol Ecol 20:1155–1165. doi: 10.1111/j.1365-294X.2010.04962.x

Yin H, Guo HB, Weston DJ, et al (2018) Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC Genomics 19:1–16. doi: 10.1186/s12864-018-4964-7

Yoshida K, Oyama-okubo N (2018) An R2R3-MYB transcription factor ODORANT1 regulates fragrance biosynthesis in lilies (*Lilium* spp.). Mol Breed 28:1–14. doi: 10.1007/s11032-018-0902-2

Yuzuak, S., Ballington, J., & Xie, D. Y. (2018). HPLC-qTOF-MS/MS-Based Profiling of Flavan-3-ols and Dimeric Proanthocyanidins in Berries of Two Muscadine Grape Hybrids FLH 13-11 and FLH 17-66. Metabolites, 8(4), 57.

Zeng L, Wang X, Kang M, et al (2017) Regulation of the rhythmic emission of plant volatiles by the circadian clock. Int J Mol Sci 18:1–10. doi: 10.3390/ijms18112408

Ziska LH, Manalo PA (1996) Increasing night temperature can reduce seed set and potential yield of tropical rice. Aust J Plant Physiol 23:791–794. doi: 10.1071/PP9960791

Zobayed SMA, Afreen F, Kozai T (2005) Temperature stress can alter the photosynthetic efficiency and secondary metabolite concentrations in St. John's wort. Plant Physiol Biochem 43:977–984. doi: 10.1016/j.plaphy.2005.07.013

Zobayed SMA, Afreen F, Kozai T (2007) Phytochemical and physiological changes in the leaves of St. John's wort plants under a water stress condition. Environ Exp Bot 59:109–116. doi: 10.1016/j.envexpbot.2005.10.002

APPENDIX

Script Source Code

Unique and Unmapped Counts

The following source code can analyze six isoform samples using “read_stat.txt” file outputs from the count abundance step of cDNA_cupcake.

```
library(ggplot2)

#input
wd2 <- readline(prompt = "path to SAMPLE directory: ")
name1 <- readline(prompt = "SAMPLE 1 prefix: ")
name2 <- readline(prompt = "SAMPLE 2 prefix: ")
name3 <- readline(prompt = "SAMPLE 3 prefix: ")
name4 <- readline(prompt = "SAMPLE 4 prefix: ")
name5 <- readline(prompt = "SAMPLE 5 prefix: ")
name6 <- readline(prompt = "SAMPLE 6 prefix: ")

#switch working directory
wd0 <- getwd()
setwd(wd2)

#sample 1 analysis
path1 <- paste(wd2, "/", name1, ".collapsed.read_stat.txt", sep = "")
stat1 <- read.delim(path1)
Unmapped <- sum(stat1$stat == "unmapped", na.rm = TRUE)
Unique <- sum(stat1$stat == "unique", na.rm = TRUE)
Count <- c(Unmapped, Unique)
Id <- c("Unmapped", "Unique")
Sample <- rep((name1), 2)
stat1.data <- data.frame(Id, Count, Sample)
write.table(stat1.data, file = name1)
total.perc <- Unique/(Unique+Unmapped)*100
print(paste(name1, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique
reads: ", total.perc))

#sample 2 analysis
txt2 <- paste(name2, "txt", sep = ".")
path2 <- paste(wd2, "/", name2, ".collapsed.read_stat.txt", sep = "")
stat2 <- read.delim(path2)
Unmapped <- sum(stat2$stat == "unmapped", na.rm = TRUE)
Unique <- sum(stat2$stat == "unique", na.rm = TRUE)
Count <- c(Unmapped, Unique)
Id <- c("Unmapped", "Unique")
Sample <- rep((name2), 2)
```

```
stat2.data <- data.frame(Id, Count, Sample)
write.table(stat2.data, file = name2)
total.perc <- Unique/(Unique+Unmapped)*100
print(paste(name2, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique reads: ", total.perc))

#sample 3 analysis
txt3 <- paste(name3, "txt", sep = ".")
path3 <- paste(wd2, "/", name3, ".collapsed.read_stat.txt", sep = "")
stat3 <- read.delim(path3)
Unmapped <- sum(stat3$stat == "unmapped", na.rm = TRUE)
Unique <- sum(stat3$stat == "unique", na.rm = TRUE)
Count <- c(Unmapped, Unique)
Id <- c("Unmapped", "Unique")
Sample <- rep((name3), 2)
stat3.data <- data.frame(Id, Count, Sample)
write.table(stat3.data, file = txt3)
total.perc <- Unique/(Unique+Unmapped)*100
print(paste(name3, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique reads: ", total.perc))

#sample 4 analysis
txt4 <- paste(name4, "txt", sep = ".")
path4 <- paste(wd2, "/", name4, ".collapsed.read_stat.txt", sep = "")
stat4 <- read.delim(path4)
Unmapped <- sum(stat4$stat == "unmapped", na.rm = TRUE)
Unique <- sum(stat4$stat == "unique", na.rm = TRUE)
Count <- c(Unmapped, Unique)
Id <- c("Unmapped", "Unique")
Sample <- rep((name4), 2)
stat4.data <- data.frame(Id, Count, Sample)
write.table(stat4.data, file = txt4)
total.perc <- Unique/(Unique+Unmapped)*100
print(paste(name4, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique reads: ", total.perc))

#sample 5 analysis
txt5 <- paste(name5, "txt", sep = ".")
path5 <- paste(wd2, "/", name5, ".collapsed.read_stat.txt", sep = "")
stat5 <- read.delim(path5)
Unmapped <- sum(stat5$stat == "unmapped", na.rm = TRUE)
Unique <- sum(stat5$stat == "unique", na.rm = TRUE)
Count <- c(Unmapped, Unique)
Id <- c("Unmapped", "Unique")
```

```
  Sample <- rep((name5), 2)
  stat5.data <- data.frame(Id, Count, Sample)
  write.table(stat5.data, file = txt5)
  total.perc <- Unique/(Unique+Unmapped)*100
  print(paste(name5, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique
  reads: ", total.perc))

  #sample 6 analysis
  txt6 <- paste(name6, "txt", sep = ".")
  path6 <- paste(wd2, "/", name6, ".collapsed.read_stat.txt", sep = "")
  stat6 <- read.delim(path6)
  Unmapped <- sum(stat6$stat == "unmapped", na.rm = TRUE)
  Unique <- sum(stat6$stat == "unique", na.rm = TRUE)
  Count <- c(Unmapped, Unique)
  Id <- c("Unmapped", "Unique")
  Sample <- rep((name6), 2)
  stat6.data <- data.frame(Id, Count, Sample)
  write.table(stat6.data, file = txt6)
  total.perc <- Unique/(Unique+Unmapped)*100
  print(paste(name6, "Count Unmapped: ", Unmapped, "Count Unique: ", Unique, "Percent Unique
  reads: ", total.perc))

  #merge and plot
  name.all <- paste(name1, name2, name3, name4, name5, name6, sep = ".")
  namex <- paste(name1, name2, name3, name4, name5, name6, "txt", sep = ".")
  all.stat.data <- rbind(as.data.frame(stat1.data), as.data.frame(stat2.data), as.data.frame(stat3.data),
  as.data.frame(stat4.data), as.data.frame(stat5.data), as.data.frame(stat6.data))
  write.table(all.stat.data, file = namex)
  all.stat.plot <- ggplot(data = all.stat.data, mapping = aes(x = Sample, y = Count)) + geom_col(aes(fill
  = Id), width = 0.3)
  all.stat.plot + coord_flip()

  #save plot
  namey <- paste(name.all, "png", sep = ".")
  ggsave(namey, width = 15, height = 4, units = c("cm"))

  #revert back to original working directory
  setwd(wd0)

  #end
print(paste("Analysis complete. All files saved to", wd2, sep = " "))
```

Venn Diagram

The following source code generates a Venn diagram comparing isoform overlap between three samples.

```
library(VennDiagram)
grid.newpage()

#input prompts
wd.2 <- readline(prompt = "path to SAMPLE directory (abundance.txt): ")
name.1 <- readline(prompt = "SAMPLE 1 prefix: ")
name.2 <- readline(prompt = "SAMPLE 2 prefix: ")
name.3 <- readline(prompt = "SAMPLE 3 prefix: ")
print(paste("Choose sample color representation in the Venn Diagram."))
col1 <- readline(prompt = "SAMPLE 1 color: ")
col2 <- readline(prompt = "SAMPLE 2 color: ")
col3 <- readline(prompt = "SAMPLE 3 color: ")

#switch working directory
wd.0 <- getwd()
setwd(wd.2)

#sample 1
path.1 <- paste(wd.2, "/", name.1, ".collapsed.abundance.txt", sep = "")
ab1 <- read.delim(path.1)
area.1 <- nrow(ab1)
pbid1 <- ab1[c(1:area.1),c(1)]
print(paste(name.1, " area1: ", area.1))

#sample 2
path.2 <- paste(wd.2, "/", name.2, ".collapsed.abundance.txt", sep = "")
ab2 <- read.delim(path.2)
area.2 <- nrow(ab2)
pbid2 <- ab2[c(1:area.2),c(1)]
print(paste(name.2, " area2: ", area.2))

#sample 3
path.3 <- paste(wd.2, "/", name.3, ".collapsed.abundance.txt", sep = "")
ab3 <- read.delim(path.3)
area.3 <- nrow(ab3)
pbid3 <- ab3[c(1:area.3),c(1)]
print(paste(name.3, " area3: ", area.3))


#intersect values
```

```
na <- length(intersect(pbid1, pbid2))
print(paste("n12 =", na, ", a sum of", name.1, "-", name.2, "overlap"))
nb <- length(intersect(pbid1, pbid3))
print(paste("n13 =", nb, ", a sum of", name.1, "-", name.3, "overlap"))
nc <- length(intersect(pbid2, pbid3))
print(paste("n23 =", nc, ", a sum of", name.2, "-", name.3, "overlap"))
n123a <- (intersect(pbid1, pbid2))
nd <- length(intersect(n123a, pbid3))
print(paste("n123 =", nd, ", a sum of", name.1, "-", name.2, "-", name.3, "overlap"))

#append data to table
all.txt <- paste(name1, name2, name3, "txt", sep = ".")
Venn <- c("area1", "area2", "area3", "n12", "n13", "n23","n123")
Value <- c(area.1, area.2, area.3, na, nb, nc, nd)
all.data <- data.frame(Venn, Value)
write.table(all.data, file = all.txt)

#print venn diagram
all.venn <- draw.triple.venn(area1 = area.1, area2 = area.2, area3 = area.3, n12 = na, n13 = nb, n23 =
nc, n123 = nd, category = c(name.1, name.2, name.3), fill = c(col1, col2, col3), lty = "blank")

#save venn diagram
all.jpg <- paste(name1, name2, name3, "jpg", sep = ".")
jpeg(all.jpg)
grid.draw(all.venn)
dev.off()

#revert to first working directory
setwd(wd.0)

#end
print(paste("Analysis complete. All files saved to", wd.2, sep = " "))
```