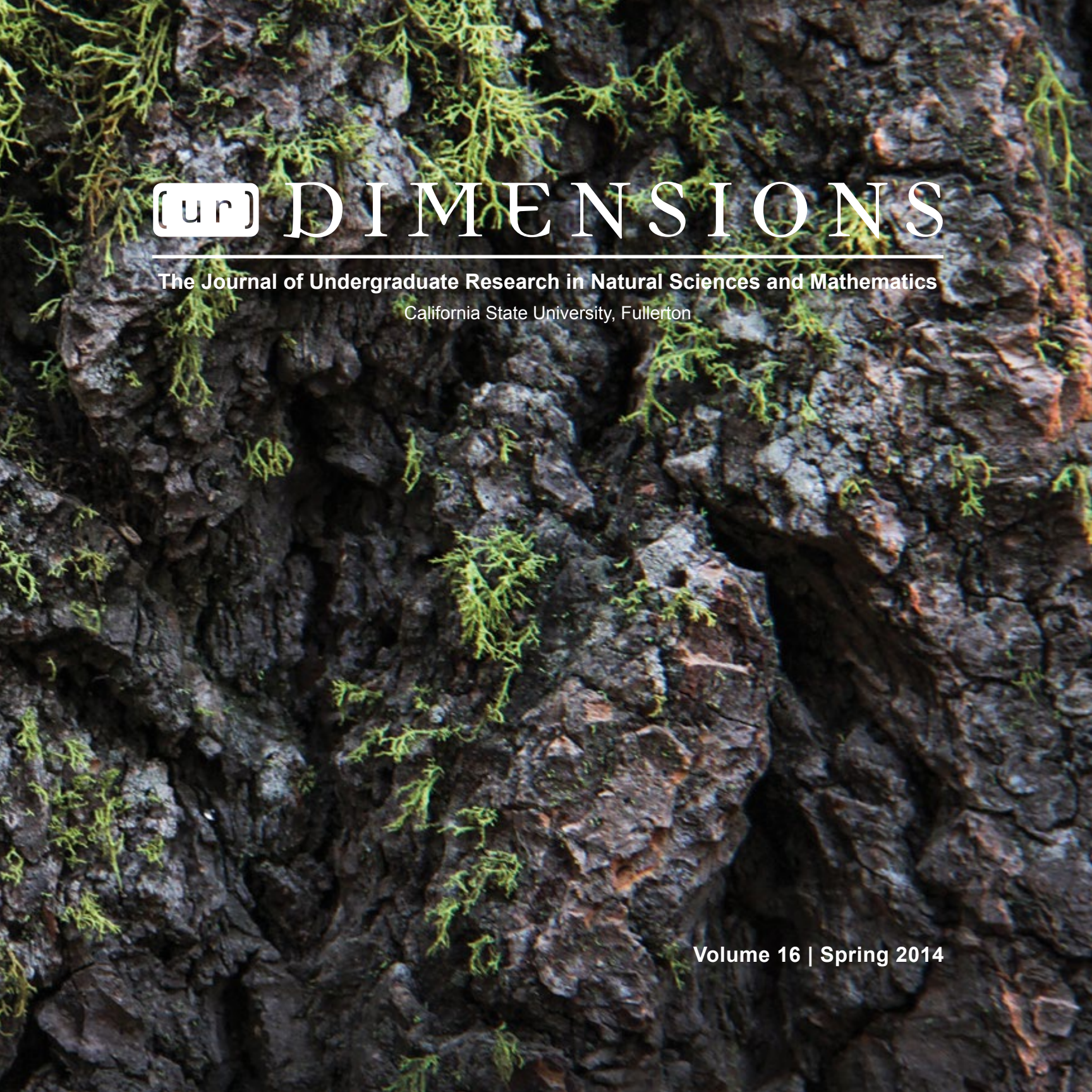


[ur] DIMENSIONS

The Journal of Undergraduate Research in Natural Sciences and Mathematics

California State University, Fullerton

Volume 16 | Spring 2014

Marks of a CSUF Graduate from the College of Natural Sciences and Mathematics

Graduates from the College of Natural Sciences and Mathematics:

Understand the basic concepts and principles of science and mathematics.

Are experienced in working collectively and collaborating to solve problems.

Communicate both orally and in writing with clarity, precision and confidence.

Are adept at using computers to do word processing, prepare spreadsheets and graphs, and use presentation software.

Posses skills in information retrieval using library resources and the Internet.

Have extensive laboratory/workshop/field experience where they utilize the scientific method to ask questions, formulate hypothesis, design experiments, conduct experiments, and analyze data.

Appreciate diverse cultures as a result of working side by side with many people in collaborative efforts in the classroom, laboratory and on research projects.

In many instances have had the opportunity to work individually with faculty in conducting

research and independent projects. In addition to attribtes of all NSM students, these students generate original data and contribute to the reseach knowledge base.

Have had the opportunity to work with very modern, sophisticated equipment including advanced computer hardware and software.

Dimensions

Dimensions: The Journal of Undergraduate Research in Natural Sciences and Mathematics is an official publication of California State University, Fullerton. Dimensions is published annually by CSUF, 800 N. St. College Blvd., Fullerton, CA 92834.

About the Cover

The national forests are one of California's most beautiful features. The dimensions of California's giant sequoias deem them to be the largest living thing on earth. The bark is extremely thick, and dynamic in depth and texture. The cover was shot at Sequoia National Forest, containing the world's most massive trees, with widths that reach as wide as 40 feet, and reach as tall as 274 feet.

Acknowledgements

Editor-In-Chief
Chris Baker - Geology

Editors
Kim Conway - Biology
Neha Ansari - Chemistry/Biochemistry
O'hara Creager - Geology
Kali (Duke) Chowdhury - Mathematics
Phillipe Rodriguez - Physics

Advisor
Amy Mattern - Assistant Dean for Student Affairs

Graphic Design
Niv Ginat - Layout Editor
Carose Le - Cover Designer

College of Natural Sciences & Mathematics
Dr. David Bowman - Interim Dean
Dr. Mark Filowitz - Associate Dean
Dr. Kathryn Dickson - Chair Department of Biological Science
Dr. Christopher Meyer - Chair Department of Chemistry and Biochemistry
Dr. Phillip Armstrong - Chair Department of Geological Sciences
Dr. Stephen Goode - Chair Department of Mathematics
Dr. James M. Feagin - Chair Department of Physics

College of the Arts
Dr. Joseph H. Arnold - Dean
Maricela Alvarado - Interim Assistant Dean for Student Affairs
John Drew - Graphic Design Professor

Special Thanks To
President Mildred García, Dean David Bowman, and **Assistant Dean Amy Mattern** for their support and dedication to Dimensions.

Biology

Chemistry & Biochemistry

Geology

Mathematics

Physics

Biographies

Determinants Of Prevalence Of Bot Fly Infestation In Thirteen-Lined Ground Squirrels In Colorado Shortgrass Steppe

Department of Biological Science, California State University, Fullerton

Kim Conway
Advisor: **Dr. Paul Stapp**

Abstract

Larvae of parasitic flies grow inside and feed upon tissues of wildlife species and therefore depend upon healthy hosts. Bot fly (*Cuterebra sp.*) larvae were discovered on thirteen-lined ground squirrels (*Ictidomys tridecemlineatus*) during long-term monitoring studies in northern Colorado. Although bot flies are common parasites of small mammals, there were no records of infestation of this squirrel species and the species of bot fly was unknown. I examined prevalence and load of bot flies in ground squirrels trapped in shrub and grassland habitats in spring and summer between 1999-2011 to determine host characteristics and environmental factors that influence patterns of infestation. I also investigated possible effects of prescribed fires in grasslands on infestation prevalence. Lastly, I used molecular genetics techniques to sequence the *COI* gene of preserved bot fly samples in an attempt to identify the species. Infested squirrels were rarely found on shrub sites and during spring trapping. Across all summers, average prevalence of infestation in grasslands was 7.9%. Infested squirrels had 1-7 bots, with 44.0% having only 1 larva. Infestation did not vary greatly with host sex, age, or weight. Prevalence was significantly higher (33.0%) in burned sites one year after a prescribed fire, and remained consistently higher on burned sites than on unburned sites. My results suggest that fires may alter the environment in ways that increase the susceptibility of squirrels to infestation or the ability of flies to infest hosts. *COI* gene sequences revealed the bot fly species to be most closely related to Cuterebra fontinella.

Introduction

Parasites harm their hosts by feeding on nutrients from inside the host's body (Catts 1982). Although parasites can sometimes directly kill their host, many parasites require healthy hosts to survive and successfully reproduce, and therefore, may not pose a direct health risk unless infestation loads are high (Slansky 2007). Bot flies (Family Oestridae), for example, spend their entire larval cycle inside mammalian hosts (Catts 1982) and benefit when their host is healthy (Slansky 2007). Bot flies of the genus Cuterebra are host-specific (Catts 1967), and typically infest small North American rodents, including chipmunks and tree squirrels (Catts 1967; Jacobson et al. 1961). Although bot fly infestation is usually not fatal, it can cause energy loss, malnutrition, and secondary infection at the site of larva emergence (Catts 1982; Slansky 2007). Higher loads of bot flies may also interfere with a host's ability to forage, escape predators, and reproduce (Catts 1982; Slansky 2007).

Host traits such as age and sex may affect the rate of bot fly infestation (Jacobson et al. 1961). In addition, habitat structure may influence parasitism rates; for example, Blair (1942) found that bot fly parasitism of rodents was higher in shrub-dominated areas than in grasslands. Prescribed burning, a common habitat management technique in grasslands that alters habitat structure (Converse et al. 2006), can also affect the abundance of both rodent hosts and their parasites. Working in Oklahoma, Boggs et al. (1991) found that bot fly parasitism in small mammals was higher in unburned areas than burned areas.

In 1999-2011, thirteen-lined ground squirrels (*Ictidomys tridecemlineatus*) that were live-trapped a part of long-term monitoring studies in northern Colorado (Stapp et al. 2008) were found to be infested with bot flies. Previous studies have determined infestation rates of bot flies on small mammals including the white-footed mouse (e.g. Clark and Kaufman 1990), gray squirrels (Jacobson et al. 1961), and chipmunks (Bergstrom 1992), but there are no reports of bot fly parasitism of thirteen-lined ground squirrels, despite the fact that these squirrels are widespread and common in Great Plains grasslands. In fact, a widely cited review by Catts (1982) specifically stated that Cuterebra infest small rodents (including tree squirrels), but not ground squirrels. For example, none of the 179

thirteen-lined ground squirrels trapped in Kansas were parasitized by bot flies (Clark and Kaufman 1990), and none of the 46 thirteen-lined ground squirrels trapped in Alberta, Canada, had bot flies either (Gummer et al. 1997). Aside from a single anecdotal reference (Lugger 1896, cited by Sabroksy 1986), there are no published records to indicate which species of bot fly infests this squirrel. Morphological traits usually can be used to identify species (Baird 1972) using a taxonomic key. However, keys for identifying bot flies are based on morphological traits of the adult fly, which live only a short time once they emerge from the soil, and are very difficult to observe and capture (Catts 1982).

For my project, I sought to better understand the interactions between thirteen-lined ground squirrels in northern Colorado and their bot fly parasites. My specific objectives were to: 1) analyze existing data from 1999-2011 on bot fly infestation of ground squirrels to identify host or environmental factors that may influence prevalence and intensity; 2) attempt to determine the species of bot fly on squirrels by analyzing samples of the late-instar larvae collected from ground squirrels using molecular genetics methods; and 3) ex tract live, late-instar larvae from ground squirrels, rear them to adulthood in the laboratory, and confirm the species of bot fly using adult morphological traits.

Methods

Data Analysis:

Information on rates of bot fly infestation of thirteen-lined ground squirrels from 1999-2011 was collected during live-trapping studies conducted as part of the Shortgrass Steppe- Long Term Ecological Research project the (SGS-LTER). The study site was the Central Plans Experimental Range (CPER), approximately 14 km north of Nunn, CO. Vegetation is characterized as shortgrass steppe, which is dominated by two perennial, warm-season short grasses (blue grama *Boutelloua gracilis*, buffalograss *Buchloe dactyloides*), although some areas also have large, woody shrubs, especially four-wing saltbush *Atriplex canescens*. Live-trapping was conducted in spring (May/June) and summer (July) each year on six 3.14-ha webs (three grassland, three shrub sites) from 1999-2011. In addition, squirrels were trapped on three additional grassland webs that were burned in the previous autumn to examine the effect of prescribed fire on squirrel populations. Bot fly prevalence was measured at one, two, three, and four years post-fire. One site (26NWSE) was burned in autumn 2007 and trapped in July each year from 2008 to 2011. Other grassland webs that were burned in 2008, 2009 or 2010 were trapped only once in July 2011.

Webs consisted of 12 100-m transects arranged in a spoke-like fashion, with extra-large Sherman live-traps every 20 m along each transect and two traps placed at the center, for a total of 62 traps. Traps were baited with peanut butter and oats and shaded with PVC to prevent mortality from heat. Traps were set at dawn and checked at mid-morning for four consecutive days in each session. Field crews collected data on sex, age, weight, and physical condition of squirrels, including presence of parasites, and marked each individual with a colored Sharpie marker to distinguish recaptures from newly caught individuals. Individuals were released unharmed at their capture locations. I used data from squirrels only on their first date of capture. Prior to analysis, juvenile and subadult squirrels were combined into one age class (young-of-year; YOY). In some instances, field crews recorded evidence of multiple bot fly warbles on each host; however, because the exact numbers were not always recorded, squirrels were categorized as having no flies, one bot fly, or more than one bot fly.

I used two metrics of bot fly infestation. First, prevalence was the proportion of individual hosts that were infested with bot flies, and was calculated for different sex, age, and weight classes as the number of bot flies divided by the total number of individual hosts in a particular class. Second, for hosts that had at least one bot fly, intensity was calculated as the proportion of infested hosts that had one fly versus more than one fly. Intensity was calculated by dividing the number of hosts that had multiple fly larvae by the total number of infested hosts; i.e., with at least one.

In late June 2013, thirteen-lined ground squirrels were live-trapped for four consecutive mornings in an attempt to capture squirrels infested with bot fly larvae so that larvae could be extracted, reared to adulthood, and identified using a morphometric key.

Laboratory Methods:

Late-instar larvae were collected in July 2011. Three larvae were dissected to obtain interior body tissue and cuticle tissue for DNA extraction. Cuticle tissue was cut in 2-mm x 2-mm squares. A clear, tube-like interior section of the larva was used, as it appeared to be the only visibly intact structure. Additional interior tissue was added to ensure approximately 0.025 grams of tissue was collected. DNA was extracted from interior and cuticle tissue of three bot fly samples using DNeasy Blood & Tissue kit and following the manufacturer's recommendations (QIAGEN®, Valencia, CA).

DNA was extracted following the manufacturer's Animal Tissue protocol, with the adjustment of incubating samples overnight and eluting with 100 µl instead of 200 µl. Three samples were eluted a second time. DNA concentration (ng/µl) was measured using NanoDrop spectrophotometer. Three interior tissue samples were chosen for further analysis because they had the highest DNA concentration of all purified samples.

Polymerase chain reaction (PCR) was used to amplify 657 bp of the cytochrome oxidase subunit I (*COI*) gene in three interior samples with the highest concentration of DNA. The *COI* gene is a region of mitochondrial DNA that is universally used in species identification. The total reaction volume was 50 µl, including 1 µl of forward primer, 1 µl of reverse primer, 3 µl of DNA sample, and 45 µl of Platinum Taq Master Mix (Invitrogen™ Life Technologies). Samples were run in a BioRad t-100 Thermocycler at 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 49°C for 40 s, and 72°C for 1 min, with a final extension of 72°C for 5 min. Genus-specific primers (LC0149of and HC02198r) and species-specific sequences were provided by Dr. Brian Weigmann and Brian Cassell from North Carolina State University. PCR products were visualized on a 2% agarose gel using gel electrophoresis.

PCR products were sequenced by SEQUETECH (Mountain View, CA). Consensus sequences were assembled using bidirectional sequences using CodonCode Aligner software (CodonCode Corporation) and aligned using clustW in MEGA 5.1 (Kumar et al. 2011). Phylogenetic analysis was conducted with maximum likelihood, maximum parsimony, and neighbor-joining models using MEGA 5.1.

Results

Analysis Of Long-Term Infestation Patterns

Between 1999 and 2011, the earliest date that bot fly warbles were detected on squirrels was 16 May, but only six observations of bot fly infestation were noted before 3 June. Prevalence of bot flies was significantly lower in May and June (1.8% and 0.3%, respectively) than in July (Fig. 1; $X_2 = 21.75$, d.f. = 2, $P < 0.001$). In addition, squirrels in grassland webs had a much higher prevalence of infestation (5.3%) than those in shrub areas (Fig. 1; $X_2 = 16.02$, d.f. = 1, $P < 0.001$). For this reason, further analyses were only conducted on data from grassland webs trapped in July.

Long-term data (1999-2011) from grassland webs in the summer indicated that prevalence was highest in 2008 (Fig. 2), with little variation in other years. On average, 7.5% of squirrels were infested in a given year. Prevalence did not vary much by weight class, although the largest numbers of bots were on YOY (< 100 g; Fig. 3). Bot fly prevalence of female adults (8.2%) was similar to that of male adults (7.0%). Likewise, prevalence rates of female YOY (7.4%) did not differ significantly from male YOY (7.1%; Fig. 4; $X_2 = 0.15$, d.f. = 3, $P = 0.985$). Intensity of infestation remained similar among different sex and age classes ($X_2 = 2.49$, d.f. = 3, $P = 0.480$), although adult females had a higher intensity (75%) than other groups (Fig. 5). For hosts where the actual number of bots was recorded, the number of larvae ranged from 1 to 7.

Combining all years of the prescribed fire study (2008-2011), bot fly prevalence was consistently higher on squirrels from the three webs that were burned in 2007 than from those on unburned grassland webs between 2008-2011 (Fig. 6; paired $t = 3.91$, d.f. = 3, $P = 0.030$). Combining results from webs based on the number of years post-fire, prevalence of bot flies was significantly higher (33%) on sites one year after being burned (Fig. 7; $X_2 = 11.98$, d.f. = 3, $P = 0.007$). After an increase in prevalence during the first year after burning, prevalence returned to the same rate (11-13%) by two years after a fire, which was the same rate as that on unburned grassland webs during the same time period (12%; Fig. 7).

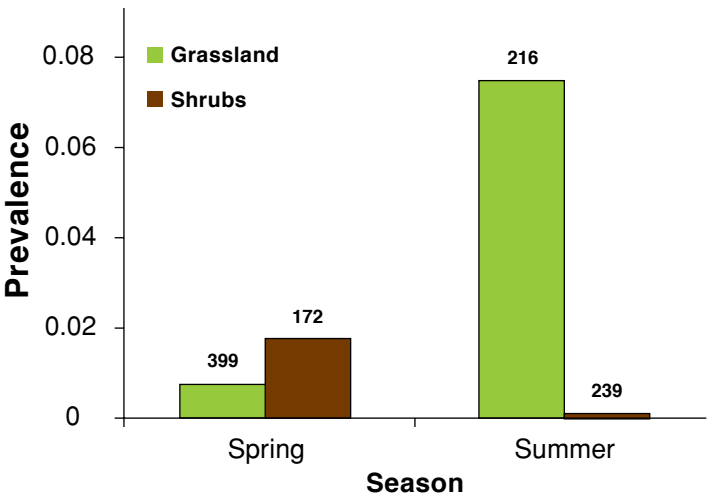


***Figure 1**. Prevalence of bot fly infestation in thirteen-lined ground squirrels by habitat type in long-term trapping webs in northern Colorado in spring and summer were highest in grasslands in summer (7.5%). Total number of hosts examined is shown above bars (Season: $X_2 = 21.75$, d.f. = 2, $P < 0.001$, Habitat: $X_2 = 16.02$, d.f. = 1, $P < 0.001$).*

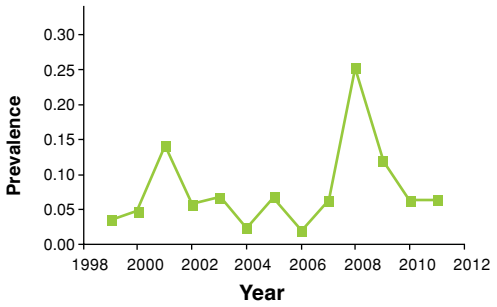


Figure 2. *Prevalence of bot fly infestation in thirteen-lined ground squirrels in long-term grassland trapping webs in northern Colorado during summer from 1999-2011. Number of hosts examined in a given year ranged from 31 to 101, with 59 hosts in 2008, the year with the highest prevalence.*

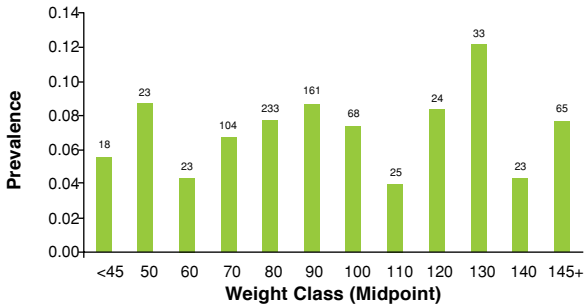


Figure 3. *Prevalence of bot fly infestation in thirteen-lined ground squirrels in long-term grassland trapping webs in northern Colorado in summer, based on data from 1999-2011. Number of hosts examined in each weight class is shown above bars. Values on the x-axis are the midpoints of the 10-g weight classes.*

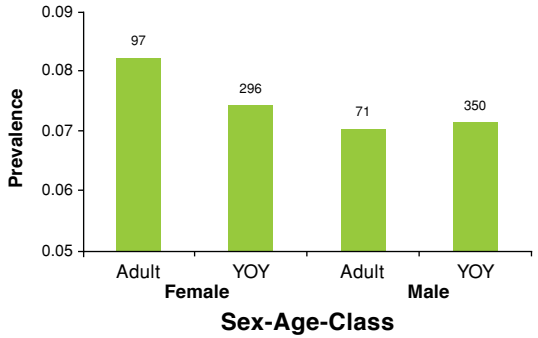


Figure 4. *Prevalence of bot fly infestation in thirteen-lined ground squirrels in the summer in long-term grassland trapping webs in northern Colorado, based on data from 1999-2011. Number of hosts examined is shown above bars (X2 = 0.15, d.f. = 3, P = 0.985).*

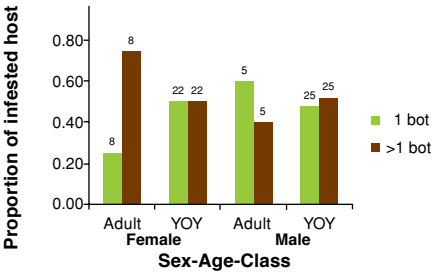


Figure 5. *Intensity of bot fly infestation in thirteen-lined ground squirrels in long-term grassland trapping webs in northern Colorado in the summer, based on data from 1999-2011. Number of infested hosts is shown above bars (X2 = 2.49, d.f. = 3, P = 0.480).*

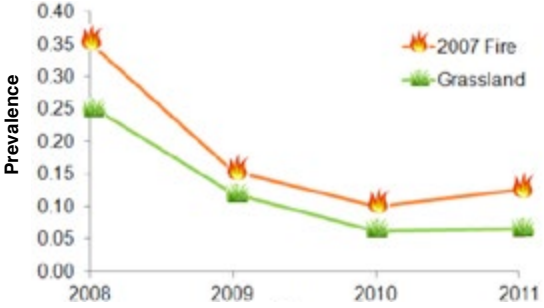


Figure 6. *Prevalence of bot fly infestation of thirteen-lined ground squirrels in three grassland trapping webs that were burned in autumn 2007, compared to three long-term grassland webs that were never burned (2008-2011; paired t = 3.91, d.f. = 3, P = 0.030).*

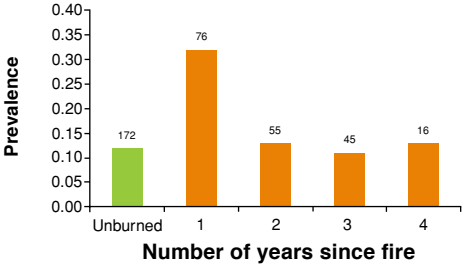


Figure 7*. Prevalence of bot fly infestation in thirteen-lined ground squirrels in northern Colorado on trapping webs in July as a function of the number of years since autumn prescribed fires. Six sites were trapped one, two, and three years post-fire, and three sites were sampled four years post-fire, from 2008-2011. Data from unburned webs were from three grassland webs trapped in July from 2008-2011 (X2 = 11.98, d.f. = 3, P = 0.007).*

Molecular Genetics Analysis

The *COI* consensus sequences revealed the three samples to be 0.21% different from each other, and 4.9% different from its closest relative *C. fontinella* (Fig. 8).

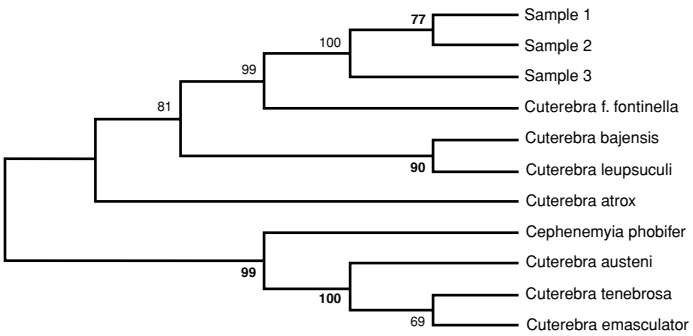


Figure 8. *Phylogenetic tree indicating C. fontinella as the closest relative to the unidentified bot fly samples (Sample 1-3). Maximum Likelihood bootstrap consensus tree using General Time Reversible Model and discrete Gamma distribution was generated using MEGA 5.1 software.*

Discussion

My study offers the first detailed description of infestation of thirteen-lined ground squirrels with bot flies. Prevalence was significantly higher in summer (July) than late spring (May, June), and in grasslands than in shrub areas, which differs from the account of Clark and Kaufman (1990) who found parasitism was higher in shrub areas than grasslands in tallgrass prairies. Prevalence did not vary much by host sex or age. Of the squirrels infested with bot flies, adult females tended to have higher loads that other sex and age classes, although this result was not significant. Prevalence varied from 2.0% to 25.4%, with an average of 7.5%, across the 13 years of sampling in grasslands. Prevalence was unusually high in 2008, although the reason for this spike is unknown. April-June 2008 was the third of three consecutive years with very low spring precipitation and represented a period during which squirrel population numbers on the site had been declining since 2006 (Fig. 2; P. Stapp, unpublished data). Moreover, relatively few squirrels were captured in May 2008 (7), compared to an average of 29.5 in other years (SD=8.7, range 19-45), so available hosts may have been scarce.

Prescribed burns seemed to affect prevalence of bot fly infestation, with prevalence significantly higher the first year after a fire. In addition, prevalence was significantly higher on three grassland webs that were burned than on unburned webs trapped at the same time. These results differ from those of Boggs et al. (2007), who found lower levels of bot fly infestation of small mammals in burned tallgrass prairie in Oklahoma. Boggs et al. (2007) argued that fire might have killed eggs and larvae belowground, or that the removal of litter by burning made the microclimate unsuitable for developing larvae. My results suggest that, in shortgrass steppe, where there is no significant litter layer, fires may cause environmental changes that increase the susceptibility of squirrels to bot fly infestation. For example, burning of grasses may alter the vegetation near burrows in a way that makes it easier for squirrels to come in contact with bot fly eggs. In addition, fires may have altered the environment in a way that increased bot fly survival, such as warming the soil. Having a warmer soil temperature during the time larvae pupate may lead to a higher number of bot flies emerging in the spring. Lastly, autumn fires may alter the suitability of habitat for oviposition by adult flies.

Because bot fly larvae could not be identified morphologically, molecular genetic testing of preserved larvae was used needed to attempt to identify the species of fly. My results indicate that, with a difference of only 0.21% between sequences, the three bot fly larvae are most likely of the same species (intra-specific variations are less than 1.0%, Hebert et al 2003). Based on the maximum likelihood trees, the larvae appear to be most closely related to *C. fontinella*, although there is a 4.9% difference between the *COI* gene sequence and that of *C. fontinella*. It is unclear, however, whether this difference is large enough to consider it a new species. Although the widely used difference in gene sequence between species is 2.0%, studies have indicated that this boundary may not be accurate in determining species divergence in insects, for which inter-specific variation can range from 1.0-30.7% (Cognato 2006). Examination of sequences of other genes, e.g. 12S, may help to resolve the species identification.

The final part of my project was to extract live bot fly larvae from squirrels, bring them back to the laboratory, and rear larvae to adulthood to observe morphological traits. Trapping success in 2013 was low, however, with only two squirrels captured in grasslands areas in late June, neither of which had bot flies. Similarly, no squirrels captured in 2012 had bot flies, although the number of hosts again was low (nine individuals in July 2012). With so few hosts captured over the past two years, it is not surprising that prevalence would be at or near zero. It is not clear if bot flies could have switched to a different, more abundant rodent host, although, to date, there are no records on bot flies on other rodents captured at CPER since 1994 (P. Stapp, unpublished data). Future work should attempt to determine the fate of the bot fly populations and to identify other possible hosts.

Acknowledgements

I thank the Faculty Development Center and Associated Students, Inc., for funding, Mark Lindquist and Kevin Meierbachtol (SGS-LTER) for collecting larvae samples and for supervising field crews, Nicole and Kaplan and SGS-LTER field crews for collecting data, and David Augustine (USDA-ARS) for coordinating prescribed burns. I thank Jeff Boettner for his extensive knowledge of bot fly biology and support of my project, and Dr. Brian Wiegmann and Brian Cassell for providing primers and species sequences. I also thank Karla Flores, whose knowledge and experience with molecular techniques made the laboratory analysis part of my project possible.

References

Baird, C.R. 1972. Development of *Cuterebra Ruficrus* (Diptera: Cuterebridae) in six species of rabbits and rodents with a morphological comparison of *C. Ruficrus* and *C. Jellisoni* third instars, *Journal of Medical Entomology*, **9**:81-85.

Bergstrom, B.J. 1992. Parapatry and encounter competition between chipmunk (Tamias) species and the hypothesized role of parasitism, *Am. Midl. Nat.* **128**:168-179.

Blair, W.F. 1942. Size of home range and notes on the life history of the woodland deer-mouse and eastern chipmunk in northern Michigan, *J. Mamm.* **23**:27-36.

Boggs, J.F., Lochmiller, R.L, McMurry, S.T., Leslie, D.M. Jr., and D.M. Engle. 1991. *Cuterebra* infestations in small-mammal communities as influenced by herbicides and fire, *J. Mamm.* **72**:322-327.

Catts, E.P. 1967. Biology of a California rodent bot fly, Cuterebra latifrons, *Journal of Medical Entomology* **4**:87-101.

Catts, E.P. 1982. Biology of new world bot flies: Cuterebridae, *Ann. Rev. Entomol.* **27**:313-338.

Clark, B.K. and D. W. Kaufman. 1990. Prevalence of botfly (Cuterebra sp.) parasitism in populations of small mammals in eastern Kansas, *Am. Midl. Nat.* **124**:22-30.

Cognato, A. 2006. Standard percent DNA sequence difference for insects does not predict species boundaries, *J. Econ. Entomol.* **99**:1037-1045.

Converse, S.J., Block, W.M. and G.C. White. 2006. Small mammal population and habitat responses to forest thinning and prescribed fire, *Forest Ecology and Management* **228**:263-273.

Gummer, D.L., Forbes, M.R., Bender, D.J., and R. M. R. Barclay. 1997. Botfly (Diptera: Oestridae) parasitism of Ord's kangaroo rats (*Dipodomys ordii*) at Suffield National Wildlife Area, Alberta, Canada, *J. Parasitol.* **83**:601-602.

Hebert, P.D.N, Ratnasingham, S., and J.R. de Waard. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species, *Proc. R. Sond. Lond.* **270**:S96-S99.

Jacobson, H.A., Hetrick, M.S. and D.C. Guynn. 1961. Prevalence of *Cuterebra emasculator* in squirrels in Mississippi, *Journal of Wildlife Diseases* **17**:79-85.

Slanksy. F. 2007. Insect/mammal associations: Effects of Cuterebrid bot fly parasites on their hosts, *Annu. Rev. Entomol.* **52**:17-30.

Stapp, P., Van Horne, B., and M.D. Lindquist. 2008. Ecology of mammals of the shortgrass steppe. Pp. 132-180 in: Ecology of the shortgrass steppe: a long-term perspective (eds. W.K. Lauenroth and I.C. Burke). Oxford Univ. Press.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and S Kumar. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution **28**: 2731-2739.

.

Comparison of Worm Cast Application Methods on Soil Water Retention and Plant Growth

Department of Biological Science, California State University, Fullerton

Tiffany Duong
Advisor: **Dr. Joel Abraham**

Abstract

Southern California experiences regular annual water deficit. Because of this, conservation agriculture has been developed to help California meet its increasing water needs. Conservation agriculture includes the use of soil amendments to increase the water retentive properties of the soil and thus, decreasing the amount of water necessary for each crop yield. Some amendments, such as worm casts, can accomplish this as well as increase plant productivity. Worm casts, plant wastes processed by worms that are traditionally used as an organic fertilizer, were used in this experiment as a soil amendment in an attempt to increase the soil water retention and plant productivity. The worm casts were added to the soil in two ways: mulch and mixed into the soil. It was found that worm casts do significantly increase soil water retention, more so when incorporated into the soil than used as mulch. During the plant productivity observation based on the same treatments, it was also found that the addition of the worm casts increase overall plant productivity, and for radishes, increases the root size instead of shoot size more so when the worm casts were used as mulch than mixed into the soil. In conclusion, when worm casts are added to soil, they are able to not only increase the soil water retention, but also increase the roots of plants, which have a direct implication to vegetation where the root is the more necessary portion of the whole plant.

Introduction

Southern California is a Mediterranean-type climate. For much of the year, there are high evaporation rates with low levels of precipitation, causing a water deficit. This annual water deficit is predicted to increase over the next 100 years (World Water Assessment Programme). Because of this annual water deficit, the water we use in California has to be imported from neighboring states at progressively higher costs through a complicated system of aqueducts. It is estimated that up to 70% of the water Californians use is imported from out of state sources which have caused many political issues between California and its neighbors (California Department of Water Resources), (World Water Assessment Programme). The cost of importing this water increases each year with a decreasing supply, so water conservation is becoming of utmost importance (World Water Assessment Programme). Not only is most of the imported water used for agriculture, estimated to be up to 80-85%, but the state of California alone produces 7.1% of the world's produce and more than half of the nation's produce (California Department of Food and Agriculture), (California Department of Water Resources). With the predicted population increases, California's water needs will only increase (World Water Assessment Programme). For these reasons, identifying methods to conserve water in agricultural practices is currently the main area of focus in water conservation.

Agriculturists have combined several methods that conserve water into what they refer to as conservation agriculture. Some common methods of conservation agriculture include efficient irrigation, genetic modification of crops, and watering crops only when needed in order to decrease costs (Kassam et al., 2012). Soil amendments are known to help with the latter. Mulching has been developed in an attempt to prolong the times in between watering by decreasing the evaporative water loss from the soil and is the process where the surface of the soil is covered by a layer of amendment which can range between varieties of substances. Popularly used soil amendments include wood chips and inorganic substances such as rocks or pebbles (Dahiya, Ingwersen, and Streck, 2007). One of the many organic amendments used as mulch are worm casts, or processed plant waste produced by worms. Typically used as an organic fertilizer in agriculture, numerous studies have been done on their ability to provide plants with plenty of macro- and micronutrients and their ability to reduce a plant's susceptibility to parasites and diseases, but not a lot has been done to study the affect worm casts have on water retention (Norgrove and Hauser, 1999).

It has been well documented that increasing the organic material in the soil directly influences the water retention of said soil (Rawls et al., 2003). There are different methods of application of soil amendments. The two main ways in which the amendments are added are through mulch or incorporating it into the soil. The different methods of application could have positive implications on plant productivity (Asawalam and Hauser, 2001). In conservation agriculture, not only do costs need to be minimized, but the benefits need to be maximized. It is a constant struggle for farmers to identify the correct combination of minimizing and maximizing, especially in arid environments such as in Southern California (Kassam et al., 2012) (Shemdoe, Van Damme, Kikula, 2009). Worm casts can aid farmers in finding this balance. They can not only retain water, but they can also increase plant productivity. This study sets out to determine if the use of worm casts can increase soil water retention and if method of application will affect the retention any differently. Also, this study sets out to determine if different methods of application will have an effect on plant productivity for maximization of crop yield.

Methods and Materials

Experiment 1: Soil Water Retention

The soil water retention levels were measured using gravimetric water content methods. The evaporative rate of water from the soil mixture was measured for four different treatments each with five replicates. The treatments include an experimental control (experimental) where the soil was not mixed and no casts were added, a procedural control (procedural) where the soil was mixed but no casts were added, the worm casts used as mulch (mulch) on top of the soil, and the casts mixed (mixed) into the soil. For this portion of the experiment, a 50% sand 50% nutrient deficient soil mix was used to ensure nothing present in the soil could affect the rate of evaporation of water other than the added worm casts. 35 mL clay pots commonly found in any variety store were used as containers. All the soil mixture as well as worm casts were dried in the oven for three days before use in the experiment. For the mixed treatment, 140 mL of the soil mix was placed in a large mixing bowl along with 35 mL of worm casts (casts), and then divided into the treatment pots, 35 mL each. For the mulched experiment, the same amount of soil mix was placed in a separate mixing bowl, 28 mL were placed in each pot, and 7 mL casts were added to the surface. For both the experimental and procedural controls, 140 mL of soil mixture was placed in separate containers, the soil used for the procedural control was agitated, and divided into the treatment pots, 35 mL each. 6 mL of water was added to each pot and were placed in a drying oven set to about 40-45° C and their weights were measured every hour for 10 consecutive hours.

Experiment 2: Plant Productivity

The second portion of the project is the observation of plant productivity based on the different application methods of the casts. The same four treatments were used, with twelve plant replicates each. The radishes were grown from seed into seedlings for two weeks in large bins using regular potting soil inside a greenhouse. The seedlings were given ample amounts of water at this stage. The seedlings were then transplanted into their individual treatment pots and grown outside for an additional four weeks when the radishes achieved maturity based on radish growing instructions. 0.6 liter plastic pots commonly found in plant nurseries were used. The initial watering of the seedlings in their treatment pots was 80 mL to account for the shock of transplantation and reduced to 40 mL for every watering period after. For the mixed experiment, 1.2 liters of casts and 4.8 liters of soil mixture were placed into a bucket, mixed thoroughly, and 0.5 liters was placed in each treatment pot along with one transplanted radish. For the mulch experiment, the same amount of soil mixture was placed into a separate bucket, then divided into its treatment pots, and 0.1 liters of casts were added to the surface of each pot after the addition of a transplanted radish. For the experimental and procedural controls, 6 liters of soil mixture was placed in a large bucket for each control, the mixture for the procedural control was agitated, and then 0.5 liters of each mixture was added into their respective treatment pots. Again, the same soil mixture of sand and nutrient deficient soil was used.

For data collection, the radishes were removed from their treatment pots and rinsed gently in water to ensure maximum recovery of the fine root hairs. The roots and shoots were separated, dried for three days in a drying oven, and their dry biomass was recorded.

Statistical Analyses:

For the water retention portion of the experiment, the data was compiled into excel and an ANOVA test was run on the last data point for each treatment. For the plant productivity portion of the experiment, a single factor ANOVA test was done on the root, shoot, and overall biomass data. A pairwise comparison was performed on the root:shoot and total biomass data to determine if the data were significantly different from each other.

Study Organism:
For the plant productivity portion of the experiment, Raphanus sativus (cherry radishes or radishes) was used. Easily found in any variety store or gardening shop, radishes are a common household vegetable. Depending on the growing conditions, radishes take up only six weeks to mature and are good companion crops because of this.

Results

In the soil water retention portion of the experiment, every hour for ten consecutive hours, the masses of the pots filled with a known amount of dirt mixture and worm casts were measured. The data were recorded and it was found that the addition of worm casts to the soil as both mulch and mixed into the soil increases the soil's water retention capabilities, shown in Figure 1. However, according to the data, it also shows that when the worm casts are added as incorporated into the soil, it significantly increases the soil's water retention more than using worm casts as mulch. It was also found that both the experimental control, where there were no worm casts added and no mixing of dirt, and the procedural control, where there were no worm casts added but there was mixing of the dirt, were not significantly different from each other.

In the observation on plant productivity portion of the experiment, radishes were grown in the same treatments as the soil water retention portion of the experiment for four weeks. As shown in Figure 2a, it was found that no matter the method of application, the addition of the worm casts significantly increased plant productivity. Upon further inspection, it was observed that it was the root of the plant, and not so much the shoots, that had increased in biomass (Figure 2b). According to the data collected and analyzed using a pairwise comparison, it was found that the procedural and experimental controls were not significant from each other. It was also found that the mulch and mixed treatments were significantly different from each other as well as from the two controls.

Conclusions

For the soil water retention portion of the experiment, it was found that the addition of the worm casts positively and significantly increased the soil water retention. This could be because worm casts are organic material. It is widely known that soil moisture and the amount of water soil is able to hold increases with increasing humus, or soil organic material (Rawls et al., 2003). One of the criteria for high quality soil is the occurrence of humus (Shemdoe, Van Damme, Kikula, 2009). In this experiment, the increased water retention ability of the sand and dirt mixture was due to the addition of the worm casts, an organic material, when there was originally no organic material. Furthermore, it was also found that mixing the worm casts into the soil significantly increased the soil water retention more than using the casts as mulch. Rawls, Pachepsky, Ritchie, Sobecki, and Bloodworth (2003) believed this is because the organic material has a water potential of its own. This means that the worm casts have a different rate of evaporation than the soil mixture would. Since the water was added after the casts were added, this could mean that the water in the mulch experiment stayed in the cast surface layer so all the water was also at the surface and more prone to evaporation. However, when the casts were mixed into the soil, the water was drawn from the surface, deeper into the soil and wasn't as readily able to evaporate, resulting in higher water retention.

For the plant productivity portion of the experiment, it was found that the overall biomass significantly increased upon addition of the casts. These findings correlate with Norgrove and Hauser (1999) who also demonstrated that the addition of worm casts significantly increased plant biomass. This could be due to the increased retentive powers the soil had due to the added organic matter. In the experiment, the radishes were water stressed to ensure that the difference between the treatments with and without the casts could be more visible. Because of this, the radishes with the casts had more water held within the soil and were able to out-perform the radishes that were not treated with worm casts. It is also possible that the increased biomass could have been the result of the added macronutrients to the soil. In the experiment, the two controls had no to little nutrients. The difference in biomasses between the controls and the two treatments could possibly have been the addition of more nutrients. This is what Norgrove and Hauser (1999) concluded as well. It was also found that the roots rather than the shoots were increasing in size in the mulch and mixed treatments contrary to what previous literature has stated (Asawalam and Hauser, 2001). They found that there was only a significant increase in dry biomass in the mulch experiment and determined that there was no significant increase in the mixed treatment. However, they studied maize and cowpea plants where the edible portion of the plant is above ground as well as most of the plant. The roots of these plants do not play a major role in the overall biomass of the plant. The root of radishes is where the plant stores its source of carbohydrates. In the case of this experiment, the root size increased because the plant was more

easily able to obtain water from the soil and as a result, was able to photosynthesize and store more energy. These data have major implications on future farming techniques for vegetation where the root is edible. Urban farmers must maximize the use of the small plots of land they are allotted. Because of the results of these data, it would seem that urban farmers can possibly have higher yields or better results while also saving water and costs.

However promising the data may seem, further research must be done on whether the biomass data are representative of the worm casts' ability to increase soil water retention or whether the data are the result of the added macro- and micronutrients. Similarly, the soil water retention study portion of the experiment was done in a controlled environment in small pots. Future research must be done on whether or not the results would be the same if more common sized pots were to be used.

Several errors could have arisen during the course of the experiment. For instance, during the water retention portion, an error could have occurred when the pots were weighed. There was a difference in mass between when the pots cooled and when they were measured right after they were removed from the oven. This error could possibly alter the results. Error could also have occurred during the plant productivity study portion as well. One error could have occurred due to the water stress put on the plants. The way in which the radishes were determined as experiencing water stress was slight wilt and depending on the amount of wilt, the amount of water the radishes were given differed. This could have resulted in too much water stress which would be detrimental to the health of the plants or it could have resulted in too little water stress and the true value of the worm casts was not actually determined. The last error that could have occurred was the consumption of the radish shoots by the larvae of an insect. This could have resulted in false data and without this error, it could have been that both the roots and shoots were increasing at a roughly equal rate.

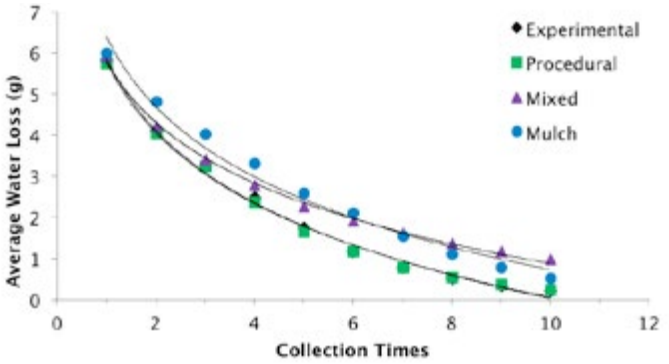


Figure 1. *Average water loss in soil treatments over time. Mulch and mixed retained significantly more water at 10 hours than did either control treatment (F= 21.17, df = 4, p < 0.001).*

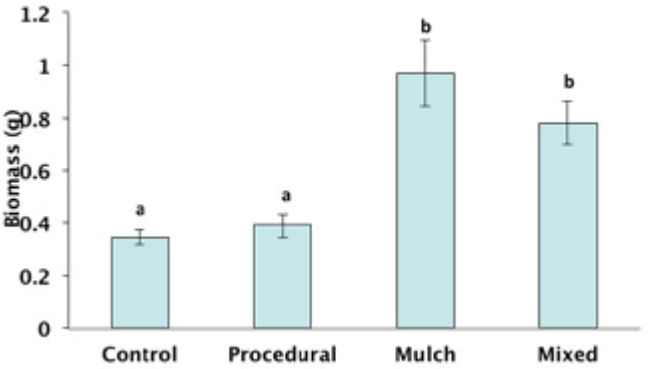


Figure 2a. *Mean total biomass of radishes across soil treatments. Radish biomass in mulch and mixed significantly greater than in control treatments (F = 14.666, df = 3, p < 0.001).*

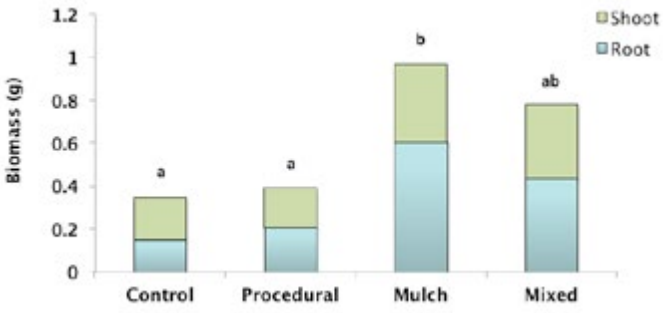


Figure 2b. *Root: shoot of radish across soil treatments. Root: shoot in mulch significantly greater than control treatments. Mixed root: shoot did not differ from other treatments (F = 6.705, df = 3, p < 0.001).*

Acknowledgements

This project was supported by the United States Department of Agriculture. Special thanks to Dr. Abraham and Dr. Johnson, the supporting staff of the California State University Greenhouse, Ed Read and Jarrett Jones, and the staff at the Fullerton Arboretum. Without their support this project would not have been possible. This project was funded by USDA NIFA HSI Grant 2011-38422-30838.

References

Asawalam, D. O., & Hauser, S. (2001). Effects of earthworm casts on crop production. *Tropical Agriculture*, *78*(2), 76-81.

California Department of Food and Agriculture. *CDFA > STATISTICS*. N.p., n.d. Web. 11 Sept. 2013. http://www.cdfa.ca.gov/Statistics/.

California Department of Water Resources. *California Department of Water Resources.* N.p., n.d. Web. 14 Sept. 2013. http://www.water.ca.gov/.

Dahiya, R., Ingwersen, J., & Streck, T. (2007). The effect of mulching and tillage on the water and temperature regimes of a loess soil: Experimental findings and modeling. Soil & Tillage Research, 96(1-2), 52-63. doi: 10.1016/j.still.2007.02.004

Kassam, A., Friedrich, T., Derpsch, R., Lahmar, R., Mrabet, R., Basch, G., . . . Serraj, R. (2012). Conservation agriculture in the dry Mediterranean climate. Field Crops Research, 132, 7-17. doi: 10.1016/j.fcr.2012.02.023

Norgrove, L., & Hauser, S. (1999). Effect of earthworm surface casts upon maize growth. Pedobiologia, 43(6), 720-723.

Rawls, W., Pachepsky, Y., Ritchie, J., Sobecki, T., & Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. Geoderma, 116(1-2), 61-62.

Shemdoe, R. S., Van Damme, P., & Kikula, I. S. (2009). Increasing crop yield in water scarce environments using locally available materials: An experience from semi-arid areas in Mpwapwa District, central Tanzania. Agricultural Water Management, 96(6), 963-968. doi: 10.1016/j.agwat.2009.01.007

World Water Assessment Programme I United Nations Educational, Scientific and Cultural Organization. World Water Assessment Programme I United Nations Educational, Scientific and Cultural Organization. N.p., n.d. Web. 19 Sept. 2013. http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/.

Effects of Sorted and Unsorted Food Waste Diets on Worm Cast Quality and Quantity

Department of Biological Science, California State University, Fullerton

Calvin Lung
Advisor: **Dr. Joel K. Abraham**

Abstract

Food waste has steadily increased over the past few decades, spurring the adoption of various waste diversion practices. One form of waste diversion gaining in popularity is vermicomposting, the use of worms to transform food waste into castings. Worm castings are an effective fertilizer. However, since food waste varies among households and over time, it is possible that the productivity of vermicompost systems, or the quality of the castings produced, are variable as well. The worms used in vermicompost systems respond poorly to some food inputs, such as citrus, onions, or meats, so the inclusion of those foods may negatively affect the productivity of those systems. However, standardization of food waste streams is time consuming and thus may reduce adoption of vermicomposting. In this study, I investigated whether food quality affects the productivity of vermicompost systems or the nutrient content of castings. I set up two vermicompost treatments (n = 7): sorted and unsorted food inputs. I measured final mass of worm castings, worm biomass, and casting nutrient content. I found no difference in the size of worm populations, casting yield, or casting quality between treatments. The results suggest that sorting food waste has little impact on the productivity of household- scale vermicomposting systems. Given the time investment needed to sort food waste streams and potential benefit of worm castings, the results may lead to higher adoption of vermicomposting as waste diversion practice.

Introduction

Approximately, 34 million tons of food waste goes into landfills each year.[1] However, because landfills are compacted daily, they often lack the oxygen and microbes necessary for waste decomposition.[2] As the demand in landfills increase, the need to convert large areas of land into landfills increases, leading to a decrease in land that can be used for beneficial purposes.

Vermicomposting is an increasingly popular alternative to food waste dispersal in many countries, including the USA, Japan, Australia and Italy.[3] Vermicomposting is the use of worms, most commonly Eisenia fetida, to break down waste into fertilizer (worm castings).[4] In the wild, E. fetida are epigeous, removing organic leaf litter, vegetation and manure.[5] This attribute makes E. fetida an ideal species for use in converting food waste into worm castings.

The castings produced by the worms are valuable as soil amendment. Worm castings have high levels of nitrogen, phosphorus and potassium, and lead to increases in plant productivity when applied to the soil.[3,6] In addition, worm castings can reduce the prevalence or severity of certain pests and diseases, and are associated with increases in beneficial soil microbes.[7–9] Some research shows that amending soil with worm castings helps reduce heavy metal uptake in plants, increase levels of soil enzymes, and increase nutrient retention in the soil.[10] In addition, producing synthetic fertilizer is an extremely energy extensive process.[11] By replacing synthetic fertilizers with worm casts, it decreases the need to consume so much energy. Given these benefits and the growing interest in home vermicomposting systems, it is important to know what factors might influence the quality of the castings produced to better understand the potential benefits.

One factor that could affect casting quality is the quality of food waste used in the system. There is some controversy among avid worm growers about the different types of food that worms can consume.[12] Most organisms have a preferred diet that will maximize growth and productivity. The majority of studies on vermicompost worms focus on optimizing the population size of worms in a bin and the quantity of food worms can consume in a day, while other studies focus on th e ability of waste to maintain a certain biomass of worm.[4,13] Another variable that depends on food waste is the variability in the quality of worm casting.

Vermicomposting is increasingly popular for home use.[3] However, given the variability in waste streams among households and within households over time, there may be variation in the nutritional quality of worm castings as a result. This study will investigate whether or not such variation exists.

Sorting trash to remove potentially harmful components may improve the consistency of quality in worm casts.[4] However, sorting food takes time and energy that may outweigh any potential benefit. Thus, it is important to understand what impact variation in food quality has on worm cast production.

This study focuses on the effects of food sorting on E. fetida population growth, casting production, and casting quality. I hypothesized that sorting food waste leads to larger populations, higher casting production, and higher casting nutrient content.

Methods

Food waste was gathered from the CSU Fullerton Gastronome two times a week, Tuesdays and Thursdays, with food scraps with one of the day's waste saved for Saturday. Waste was gathered and separated into two different categories. One category included unsorted trash while waste from the other included an ideal mix of fruits and vegetables. Although worms are quite robust, there may be some waste, such as meats, grains and fish, which may prevent them from producing optimum castings. As a result, the sorted category only contained food that is found in their ideal diet. Each category was then chopped into small pieces and was placed in its respective tray of soil and worms.[14]

A total of fourteen bins were used. Seven replicates were performed for each treatment and each sample required two trays. The trays were placed in the shade at the Fullerton Greenhouse to prevent desiccation and overheating. Although the age of the worms factor into how much casts are produced on a daily basis, all worms will be obtained from one source so that there will be no significant variation in relative age.[15] The optimal amount of worms for each bin was 1.60 kg worm/m2 and optimal amount of food was provided at 0.75 kg feed/kg-worm/day.4 Food waste was chopped into fine pieces, estimated to be no larger than one cubic centimeter in size. 100 grams of food waste were measured into each bin three times a week. There was at least one day in between feedings in order to allow the worms to breakdown the food.

Biomass and Population Dynamics Sampling

The amount of worm casts produced is based off the amount of food consumed by E. fetida, and as a result, biomass is a good indicator of worm cast production. The biomass also describes how the worms fare in each type of environment. The biomass was measured by filling two 400 mL beakers of substrate from alternate corners of the bin and the worms and eggs were removed. The worms were sprayed with a small amount of water to remove the soil and then blotted with filter paper for the dry weight. Two worm biomass and egg samples were taken every other week for the duration of the study.

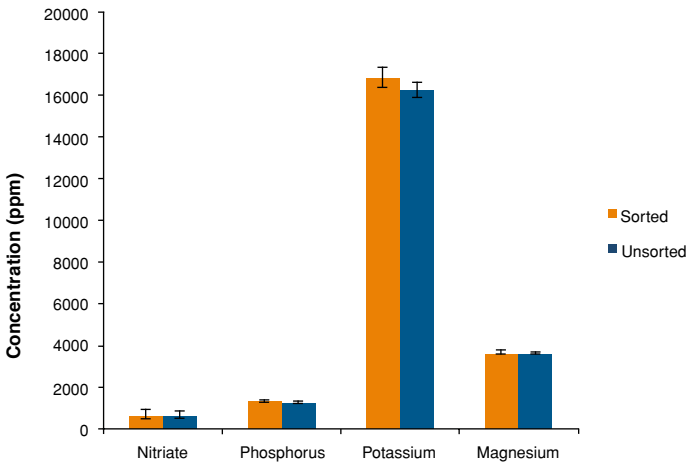


Figure 1. *The average concentration of nitrogen, phosphorus, potassium, and magnesium did not differ between the sorted and unsorted treatments. Orange bars represent the sorted treatment group and the blue represents the unsorted treatment group. Error bars signify standard error (p>0.05).*

Soil Analysis

At the end of the study, one cup of castings was taken from each bin. Each of the randomly selected samples was dried in an incubator for nine days at 35°C. Four randomly selected samples from each treatment were sent to the Soil and Plant Tissue Testing Laboratory at the University of Massachusetts for soil analysis. One sample from the sorted treatment was removed because it was an extreme outlier (order of magnitude difference in nutrient content). Each sample was analyzed for four key nutrients, (Phosphorous, Potassium, Magnesium and Nitrogen), in the samples.

The data were analyzed using StatPlus:mac V5 and Microsoft Excel 2011 for Mac.

Results

There were no differences in nutrient levels (N, P, K, and M) between the two treatment groups (p>0.05) (Figure 1). Over the course of ten weeks, two out of the five measurements taken showed a significant difference in the biomass of the worms (Figure 2). There was no significant difference in total worm biomass between the two treatment groups (P>0.05) (Figure 3). There was no significant difference found in the total mass of castings (p>0.05) (Figure 4).

Discussion

In the two different types of diets tested, there appears to be no significant difference between the two treatments groups, in both quality and quantity of worm casts. All of the nutrients measured, nitrogen, phosphorus, potassium and magnesium levels were insignificantly different when analyzed. The biomass of the worms was also not significantly different. The culmination of these results dictates that there is no reason to separate out food types for worm cast production. The two different types of diet had no effect in any of the variables measured and there were no distinct differences between the two groups.

While there were clearly two different treatment groups in this experiment, this research does not exactly quantify the difference in the two treatment groups. As a result, the similar results obtained in this experiment may be due to insignificant variation or quantity of food. Further research should investigate if large amounts of unsorted food would cause a larger difference in results e.g. changes in microbial community.

As landfills fills up, it is necessary to utilize the space in a more efficient manner. Previously, the main focus on waste divergent strategies were on landfill design and optimizing the amount of decomposition.[2,16] However, another way to tackle this problem is to divert waste from the landfills. Currently, only 24% of municipal solid waste is recycled and composted and vermicomposting helps in minimizing food waste in landfills.[17] In addition, vermicompost has the potential to convert waste into a novel income stream. One of the current obstacles in vermicomposting is the time and energy that was previously thought to be required in managing food waste.

Based on the results of this study, food sorting does not appear to have a measureable impact on casting production or quality. If true, it is likely that the additional effort put into sorting food for household scale vermicompost systems does not yield any benefits. This may increase vermicompost adoption and participation rates of households, thereby reducing load on landfill.[18]

Encouraging sustainability, even on a small scale, plays a larger role by minimizing waste and decreasing a need for synthetic fertilizers. These findings promote the ease and benefits of vermicompost.

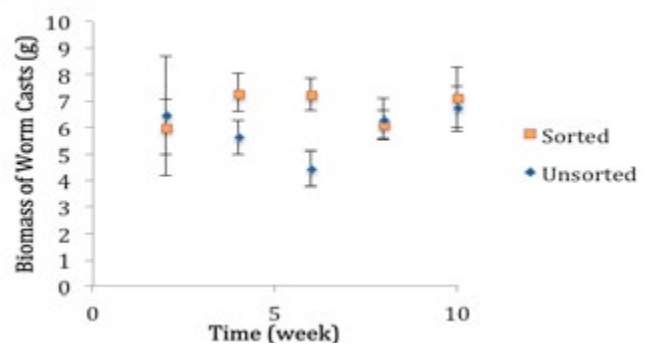


Figure 2. *The average biomass of each sample group over a series of weeks. Data was collected every other week. Error bars signify standard error.*

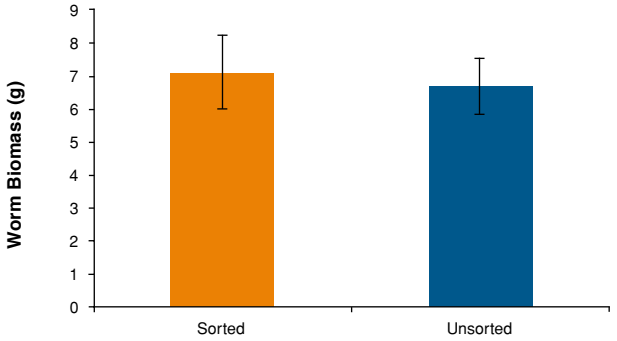


Figure 3. *A graph depicting the difference in average mass of worm casts between the two sample groups. Error bars signify confidence intervals (p>0.05).*

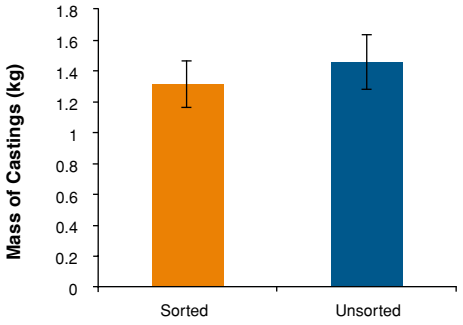


Figure 4. *The average mass of castings collected over a ten-week period. Error bars signify confidence intervals and there were no significant difference between the two treatment groups (p>0.05).*

References

1. Food Recovery Challenge. (2014). at <http://www.epa.gov/smm/foodrecovery/>

2. Read, A. D., Hudgins, M. & Phillips, P. Aerobic landfill for test sustainable cells and waste their disposal implications. Geogr. J. 167, 235–247 (2014).

3. Businelli, M., Perucci, P., Patumi, M., Giusquiani, P. L. & Ash, C. Chemical composition and enzymic activity of some worm casts. Plant Soil 80, 417–422 (1984).

4. Ndegwa, P. M., Thompson, S. A. & Das, K. C. Effects of stocking density and feeding rate on vermicomposting of biosolids. 71, 5–12 (2000).

5. Aira, M., Monroy, F. & Domínguez, J. Microbial Ecology Eisenia fetida (Oligochaeta : Lumbricidae) Modifies the Structure and Physiological Capabilities of Microbial Communities Carbon Mineralization During Vermicomposting Improving of Pig Manure. (2013).

6. Bachman, G. R. & Metzger, J. D. Growth of bedding plants in commercial potting substrate amended with vermicompost. Bioresour. Technol. 99, 3155–61 (2008).

7. Pulleman, M. ., Six, J., Uyl, a., Marinissen, J. C. Y. & Jongmans, a. G. Earthworms and management affect organic matter incorporation and microaggregate formation in agricultural soils. Appl. Soil Ecol. 29, 1–15 (2005).

8. Theunissen, J., Ndakidemi, P. A. & Laubscher, C. P. Potential of vermicompost produced from plant waste on the growth and nutrient status in vegetable production. Int. J. Phys. Sci. 5, 1964–1973 (2010).

9. Nechitaylo, T. Y. et al. Effect of the earthworms Lumbricus terrestris and Aporrectodea caliginosa on bacterial diversity in soil. Microb. Ecol. 59, 574–87 (2010).

10. Sharma, S., Pradhan, K., Satya, S. & Vasudevan, P. Potentiality of Earthworms for Waste Management and in Other Uses – A Review. 1, 4–16 (2005).

12. Vaneeckhaute, C. et al. Closing the nutrient cycle by using bio-digestion waste derivatives as synthetic fertilizer substitutes: A field experiment. Biomass and Bioenergy 55, 175–189 (2013).

13. Frederickson, J., Howell, G. & Hobson, A. M. Effect of pre-composting and vermicomposting on compost characteristics. Eur. J. Soil Biol. 43, S320–S326 (2007).

14. Haimi, J. & Huhta, V. Capacity of various organic residues to support adequate earthworm biomass for vermicomposting. Biol. Fertil. Soils 23–27 (1986).

15. Jain, K., Singh, J. & Gupta, S. K. Development of a modified vermireactor for efficient vermicomposting: a laboratory study. Bioresour. Technol. 90, 335–337 (2003).

16. Mba, C. Biomass and vermicompost production by the earthworm. Rev. Biol. Trop. (1988).

17. Barlaz, M. A. Carbon storage during biodegradation of municipal solid waste components in laboratory-scale landfills. Glob. Biochem. cycles 373–380 (1998).

18. Haaren, R. Van et al. The State of Garbage in America. Biocycle (2010).

19. Brehm, J. W. The Intensity of Motivation. Annu. Rev. Psychol. 109–131 (1989).

Effects Of Surface Drip Irrigation Compared To Sub Surface Irrigation On The Yield Of Peppers

Plants & H2O Lab
Department of Biological Science, California State University, Fullerton

Mirian Morua
Advisor: Dr. Jocken Schenk

Abstract

Conservation of irrigation water is a major concern for urban agriculture in semi-arid regions, such as southern California. A field experiment was carried out in the Fullerton Arboretum February through October 2013 to compare the efficiency of different drip irrigation systems, surface irrigation and subsurface irrigation for growing arbol and poblano peppers. The subsurface irrigation system was buried six inches deep into the soil, while the other was placed on the soil for surface drip irrigation. It was hypothesized that subsurface irrigation would be the most efficient system because of reduced evaporation and water delivery directly to the root zone. The results of this experiment revealed that the highest productivity of peppers was seen for plants under subsurface irrigation. Soil moisture contents for subsurface irrigation varied less over time than those for surface irrigation. This suggests that there was less evaporation and more water retention in the soil under subsurface irrigation, which benefited the plant growth and productivity. Lastly, stomatal conductance for arbol peppers was higher under subsurface irrigation, indicating lower water stress. Thus, it was concluded that subsurface irrigation is a better system for conserving water and maintaining high pepper yields.

Introduction

Southern California is suffering from limited water supply due to high evaporation and low precipitation coupled with frequent droughts (McDonald, 2007). It is estimated that half of the water used in southern California is for irrigation (Baum, 2012). It is necessary to optimize water usage especially in agricultural applications in order to carry the growing populations. Moreover, there are environmental concerns related to irrigation including depletion of water source and soil erosion. Therefore, it is crucial to present a low cost, efficient, and resourceful irrigation model that can be used and applied in urban settings such as residential homes, school gardens, or agriculture resource centers. Appropriate irrigation models should be researched to demonstrate to the general public how easy it is to practice efficient water management and conservation.

About 39% of all fresh water in the United States goes to irrigation; of which mostly is used for Furrow irrigation. Furrow irrigation is irrigation by water run in furrows between crop rows. Most farmers practice furrow because it can be recycled back to a major source of water to minimize water waste. However, furrow irrigation tends to not be uniform and labor is wasted in forming the canals, furrow ends, and borders. The lack of uniformity directly affects the distribution of water irrigation because it can result in drainage and increase of salinity in the soil. Since the 1990's, alternative methods of deficit irrigation have been used such as micro-jet irrigation, drip irrigation, and subsurface irrigation. English et al. (1990) urged the usage of deficit irrigation in order to optimize strategies to sustain water deficit and reduction of productivity yield. Furthermore, they suggested that increasing profits and crop production could be achieved if deficit irrigation was used because irrigation is limited within vegetative and late ripening stages in order to optimize water use efficiency.

Snyder et al. (2008) surveyed the irrigation methods used in California in 2001. The survey was mailed to 10,000 growers in California that were randomly selected from the California Department of Food and Agriculture database. This study concluded that there was an increase of 15% to 31% of orchards mostly irrigated by low volume irrigation such as drip and micro-sprinkler irrigation due to an increase of 33% from 1972 to 2002. There was also a 31% increase of the use of surface irrigation methods in land use. The authors noted that sprinkler usage has decreased in orchards and vineyards but still used in vegetable crops.

Based on Snyder et al. it was suggested that the most popular irrigation method being used for land use was drip, surface, and

micro-jet irrigation. There is a consensus that sprinkler systems are no longer used in orchard agriculture but may be used for vegetable crops. Additional research must be made in order to assess which irrigation system is more efficient in terms of productivity and water conservation. For the purposes of this study, surface drip irrigation (SDI) and subsurface irrigation (SSDI) are the two methods that are investigated.

SDI is commonly used for more than one land use purpose; it can be used with orchards and vegetable crops. Surface drip irrigation involves dripping of water slowly to the roots of plants on the soil surface through emitters that control water flow. SDI is the network of drip tape on the surface that carries a low flow of water under low pressure to plants. It is adaptable and changeable because it can be removed and manipulated to fit any crop. Another advantage is that it minimizes water loss through runoff.

SSDI is the application of drip line polyethylene tubing with built in emitters located beneath the soil to pump water under low pressure. SSDI involves temporarily buried drip tape located below or at the plant roots. It involves water delivery directly to the root systems of the plants; however, it is more permanent because perforated pipes are installed beneath bed to control water flow and drainage. As a result it expected to have no water loss due to evaporation, run off, or wind drift.

The purpose of this experiment was to compare these two irrigation methods in order to investigate a difference between each mechanism and access water management and productivity yields. The efficiency will be measured through productivity yield, stomatal conductance, and soil moisture retention. It was hypothesized that subsurface irrigation would be the most efficient in conserving water (soil retention) and have highest productivity yield.

Methods

Expirmental Site

The field experiment was carried out at the Fullerton Arboretum in California State University, Fullerton. To investigate the efficiency of different irrigation systems, the techline EZ (12 mm dripper line) was used under different irrigation conditions. The irrigation water source was obtained from the Arboretum's water supply.

Expirmental Procedure

Pepper seeds were ordered from the Chile Pepper Research Institute of the University of New Mexico. Germination of seeds was done using two multi-block sewing containers with 200 cells, used to plant 150 seeds of each pepper variety. Each cell was filled with starting compressed soil, sunshine#1 which contained important nutrients (i.e nitrogen) purchased from McConkeyCo in Anaheim. One seed was added per cell to optimize growth of plants. Cells were kept under fluorescent light for 24 hours at room temperature and watered twice a week. Germination was seen in early February and the seed was transplanted into larger pots in early March (Fig A, Appendix). These pots were kept inside the greenhouse until they began to mature. At the end of March, the pots were moved to an open greenhouse where they were allowed to acclimate to higher temperatures and sunlight (Figure b, Appendix). Also, the plants were watered with soluble fertilizer bi-weekly to provide them with nutrients. In early April, the individual pots (200+) were moved outside the greenhouse to acclimate to stronger sunlight, wind, and higher temperatures. Peppers (Capsicum annuum, two varieties of peppers: chile de arbol and poblano) were transplanted at the end of May in the Fullerton Arboretum farm site. The area being used was 30ft by 30ft and contained 11 prepared soil beds. The distance between the rows was 27 inches apart and treatments were divided by rows. A total of 5 rows were used for each treatment as replicates for the irrigation method. For example, 5 rows had the same set up that displayed the dripper line on the surface of the bed. The last five rows obtained the dripper line 6 inches deep into the bed to display the underground subsurface treatment.

Expirmental Treatment

The experiment consisted of two irrigation methods, SDI and SSDI on two different types of peppers. Each treatment was replicated four times with random distribution of pepper plants on each bed (n=20 count). Each replicate or row measured 30 feet in length and 15 inches in width. Both treatments were equipped with separate pipelines connecting to separate pressure regulators and pipelines that connected to the main water source. Each row was irrigated by a single lateral pipeline connected to the main water source. The techline EZ was placed in the surface for the surface treatment and buried underground for the subsurface treatment (Figure C, Appendix). The techline EZ was one quarter inch thin and was placed on the bed 12 inches apart. Drip irrigation pressure regulators were required in order to ensure the operating water pressure rating for drip components does not exceed manufacturers recommended operating pressure. There were two pressure regulators used to separate the treatments at 45 psi. The same amount of water was

delivered twice a week in the summer and once a week in the fall time. The total amount of water delivered was 12 gallons.

Measurements

Stomatal Conductance (mmol/m2s) was collected using a steady state leaf porometer(Model SC-1,Decagon Devices Inc). Measurements were collected once a week for 10 plants for each replicate (*n*=5) from July to September. Soil moisture content (pct) was collected by placing soil moisture probes (EC-10, Decagon Devices Inc.) at two inches in depth for surface treatment, and 6.5 for subsurface treatment. Probes were installed upstream, middle, and downstream of each replicate's' bed for each treatment: upstream described the probe closer to the balb valve, the middle probe was placed to describe the moisture in the middle of the bed. While the downstream probe described the moisture at the end of the row near the kink, where flushing occurred. Measurements were done weekly to measure percent of water content (pct)(cm/m) in the soil to assess the amount of water retained after irrigation (Model ECH2O, Decagon Devices Inc.). Lastly, productivity was measured using the Salter Brecknell scale (capacity 5 lbs). Water stress was evaluated based on the visible indications of reduced growth, delayed maturity, and reduced crop yield per treatment replicates.

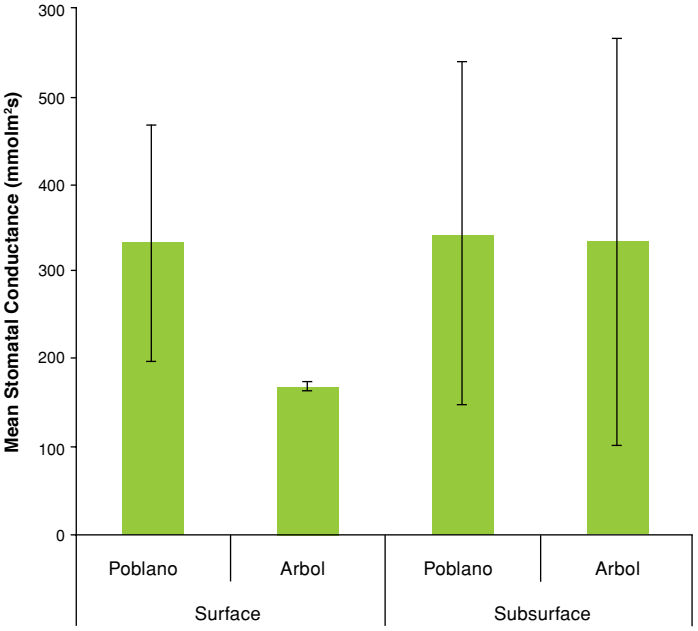


Figure 1: *Mean stomatal conductance for arbol peppers was higher under subsurface irrigation, indicating lower water stress (m=333.1 mmol/m²s) compared to the surface irrigation value. Statistical analysis indicates significant difference between arbol peppers (p=0.033). 95% confidence intervals were established as error bars.*

Results

For three months the pepper plants were allowed to grow and mature at the Fullerton Arboretum site. Starting in July when they were first transplanted, plants were maintained and watered at the same time and consistency. Based on weather reports from the Fullerton Municipal Airport weather station, the average humidity in Fullerton, California for the months of July through October were ranging from 43 % to 93% humid. October reached very humid conditions rising to almost 88%. During the treatment trials, weather was very dry and at hot temperatures thus plants were under very dry conditions.

Stomatal Conductance:

The mean for each replicate stomatal conductance per pepper variety was obtained as a direct reflection for each treatment, SDI and SSDI. A high conductance mean indicated that the stoma for each pepper variety was open more frequently for the subsurface treatment. Therefore, there was a higher stomatal conductance observed for the subsurface treatment compared to the surface treatment for both arbol (mean=333.11 mmol/m2s) and poblano peppers (mean=341.849 mmol/m2s). A t-test indicated that the values were not significantly different when comparing the poblano peppers of each treatment (surface vs. subsurface) (p=0.35). However, there was a significant difference between the arbol peppers of each treatment for stomatal conductance (p=0.0325).

Soil Moisture:

The results demonstrated that soil moisture retention was observed to be more consistent in the SSDI treatment because there was a trend that showed similar soil moisture retention (6in depth) between each replicate at upstream, middle, and downstream placements (Figure 2). When comparing soil moisture retention in the SDI (2in depth) there was no obvious trend observed upstream, middle or downstream in the mean replicate values (Figure 3). ANOVA demonstrated there was a significant difference between all the soil moisture groups for surface and subsurface values (p=0.011). However, the post hoc of multiple comparisons for each group at each treatment was not significantly different (p>.05).

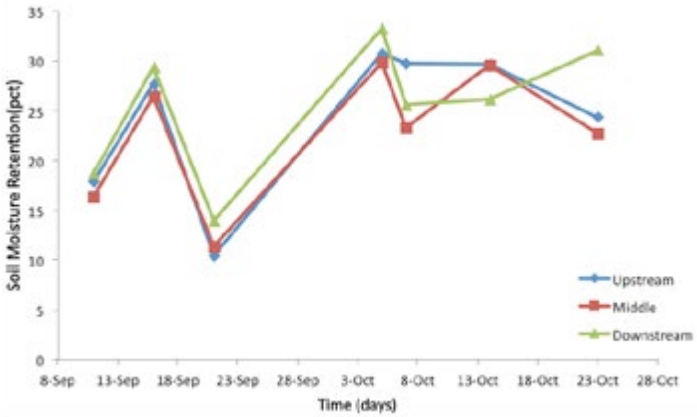


Figure 2: *Soil moisture (pct) was averaged for each row to represent the surface treatment, upstream (blue), middle (red), and downstream (green) of the row. There is no obvious trend observed, thus no indication of similar soil moisture retention throughout the treatment. ANOVA indicates there is a significant difference between all groups in surface and subsurface sections (p=0.011).*

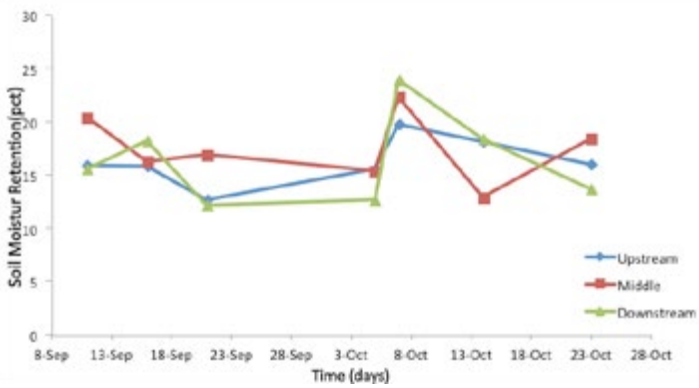


Figure 3: *The soil moisture mean was obtained for all four replicates to represent the subsurface treatment, upstream (blue), middle (red), and downstream (green) of the rows. There is an obvious trend that indicates similar water distribution. ANOVA demonstrated that there was a significant difference between all groups for surface and subsurface irrigation (p=0.011).*

Productivity & Qualitative Analysis:

Throughout the experiment, productivity was witnessed by observing physical stress of the plant, maturity, and relative growth of the plants. When comparing methods, the surface irrigation method, which involved the dripper line on the surface, had more flower buds compared to the subsurface (Appendix, Figure D). In addition, surface had more arbol pepper count and stronger pepper coloration compared to the subsurface. Overall, the peppers under surface irrigation seemed to be doing better compared to the subsurface method. However, results indicated differently because overall pepper productivity was continuously higher under subsurface irrigation (224 lbs) compared to the surface irrigation (217 lbs) method throughout time (Figure 4). A t-test was used to compare the end productivity of each treatment where it showed no significant difference between them (p=0.33).

The goal of this experiment was to investigate which irrigation method was better in conserving water while maintaining productivity. It was hypothesized that the subsurface irrigation would be a better choice for water conservation because the irrigation line was underground leading to reduced evaporation and higher producitivity because water was delivered under low pressure

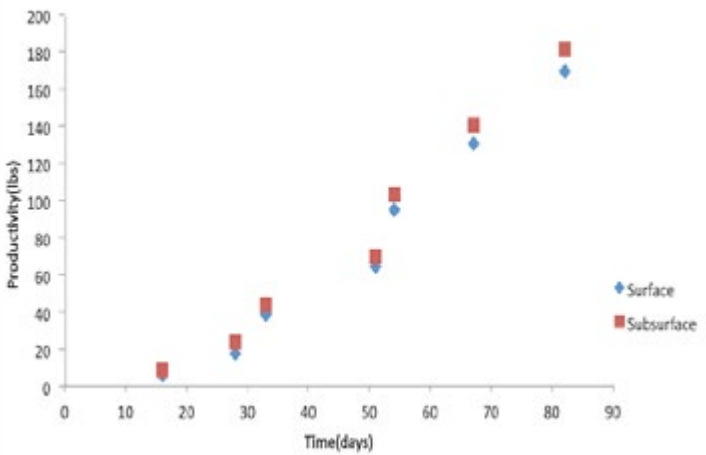


Figure 4: *Productivity was measured in pounds from July to September (t=90 days). Subsurface (224lbs) had a higher productivity over time compared to the surface (217lbs) treatment. Statistical analysis from t-test between groups showed no significant difference (p=0.33).*

directly to the roots. Although surface irrigation also delivered water under low pressure directly, risks of evaporation and run off had a higher chance to occur. However, it was an interesting method to investigate because it required less labor, better accessibility, and water was still delivered at a close proximity to the plant.

Based on the results obtained from this study, it is recommended that subsurface irrigation is the better system to use when insufficient water is available because there was a more uniform soil moisture

retention within this system. The results demonstrated an obvious trend of soil moisture content (pct) throughout each replicate and section of the subsurface system treatment, however there was no obvious trend to indicate water was delivered uniformly throughout in the surface treatment. Therefore, this suggests that subsurface irrigation allowed more water to be retained within the root systems of the plants. This ensures the plant to have water available under drought conditions to prevent water stress, ensure growth and maturity, and provide higher productivity yields. The subsurface dripper line was buried 6.5 inches below the soil surface, and root systems were rooted in deep close to the dripper line. At this depth, soil moisture retention was higher because of minimum evaporation loss with this method. It was expected to find a higher percent water content downstream of the row because that was the area where water was flushed out of the irrigation system. These observations were also reflected on the roots of the plants of each system when observed. The root system for subsurface were deep rooted compared to shallow roots in the surface method.

Results compare to previous research findings because where they also found higher soil moisture content in subsurface irrigation. Kheira and Abdrabbo (2009) investigated different decifit irrigation systems of corn in the nile river. The authors investigated three systems of irrigations such as furrow irrigation using gated pipes technique, subsurface drip irrigation system, and surface drip irrigation. They concluded that subsurface irrigation was the best option for drought conditions because it displayed higher efficiency compared to the other irrigation systems (Kheira and Abdraddo, 2009). They compared the soil moisture distribution before irrigation to find higher soil moisture retention at 60 cm compared to the other treatments (Kheira and Abdraddo, 2009). Moreover, they discovered that the distribution of water was more uniform for all treatments in the subsurface for the corn irrigated with the subsurface treatment (Kheira and Abdraddo, 2009). Therefore, subsurface irrigation is a better irrigation system because water is conserved more efficiently and thus a better system for high drought areas like southern California.

Due to more water being available for the plant's subsurface root system, it was expected to find that productivity was higher in plants under subsurface irrigation compared to the plants under surface dripper irrigation. There was a higher productivity of pepper yield throughout time in the subsurface system. This suggests that having higher soil retention near root systems ensures plants to grow, and thus reflects productivity of the plant. Although, at times pepper plants under subsurface irrigation seemed to wither and break faster, it was due to the faster growth and higher productivity of the plants. Thus, it was concluded that the subsurface irrigation system is a better method to use because it ensures conservation of water while promoting high productivity yields.

In addition, stomatal conductance or the state of the stoma was observed in order to see how frequently the stomas were open. Results indicated that stomas were opened more frequently in plants under subsurface irrigation treatment. Having plants with their stomas open suggests that plants are exchanging gases with the environment at a higher rate. Since stomatal opening and closing is a result of changes in the turgor of guard cells and epidermal cells, stomtatal state to reduced leaf water can be assumed. Tardieu and Davies (!993), demonstrated that stomatal conductance is regulated by leaf water status of plants in drying soil. That the opening of stoma is controlled by chemical messages from dehydrating roots (Tardieu and Davies, 1993). The authors results agree with our investigation of stomatal conductance to suggest that the reason stomatal conductance was lower in surface irrigation system was because the roots may have been dehydrated due to a dryer soil. These findings agree with the random uniformity of the soil moisture content (Figure 2&3).

It was concluded that subsurface irrigation is a more efficient irrigation system compared to surface drip irrigation because it conserved water while maintaining high productivity of pepper yield. Based on these findings, it is recommended for areas with high drought such as in southern California to consider using subsurface irrigation because it ensures plant growth and minimum evaporation of water.

SSDI is a good method to optimize water use because water can be conserved, supported by our findings of soil moisture retention. Although this system does require more labor than surface irrigation, it makes a larger impact on water conservation and productivity. Future studies should include measuring the resistance and transpiration rate per plant variet under irrigation methods which would provide more direct information on the plant exchanges of gases. Measurements of leaf water potential and root water potential can also be done to obtain information on water flow between the plants. Improvement on this experimental design can be done by using the same pepper variety on each sub plot and not randomizing the peppers to make a true replicate.

References

Baum-Haley (2012) Water Use Efficiency Programs Specialist Municipal Water District of Orange County

English, M. J., J. T. Musick, and V. V. N. Murty. 1990. Deficit irrigation. Ch. 17. In Management of Farm Irrigation Systems, (Eds.) ASAE.

Kheira, A. and Abdrabbo, A. 2003. Comparison among Different Irrigation Systems for Deficit Irrigated Corn in the Nile Valley. Agricultural Engineering International the CIGR Ejournal (5- 22).

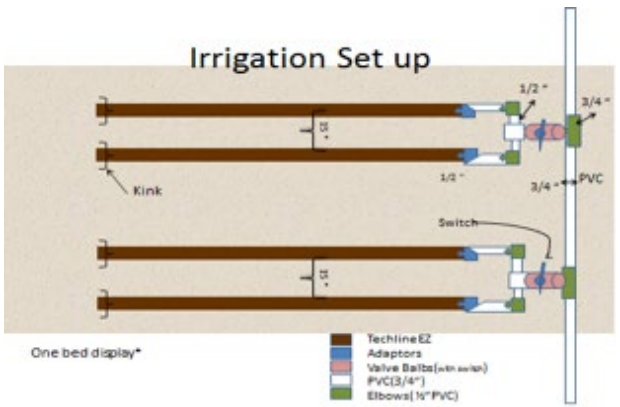
MacDonald, G.M., 2007. Severe and sustained drought in southern California and the West: present conditions and insights from the past on causes and impacts. Quaternary International 173–174, 87–100.

Snyder, R.L. Orang, M.N. and Matyac, J.S. 2008. Survey of irrigation methods in California in 2001. Journal of Irrigation and Drainage Engineering. 134.1 (96-100).

Tardieu, J. Davies, W. 1993. Integration on of hydraulic and chemical signaling in the control of stomatal conductance and water status of droughted plants. Journal of plant, cell and Environment. 16:4

Appendix

A: *Germination and growth for peppers in Biology Greenhouse under fluorescent light for 24 hours.*



C: *The irrigation was set up according to the image (right) which included the drip line, adaptors, balb valve, pressure regulator, and PVC*

B: *Growth for peppers in biology greenhouse for acclimation to the sun before transplantation to the field.*

D: *Demonstrates productivity growth of peppers from surface to subsurface (left to right) indicates differences in coloration and reveals overall growth.*

Morphological and Genetic Identification of California Pipefishes (Syngnathidae)

Department of Biological Science, California State University, Fullerton

C.A. Rice, D.J. Eernisse, and K.L. Forsgren

Abstract

California pipefishes are highly diverse and particularly difficult to consistently identify. The widely used Miller and Lea Guide To The Coastal Fishes of California does not include all of the recognized Californian province species. We reexamined the key morphological traits emphasized in pipefish keys seeking to improve the diagnostic separation of pipefish species. Our data indicate that certain features, such as barred markings or a truncated snout, are reliable in separating three recognized species in California; kelp pipefish (*Syngnathus californiensis*), barred pipefish (*S. auliscus*), and snubnose pipefish (*Cosmocampus arctus*). In contrast, a combination of morphological and mitochondrial 16S and COI analyses have so far not supported three currently recognized species as distinct: the bay pipefish (*Syngnathus leptorhynchus*), and barcheek pipefish (*S. exilis*), and all could be synonymous with *S. californiensis*. Future work will further test these conclusions and our goal is to produce a more useful dichotomous key to California pipefishes. We expect to add more pipefish localities and species sequence comparisons and extend what we learn to also better characterize the identification of juveniles. The results from this study will be beneficial to fishery biologists working in the field to more effectively identify pipefishes.

Investigating Defense Responses of *Nicotiana benthamiana* Involving the 14-3-3 Gene Family Using Virus-induced Gene Silencing (VIGS)

Department of Biological Science, California State University, Fullerton

Jennifer Spencer

Introduction

Plants possess multiple tiers of immunity that act to protect the plant cells against various types of invading pathogens including bacteria, viruses, fungi, nematodes, aphids and oomycetes (Jones and Dangl, 2006). The first line of defense that provides plants with protection involves physical and chemical barriers to protect against infection. These types of barriers include the cell wall and waxy cuticle layer covered in anti-microbial compounds that surround the cell (Dangl and Jones, 2001).

If pathogens are able to breach this first line of defense, the plants innate, or basal, immune system responds through recognition of conserved molecules associated with the pathogen known as pathogen-associated molecular patterns (PAMPs) using transmembrane pattern recognition receptors (PRRs) located on the cell surface (Boller and He, 2009). PAMPs are essential to the ultimate survival of the pathogen, such as bacterial flagellin or viral double- stranded RNA (dsRNA), and are therefore constitutively expressed and not easily altered (Akira et al., 2006; Jones and Dangl, 2006). These unchanging characteristics of PAMPs allow their continued recognition by plants. Recognition of PAMPs alerts the plant to the presence of pathogens and leads to PAMP-triggered immunity (PTI), which acts to arrest pathogen colonization in the plant (Akira et al., 2006; Jones and Dangl, 2006).

Pathogens have evolved methods of counteracting PTI through proteins, commonly referred to as effectors. Bacterial pathogens inject effectors directly into plant cells through a type III secretion system (T_3SS) while other pathogens use alternate methods (Salmond and Reeves, 1993). Effectors have been shown to target PRRs and negatively affect downstream signaling or host vesicle transport in plant cells, leading to heightened virulence of the pathogen and interfering with plant PTI to cause effector-triggered susceptibility (ETS) (Jones and Dangl, 2006).

To protect themselves from infection caused by pathogen injected effectors and ETS, plants have evolved R genes that produce resistance – or R proteins. R proteins are typically classified by their conserved domains, specifically the nucleotide binding (NB) and leucine rich repeat (LRR) domains with some R proteins containing a coiled-coil (CC) domain or toll/interleukin-1 receptor (TIR) cytoplasmic domain at their N terminus (Jones and Dangl, 2006). According to the guard hypothesis, R proteins indirectly recognize pathogen avirulence (Avr) proteins, termed effectors, by monitoring effectors cellular targets in order to activate and elicit an immune response to counteract their effects (Jones and Dangl, 2006). The concept known as the gene-for-gene hypothesis is based on the idea that disease resistance in plants is mediated through two complementary genes including the *Avr* pathogenic gene and the host plant R gene (Van Der Biezen and Jones, 1998). Through the elicitation of a localized programmed cell death (PCD), commonly known as a hypersensitive response (HR), the plant prevents the spread of infection to healthy cells (Jones and Dangl, 2006). Recognition of Avr proteins by R proteins to trigger HR is known as effector-triggered immunity (ETI) (Dangl and Jones, 2001).

The 14-3-3 gene family is highly conserved across eukaryotes and encodes multifunctional proteins acting in a range of cellular regulatory processes. 14-3-3 proteins have been previously shown to serve as linkers and adapters between sensing and activation in cell signaling through dimerization or by binding to functionally diverse proteins that influence myriad cellular processes (Denison et al., 2011). In plant species, 14-3-3 genes have been shown to act in response to biotic stress in addition to abiotic stress, growth and division, metabolism, hormone pathways and responses to light (Denison et al., 2011). Several 14-3-3 proteins have been implicated in disease resistance responses to pathogens that act as biotic stress. In the case of the N protein recognizing Tobacco Mosaic Virus in *Nicotiana tabacum*, 14-3-3 isoforms interact with the TIR cytoplasmic

domain, NBS and LRR domains of this R protein (Konagaya et al., 2004). Species of Arabidopsis have also demonstrated disease resistance caused by 14-3-3 genes in the case of the powdery mildew fungal infection. Low expression of 14-3-3s have been associated with susceptibility and decreased resistance to powdery mildew fungus, while over- expression of 14-3-3 genes elicit HR potentially due to an interaction between 14-3-3 proteins and the R protein RPW8.2 in order to increase resistance (Denison et al., 2011).

Within *Solanum lycopersicum* (tomato), 14-3-3 gene products have been shown to act as a receptor for the fungal toxin Fusicoccin (FC), and treatments with FC induced higher levels of 14-3-3 proteins (Moorhead et al., 1996). The mitogen-activated protein kinase (MAPK) pathway has been shown to initiate a PCD or HR in tomato carrying the R protein Pto when infected with the pathogen *Pseudomonas syringae* (Oh et al., 2010); it was hypothesized that 14-3-3 proteins play a role in the MAPK pathway by stabilizing the MAPKKKα protein and activating down- stream signaling cascades that lead to PCD in plant tissues (Oh et al., 2010). However, it was demonstrated that PCD induction was not dependent on 14-3-3 protein interactions with MAPKKKα, but may help to bring MAPK and MAPKKKα together in the cell and aid in their overall interaction (Oh et al., 2010). Although some 14-3-3 proteins have demonstrated a possible role in disease resistance to fungal or bacterial pathogens in tomato, the 14-3-3 family of proteins has not been studied intensively in the context of resistance to the full spectrum of plant pathogens, including viruses.

Previously, an interaction between *N. benthamiana* 14-3-3 homologs and the Tm2-2 resistance protein from *S. lycopersicum* was found biochemically (Sobhanian, 2011). Tm2-2 confers resistance to the Tobacco mosaic virus (TMV) 30K movement protein in tomato species. The function of 14-3-3 genes in plant immunity through Tm2-2 will be investigated through silencing of endogenous *N. benthamiana* 14-3-3 homologs in a transgenic line expressing immune receptor Tm2-2. *N. benthamiana* was chosen specifically for this experiment as the genome was recently sequenced, it is thought to possess the same number of isoforms as tomato and has shown high efficiency in gene silencing using the virus-induced gene silencing (VIGS) method. Additionally, the Tm2-2 gene is functional within *N. benthamiana* because it is in the same family as *S. lycopersicum.*

To better understand what the overall function of a specific gene may be within the cell, the gene can be turned off and its function muted to observe phenotypic differences in its absence. In plants, gene expression can be blocked by the use of VIGS. VIGS uses the mechanism of posttranscriptional gene silencing (PTGS), an endogenous defense mechanism that has evolved to protect cells against infection by targeting foreign RNAs for degradation (Purkayastha and Dasgupta, 2009). In PTGS, the cell recognizes foreign double-stranded RNAs (dsRNAs) through the Dicer protein, which cleaves the dsRNAs into small interfering RNAs (siRNAs) that are subsequently loaded into the RNA-induced silencing complex (RISC) (Purkayastha and Dasgupta, 2009). RISC is able to recognize foreign RNAs using siRNA complementarity to other RNA molecules and targets them for degradation by cleaving the base- paired region. Using VIGS, a fragment of the plant gene of interest is cloned into viral vectors to become equivalent to foreign RNA when the recombinant viruses are used to infect plant tissue. The RISC complex targets RNAs that complementary base-pair with the recombinant virus-derived siRNAs, including both viral RNAs and endogenous mRNAs encoding the gene of interest produced by the cell (Purkayastha and Dasgupta, 2009). RISC will recognize the specified gene transcripts as foreign even if they are normally found endogenously in the cell and cleave them, causing their elimination. Inhibition of gene expression through the destruction of targeted mRNAs is sometimes referred to as gene knockdown since expression is greatly reduced but not eliminated (as in a gene knockout) (Purkayastha and Dasgupta, 2009). Although VIGS has been attempted in other plant species, such as tomato, it has been highly inefficient in silencing targeted genes. However, *Nicotiana benthamiana* has shown notably efficient gene knockdown using VIGS, making it a model plant to use this type of silencing to study gene function (Velásquez et al., 2009).

Preceding VIGS, plasmid constructs carrying the 14-3-3 genes need to be created and validated to ensure the correct genes will be silenced. This report will focus on the phylogeny of *S. lycopersicum* and *N. benthamiana* 14-3-3 orthologs, amplification and creation of the 14-3-3 homolog set and the subsequent verification of their cloning. The set of *N. benthamiana* 14-3-3 homologs will be used to create a silencing library for gene knockdown using VIGS. By investigating potential roles of 14-3-3 genes in resistance to pathogens within *N. benthamiana*, we will be able to ultimately infer about functions of 14-3-3 genes in future experiments.

Materials and Methods

RNA Extraction

Total RNA from Nicotiana benthamiana was isolated by RNA extraction from leaves as outlined in the protocols of the RNeasy Plant Mini Kit (Qiagen). Briefly, leaf disks weighing an approximate total of 100 mg were punched out of leaves, flash frozen in liquid nitrogen and ground with a mortar. The frozen leaf powder was homogenized in 450 µl buffer RLT containing beta-mercaptoethanol by vortexing before being transferred to a spin column collection tube. The spin column was centrifuged for 2 minutes at 10,000 rpm and the flow-through was transferred to a clean tube. Absolute ethanol (0.5 volumes of the leaf lysate) was added and mixed by pipetting. The lysates were transferred to the new spin column, centrifuged for 15 seconds at 10,000 rpm and the flow-through discarded. The spin column was washed once with 700 µl buffer RWI and twice with 500 µl buffer RPE with centrifugation for 15 seconds at 10,000 rpm, followed by for a drying spin of 2 minutes at 10,000 rpm to remove residual wash solution. RNA was eluted from the column in 50 µl RNase-free water. The RNA concentration in ng/µl and 260 nm/280 nm ratio were determined using a NanoDrop ND-1000 spectrophotometer.

Primer Design

Forward and reverse primers were designed for ten 14-3-3 isoform sequences from Nicotiana benthamiana using the Primer 3 function from the JustBio online analysis tools (www.justbio.com) and incorporated restriction enzyme cut sites into their 5′ ends.

RT-PCR

To create cDNA from the extracted RNA, reverse transcription-polymerase chain reaction (RT-PCR) was performed using random hexamer primers and the SuperScript III Reverse Transcriptase Kit (Invitrogen). The combination of 8 µl RNA, 1 µl random hexamers (50 ng/µl) and 1 µl dNTP mix (10 mM) was incubated at 65°C for 5 minutes and chilled on ice for 1 minute. cDNA synthesis mix (2 µl 10X RT buffer, 4 µl 25 mM $MgCl_2$, 3 µl 0.1 M DTT, 1 µl 40 U/µl RNaseOUT and 1 µl 200 U/µl SuperScript III RT) was added to the RNA/primer mixture and incubated for 10 minutes at 25°C, followed by 50 minutes at 50°C. The reverse transcription reaction was terminated by heating at 85°C for 5 minutes and then chilling on ice. RNaseH (1 unit) was used to digest RNA template at 37°C for 20 minutes.

PCR was performed to amplify the targeted 14-3-3 sequences using the MJ Mini BioRad thermocycler and Herculase enzyme. The cDNA was mixed with 10 µl 5X Herculase II reaction buffer, 34.5 µl nanopure water, 0.5 µl 40 mM dNTP mix, 1.25 µl 10 µM forward primer, 1.25 µl 10 µM reverse primer, 1 µl DMSO and 0.5 µl Herculase II fusion DNA polymerase and ran in the thermocycler with the following program: denaturation of the DNA for 2 minutes at 98°C, 40 cycles of 30 seconds at 98°C, 30 seconds at 48°C and 1 minute at 72°C, followed by 3 minutes at 72°C.

Gel electrophoresis was performed at 125V using 0.8% agarose gels in 1X TBE buffer with 1% ethidium bromide (Fisher BioReagents) for visualization. The 100bp Low Scale DNA Ladder (Fisher Scientific) was used for size estimation of PCR products. Pictures were taken with the Gel Logic 100 Imaging System using an ultraviolet transilluminator. DNA bands were excised with a razor blade and extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen). The gel slices were weighed and melted in 3 volumes buffer QG at 50°C for 10 minutes. After mixing with 1 gel volume of isopropanol, the dissolved gel was transferred to a spin column, centrifuged for 1 minute at 10,000 rpm, and the flow-through discarded. This process was repeated once with 500 µl QG buffer and twice with 700 µl PE wash buffer before centrifuging an additional minute at 13,000 rpm and transferring the top of the spin column to a clean microcentrifuge tube. Lastly, DNA was eluted from the column in 50 µl EB buffer and stored at 4°C.

14-3-3	Sequence (5' to 3')	Restriction Sites	Tm (50mM NaCl) °C
NbFTT1 For	CCTAGGGCCCTGATTGGGT GAGTAGGAG	AvrII, ApaI	64.63103202
NbFTT1 Rev	CTGCAGGGATCCATGTGAT GGGATTTTGAAC	PstI, BamHI	62.51319076
NbFTT2 For	CTGCAGGGATCCCTGATTG GGTGAGTAGGAG	PstI, BamHI	65.48786401
NbFTT2 Rev	GAATTCGGGCCCATGTGAT GGGATTTTGAAC	EcoRI, ApaI	62.97302754
NbFTT3 For	GAATTCGGATCCCCTTAGA ATCACCAGATGC	EcoRI, BamHI	61.65791361
NbFTT3 Rev	CCTAGGGTCGACGTTGAAT TCATGGAAAAGG	AvrII, SalI	61.92139452
NbFTT4 For	CCTAGGGTCGACTTCTAGG AGAGCATCTTGG	AvrII, SalI	63.3300601
NbFTT4 Rev	GATATCGGATCCCCCAGAA GTCAGAACACTC	EcoRV, BamHI	62.30036992
NbFTT5 For	CCTAGGGTCGACCTGATTG GGTGAGTAGGAG	AvrII, SalI	64.80954167
NbFTT5 Rev	GGGCCCGAATTCATGTGAT GGGATTTTGAAC	ApaI, EcoRI	62.97302754
NbFTT6 For	GGGCCCGTCGACGAGTTCA TGGAGAAAGTCG	ApaI, SalI	67.4815255
NbFTT6 Rev	CTGCAGCTCGAGCTCAAGT ACCTCTTTCAGC	PstI, XhoI	64.64588197
NbFTT7 For	GGGCCCGTCGACAGAAGTT CAATGCAAGTCC	ApaI, SalI	66.78725255
NbFTT7 Rev	CTGCAGCTCGAGAATGAAG ACGGTTGCTAAG	PstI, XhoI	63.38719741
NbFTT8 For	GGGCCCGTCGACGCCTTAT AGGATGTGTTGG	ApaI, SalI	67.66004745
NbFTT8 Rev	CTGCAGGGATCCCGGAATA TCCTACACAACC	PstI, BamHI	64.33584513
NbFTT9 For	GGGCCCGTCGACTCCTCTAT TGAGCAGAAGG	ApaI, SalI	67.05961994
NbFTT9 Rev	CCTAGGGGATCCCGGAATA TCCTACACAACC	AvrII, BamHI	63.65083697
NbFTT10 For	GGGCCCGTCGACGCCTTAT AGGATGTGTTGG	ApaI, SalI	67.66004745
NbFTT10 Rev	CTGCAGGGATCCCGGAATA TCCTACACAACC	PstI, BamHI	64.33584513

Table 1: *Sequences, restriction enzyme sites and Tm values for all 14-3-3 isoform primers.*

Cloning into pGEM-T

In preparation to clone 14-3-3 genes into pGEM-T vectors, an A-tailing reaction was performed to add a single adenine nucleotide to the 5′ ends of the 14-3-3 PCR products. This was done by combining 7 µl of the PCR product from the gel extraction with 1 µl 10X Thermopol buffer, 0.5 µl dATPs, 1 µl Taq DNA Polymerase and 0.5 µl nanopure water (New England BioLabs). The reaction mixture was incubated at 70°C for 30 minutes and then ligated into vector pGEM-T (Promega). The ligation reaction was performed by combining 3.5 µl of the A-tailing product, 5 µl 2X ligation buffer, 1 µl pGEM-T vector and 1 µl T_4 DNA ligase (New England BioLabs) and incubating the mixture at 4°C overnight.

Transformation Into E-Coli

Ligation reactions were transformed into *E. coli* by mixing 50 µl competent *E. coli* cells with 5 µl of the pGEM-T ligation reactions and incubating on ice for 10 minutes. Bacteria were heat shocked in a 37°C water bath for 90 seconds and snap cooled on ice for 1 minute before addition of 450 µl of LB to each tube and recovery at 37°C for 30 minutes. The transformed *E. coli* were spread on LB agar plates containing 100 µg/mL ampicillin and 100 µg/mL tetracycline. Bacterial plates were incubated at 37°C overnight.

A PCR screen for bacteria containing pGEM-T vector with inserts was performed on white colonies using the MJ Mini BioRad thermocycler DreamTaq program. Reactions containing 2 µl each of forward and reverse M13 primers, 6 µl nanopure water and 10 µl 2X Dream Taq polymerase (Thermo Scientific) master mix were set up before picking *E. coli* from the transformation plates, spotting onto a new LB agar plate containing 100 µg/mL ampicillin and 100 µg/mL tetracycline and mixed into the PCR reactions using sterile toothpicks. PCR reactions were performed with an initial denaturation at 98°C for 3 minutes and 40 cycles as follows: 1 minute at 98°C, 15 seconds at 98°C, 30 seconds at 50°C and 72°C for 1 minute, with an additional 5 minutes at 72°C after the cycles ended.

Gel electrophoresis was performed on the PCR screen products as described above. Clones containing inserts with the correct size predicted for the 14-3-3 isoform cDNA fragments were incubated in LB media containing 100 µg/mL ampicillin and 100 µg/mL tetracycline overnight at 37°C with shaking. Glycerol stocks were created by adding 300 µl of 50% glycerol to 700 µl of the bacterial culture and freezing at -80°C. To isolate plasmid DNA, the remaining bacterial cultures containing the pGEM-T 14-3-3 clones were centrifuged at 4,400 rpm for 15 minutes, the supernatant discarded and the pellets were frozen before use.

pGEM-T PLASMID ISOLATION

The isolation of plasmid DNA of pGEM-T clones containing the 14-3-3 cDNA fragments was performed as according to the directions outlined in the QIAprep Spin Miniprep Kit (Qiagen). Briefly, the bacterial pellets were resuspended with 250 µl P1 buffer and transferred to a sterile microcentrifuge tube. 250 µl P2 and 300 µl N_3 buffers were sequentially added to the bacteria, inverting between each addition, and the lysate was centrifuged for 10 minutes at 13,000 rpm. The supernatant was transferred to a spin column, centrifuged for 1 minute, and the flow-through discarded. The spin column was washed with 500 µl PB buffer and twice with 500 µl PE buffer and centrifuged an additional 2 minutes for a dry spin. Plasmid DNA was eluted from the spin columns in 50 µl EB buffer. The NanoDrop ND-1000 spectrophotometer was used to determine the 260 nm/280 nm ratio as well as the DNA concentration in ng/µl.

DNA Sequencing/NCBI Blast Search

The pGEM-T clones were sent to Eton Biosciences Inc., in San Diego, CA for DNA sequencing. Sequence identities and similarities to known 14-3-3 tomato genes (TFTs) were determined using the Basic Local Alignment Search Tool (BLAST) from the National Center of Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov).

Phylogenetic Tree And Alignment Construction

NbFTT and TFT sequences were aligned for phylogenetic analysis using ClustalW2 to create a preliminary sequence alignment for further similarity comparison (Larkin et al., 2007). The initial Clustal alignment in conjunction with Seaview software was subsequently used to construct a Parsimony phylogenetic tree for similarities in the amino acid sequences (Gouy et al., 2010). Lastly, the BoxShade server 3.21 was used to create shaded sequence alignments to better visualize relatedness of the NbFTTs and TFTs (www.ch.embnet.org/software/BOX_form.html).

Results

To investigate potential relatedness of 14-3-3 homologs in *Nicotiana benthamiana* to their previously studied orthologs from *Solanum lycopersicum* (tomato), the sequences were obtained from the Solgenomics Network genomic databases and a phylogenetic analysis of the two families of proteins was conducted (Bombarely et al., 2012). Evolutionary relatedness of *S. lycopersicum* 14-3-3 proteins, termed TFTs, and N. benthamiana 14-3-3 homologs is demonstrated in the phylogenetic tree (Figure 1). *N. benthamiana* 14-3-3 homologs, referred

to as NbFTTs, were given number designations based on the initial Basic Local Alignment Search Tool (BLAST) search of the *N. benthamiana* genome database with the TFT protein sequences; the best matching NbFTT was named based on the TFT used in the search. The oldest node separating the clusters indicates that the orthologs evolved from a single common ancestor forming two main clades. The branching off points between the two groups show that NbFTT and TFT homologs 3, 7, 8 and 10 are most closely related to each other and paired together in the same clade, confirming their named designations as accurate. Interestingly, TFT1 and TFT2 evolved separately from NbFTT1 and NbFTT2 and reside in various clades, stemming from their most recent common ancestor. TFT4 and NbFTT4 are shown to be the most distantly related from each other of the homologs paired by their numerical designations, despite their initial match, and reside in separate clades. Although TFT6 and TFT9 are closely related to their *N. benthamiana* orthologs, NbFTT6 and NbFTT9, the branch distances of the tree show that NbFTT6 and NbFTT9 evolved much later than TFT6 or TFT9.

A boxshade sequence alignment of NbFTT and TFT orthologs further demonstrated the evolutionary relatedness between the two groups (Figure 2). TFTs and NbFTTs show common residues with each other and demonstrate blocks of amino acid similarities distributed at both the N-terminal and C-terminal ends of the proteins. Additionally, a search of the conserved amino acid sequences using the BLAST database shows high conservation among the 14-3-3 gene superfamily as well as multiple peptide binding sites. Further supporting their arrangement in the phylogenetic tree, TFT6, NbFTT6 and TFT5 show groupings of amino acids around amino acid residues 7 and 61 that are not shared by members of other clades. TFT3 and NbFTT3 also demonstrate similar sequence similarities not seen in other TFTs or NbFTTs near amino acid position 34. NbFTT6 and NbFTT9 show less conserved amino acids throughout the entirety of their sequences, starting at amino acid residue 161 for NbFTT6 and 154 for NbFTT9. As similarly seen in the distance away from their most closely related orthologs in the phylogenetic tree, the differences in amino acid sequence of NbFTT6 and NbFTT9 further supports their more recent evolution from the last common ancestors they share with TFT5 and 6 or NbFTT4, respectively. Additionally, many N. benthamiana 14-3-3 proteins show large insertions in their amino acid sequences. The splice donor and splice acceptor sites surrounding introns may not be accurate as a result of the protein-coding sequences being pieced together in silico from the genomic DNA sequence and the insertions may not reflect real coding sequence insertions.

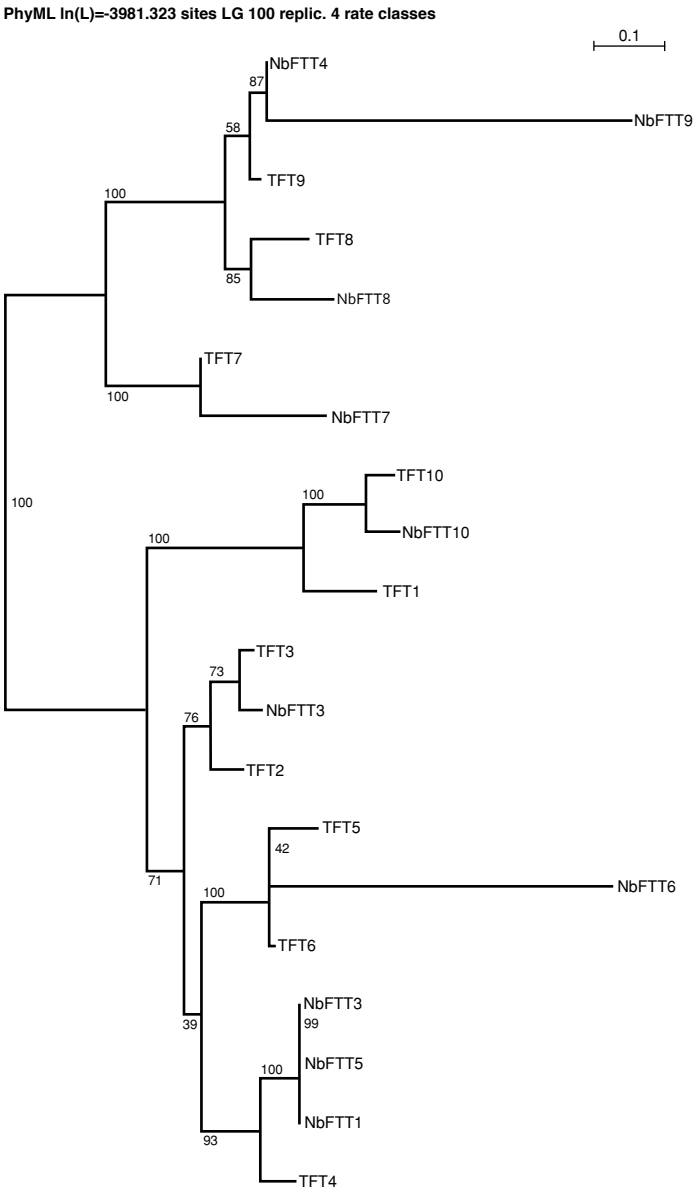
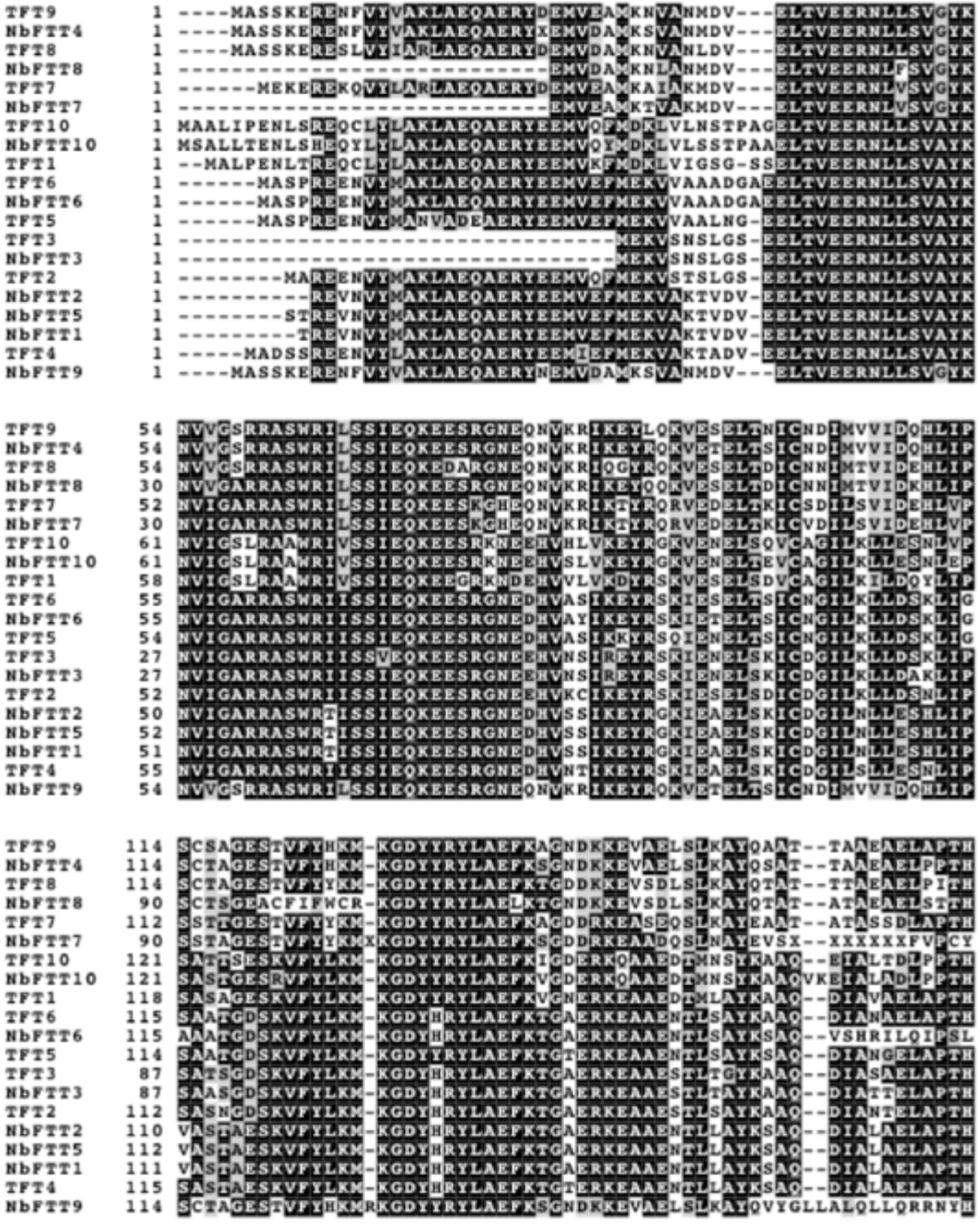


Figure 1. Phylogenetic tree demonstrating relatedness of 14-3-3 orthologs. *The phylogenetic tree was constructed using Seaview software with PhyML tree function, 100 replicates and showing bootstrap values at branch nodes. S. lycopersicum 14-3-3 isoforms are indicated as TFTs and N. benthamiana 14-3-3 isoforms indicated as NbFTTs.*


TFT9 1 ----MASSKEREWFVIVAKLAEQAERYDEMVEAMKNVANMDV---ELTVEERNLLSVGYK
NbFTT4 1 ----MASSKEREWFVIVAKLAEQAERYXEMVDAMKSVANMDV---ELTVEERNLLSVGYK
TFT8 1 ----MASSKERESLVIIARLAEQAERYDEMVDAMKNVANLDV---ELTVEERNLLSVGYK
NbFTT8 1 ----------MVDAMKNLANMDV---ELTVEERNLFSVGYK
TFT7 1 ------MEKEREKQVYLARLAEQAERYDEMVEAMKAIANMDV---ELTVEERNLVSVGYK
NbFTT7 1 ----------EMVEAMKTVAKMDV---ELTVEERNLVSVGYK
TFT10 1 MAALIPENLSREQCLYLAKLAEQAERYEEMVQFMDKLVLNSTPAGELTVEERNLLSVAYK
NbFTT10 1 MSALLTENLSHEQYLYLAKLAEQAERYEEMVQYMDKLVLSSTPAAELTVEERNLLSVAYK
TFT1 1 --MALPENLTREQCLYLAKLAEQAERYEEMVKFMDKLVIGSG-SSELTVEERNLLSVAYK
TFT6 1 ------MASPREENVYMAKLAEQAERYEEMVEFMEKVVAAADGAEELTVEERNLLSVAYK
NbFTT6 1 ------MASPREENVYMAKLAEQAERYEEMVEFMEKVVAAADGAEELTVEERNLLSVAYK
TFT5 1 ------MASPREENVYMANVADEAERYEEMVEFMEKVVAALNG-EELTVEERNLLSVAYK
TFT3 1 ----------------------------------MEKVSNSLGS-EELTVEERNLLSVAYK
NbFTT3 1 ----------------------------------MEKVSNSLGS-EELTVEERNLLSVAYK
TFT2 1 --------MAREENVYMAKLAEQAERYEEMVQFMEKVSTSLGS-EELTVEERNLLSVAYK
NbFTT2 1 ----------REVNVYMAKLAEQAERYEEMVEFMEKVAKTVDV-EELTVEERNLLSVAYK
NbFTT5 1 --------STREVNVYMAKLAEQAERYEEMVEFMEKVAKTVDV-EELTVEERNLLSVAYK
NbFTT1 1 ---------TREVNVYMAKLAEQAERYEEMVEFMEKVAKTVDV-EELTVEERNLLSVAYK
TFT4 1 -----MADSSREENVYLAKLAEQAERYEEMIEFMEKVAKTADV-EELTVEERNLLSVAYK
NbFTT9 1 ----MASSKERENFVIVAKLAEQAERYNEMVDAMKSVANMDV---ELTVEERNLLSVGYK
TFT9 54 NVVGSRRASWRILSSIEQKEESRGNEQNVKRIKEYLQKVESELTNICNDHMVVIDQHLIP
NbFTT4 54 NVVGSRRASWRILSSIEQKEESRGNEQNVKRIKEYRQKVETELTSICNDHMVVIDQHLIP
TFT8 54 NVVGSRRASWRILSSIEQKEDARGNEQNVKRIQGYRQKVESELTDICNNHMTVIDEHLIP
NbFTT8 30 NVVGARRASWRILSSIEQKEESRGNEQNVKRIKEYQQKVESELTDICNNHMTVIDKHLIP
TFT7 52 NVIGARRASWRILSSIEQKEESKGHEQNVKRIKTYRQRVEDELTKICSDHLSVIDEHLVP
NbFTT7 30 NVIGARRASWRILSSIEQKEESKGHEQNVKRIKTYRQRVEDELTKICVDHLSVIDEHLVP
TFT10 61 NVIGSLRAAWRIVSSIEQKEESRKNEEHVHLVKEYRGKVENELSQVCAGILKLLESNLVP
NbFTT10 61 NVIGSLRAAWRIVSSIEQKEESRKNEEHVSLVKEYRGKVENELTEVCAGILKLLESNLEP
TFT1 58 NVIGSLRAAWRIVSSIEQKEEGRKNDEHVVLVKDYRSKVESELSDVCAGILKILDQYLIP
TFT6 55 NVIGARRASWRIISSIEQKEESRGNEDHVASIKEYRSKIESELTSICNGILKLLDSKLIG
NbFTT6 55 NVIGARRASWRIISSIEQKEESRGNEDHVAYIKEYRSKIETELTSICNGILKLLDSKLIG
TFT5 54 NVIGARRASWRIISSIEQKEESRGNEDHVASIKKYRSQIENELTSICNGILKLLDSKLIG
TFT3 27 NVIGARRASWRIISSVEQKEESRGNEEHVNSIREYRSKIENELSKICDGILKLLDSKLIP
NbFTT3 27 NVIGARRASWRIISSIEQKEESRGNEEHVNSIREYRSKIENELSKICDGILKLLDAKLIP
TFT2 52 NVIGARRASWRIISSIEQKEESRGNEEHVKCIKEYRSKIESELSDICDGILKLLDSNLIP
NbFTT2 50 NVIGARRASWRTISSIEQKEESRGNEDHVSSIKEYRGKIEAELSKICDGILNLLESHLIP
NbFTT5 52 NVIGARRASWRTISSIEQKEESRGNEDHVSSIKEYRGKIEAELSKICDGILNLLESHLIP
NbFTT1 51 NVIGARRASWRTISSIEQKEESRGNEDHVSSIKEYRGKIEAELSKICDGILNLLESHLIP
TFT4 55 NVIGARRASWRIISSIEQKEESRGNEDHVNTIKEYRSKIEAELSKICDGILSLLESNLIP
NbFTT9 54 NVVGSRRASWRILSSIEQKEESRGNEQNVKRIKEYRQKVETELTSICNDHMVVIDQHLIP
TFT9 114 SCSAGESTVFYHKM-KGDIYRYLAEFKAGNDKKEVAELSLKAYQAAT--TAAEAELAPTH
NbFTT4 114 SCTAGESTVFYHKM-KGDIYRYLAEFKSGNDKKEVAELSLKAYQSAT--TAAEAELPPTH
TFT8 114 SCTAGESTVFYYKM-KGDIYRYLAEFKTGDDKKEVSDLSLKAYQTAT--TTAEAELPITH
NbFTT8 90 SCTSGEACFIFWCR-KGDIYRYLAELKTGNDKKEVSDLSLKAYQTAT--ATAEAELSTTH
TFT7 112 SSTTGESTVFYYKM-KGDIYRYLAEFKAGDDRKEASEQSLKAYEAAT--ATASSDLAPTH
NbFTT7 90 SSTAGESTVFYYKMX KGDIYRYLAEFKSGDDRKEAADQSLNAYEVSX--XXXXXXFVPCY
TFT10 121 SATTSESKVFYLKM-KGDIYRYLAEFKIGDERKQAAEDTMNSYKAAQ--EHALTDLPPTH
NbFTT10 121 SASTGESRVFYLKM-KGDIYRYLAEFKVGDERKQAAEDTMNSYKAAQVKEHALNDLPPTH
TFT1 118 SASAGESKVFYLKM-KGDIYRYLAEFKVGNERKEAAEDTMLAYKAAQ--DIAVAELAPTH
TFT6 115 SAATGDSKVFYLKM-KGDIYRYLAEFKTGAERKEAAENTLSAYKAAQ--DIANAELAPTH
NbFTT6 115 AAATGDSKVFYLKM-KGDIYHRYLAEFKTGAERKEAAENTLSAYKSAQ--VSHRILQIPSL
TFT5 114 SAATGDSKVFYLKM-KGDIYRYLAEFKTGTERKEAAENTLSAYKSAQ--DIANGELAPTH
TFT3 87 SATSGDSKVFYLKM-KGDYHRYLAEFKTGAERKEAAESTLTGYKAAQ--DIASAELAPTH
NbFTT3 87 SAASGDSKVFYLKM-KGDYHRYLAEFKTGAERKEAAESTLTAYKAAQ--DIATTELAPTH
TFT2 112 SASNGDSKVFYLKM-KGDYHRYLAEFKTGAERKEAAESTLSAYKAAQ--DIANTELAPTH
NbFTT2 110 VASTAESKVFYLKM-KGDYHRYLAEFKTGAERKEAAENTLLAYKSAQ--DIALAELAPTH
NbFTT5 112 VASTAESKVFYLKM-KGDYHRYLAEFKTGAERKEAAENTLLAYKSAQ--DIALAELAPTH
NbFTT1 111 VASTAESKVFYLKM-KGDYHRYLAEFKTGAERKEAAENTLLAYKSAQ--DIALAELAPTH
TFT4 115 SASTAESKVFYLKM-KGDYHRYLAEFKTGTERKEAAENTLLAYKSAQ--DIALAELAPTH
NbFTT9 114 SCTAGESTVFYHKMRKGDIYRYLAEFKSGNDKKEVAELSLKAYQVYGLLALQLLQRRNYH

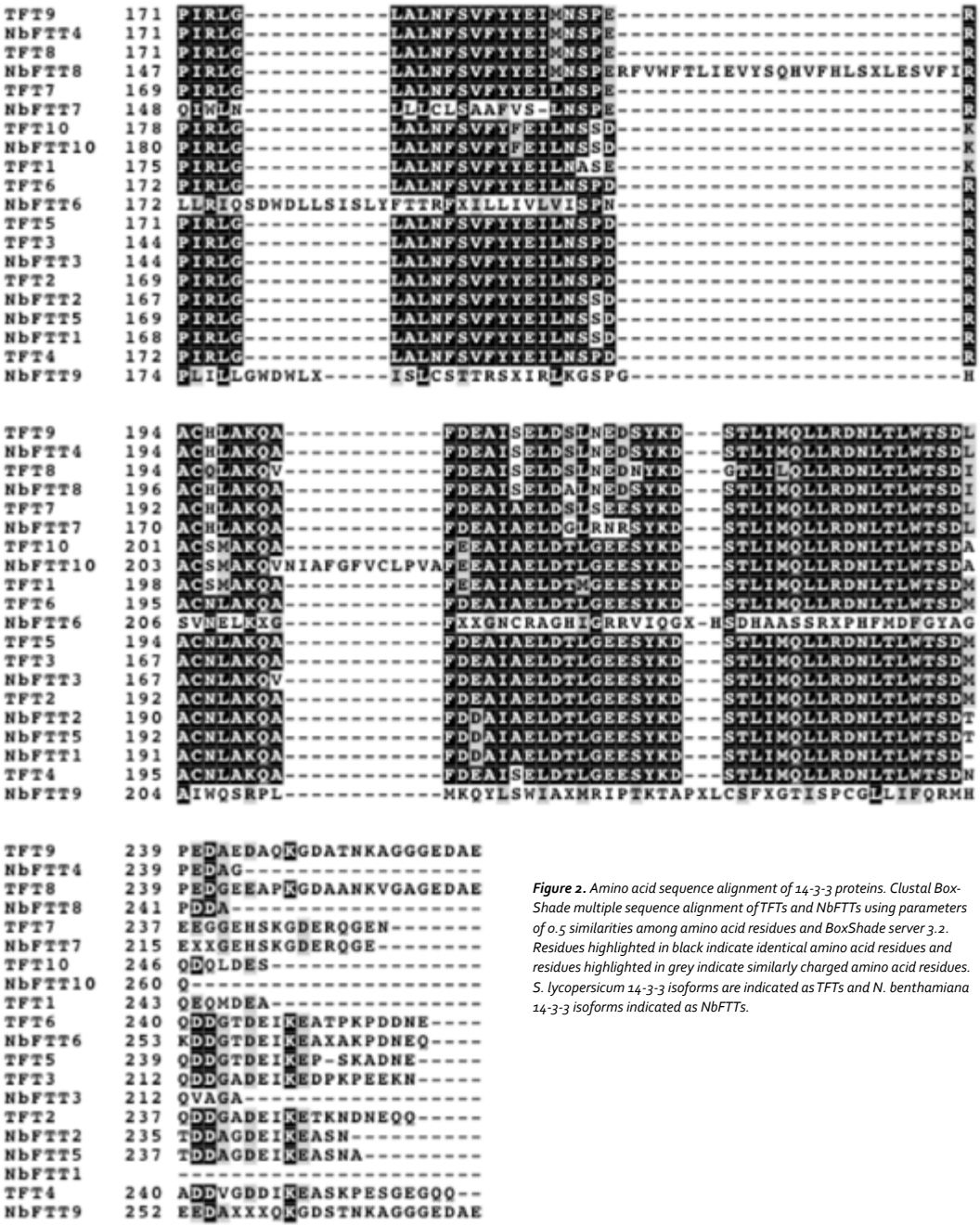


Figure 2. *Amino acid sequence alignment of 14-3-3 proteins. Clustal BoxShade multiple sequence alignment of TFTs and NbFTTs using parameters of 0.5 similarities among amino acid residues and BoxShade server 3.2. Residues highlighted in black indicate identical amino acid residues and residues highlighted in grey indicate similarly charged amino acid residues. S. lycopersicum 14-3-3 isoforms are indicated as TFTs and N. benthamiana 14-3-3 isoforms indicated as NbFTTs.*

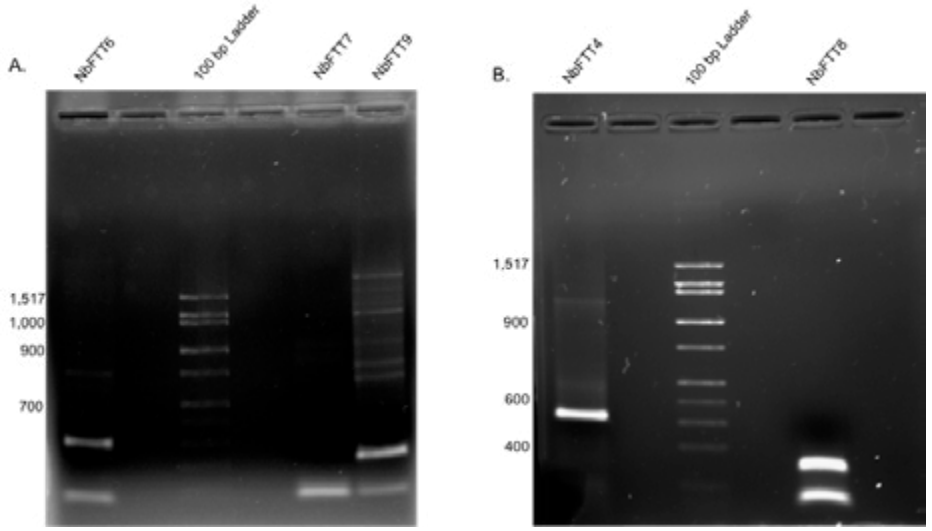


Figure 3. *RT-PCR amplification of N. benthamiana 14-3-3 cDNA. cDNA was separated by gel electrophoresis and visualized on 0.8% agarose gels. A. NbFTT6 and NbFTT9 with bands shown around 600bp and 400bp were both extracted for cloning and sequencing. The single NbFTT7 band around 400bp was also excised. B. NbFTT8 with bands shown around 400bp and 300bp were both extracted for cloning and sequencing and single NbFTT4 band around 500bp was also excised. DNA bands were imaged with the Gel Logic 100 Imaging System, excised and purified with the Qiagen gel extraction kit.*

Amplification of N. benthamiana 14-3-3 DNA.

In order to create a library for use in silencing of the NbFTT homologs, primers were designed for use in reverse transcription-polymerase chain reaction (RT-PCR) to amplify NbFTT cDNA fragments from N. benthamiana RNA. Primers were designed to contain restriction enzyme cut sites for subsequent cloning into the silencing vector. RT-PCR gel electrophoresis confirmed the amplification of cDNA of 14-3-3 homologs of the expected sizes. Several 14-3-3 homologs, including NbFTT1, NbFTT2, NbFTT3, NbFTT5 and NbFTT10, were previously amplified using RT-PCR, separated by gel electrophoresis and extracted for further use in cloning and creation of the silencing library. For this study, NbFTT6, NbFTT7, NbFTT9, NbFTT4 and NbFTT8 were reverse transcribed and amplified from Nicotiana benthamiana leaves (Figure 3). Due to the presence of higher bands around 500-600bp and lower bands around 300-400bp, the additional bands were also extracted from the agarose gel in case the different sized PCR products represented alternative splice variants and the PCR products were designated as NbFTT#up or NbFTT#down. Each NbFTT RT-PCR product confirms the expression of 14-3-3 homologs in N. benthamiana leaves, which is promising for adequate silencing of 14-3-3 genes in future experiments.

Cloning of N. benthamiana 14-3-3 isoforms

In order to complete creation the 14-3-3 homolog silencing library, 14-3-3 cDNA fragments were cloned into the pGEM-T vector for future subcloning into the silencing vector. After the amplification and isolation of 14-3-3 homolog cDNA by RT-PCR and gel extraction, the cDNA fragments were cloned into the pGEM-T vector using A-tailing reactions, ligation reactions and transformation into Escherichia coli. Transformation of the correct 14-3-3 homolog fragments needed to be validated before further use in silencing of NbFTT homologs in N. benthamiana leaves. Validation of 14-3-3 isoform clones were done by using PCR screening with the primers designed for the vector and separated by gel electrophoresis (Figure 4). After validation from PCR screening, clones of NbFTT3 (lanes 3 and 4), NbFTT4 (lane 3), NbFTT6 (lanes 6 and 9) and NbFTT10 (lane 7) were selected for DNA sequencing for final confirmation of the correct 14-3-3 homolog insert sequences. DNA sequencing confirmed cloning of fragments of NbFTT3 (lanes 3 and 4), NbFTT4 (lane 3) and NbFTT10 (lane 7). Several 14-3-3 homologs, including NbFTT1, NbFTT2 and NbFTT5, were previously validated and confirmed using this method. Clones of NbFTT6, NbFTT7 and NbFTT9 were amplified in E. coli and sequenced directly for confirmation of correct gene fragment inserts. While most NbFTTs were successfully amplified and cloned, DNA sequencing revealed that the primers designed for the amplification of NbFTT8 did not amplify the targeted 14-3-3 gene and instead amplified another gene. New primers need to be designed to target a different region of NbFTT8 in order to amplify and clone a fragment from the intended sequence. Glycerol stocks of all sequenced and confirmed 14-3-3 homologs were retained for further use in subcloning for VIGS studies.

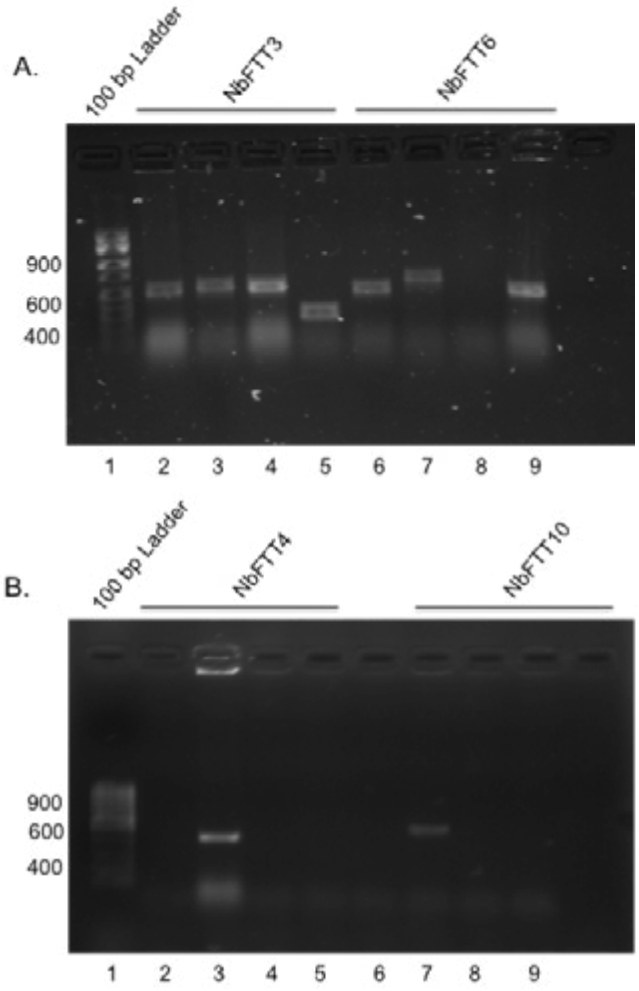


Figure 4. *Screening of N. benthamiana 14-3-3 DNA inserts. PCR screen products separated by gel electrophoresis and visualized on 0.8% agarose gels. A. NbFTT3 (lanes 3 and 4) and NbFTT6 (lanes 6 and 9) clones were selected for validation of the insert. B. NbFTT4 (lane 3) and NbFTT10 (lane 7) clones were selected for validation of the insert. DNA bands were imaged with the Gel Logic 100 Imaging System.*

DISCUSSION

The role of tomato 14-3-3 homologs has been previously described in plant immunity through involvement in the MAPK pathway and a putative interaction observed between the tomato Tm2-2 resistance protein (Denison et al., 2011) and a NbFTT homolog (Sobhanian, 2011). Investigating the evolutionary relationship between TFTs and NbFTTs could prove beneficial in determining the potential role of NbFTTs in plant immunity. Although the N. benthamiana 14-3-3 homologs were named based on their retrieval from the genome database as matches to the TFT sequences, the phylogenetic tree and sequence alignment show the designated names for some NbFTT homologs did not correspond to their most closely related TFT homologs. The differences in phylogeny between TFTs and NbFTTs could be attributed to gene duplication among the NbFTTs and contribute towards functional redundancy of several NbFTT homologs. After silencing and testing each NbFTT homolog's interaction with Tm2-2 individually in multiple N. benthamiana plants, NbFTT members of clades, including NbFTT2, NbFTT5 and NbFTT1, will be silenced together in one N. benthamiana plant. Silencing multiple homologs simultaneously would down-regulate NbFTT proteins together that are likely functionally redundant and provide a clearer assessment of the putative interaction between NbFTT homologs and Tm2-2 (Sobhanian, 2011). However, genetic duplication could have attributed to diversification of the NbFTT genes and resulted in differences in phylogeny. The incongruity between the designations of NbFTTs and their closest evolutionary match means that the N. benthamiana sequences need to be renamed accordingly in future experiments.

Although NbFTT8 was not correctly amplified, the remaining nine homologs were successfully cloned into pGEM-T to partially complete the set of NbFTTs. Primers were redesigned for future amplification and cloning of NbFTT8 in order to complete cloning of all ten homologs. After completion of the set of N. benthamiana 14-3-3 homolog pGEM-T clones, the ten NbFTT homologs will be used to create a silencing library of clones in the pTV viral silencing vector. These clones will be subsequently transformed into A. tumefaciens in the future for agroinfiltration and use in VIGS experiments. Interactions between 14-3-3 homologs and the Tm2-2 resistance protein will be assessed through the presence or absence of HR when transgenic N. benthamiana expressing Tm2-2 is challenged with the 30K movement protein. Silencing of 14-3-3 homologs essential for Tm2-2 function could potentially prevent HR in transgenic N. benthamiana in the presence of movement protein and substantiate a role for one or several 14-3-3 proteins in Tm2-2-mediate immune responses.

In some species of Arabidopsis, low expression of 14-3-3 proteins has been previously associated with susceptibility and decreased resistance to powdery mildew fungus (Denison et al., 2011). Conversely, over-expression of Arabidopsis 14-3-3 genes elicited HR, potentially due to an interaction between 14-3-3 proteins and the R protein, RPW8.2, thereby increasing resistance (Denison et al., 2011). Additional future experiments will address the latter of the two cases and

involve over-expression of tomato 14-3-3 genes and co-expression of Tm2-2 in wild- type N. benthamiana leaves. When challenged with the 30K movement protein, differences in HR on the leaves over-expressing 14-3-3 genes and co-expressing Tm2-2 could further indicate an interaction between 14-3-3 proteins and the Tm2-2 resistance protein. The initiation of faster, more robust HR could indicate that 14-3-3 proteins function in Tm2-2-mediated immune responses. Additionally, co-immunoprecipitation experiments with 14-3-3 proteins and the CC domain of Tm2-2 will further investigate these putative interactions.

References

Akira, S., Uemastu, S., and Takeuchi, O. 2006. Pathogen recognition and innate immunity. Cell, 124, 783-801.

Boller, T., He, S.Y. 2009. Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. Science, 324, 742-744.

Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A. and Martin G.B. 2012. A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. Molecular Plant-Microbe Interactions, 25, 1523-1530.

Dangl, J.L. and Jones, J.D. 2001. Plant pathogens and integrated defense responses to infection. Nature, 18, 826-833.

Denison, Fiona C., Paul, Anne-Lisa, Zupanska, Agata K. and Robert J. Ferl. 2011. 14-3-3 proteins in plant physiology. Seminars in Cell & Developmental Biology, 22, 720–727.

Gouy, M., Guindon, S. and Gascuel, O. 2010. Molecular Biology and Evolution, 27, 221-224.

Jones, J.D., and Dangl J.L. 2006. The Plant Immune System. Nature 444.7117, 323-29.

Konagaya, K., Matsushita, Y., Kasahara, M., & Nyunoya, H. 2004. Members of 14-3-3 protein isoforms interacting with the resistance gene product N and the elicitor of Tobacco mosaic virus. Journal Of General Plant Pathology, 70(4), 221-231.

Larkin, M.A, Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. 2007. Bioinformatics, 23, 2947-2948.

Moorhead, G., Douglas, P., Morrice, N., Scarabel, M., Aitken, A., & Mackintosh, C. 1996. Phosphorylated nitrate reductase from spinach leaves is inhibited by 14-3-3 proteins and activated by fusicoccin. Current Biology, 6(9), 1104-1113.

Oh, C., Pedley, K.F. and Martin, G.B. 2010. Tomato 14-3-3 Protein 7 Positively Regulates Immunity-Associated Programmed Cell Death by Enhancing Protein Abundance and Signaling Ability of MAPKKK α. The Plant Cell, 22(1), 260-272.

Purkayastha, A., and Dasgupta, I. 2009. Virus-induced gene silencing: A versatile tool for discovery of gene functions in plants. Plant Physiology and Biochemistry, 47, 967–976.

Ratcliff, F., Martin-Hernandez, A.M.and Baulcombe, D.C. 2001. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J., 25, 237-245.

Salmond, G.P., and Reeves, P.J. 1993. Membrane traffic wardens and protein secretion in Gram- negative bacteria. Trends Biochemical Sciences, 18, 7-12.

Sobhanian, S. 2011. Contribution of host proteins toward R-protein mediated resistance in Solanaceae family. Graduate Thesis, California State University, Fullerton.

Velásquez, A.C., Chakravarthy, S. and Martin, G.B. 2009.Virus-induced Gene Silencing (VIGS) in Nicotiana benthamiana and Tomato. Journal of Visualized Experiments, 28, 1292.

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

Development of Micelles Containing Self-Immolative Polymers for Applications in Drug Delivery

Department of Chemistry and Biochemistry, California State University, Fullerton
Department of Chemistry, University of Washington

Neha F. Ansari, Gregory I. Peterson, Andrew J. Boydston

Abstract

Self-immolative polymers (SIPs) have proven to be an innovative system for the controlled release of small molecules. Most SIPs have chain end triggers that once activated, can lead to a head-to-tail depolymerization of the polymer backbone. Each monomer is then released sequentially in a timely and consistent manner, concomitant with release of the output molecule (e.g., drug or reporter molecules) from the monomer. However, the SIP's hydrophobic properties are problematic for its implementation into biological applications, thus making it necessary to further functionalize the polymer for water solubility. This can be achieved by attaching a hydrophilic polymer to the SIP. As a result, the amphiphilic diblock copolymer can form micelles which are of interest in the field of drug delivery. In this study, we examine the multiple syntheses that lead to the formation of the SIP-containing amphiphilic diblock copolymers. Through these experiments, we aspire to learn more about the creation and stability of micelles from SIP diblock copolymers and ultimately aim to use this as a platform to demonstrate thermally activated SIPs in a novel drug delivery system.

The Synthesis of Anhydrous Cyclopropanol

Department of Chemistry and Biochemistry, California State University, Fullerton

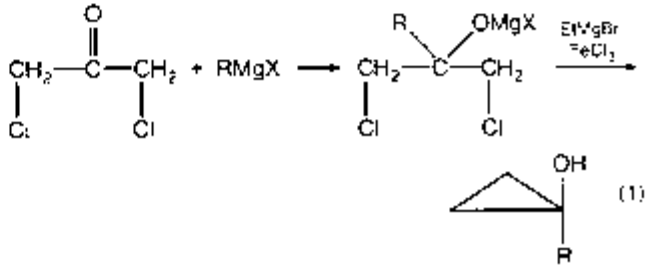
Jennifer Castillo
***Advisor*: Dr. Rogers**

Abstract

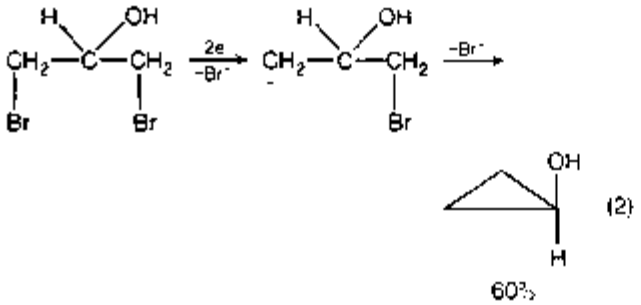
A strategy for the synthesis of anhydrous cyclopropanol is being developed from a commercially-available precursor. The core of the method involves protecting 1,3-dichloro-2-propanol with a suitable silylating agent such as t-butyldimethylsilyl chloride to form a precursor compound using Schlenk techniques. Subsequent 1,3-reductive elimination to close the three membered ring, followed by removal of the protecting group under anhydrous conditions, should result in an accessible gram-scale preparative method for cyclopropanol. Formation of the protected alcohol was confirmed by infrared spectroscopy and 1H-NMR, which showed possible conformational isomers.

Introduction

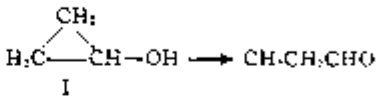
Cyclopropanol is not an unknown compound; however, it has not been prepared and isolated in gram-sized quantities as a pure alcohol. Main commercial use of cyclopropanol is in the formation of esters, which is used in preparations of pharmaceuticals, flavors, fragrances, and pesticides. Also, the chemical properties of cyclopropanol are more diverse than those of enols or enolates. Another useful property is that cyclopropanols are able to undergo synthetically useful transformations where the three-carbon ring is retained or cleaved. Cleavage causes the cyclopropanols to act as equivalents of enolates or corresponding allylic derivatives.[1] Many methods to synthesize cyclopropanol have been attempted, however, these methods produced cyclopropanol containing other substituent groups. One such simple method of preparation involves the use of Grignard addition to 1,3-dichloroacetone, as shown in Equation 1.[2] Cyclopropanol was first synthesized in pure form accidently in 1942 by Cottle and Magrane. This method involved the reaction of epichlorohydirn with magnesium bromide, ethylmagnesium bromide, and ferric chloride.[2]



Modification of this method involves the reaction of epichlorohydrin with $MgBr_2$ to form the magnesium salt of 1-bromo-3-chloro-2-propanol. If the salt is treated with a Grignard reagent in the presence of ferric chloride, ring closure occurs.[3] However, the problem with this method is that the cyclopropanols have to avoid contact with acids or bases in the presence of protic solvents.[3] Gerdil reported a further modification of the 1,3-dihalo-2-propanol method where dehalogenation is accomplished by an electrochemical process giving good yield of cyclopropanol[4] seen in equation 2.



Also, 1,3-dihalides can be converted into cyclopropanes by dehalogenation with magnesium.[5] Attempts to purify cyclopropanol have been unsuccessful. Cottle and Magrane devised an isolation method for cyclopropanol using fractional distillation, however, they found that their experiments gave cyclopropanol fractions that contained halogen, possibly as epichlorohydrin, and fractional distillation or chemical treatment could not remove these fractions without damage to the alcohol.[6] A later experiment's sample purified by Stahl and Cottle resulted in a Zerewitinoff value that indicated an 87% content of cyclopropanol. Also, cyclopropanol's change to propionaldehyde was obtained when the alcohol was treated with potassiumcarbonate.7 Furthermore, cyclopropanol readily rearranges to propionaldehyde especially in basic solution[7], seen in equation 3.



Experiments conducted by Roberts and Chambers also concluded analytically pure cyclopropanol was not obtained in repeated Magrane and Cottle methods and rearranged easily to propionaldehyde.[8]

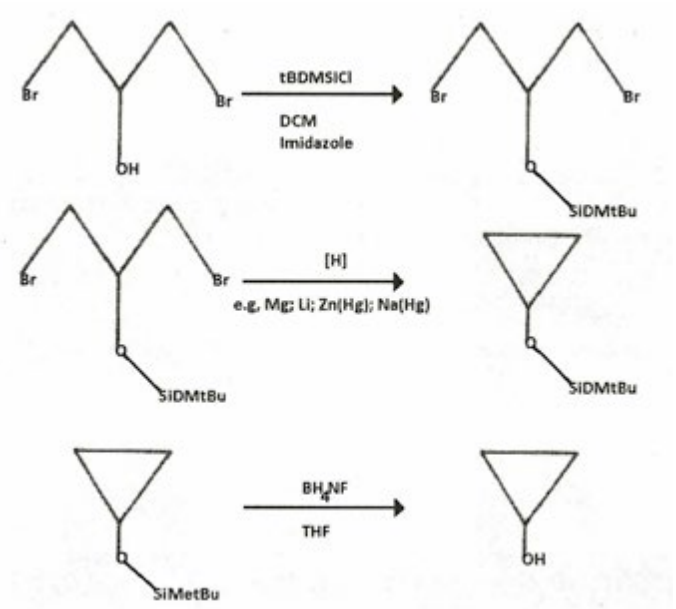
Another method in the formation of cylcopropanol is by cleavage of esters and ethers. Generally, basic cleavage of a cyclopropyl ester produces cyclopropanol in good yield, however Paukstelis and Kao have reported an instance where skeletal rearrangement takes precedence over the formation of alcohol.[3] Preparing cyclopropanols using ethers was first developed by Schollkopf and his co-workers. The problem with this method, however, is finding an R group that can be removed without disrupting the three-membered ring.[3]

One of the most well-known transformations of cyclopropane derivatives is the cyclopropyl to allyl rearrangement and is widely used in the preparation of allylic compounds from halogenocyclopropanes.[1]

It is important to use a protecting group on the alcohol before ring closure can occur. The use of a tert-Butyldimethylsilyl group is advantageous in that it does not have a chiral center and is both stable and applicable in the protection of alcohol.[9] Furthermore, silyl ethers are easily prepared, show resistance to oxidation, good thermal stability, low viscosity, and are easily recoverable from their parent compound.[10]

Known physical properties of cyclopropanol include index of refraction (1.526), molar volume $(50.2 \pm 3 \mathrm{cm3})$, flash point $(22.2 \pm 10.9 ℃)$, boiling-point $(90.4 \pm 8.0 ℃ \text{ at } 760 \text{ mmHg})$, densit (1.2±0.1 g/cm3), and vapor pressure $(34.4 \pm 0.3 \text{ mmHg at } 25 ℃)$.[11]

Known physical properties of cyclopropanol include index of refraction (1.526), molar volume $(50.2 \pm 3 \mathrm{cm3})$, flash point $(22.2 \pm 10.9 ℃)$, boiling-point $(90.4 \pm 8.0 ℃ \text{ at } 760 \text{ mmHg})$, density $(1.2 \pm 0.1 \text{ g/cm3})$, and vapor pressure $(34.4 \pm 0.3 \text{ mmHg at } 25 ℃)$.[11]



Results

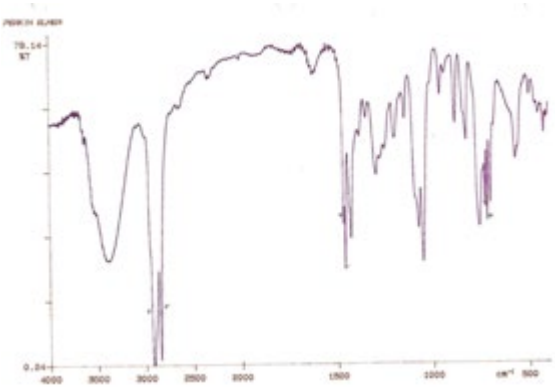


Figure 1: *IR of starting material 1,3-dichloro-2-propanol.*

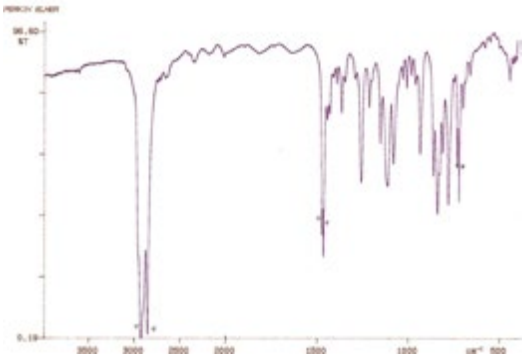


Figure 2: *IR indicating the –OH functional group is no longer present.*

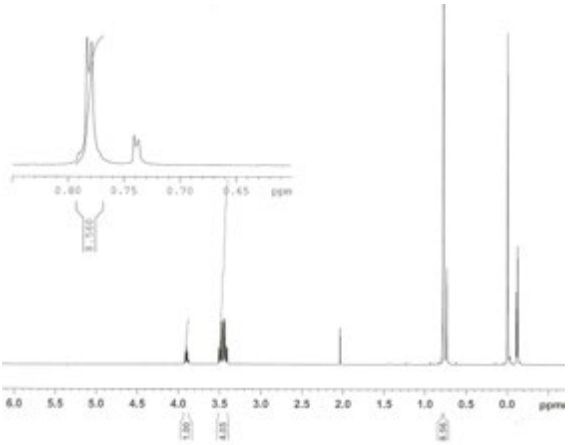


Figure 3: *NMR with a doublet around 0.8ppm in crude product*

Experimental

Dichloromethane, 50 mL, and a magnetic stir-bar were added to a 250-mL round bottom flask. Imidazole, 3.3318 g (16.26mmol), dissolved in 10 mL of dichloromethane was added to the flask. Approximately 2.35 mL (25mmol) of 1,3-dichloro-2-propanol was added to the flask using an appropriate syringe. t-Butyldimethylsilyl chloride (TBDMSiCl), 8.1556 g (18mmol) dissolved in 10 mL of dichloromethane, was transferred under argon to an addition funnel along with an additional 30 mL of dichloromethane. The flask was placed under a slight positive pressure of argon. The TBDMSiCl solution was added drop-wise over a period of 48 hours at room temperature. Vacuum filtration was performed to separate the precipitate. An extraction was performed three times with deionized water and magnesium sulfate was added to the organic liquid. The solution dried for 24 hours. An IR spectrum was taken, using a Perkin Elmer 1500 Series FTIR, (neat liquid on a polyethylene IR card) to check for any trace of alcohol; IR confirmed none was present. Vacuum filtration was performed to remove the drying agent. A rotary evaporator was used to remove the solvent. The product recovered was clear. An IR spectrum of product was taken as well as an ^{1}H-NMR, which showed volatile impurities; therefore, rotovap was performed a second time. ^{1}H-NMR of this purified product was taken, spectrum seen in Figure 3. Each ^{1}H-NMR taken had TMS as the solvent.

Discussion

A white precipitate and yellowish liquid were formed during the reaction of the alcohol with the silylating agent. After vacuum filtration, IR confirmed no alcohol was present in the solution, as seen in Figure 2. Also, the doublet present around 3000 cm^{-1} and second doublet around 1500 cm-1 represent the presence of TBDMSiCl and are consistent with those found in literature. ^{1}H- NMR of the crude protected alcohol showed a doublet was present at 0.76 ppm with an integration number of 8.560 as shown in Figure 3. It is believed that this doublet is actually two singlets representing conformational isomers, which is consistent with restricted rotation due to the bulky *t*-butyl group.

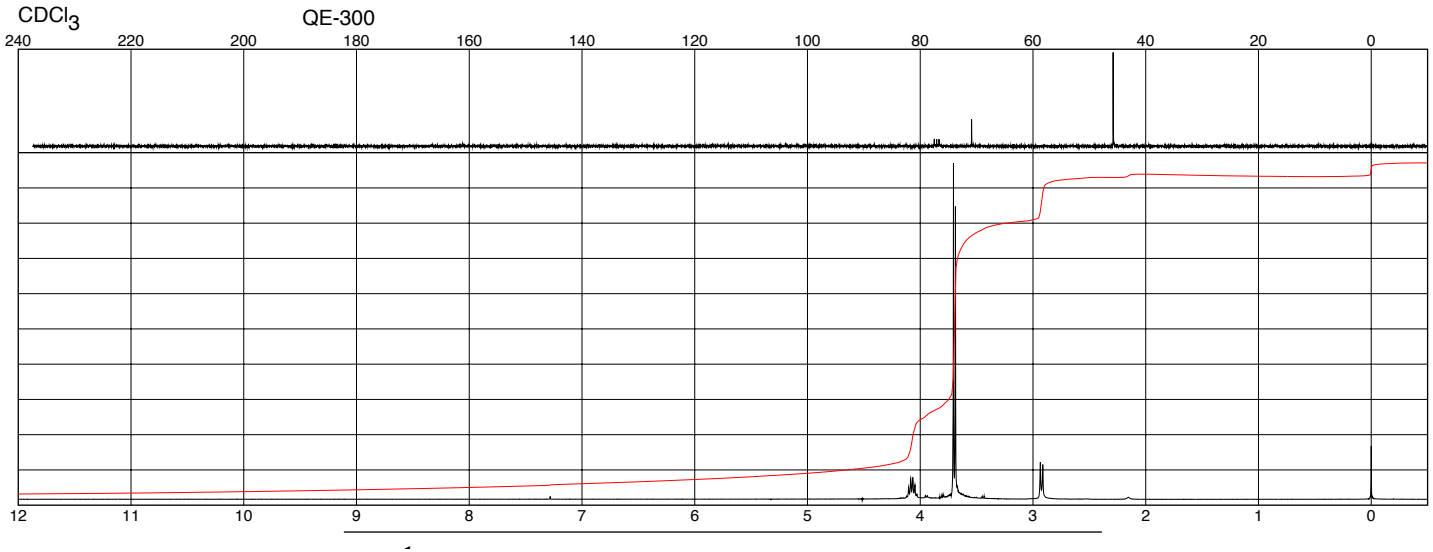


Figure 4: *Literature 1H-NMR of 1,3-dichloro-2-propanol*

The remaining features of the spectrum are consistent with the target compound. Peaks present around 4.00, 3.50, and 2.00 ppm indicate the presence of TBDMSiCl. Comparing Figures 3 and 4, differences are noticed in peak positions.

Future work will determine if the desired product was formed. Firstly, vacuum distillation or chromatographic purification of the protected alcohol product must be done. ^{13}C-NMR will check for impurities. Various reducing agents can be used such as magnesium, lithium, sodium amalgam, or zinc amalgam to form the ring closure. After ring closure, the protecting group must be removed. Removal can be done with tetrabutylammonium fluoride in THF under anhydrous conditions. It is important to note that none of these steps involves the use of water.

Conclusion

A Schlenk technique resulted in the protection of the alcohol group in 1,3-dichloro-2-propanol using silylating agent *t*-butyldimethylsilyl chloride. Comparison between literature values and experimental values in FTIR and ^{1}H-NMR confirmed the protection. Future work including vacuum distillation, ring closure using reducing agents, and removal of protecting group through the use of tetrabutylammoinium fluoride in THF under anhydrous conditions can possibly form anhydrous cyclopropanol.

References

1. Kulinkovich, O.G. The Chemistry of Cyclopropanols. J. Am. Chem. Soc. 2002. 103, 2597-2632.

2. DePuy, C.H.; Mahoney L.R.. The Chemistry of Cyclopropanols. I. The Hydrolysis of Cyclopropyl Acetate and the Synthesis of Cyclopropanol. J. Am. Chem. Soc. 1963. 86 (13) 2653-2657.

3. DePuy, C.H.; Gibson D.H. Cyclopropanol Chemistry. J.Am. Chem. Soc. 1973. 1974 (6) 605-623.

4. Gerdil, R. Helv. Chim. Acta.1970. 53 2100.

5. Cudre, Y.; Fernandez-Zumel, M.A.; Risse, J.; Severin, K. Synthesis of Tri-fluoromethyl- Substituted Cyclopropanes via Sequential Kharasch Dehalogenation Reactions. Org. Lett. 2012. 14 (12) 3060-3063.

6. Cottle, D.L.; Magrane J.K. The Reaction of Epichlorohydrin with the Grignard Reagent. J. Am. Chem Soc. 1941. 65 483-487.

7. Cottle, D.L.; Stahl G. W. The Reaction of Epichlorohydrin with the Grignard Reagent. Some Derivatives of Cyclopropanol. J. Am. Chem. Soc. 1943. 65 (9) 1782-1783.

8. Chambers, C.V.; Roberts J.D. Small Ring Compounds. VI. Cyclopropanol, Cyclopropyl Bromide and Cyclopropylamine. J. Am. Chem. Soc. 1951. 73 (7) 3176-3179.

9. Corey, E.J.; Venkateswarlu, A. Protection of Hydroxyl Groups as tert-Butyldimethylsilyl Derivatives. J. Am. Chem. Soc. 1972. 94 (17) 6190-6191.

10. Ballini, R.; Bigi, F.; Bosica G.; Maggi, R.; Righi P.; Satori, G. Protection (and Deprotection) of Functional Groups in Organic Synthesis by Heterogeneous Catalysis. Chem. Rev. 2004. 104 (1) 199-250.

11. Chemider. http://www.chemspider.com/Chemical-Structure. 109961.html (accessed 12/9/13).

The Intrinsically Disordered Protein Stathmin Exists as an Oligomer in Solution, as Measured by Static Light-scattering and Dipolar Broadening EPR Spectroscopy.

Department of Chemistry and Biochemistry, California State University, Fullerton

Ashley J. Chui, Katherina C. Chua, Jesus M. Mejia, and Michael D. Bridges

Abstract

Intrinsically disordered proteins (IDPs) are an interesting class of highly dynamic, typically regulatory, proteins. They lack a native three-dimensional fold, but often acquire a stable, ordered structure upon interaction with a binding partner. Various IDPs have been reported to exist in cells as ordered oligomers or disordered aggregates, often resulting in disease. In particular, stathmin is a regulatory IDP involved in the disassembly of cytoskeletal microtubules. As such, it is essential for proper cell function (i.e., processes coordinating the cell cycle, maintaining cell shape, etc.); improper regulation of stathmin activity has been linked to neurodegenerative diseases, mental disorders, and various cancers. It is thus important to study the solution-phase structure and conformational dynamics of stathmin, as they likely emulate the protein's behavior in cellular environments. Upon obtaining preliminary Native-PAGE results that indicated multiple stathmin complexes of varying mass, we hypothesized that it may exist as an oligomer in solution, which contradicts previous observations of a purely monomeric state as seen by analytical ultracentrifugation. We thus obtained static multi-angle light scattering data for stathmin solutions of varying concentrations, which show distinct concentration-dependent variation on measured particle size, as expected from an equilibrium system of monomers and oligomers. To investigate this further, we then performed site-directed spin labeling (SDSL) electron paramagnetic resonance (EPR) spectroscopy on multiple singly- and doubly-labeled stathmin mutants. Interestingly, the resulting EPR data all exhibit 'complex' (multicomponent) spectra, which could be easily deconvoluted and interpreted. In each case, one spectral component exhibited the predicted high-mobility – but ordered – state of a nitroxide side chain tethered to a stable alpha helix. The other spectral component, on the other hand, was substantially broadened due to the dipolar interaction, implying the close proximity of two or more spin labels, likely due to dimerization or higher-order oligomerization. Upon dilution of the spin-labeled proteins with unlabeled wild-type stathmin, the spectral weight of the dipolar-broadened component was significantly reduced. The data collectively presented herein support our hypothesis that stathmin exists as an oligomer in solution. These results have important implications on our understanding of the conformational dynamics of this IDP and the roles that oligomerized or aggregated IDPs have in diseases.

Michaelis-Menten Kinetics of TEV Protease as Observed by Time-domain EPR Spectroscopy

Department of Chemistry and Biochemistry, California State University, Fullerton

Estefania Larrosa and Michael D. Bridges

Abstract

Here we report the utility of time-domain electron paramagnetic resonance (EPR) spectroscopy for the direct detection of tobacco etch virus (TEV) protease activity. Our target substrate for monitoring the action of the protease was a fusion protein construct consisting of maltose binding protein linked to the intrinsically disordered protein 'stathmin' by the TEV protease-specific amino acid sequence (ENLYFQG). We used site-directed mutagenesis and nitroxide spin labeling to probe local motion at specific sites in the stathmin portion of the fusion construct (i.e., at residues 3, 12, 54, 74, 91, 113, and 146). Spin labeling involves the attachment of an EPR-active nitroxide side chain to a cysteine residue and the relative mobility of the spin label is directly reported via its exhibited EPR spectrum. As such, we were able to monitor the cleavage of the fusion construct by an increase in the observed motion of the stathmin-bound nitroxide. The spectra recorded over the course of the proteolysis reaction were globally fit by using a simple Michaelis-Menten model. K_m and k_{cat} were extracted from these fits and were compared to previously published values. While the K_m values $(0.103 \pm 0.011\ mM)$ recorded for most mutants spin labeled in the C-terminal end or middle of the protein agreed well with those in the literature, their k_{cat} values $(0.028 \pm 0.005\ s^{-1})$ were consistently lower, approximately one-eighth of the published values. This might suggest that the intrinsically disordered protein portion of the fusion construct interacts with the protease while bound to it, interfering with the expected catalytic activity though not affecting substrate binding. Another interesting result is that the presence of the neutral-polar spin label close to the peptide 'cut site' itself (i.e., a nitroxide at residue 3 in stathmin) appears to dramatically increase the binding affinity of the peptide to the protease, as evidenced by a notably smaller K_m value $(0.022 \pm 0.003\ mM)$. This suggests that TEV protease activity is strongly dependent on not just the sequence of peptide cut-site, but also the polarity and charge make-up of nearby residues.

Geochemical Correlation of Basalts in Northern Deep Springs Valley, California, by X-Ray Fluorescence Spectroscopy (XRF)

Department of Geological Sciences, California State University, Fullerton

Aaron Justin Case
Advisor: **Jeffrey R. Knott, Ph.D.**

Abstract

In northern Deep Springs Valley (DSV), between Owens Valley and Death Valley, California, Miocene-Pliocene-age, olivine basalts lie on the valley floor and atop the adjacent White/Inyo Mountains to the west and the Deep Springs Range to the east. Previous geologic mapping shows the DSV basalt flows and the Last Chance Range (LCR) basalts found to the southeast are the same geologic unit with a source in the White/Inyo Mountains. The basalts in northern DSV are offset ~400 m by the Deep Springs fault and have a K/Ar age of 10.8 Ma.

To determine if the olivine basalts found in the region are all from the same source, four samples were collected in a linear pattern from west to east across northern DSV. The samples were powdered and analyzed for major and trace element composition by X-Ray Fluorescence spectrometer (XRF).

Trace-element plots (e.g., Ba, Nb, Zr, Y, Ce, etc.) show that the DSV basalts are similar and are likely from the same source; however, the composition of the DSV basalts are distinct from the LCR basalts. I interpret this data by presenting the DSV and LCR basalts have different sources and should not be mapped as the same geologic unit. The likely source of the DVS basalts is in the White/Inyo Range. The geochemical correlation shows that the DSV basalts flowed NW to SE in a paleochannel 10.8 Ma and that DSV did not exist at that time.

Abstract

Questions exist about the possible dispersal pathways of various ancient species such as the pupfish (Knott et al., 2008; Phillips, 2008; Echelle, 2008). Early hypotheses suggested that pupfish dispersed along the ancient rivers and lakes of eastern California about 20,000 years ago (Blackwelder, 1933); however, biological studies show that this is an insufficiently short time period and a more realistic time is about 4 Ma (Smith et al., 2002). A 4,000,000 year time frame allows for plate tectonic motions to impact migration pathways. Reheis and Sawyer (1997) hypothesized that basalt flows of the White Mountains occupied ancient river channels and flowed possibly as far southeast as Death Valley. Basalt flows found atop the White Mountains, Deep Springs Range (DSR) and Last Chance Range (LCR) are all mapped as the same geologic unit (Tb; McKee and Nelson, 1967; Wrucke and Corbett, 1990) presenting the possibility that one flow emanated from the White Mountains and possibly reached Death Valley, thus showing the possible ancient river trace that would connect pupfish populations.

In this study, I present geochemical analyses of the basalt flows of the northern DSR. These basalts are the first key link between the White Mountains and Death Valley. I will then compare these data with basalts atop the LCR to the southwest to try and show that these are or are not the same basalt flow and potential pathway for pupfish dispersal.

Figure 1: *Eastern California/ Southern Nevada area showing mountain ranges and valleys. The box is the research area.*

Location

Deep Springs Valley (DSV) is located about 37±1 kilometers (~23±1 miles) east of Bishop, California. DSV is a 25 kilometers long closed basin between the southern White Mountains and the northern Inyo Mountains (Figure 1). The olivine basalts crop out in the north end of DSV (Figure 2). The east side of DSV is bounded by the DSR. At the base of the DSR is the north-northeast striking, normal slip Deep Springs fault zone (Reheis and Sawyer, 1997).

Background

Observations indicate that ancient basalts flowed downhill along ancient riverbeds (Dalrymple, 1963). Dalrymple (1963) determined by whole-rock K/Ar that the olivine basalts with locally scoriaceous features in the DSV and the White Mountains are 10.8±0.1 Ma. A tuff below the basalt yields an age of 10.8±0.1 Ma.

McKee and Nelson (1967) mapped olivine basalt flows (Tb) in the northern DSV, DSR and LCR (Figure 3). These flows span from the Inyo Mountains east to Piper Mountain in the DSR. McKee and Nelson (1967) show the basalt offset by strands (west side down) normal faults of the Deep Springs fault zone (Figure 4). Most of these normal faults do not offset the older alluvium (Qoa); however, one north-trending normal fault produces a scarp in the Qoa. All of the faults are buried by the younger alluvium (Qa) and alluvial fan (Qf) deposits. Thus, the Deep Springs fault zone began sometime after 10.8 Ma, but has not produced ground rupture since deposition of the Qoa.

Krauskopf (1971) listed a 4.8 Ma K/Ar age for olivine basalts with locally scoriaceous features (Tb); however, this basalt is actually not in the White Mountains, but across Fish Lake Valley to the northeast. Krauskopf mapped a basalt overlying rhyolite tuff near Cottonwood Creek in the White Mountains, which is the likely location of Dalrymple's (1963) 10.8 Ma basalt.

Reheis and Sawyer (1997) proposed that basalts of the Deep Springs/Eureka Valley area flowed from the White/Inyo Mountains to the LCR and had the same source. They wrote "major oxide and trace elements analyses indicate that the basalts are from the same sequence of flows or at least share a common parentage" (Reheis and Sawyer 1997, pg. 284). They showed paleo-channels locations and report the 10.8±0.1 Ma K/Ar age of Dalrymple (1963) for the basalt at Piper Mountain. Correlation of the Piper Mountain basalt with those dated in the White Mountains by Dalrymple (1963) is reasonable; however, McKee and Nelson (1967) mapped sedimentary rocks below the Piper Mountain basalt, not a rhyolite like Dalrymple (1963). One drawback to the correlation of basalts in the White/Inyo

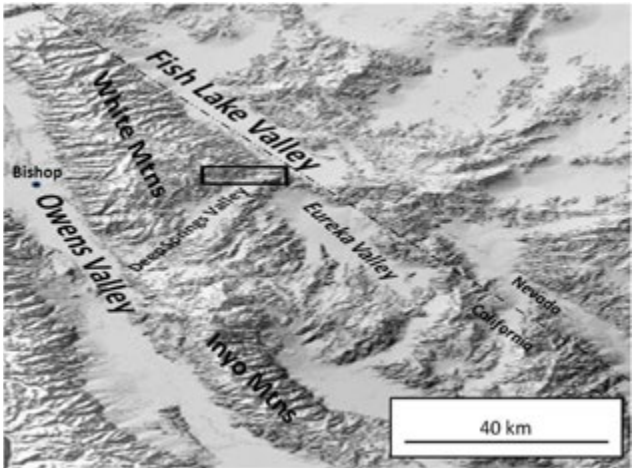


Figure 2: *Deep Springs Valley area showing Mountain Ranges and Valleys zoomed in. The box is the research area.*

Range is that there are no published geochemical analyses of the White Mountain basalts.

Ormerod et al. (1988; 1991) showed that basalts of the Great Basin have a unique ratio of Zr/Ba. They used these two elements rather than Nb because the concentration of Nb is relatively low and subject to analytical error. Zr and Ba are both relatively immobile elements in basalt magmas and Ormerod et al. (1988; 1991) showed that the ratio is consistent and indicative of magma source. They found that basalts with Zr/Ba ratios <0.2 are greater than 5 Ma at the latitude of Bishop and indicate a lithospheric contamination. In contrast, basalts with a Zr/Ba ratio >0.2 are younger than 5 Ma and indicate an asthenosphere source.

Kempton et al. (1991), also working on Great Basin basalts, found that there was significant variation in Ce and Y concentrations. They used the ratio of Ce/Y to illustrate the fractionation of rare-earth elements in various magma sources.

Pluhar et al. (2005) showed that X-ray Fluorescence spectroscopy (XRF) effectively sorted out different basalt flows of the Coso Volcanic field, California, just southwest of Deep Springs Valley and part of the Great Basin. Many of the Coso basalts were thought to be the same; however, Pluhar et al. used Harker diagrams and plots of Rb/Nb vs. Sr/Zr to segregate different basalts from the same source. Manoukian (2012) analyzed basalts from the LCR by XRF as well. Miller and Wrucke (1995) reported a K/Ar age of 5.5 Ma for the basalt analyzed by Manoukian.

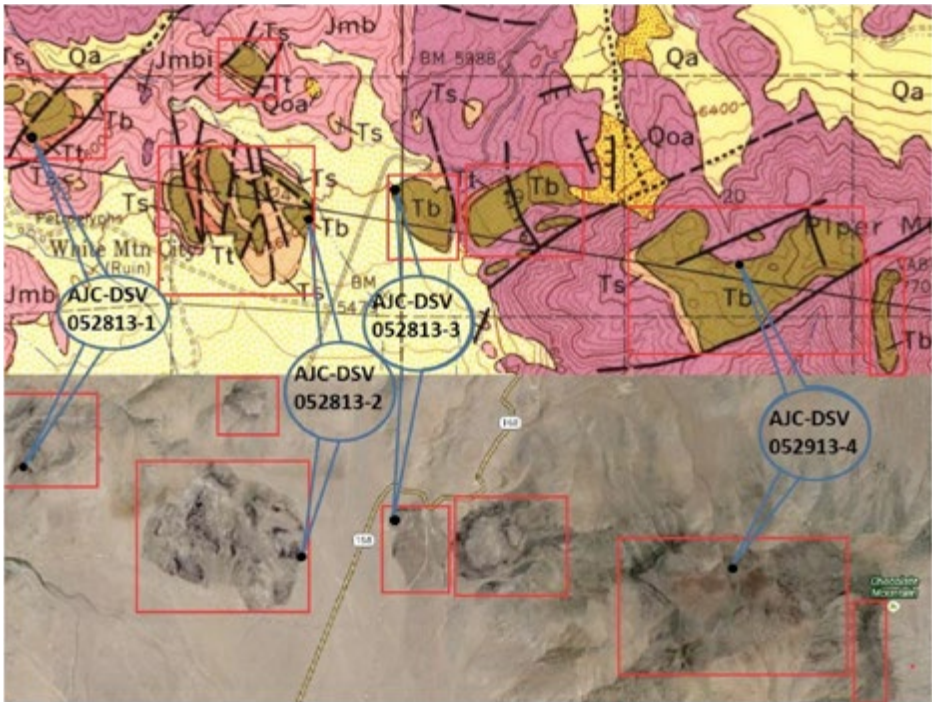


Figure 3: *The upper part isa portion of the Soldier Pass 15′ quadrangle geologic map by McKee and Nelson (1967). The lower portion is a Google Earth image of the same area. The unit Tb in the red boxed area is olivine basalt and the sample locations are shown on both. The cross section along A-A′ is shown in Figure 4.*

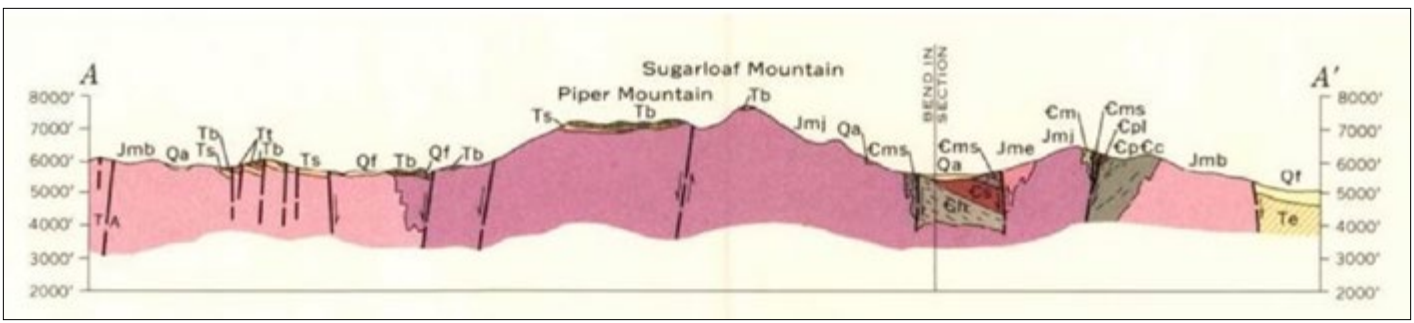


Figure 4: *Cross section of northern Deep Springs Valley by McKee and Nelson (1967).*

Figure 5: *Photograph (above) of outcrop where AJC-DSV-0528-13-2 was collected. In the low background is the location of sample 3 with the location of sample 4 on the skyline atop Piper Mountain. At left is a close-up of sample 2 location.*

Objective

I will test the following hypotheses by completing XRF analysis on basalt samples of northern DSV:

1. If the olivine basalts in northern DSV are the same, then they should all have the same geochemical composition.
2. If the olivine basalts in northern DSV are different, then they should all have different geochemical compositions.
3. If the olivine basalts in northern DSV and the LCR have the same source, then the geochemical compositions would be the same proving that the olivine basalts came from the LCR.
4. If the olivine basalts in northern DSV and the LCR have different sources, then the geochemical compositions would be different which would prove the source had a different origin such as the White Mountains located in the northwest.

Methods

I collected basalt samples (Figure 5) from four locations in northern Deep Springs Valley. Each sample will be described in the field and the location recorded using a hand-held GPS (Table 1.)

I used XRF to determine the geochemical composition of each sample collected. The XRF uses x-rays to excite the electrons of the elements. When the electrons become excited, they jump an electron shell and when returning back to its original state, energy is released as light. The amount and wavelength of the light indicates the quantity and types of major and trace elements present. The data was analyzed and data reduction completed at Pomona College.

Table 1: Universal Transverse Mercator coordinates of samples collected. Coordinates were determined with a hand-held GPS receiver and are in sections 11S

AJC-DSV-052813-1	AJC-DSV-052813-2	AJC-DSV-052813-3	AJC-DSV-052913-4
411661E	413804E	414430E	416827E
4141290N	4140912N	4141141N	4140683N

Results and Discussion

The main results are XRF analysis of whole-rock samples (Table 2). The data collected is relatively well behaved. One sample (AJC-DSV-052813-4) was randomly chosen to be run twice to measure accuracy. For the duplicate samples, the Relative Standard Deviation (RSD) ranged from 0.0% to 254%. The highest RSD was for Ta at 491.9%. The concentrations of Ta for all five AJC samples ranged from -2.5 ppm to 2.9 ppm The negative values clearly indicate poor data, as a result, Ta was no longer considered for data analysis.

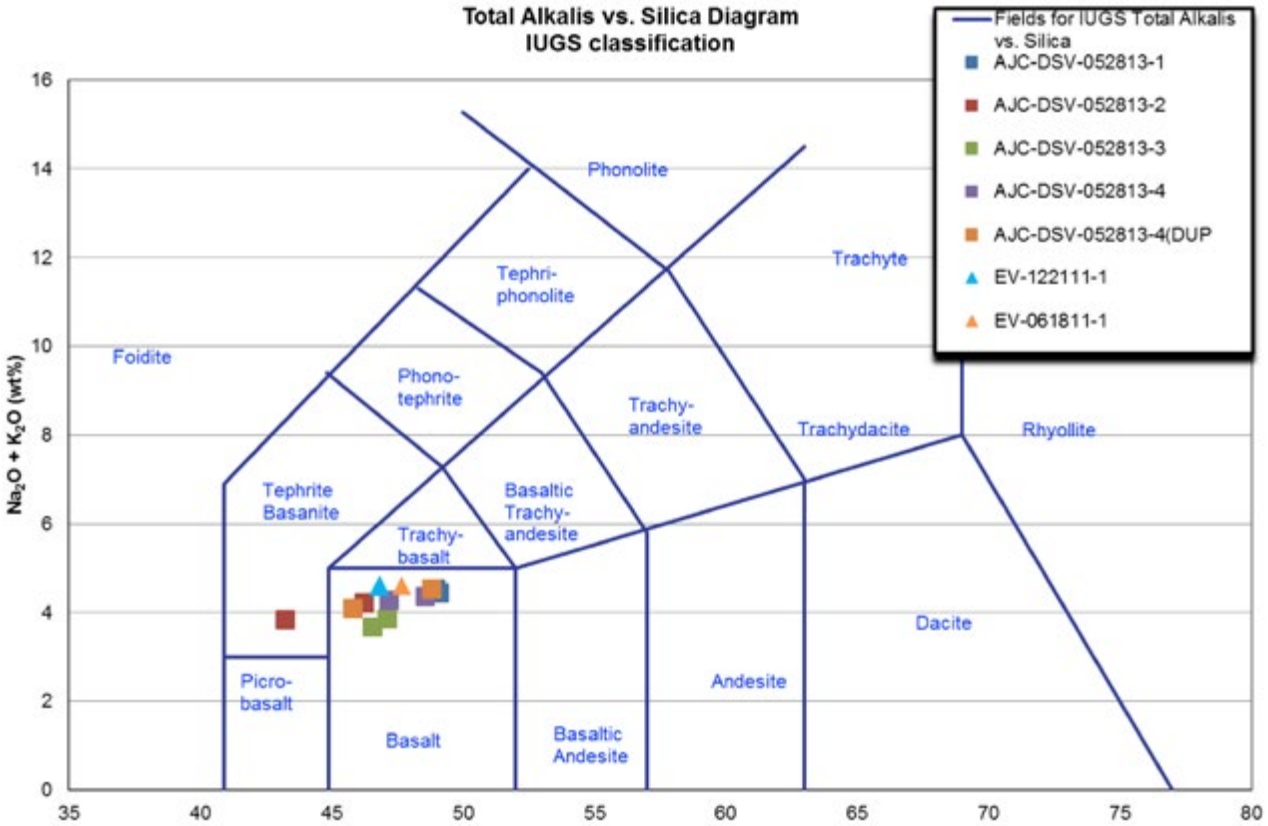
Beside Ta, U and Th also had RSD results of 47% and 70%, respectively. The concentration of U ranged from 0.7 to 4.9 ppm whereas the concentration of Th ranged from 0.9 to 4.3 ppm. These concentrations are very low and the range produces a mean with a high standard deviation. The explanation for the wide range of concentration of U and Th is probably related to the large atomic radius that may result in excessive interferences.

Excluding Ta, U and Th, the RSDs for the remaining elements and oxides for the duplicate samples range from 10-15% (Sm, Nd, Pb), 5-10% (Cr, Cu, Ni, Zn, Zr, Ce, La, Nb) and less than 5% for the remaining elements and oxides. In general, I regarded RSDs less than 10% for the duplicate analyses as acceptable and indicate that the particular concentration of that element or oxide was acceptable for data evaluation.

The Total Alkali Silica (TAS) diagrams (Figure 6) distinguish igneous rocks geochemically using the Silica (Si) and Total Alkali (K_2O + Na_2O). The DSV samples plot in the basalt field.

Sample	AJC-DSV-052813-1	AJC-DSV-052813-1	AJC-DSV-052813-2	AJC-DSV-052813-2	AJC-DSV-052813-3	AJC-DSV-052813-3	AJC-DSV-052813-4	AJC-DSV-052813-4	AJC-DSV-052813-4(DUP	AJC-DSV-052813-4(DUP	EV-122111-1	EV-061811-1
SiO2 (%)	48.97	49.1	46.24	43.24	47.11	46.55	48.58	47.2	48.81	45.82	46.85	47.67
Al2O3 (%)	15.42	15.43	13.97	13.06	15.1	14.91	16.28	15.81	16.33	15.37	15.03	15.24
Fe2O3 (%)	9.69	9.72	12.38	11.22	9.51	9.35	9.24	8.92	9.37	8.61	10.92	10.73
MgO (%)	7.28	7.31	8.21	7.66	9.34	9.21	7.28	7.12	7.29	6.87	7.4	7.86
MnO (%)	0.15	0.15	0.18	0.16	0.16	0.15	0.15	0.14	0.15	0.14	0.16	0.16
CaO (%)	10.2	10.22	10.79	10	10.92	10.79	9.87	9.63	9.86	9.17	10.18	9.18
Na2O (%)	3.03	2.96	2.78	2.53	2.47	2.32	2.72	2.67	2.88	2.57	2.83	2.83
Ba	1008.4	1024.2	1701.5	1569.7	1038.2	1019.4	1048.5	1007	1055.5	971.7	1072.6	1053.05
Rb	27.5	28.8	28.4	32.4	25.6	26	33.2	34.2	30.8	35.1	30.9	31.79
K2O (%)	1.5	1.49	1.44	1.31	1.39	1.35	1.64	1.6	1.65	1.52	1.75	1.76
Nb	10.3	10	27.4	24.4	9.1	9.4	10.7	8.6	9.1	7.8	20.2	21.64
La	40.3	37.1	69.3	54.6	52.2	50.9	45.9	43.7	45.4	49.4	50	51.49
Ce	81.1	77.1	134.9	112.5	105.4	103.4	94.3	92.8	97.5	102.4	113.6	106.74
Sr	929	933.1	1291.6	1165.1	1101.5	1078.2	1078.5	1051.1	1080.1	982	945.6	927.71
Nd	39	25.5	62.6	55.1	42.3	45.3	45.6	43.6	45.8	53.3	47.5	49.23
P2O5 (%)	0.52	0.52	0.96	0.9	0.53	0.53	0.53	0.51	0.54	0.51	0.62	0.6
TiO2 (%)	1.37	1.37	2.33	2.14	1.19	1.17	1.25	1.22	1.25	1.15	1.61	1.63
Sm	6.3	6.3	9.2	7.7	7.3	8.9	8.2	7.2	7.2	8.4	4.7	6.45
Zr	188.7	188.1	252.2	222.9	174.7	170.2	186.1	179.8	187.5	166.7	248.3	249.39
Hf	3.8	3.6	4.6	3.8	3.8	3.1	4.1	3.4	3.4	3.2	4.4	5.1
Y	19.9	19.9	29.6	29.2	22.5	21.9	22.1	22.3	22.6	23.2	22.4	22.29
Zn	94.2	94.9	125.5	107.5	100.2	98.2	102.1	95.5	98.8	83.9	105.8	105.91
V	209.6	210.3	237.3	210.1	217.5	209.6	205.5	195.2	208.2	193.4	224.2	215.64
U	0.8	0.7	1.9	4.9	-2	1.2	0.3	2	2.9	1	2.9	0.53
Th	1	0.9	1.6	4.3	-0.2	2.9	2.3	1.8	2.2	0.6	2	2.1
Ta	1.9	2.9	-0.9	-2.5	-1.2	-1.6	1.2	2.8	2.3	-0.8	-0.3	-0.54
Sc	25.9	25.9	28.4	27.7	28.5	25.4	25.6	23	26.4	22.3	24	24.06
Pr	6	10.4	15.3	9.6	7.1	10.8	13.2	10	8.9	9.6	11.3	11.51
Pb	4.8	2.8	-2.5	0.8	-1.6	0	2.3	4.2	5.8	5.1	10.7	10.2
Ni	204.4	206.2	175.6	157.4	170.1	165.2	128.1	126.8	126	112.7	147.7	141.16
Ga	18.5	18.6	19.8	16.7	18.5	16.9	18	17.7	18.5	16.6	19.6	19.43
Cu	41.2	42.5	49.1	40.9	37.8	37.2	36.3	35.1	38.1	31.5	61.3	55.66
Cr	407.5	410.9	340.8	320.7	497.5	487.9	343.2	343.5	342.4	315.3	296.5	282.43

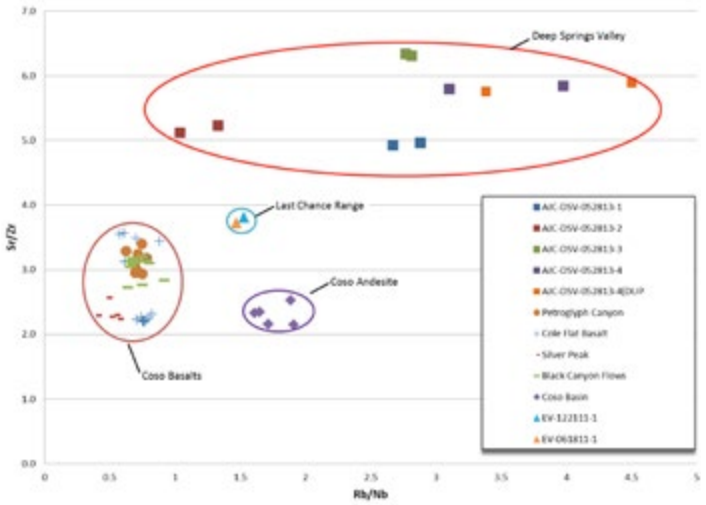
Table 2: *X-Ray Fluorescence Data. Concentrations are in parts per million unless labeled as weight percent (%).*

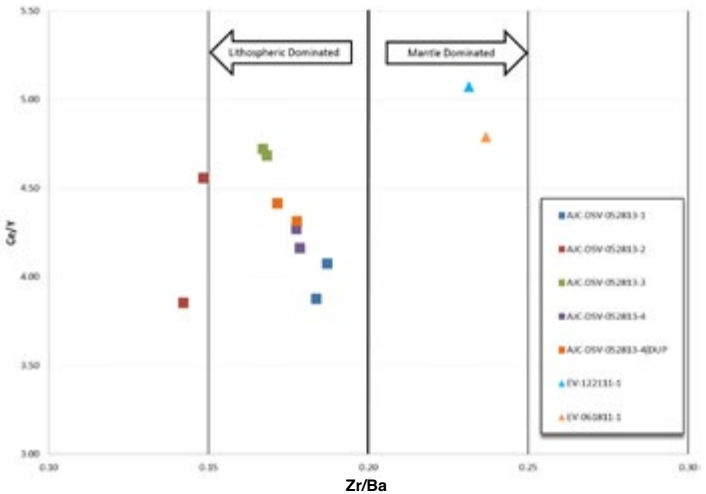


Pluhar et al. (2005) used ratios of Rb/Nb and Sr/Zr for "lava fingerprinting". A plot of Rb/Nb vs. Sr/Zr shows that the DSV basalts are distinct from the Coso basalts (Figure 7). The DSV basalts show quite a range of Rb/Nb ratios. This is likely the result of the relatively low Nb concentration as described by Ormerod et al. (1988). Ormerod et al. (1988; 1991) and Kempton et al. (1991) both noted that the unusually low concentration of Nb led to potential analytical errors. They found that ratios of Zr/Ba and Ce/Y adequately distinguished basalt flows of differing age and tectonic setting.

A plot of Zr/Ba vs. Ce/Y shows that the 10.8 Ma DSV basalts have a ratio of Zr/Ba <0.2 (Figure 8). This is consistent with Ormerod et al.'s hypothesis that basalts >5 Ma have Zr/Ba ratios <0.2 as a result of lithospheric contamination by the subducted Farallon plate. Plotting Manoukian's XRF results on the Rb/Nb vs. Sr/Zr diagram shows that the LCR basalts plot in a separate cluster from both the Coso and DSV basalts (Figure7). Plotting the LCR basalts on the Zr/Ba vs. Ce/Y diagrams (Figure 8) shows that the LCR basalts have a Zr/Ba ratio >0.2. A ratio of >0.2 is attributed to mantle magmas <5 Ma and a asthenosphere source. This >0.2 Zr/Ba ratio conflicts with Miller and Wrucke's (1995) whole-rock 5.5 Ma K/Ar age. One possible explanation is that Ormerod et al.'s observations are inconsistent. Alternatively, the K/Ar date is incorrect. The latter is more likely considering that a preliminary 40Ar/39Ar date (sanidine) on a rhyolite tuff that underlies the basalt is ~3.5 Ma.

The linear trend of Ce/Y shows that AJC-DSV-052813-1, -3 & -4 are from the same flow (Figure 8). The offset of AJC-DSV-052813-2 from the others suggests that this was separate flow or some process impacted the Zr/Ba ratio as the basalt cooled.





Conclusions, Significance, and Future Works

Geochemical data (Zr/Ba vs. Ce/Y) show that the DSV basalts are different from the LCR basalts in a similar geomorphic position atop each range. The Zr/Ba ratio indicates that DSV basalts erupted prior to 5 Ma whereas the LCR basalts erupted after 5 Ma or after the Mendocino triple junction move north of Bishop. This different Zr/Ba ratio indicates that the DSV basalts and LCR basalts had different magma sources.

Because the olivine basalts in the northern DSV are the same, then when the flows erupted, there was a topographic low (e.g. ancient river channels) that allowed the basalts to flow southeast from the White/ Inyo mountains to the Deep Springs Range. That would indicate that Deep Springs Valley and the Deep Springs Range did not exist. If this is true, another hypothesis arises in which we can calculate the minimum slip rate of the Deep Springs fault zone by taking the ~400m offset of the basalt flow and the 10.8±0.1Ma age to determine an estimated minimum 0.04 mm/yr slip rate.

These data show that there are different basalts of differing age in the Deep Springs/Eureka/Last Chance region. Correlation of the DSV basalts indicates that a flow from the White Mountains traveled to at least Piper Mountain in the DSR and shows that there was a possible river channel that connected the White Mountains with the LCR to the southeast. As to whether the channel extended further east to Death Valley or turned south is unknown; however, this potential dispersal pathway is not eliminated at this time.

Future work should include sampling of additional basalts found to the northwest and southeast of DSV. Some new geochronology of the basalts should also be done to improve the existing K/Ar chronology. Additional analytical data, such as mass spectrometry, should be done to improve the precision and accuracy of the geochemical character of the rocks.

References

Blackwelder, E., 1933, Lake Manly, an extinct lake of Death Valley: Geographical Review, v. 23, p. 464-471.

Dalrymple, G. B., 1963, Potassium-argon dates of some Cenozoic volcanic rocks of the Sierra Nevada, California: Geological Society of America Bulletin, v. 74, p. 379–390.

Echelle, A., 2008, The western North American pupfish clade (Cyprinodontidae: Cyprinidon): Mitochondrial DNA divergence and drainage history, in Reheis, M. C., Hershler, R., and Miller, D. M., eds., Late Cenozoic drainage history of the southwestern Great Basin and lower Colorado River region: Geologic and Biotic perspectives, Volume Special Paper 439: Boulder, CO, Geological Society of America, p. 27-38.

Kempton, P. D., Fitton, J. G., Hawkesworth, C. J., and Ormerod, D. S., 1991, Isotopic and trace element constraints on the composition and evolution of the lithosphere beneath the southwestern United States: Journal of Geophysical Research, v. 96, no. B8, p. 13,713-713,735.

Knott, J. R., Machette, M. N., Klinger, R. E., Sarna-Wojcicki, A. M., Liddicoat, J. C., Tinsley, J. C., David, B. T., and Ebbs, V. M., 2008, Reconstructing late Pliocene - middle Pleistocene Death Valley lakes and river systems as a test of pupfish (Cyprinodontidae) dispersal hypotheses, in Reheis, M. C., Hershler, R., and Miller, D. M., eds., Late Cenozoic drainage history of teh southwestern Great Basin and lower Colorado river region: Geologic and biotic perspectives: Boulder, Colorado, Geological Society of America Special Paper 439, p. 1-26.

Krauskopf, K. B., 1971, Geologic map of the Mt. Barcroft Quadrangle, California–Nevada: U.S. Geological Survey Geologic Quadrangle Map GQ-960, scale 1:62 500.

Manoukian, D. N., June 7, 2012, Geology of Cenozoic Deposits on top of the Last Chance Range, Death Valley National Park, California. A Thesis Presented to the Faculty of California State University, Fullerton. Department of Geological Sciences. Faculty Advisor Jeffrey R. Knott, Ph.D.

McKee, E.H. and Nelson, C.A., 1967, Geologic map of the Soldier Pass quadrangle, California and Nevada: U.S. Geological Survey, Geologic Quadrangle Map GQ-654, scale 1:62,500

Miller R.J. and Wrucke C.T., 1995, Age, Chemistry, and Geologic Implications of Tertiary Volcanic Rocks in the Last Chance Range and part of the Saline Range, Northern Death Valley Region, Califnornia. U.S. Geological Survey, Menlo Park, California 94025. ISOCHRON/WEST, no. 62, May 1995. p. 30-36.

Ormerod, D. S., Hawkesworth, C. J., Rogers, N. W., Leeman, W. P., and Menzies, M. A., 1988, Tectonic and magmatic transitions in the Western Great Basin, USA: Nature, v. 333, p. 349-353.

Ormerod, D. S., Rogers, N. W., and Hawkesworth, C. J., 1991, Melting in the lithospheric mantle: Inverse modeling of alkali-olivine basalts from the Big Pine Volcanic Field, California: Contributions to Mineralogy and Petrology, v. 108, p. 305-317.

Phillips, F., 2008, Geological and hydrological history of the paleo-Owens River drainage since the Miocene, in Reheis, M. C., Hershler, R., and Miller, D. M., eds., Late Cenozoic drainage history of teh southwestern Great Basin and lower Colorado river region: Geologic and biotic perspectives: Boulder, Colorado, Geological Society of America Special Paper 439, p. 115-150..

Pluhar et al., Glen, J.M., Monastero, F.C., Tanner, S.B., 2005, Lava fingerprinting using paleomagmatism and innovative X-ray fluoresence spectroscopy: A case study from the Coso volcanic field, California. Vol. 6, Issue 4.

Reheis, M. C. Sawyer, T. L. March, 1997,Late Cenozoic history and slip rates of the Fish Lake Valley, Emigrant Peak, and Deep Springs fault zones, Nevada and California. Geological Society of America (GSA) : Boulder, CO, United States Vol. 109, Issue 3, pp. 280-299.

Smith, G. R., Dowling, T. E., Gobalet, K. W., Lugaski, T., Shiozawa, D. K., and Evans, R. P., 2002, Biogeography and timing of evolutionary events among Great Basin fishes, in Hershler, R., Madsen, D. B., and Currey, D. R., eds., Great Basin Aquatic Systems History, Volume 33: Washington, D.C., Smithsonian Institution Press, p. 175-234.

Wrucke, C. T., and Corbett, K. P., 1990, Geologic map of the Last Chance Quadrangle, California: U.S. Geological Survey Open File Report 90-647-A, scale 1:62,500.

Possible Shell-Hash Tsunami Deposit at the Los Penasquitos Marsh, San Diego County, CA.

Department of Geological Sciences, California State University, Fullerton

Jeremy Cordova
Advisor: Dr. Brady Rhodes

Abstract

The Los Penasquitos Marsh is one of a series of coastal wetlands between San Diego and Orange County that formed within stream valleys that flooded and filled with sediment during Holocene sea-level rise. In order to test the hypothesis that these wetlands contain prehistoric tsunami deposits, within the normal wetland sediments, 21 reconnaissance cores between 48 and 321cm in length were collected and described in the field. Nearly all the sediment cores contained a peaty layer in the top 5-25cm, underlain by dark brown and black mud, above interbedded fine-medium gray sand and mud. The stratigraphy in the cores is consistent with the complete infilling of a lagoon behind a baymouth bar of sand, a 'drowned river valley'. Five of the cores taken, ranging from 1.0-1.5 km inland from the present day beach, intersect a distinctive 0.5 – 12.0 cm-thick shell-hash and muddy sand layer between 233 and 280cm depth.

Based on this discovery we collected a 285cm long 5cm diameter piston core to further analyze this possible tsunami deposit. In this larger core the 10 cm-thick shell hash layer consists of angular shell fragments up to 3cm in size, in a muddy sandy matrix that includes the following genera: Mitrella, Venus, Spirotropis, Pecten, Nassarius, and an unidentified oyster. This fossil assemblage suggests a quiet water marine source for the shell hash debris– from the lagoon and/or offshore, not typical of low energy fine grained comparatively well sorted wetland sediments. The core was analyzed for loss on ignition (LOI) at both 550° and 950°C and magnetic susceptibility (ms). The LOI550 data is unremarkable, and the LOI950 data shows an expected spike in mass percent of carbonates within the shell-hash layer. The ms data shows low values for the lagoonal muds and sands, but a pronounced spike within the shell hash layer, where the average reading more than doubles from 2.0 to 4.6. We hypothesize that the anomalously high ms value for the shell hash layer indicates a substantial component from an offshore source, where heavier magnetic minerals accumulated seaward of the baymouth bar, but were subsequently swept back inland into the lagoon. If correct this layer may represent a large-wave event, either storm or tsunami. Three C-14 dates on shell fragments cluster between 1380-1420 yrs BP.

Paleotsumani Research at the Seal Beach Wetlands, Seal Beach, CA

Department of Geological Sciences, California State University, Fullerton

D'lisa O. Creager, Brady P. Rhodes, Matthew E. Kirby, and Robert J. Leeper

Abstract

The Seal Beach Marsh is located inside the Seal Beach Naval Weapons Station in north Orange County, CA. The wetland formed as a result of flooding and infilling of topographic lowlands during early Holocene sea-level rise. The Seal Beach marsh may contain a record of prehistoric tsunami and other paleoseismic data because the marsh is a low-energy depositional environment and historic anthropogenic disturbance is limited. To test if the marsh has a record of tsunami, sixteen reconnaissance gouge cores between 150 and 240 cm in length were collected and described in the field. The reconnaissance cores showed peaty organic layers interbedded with mud and sand. To investigate the stratigraphy at greater depths, a 377-cm vibracore was collected. Preliminary analyses of the vibracore show the top 15 cm is modern marsh. From 15 to 107 cm below land surface (bls), peaty mud and mud of varying thicknesses are interbedded. At 118 cm bls, a 10-cm thick sand layer covers mud at a sharp irregular contact. A 10-cm sand layer with an irregular basal contact at 137-cm bls covers peaty mud that consists of 50% organic matter. Peaty mud transitions to mud at 140 cm bls. Alternating mud and muddy sand layers of varying thickness continue to 246 cm bls. At 250 cm bls, a 2-cm thick mud layer caps a muddy peat layer. Mud at 270 cm bls extends down until a sharp irregular contact is made with a sand layer at 356 cm. A 21-cm thick sand layer marks the base of sediment recovery in the vibracore. The core was analyzed for loss on ignition (LOI) at 550°C (% total organic matter) and 950°C (% total carbonate) as well as magnetic susceptibility (CHI) at 1-cm intervals. These analyses confirm the existence of several organic-rich zones alternating with organic-poor mud. Our working hypothesis is that these peaty layers represent repeated subsidence of the marsh, perhaps related to seismic activity on the Newport-Inglewood Fault zone.

Searching for a Paleotsunami Record at Seal Beach, Southern California

Department of Geological Sciences, California State University, Fullerton

Garcia, Dylan J., Kirby, Matthew E., Rhodes, Brady P., Leeper, Robert J.

Abstract

There has been a substantial amount of research concerning southern California's potential and vulnerability to tsunami activity. Near-field or local tsunamis can be generated by submarine landslides and/or earthquakes. As many as 20 tsunamis have breached the coast of California within the past two centuries, but none have been reported since 1800 (Eisner et al., 2001). One of the earliest is thought to be the Santa Barbara, southern California tsunami of 1812, which damaged over 60km of Santa Barbara's coast (Toppozada et al., 1981; Lander et al., 1993). Southern California need only generate a M > 6.5 earthquake along one of the offshore fault lines to create a near-field tsunami (Mcculloch, 1985). Seal Beach lies above an active right-lateral strike-slip fault known as the Newport-Inglewood fault zone (NIFZ) (Vedder, 1975; Wright, 1991; Grant et al., 1999). The Seal Beach environment is a known lowland trough and susceptible to fault zone activity causing this localized depression. Today, the coast is heavily developed and populated thus making tsunami research important.

A 344cm sediment core extracted from the subsurface of the intertidal salt marsh at Seal Beach, Orange County, southern California was examined to determine if there is any evidence for past tsunamis. The objective of this study is to identify paleotsunami units in southern California's marsh stratigraphy and to compare the sedimentological data to that of others extracted from southern California, and elsewhere. Tsunami deposits are identifiable by the types of deposits left in the sediment. High energy transportation is usually propagated by tsunamis or storms capable of carrying fairly large (≤0.5 m diameter) shells, coral, and sand in low energy environments such as swamps, mangroves, and marshes. Discerning the two has been the focus of several studies (Nanayama et al., 2000; Goff and McFadgen, 2004; Tuttle et al., 2004; Morton et al., 2007). Because tsunamis can travel several hundreds of meters inland, they are capable of leaving fairly thick deposits, some of which have been calculated up to 25cm on the Pacific coast. The coarser material tends to settle first, followed by the finer material due to wave deceleration (Morton et al., 2007). Grain size characterized by tsunamis can range from mud to boulders, but are commonly sand sized particles (Kawata et al., 1999). Deposits can be mixed under peat, given the wetland environment. Single or several beds of normally graded sand, mud laminae, and coral and shell debris all suggest tsunami deposits (Rhodes et al., 2011). Furthermore, many tsunami deposits thin landward.

Core sediment was attained using a vibracore, which penetrated the subsurface to a depth of 344cm. The core was split, described, photographed, sampled, capped, and stored in a cold storage facility at California State University, Fullerton. Multi-proxy methods were used to describe the core including: descriptive analysis, magnetic susceptibility (CHI), and loss on ignition (LOI) at 550°C (% total organic matter) and 950°C (% total carbonate). The stratigraphic record preserved in the subsurface sediments of this coastal wetland does not show definitive evidence of rapid debris transport. We see none of the indicators that characterize a tsunami deposit. Alternating sand and mud layers were identified both visually and through magnetic susceptibility. These sand layers do not necessarily indicate tsunamis, but perhaps a dynamic environment of materials washing into Seal Beach. High CHI values (above 5×10^{-7} $m^3 kg^{-1}$) represent a sandy or muddy magnetite-rich environment from a terrestrial source, such as a beach, stream flow, or flooding. There is an unusual sand intrusion feature (Unit I) between 202cm - 215cm depth. Unit I is an interesting interface of mud and sand, with softer overlying sediment surrounding a denser sand injection. This feature may have been due to subsidence caused by NIFZ activity. The structure could also be a result of liquefaction. It closely represents that of a sand or clastic dike, which are associated with storm events and earthquakes. Another common feature in stratified sands are flame

structures, which resemble this intrusion. These flame structures are formed from seismic activity and density variation between the stratified layers. Further research may characterize this structure and thus the causal event.

References

Eisner, R.K., J.C. Borrero and C.E. Synolakis, 2001. Inundation maps for the State of California. In Proceedings of the International Tsunami Symposium 2001 (ITS 2001) (on CDROM), NTHMP Review Session, R-4, Seattle, WA, 7–10 August 2001, 67-8.

Goff, J., McFadgen, B.G.C.-G.C., 2004. Sedimentary differences between the 2002 Easter storm and the 15th-century Okoropunga tsunami, southeastern North Island, New Zealand. Marine Geology 204, 235–250.

Grant, L. B., K. J. Mueller, E. M. Gath, H. Cheng, R. L. Edwards, R. Munro, and G. L. Kennedy, 1999. Late Quaternary uplift and earthquake potential of the San Joaquin Hills, southern Los Angeles basin, California, Geology 27, 1031–1034.

Kawata, Y., B. Benson, J.C. Borrero, J.L. Borrero, H.L. Davies, W.P. deLange, F. Imamura, H. Letz, J. Nott, and C.E. Synolakis, 1999. Tsunami in Papua New Guinea was as Intense as First Thought, Eos Trans. AGU, 80, # 9, 101, 104-105.

Lander, J., P.Lockridge, and M. Kozuch (Eds.), 1993, Tsunamis Affecting the West Coast of the United States, 242 pp., U.S. Department of Commerce, NOAA.

Mcculloch, D. S., 1985. Evaluating tsunami potential. In Evaluating Earthquake Hazards in the Los Angeles Region (J. I. Ziony, ed) U.S. Geolog. Survey Prof. Paper 1360, pp. 375–413.

Morton, R.A., Gelfenbaum, G. and Jaffe, B.E. 2007. Physical criteria for distinguishing sandy tsunami and storm deposits using modern examples, Sedimentary Geology, 200, 184–207.

Nanayama, F., Shigeno, K., Satake, K., Shimokawa, K., Koitabashi, S., Miyasaka, S., Ishii, M., 2000. Sedimentary differences between the 1993 Hokkaido-nansei-oki tsunami and the 1959 Miyakojima typhoon at Taisei, southwestern Hokkaido, northern Japan. In: Shiki, T., Cita, M.B., Gorsline, D.S. (Eds.), 15th International Sedimentological Congress. Elsevier, pp. 255–264.

Rhodes B.P., M.E. Kirby, K Jankaew, M Choowong, 2011. Evidence for a mid-Holocene tsunami deposit along the Andaman coast of Thailand preserved in a mangrove environment. Mar. Geol. 282: 255-267.

Ratcliff, F., Martin-Hernandez, A.M.and Baulcombe, D.C. 2001. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J., 25, 237-245.

Toppozada, T., C. Real, and D. Parke, 1981. Preparation of isoseismal maps and summaries of pre– 1900 California Eqs., CDMG # 81-11SAC, p. 34, 136–140.

Tuttle, M.P., Ruffman, A.A.T., Jeter, H., 2004. Distinguishing tsunami from storm deposits in Eastern North America; the 1929 Grand Banks tsunami versus the 1991 Halloween storm. Seismological Research Letters 75, 117–131.

Vedder, J. G., 1975. Revised geologic map, structure sections and well table, San Joaquin Hills-San Juan Capistrano area, California, U.S.Geol. Surv. Open-File Rept. 75-552, scale 1:24,000, 3 pp.

Wright, C., Mella, A., 1963. Modifications to the soil pattern of Southcentral Chile resulting from seismic and associated phenomena during the period May to August 1960. Bulletin Seismological Society of America 53, 1367–1402.

Metasomatism of the Bird Spring Formation Near Slaughterhouse Springs, California

Department of Geological Sciences, California State University, Fullerton

Taylor Kennedy
Advisor: **Dr. Brandon Browne, Dr. William Laton**

Abstract

The Bird Spring Formation has gone through two separate metamorphism events; of faulting-induced metasomatism and contact metamorphism. A coarse grained white cream dolomite has gone through metasomatism and metamorphism is compared to a white coarse grained marble with grossular garnet and epidote that has gone through metasomatism and a different grade of metamorphism. Both of these are compared to a coarse grained white gray surface of marble, which has gone through metamorphism. Most samples contained high amounts of calcium. Few samples contained iron or magnesium. While these units do not have high abundances of precious elements, they still give insight into crystal formations. The grossular garnet is a prime spot for the element vanadium. Another element, strontium, does not fit into the chemical structure of grossular garnet and is located within the other samples. These findings determine the chemical structures that these elements prefer.

Introduction

Understanding how ore deposits develop is vital to our ability to locate and develop mineral resources. In particular, ore deposits containing precious metals like gold, copper, and lead, play a major role in our economy. In this study, I characterize the ore mineralization of metasomatized Pennsylvanian Bird Spring Formation (Pbs) where in contact with the (1) high angle Slaughterhouse Fault and (2) Mid Hills quartz monzonite of the Cretaceous Teutonia batholith in the New York Mountains near Slaughterhouse Springs, California. Comparing the mineralization of the Bird Springs Formation as a result of faulting-induced metasomatism and contact metamorphism will yield valuable insight into our understanding of how metasomatic ore deposits are related to contrasting geological phenomena. My field area is located along the Slaughterhouse Fault in the New York Mountains in northeast San Bernardino California, south of Ivanpah road from I-15(See Figure 1).

Geographical History

The geologic history of this region is best summarized by Burchfiel and Davis (1977) and Miller and Wooden (1993), and their results will be summarized here. The underlying bedrock of the New York Mountains is Precambrian gneiss (Pcgn). This unit consists of both meta-igneous and metasedimentary rocks with potassium feldspar augens.

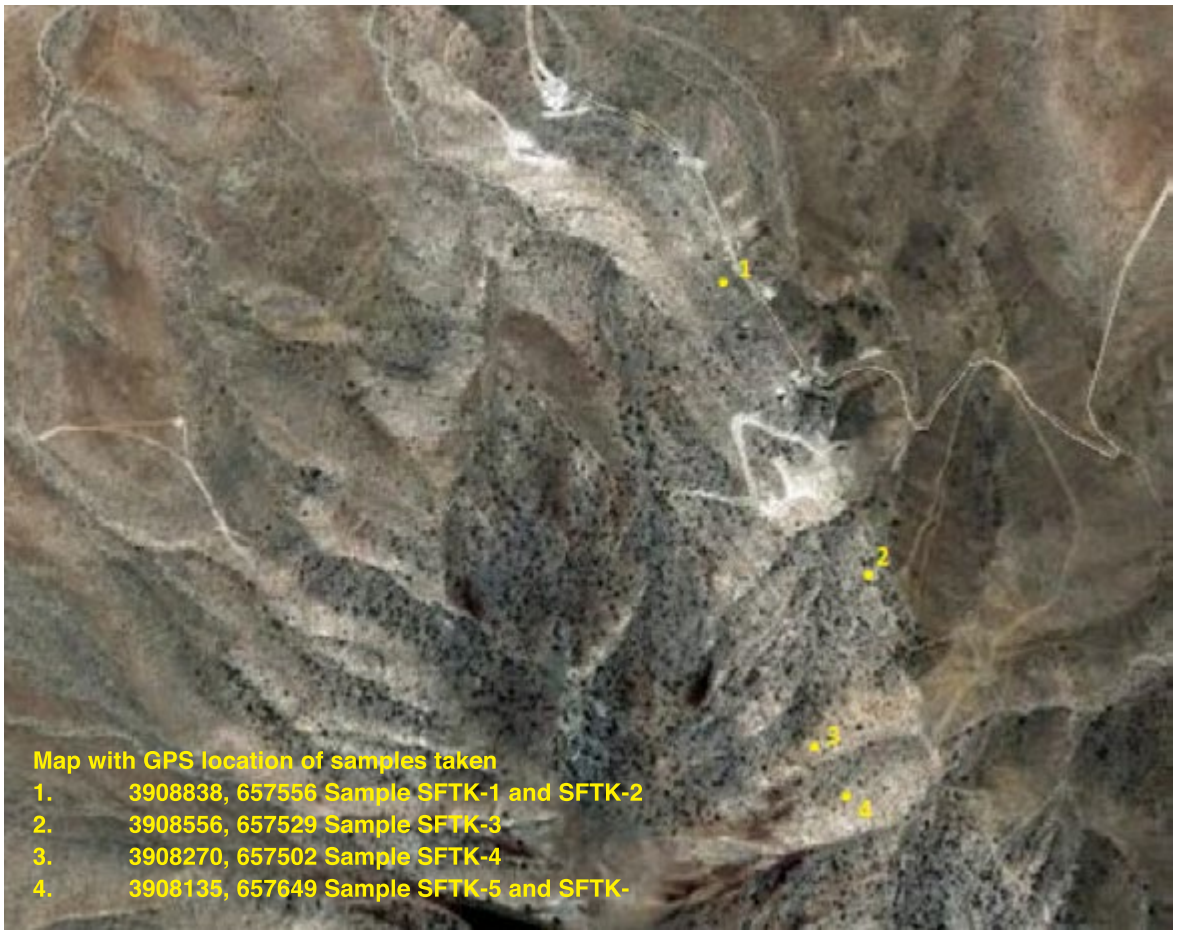
Unconformably overlying gneiss is the Cambrian Tapeats Sandstone (Ct), which is overlain by the Cambrian Bright Angel Shale (Cba), which is interbedded with calc-silicate and pelitic hornfels due to localized metamorphism.

The Cambrian Bonanza King Formation (Cbz) unconformable overlies the Bright Angel Shale, and is subdivided into the lower Papoose Lake member (Cbl) and the Banded Mountain member (Cbu). The lower Papoose Lake is interbeded with calcite and dolomite that has locally been metamorphosed to marble.

Overlying the Bonanza King is the Cambiran Nopah Formation (Cn), which is a coarse-grained white dolomite unit underlain by shale known as the Dunderberg Shale Member (Cd). Atop the Nopah Formation is the Devonian Sultan Limestone (Ds), which contains a lower member that is interbedded calcite and dolomite marble believed to be the metamorphosed equivalent of the Valentine Member (Dsv) found in other areas (REF). It has an upper member called Crystal Pass (Dscp), which is a white limestone that has been metamorphosed to coarse grained marble.

Lying nicely on top of Crystal Pass is the Monte Cristo Limestone (Mm). It has three members, from lower to upper: Dawn (Mmd), Anchor (Mma), and Buillion (Mmb). The Dawn and Anchor members have been recrystallized into marble, but the Anchor contains chert nodules. Buillion is calcite and marble with very few chert nodules.

The Bird Spring Formation overlies the Monte Cristo Limestone. Bird Spring is limestone and dolomite, and has been metamorphosed to low grade marble in many places.



Its exact date of deposition is unknown but it is restricted to the upper Mississippian to Pennsylvanian and possibly the Permian.

Unconformable overlaying the Bird Spring is a calc-sillicate rock (Mcs). This unit has been proposed to have been deposited during the Mesozoic, as well as is the next unit in the sequence: volcanic and sedimentary rocks (Mmvs). Atop of these unknown units we have the Mesozoic Sedimentary Rocks of Sagamore Canyon. This unit correlates with the nonmarine environment of the south-eastern California region.

Finally an intrusion of magma has caused some of the metamorphism in the area. This magma is now coarse grained quartz monzonite to monzonite with phenocrysts of potassium feldspar (Mp). Additionally, the Slaughterhouse Fault ripped across the area and is slowly overturning the units. It is also responsible for the rest of the metamorphism of the units.

Following this faulting is the Tertiary volcanic rocks (Tv). This is a unit of volcanic breccias that is associated with a late Miocene-Pliocene volcanic event (Burchfiel and Davis, 1977).

The last unit to be deposited in the New York Mountains is alluvium units in the Quaternary; an older unit (Qoal) and a younger unit (Qal).

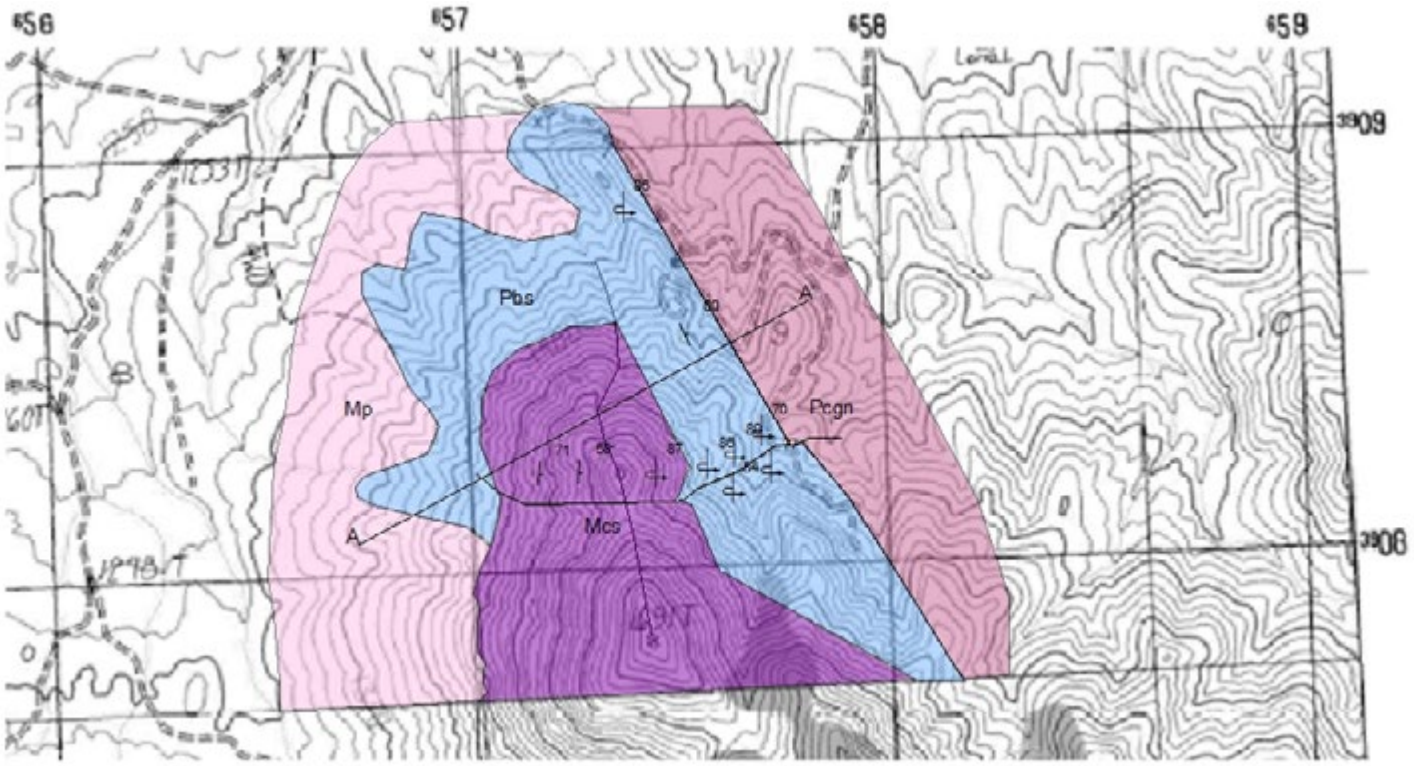


Figure 2A. *Geologic Map (U.S. Geologic Survey 1983)*

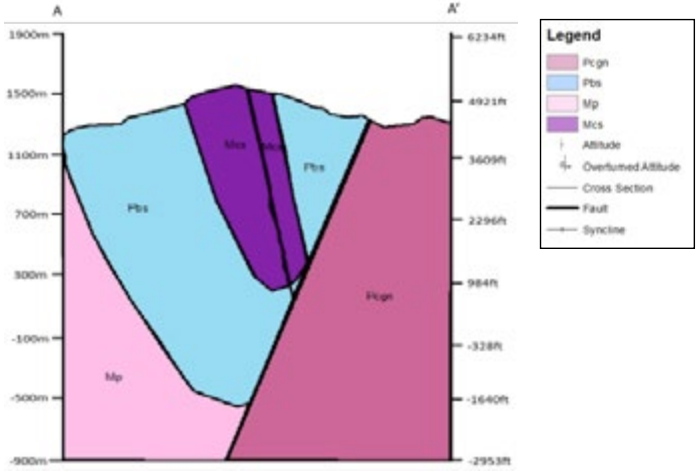


Figure 2B. *A-A` Cross Section*

Sampling Strategy and Analytical Method

In order to compare the metamorphism, samples were selected from two sampling sights; one on the western flank of the Bird Spring Formation where it has been intruded by Cretaceous quartz monzonite and the other along the northern flank where the Bird Spring Formation is in contact with the Slaughterhouse fault. These samples were analyzed for mineral content by taking the average of X-ray fluorescence scans of each sample (See Table 1). While collecting, a ~1.5 square kilometers (km2) geological map of the area, mapping the contacts between rock units, the extent and style of the metamorphism and metasomatism (see Figure 2A), and geologic structures such as folds and faults (see Figure 2B) was completed.

	SFTK-1	SFTK-3	SFTK-4	SFTK-5	SFTK-6
SiO_2	26.04	36.27	7.32	13.87	14.12
TiO_2	0.07	0.1	0.13	0.2	0.20
Al_2O_3	1.06	8.13	3.21	7.16	7.27
Fe_2O_3	0.21	18.16	1.18	0.95	0.93
MgO	19.02	0.13	0.34	0.38	0.39
CaO	29.74	33.42	52.25	48.67	48.55
K_2O	0.41	-	0.01	0.05	0.05
Na_2O	-	-	-	-	-
P_2O5	0.1	0.02	0.04	0.03	0.03
Zr	159.23	19.77	30.8	36.57	37.23
Zn	-	50.17	30.93	34.77	35.9
Y	7.4	-	4.7	5.47	5
V	8.3	199.37	49.27	86.5	87.5
U	3.3	3.23	2.97	3.13	2.77
Th	4.4	3.3	5.7	6.03	5.63
Sr	94.4	9.9	133.47	106.53	106.07
Rb	24.87	-	5.03	3.83	3.63
Sc	27.17	28.03	43.97	40.27	46.03
Ba	55.83	66.63	14.63	36.8	29.40
Pb	4.37	13.8	18.13	18.43	13.8
Ni	3.9	0.8	10.77	9.9	10.2
Cu	1.27	25.83	12.47	11.7	11.73
Cr	17.6	3.1	19.03	42.4	42.57
Ce	28.57	7.17	23.9	37.7	38.27

Table 1. *Average geochemical analysis of three separate scans of samples. SFTK-2 was not able to get an accurate reading and was rejected.*

Mineral Chemistry

SFTK-1, a coarse grained white cream dolomite collected near the north end of the fault, has gone through metamorphism and metasomatism. SFTK-3, found near the north center of the fault, a white coarse grained marble with spots of green and brown, has gone through primarily metasomatism and a different grade of metamorphism. These spots gain there color from the high amounts of grossular garnet and epidote within. SFTK-4, SFTK-5, SFTK-6 are a coarse grained white gray marble, collected along the contact between the Cretaceous Teutonia batholith and Bird Spring Formation. These samples have gone through metamorphism.

The data shows that this unit does not possess a large abundance of precious elements. It also shows consistency in SFTK-4, SFTK-5 and SFTK-6(see Figure 5). These metamorphosed samples are very high in calcium as the original limestone should be. These samples are higher in most of the precious metals compared to SFTK-1. The first metamorphism event would have made Bird Spring formation similar in content. The only differences would be based on proximity to the contact. The metasomatism event only affects a small part of the Bird Spring formation making the differences in SFTK-1 and SFTK-3. It does show that SFTK-1 contains similar amounts of magnesium and calcium with very high amounts of zircon and strontium. This holds bearing to SFTK-3 that was also found near the fault. SFTK-3 hardly contains any magnesium, strontium, and zircon, but it does have all the iron and vanadium. During the metasomatism, elements like iron and vanadium relocated within the grossular garnet. Other elements like strontium did not fit into the chemical structure of grossular garnet, so it left and went to the rest of the unit instead. Both samples have undergone the same metasomatism but have very different chemical compositions. The other thing of note is how there is not any sodium in any of the samples. This most likely comes from the original rocks being limestone and not having much sodium to begin with.

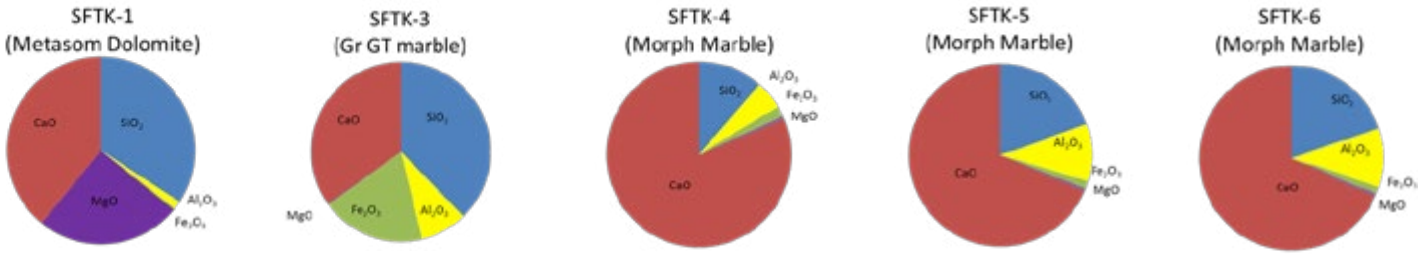


Figure 5. *Percent abundance in regards to major oxides*

Conclusion

The Bird Spring Formation does not contain much in the way of precious elements, but it does show a difference in ore development in metasomatism and metamorphism. The different events have allowed for the elements to reallocate themselves in more favorable arrangements. Vanadium favored the grossular garnet but strontium did not. Bird Spring formation is 300 meters (m) thick and only a small amount was sampled to determine a correlation. By analyzing different metamorphic events it could be possible to determine which elements and how they will relocate.

References

Burchfiel, B., & Davis, G. (1977). Geology of the Sagamore Canyon-Slaughterhouse Spring Area, New York Mountains, California. Geological Society of America Bulletin, 88(11), 1623-1640

Miller, D., & Wooden, J. (1993). Geologic Map of the New York Mountains area, California and Nevada. U.S. Geological Survey. 93-198

Google Earth (4 Nov. 2006). Slaughterhouse Springs, Ca. Google Earth. Google. Retrieved March 25, 2013.

U.S. Geological Survey, (1983). Ivanpah Quadrangle California-San Bernardino Co. 7.5 Minute Series (Topographic).

Acknowledgements

Thank you to California State University, Pomona and Jon Harris for the use of their XRF Machine. Thanks to my father, James Kennedy, for keeping me company for the drive to my field area. Special thanks to my advisor Dr. Brandon Browne for continuing to work with me even though he no longer needed to. Thanks to Dr. William Laton for extra advising.

Understanding the Orientation of the Sierra Nevada Frontal Fault System in the Vicinity of Lone Pine and Independence, CA

Department of Geological Sciences, California State University, Fullerton

Garrett Mottle

Abstract

Owens Valley in eastern California is a deep graben located between the Inyo and White Mountains to the east and by the Sierra Nevada Mountains to the west. The boundary between Owens Valley and the Sierra Nevada Mountains is the Sierra Nevada Frontal Fault System (SNFFS), which is a system of normal faults formed by crustal extension of the Basin and Range province. It generally is assumed in previous studies that the SNFFS normal faults dip steeply at ~60 degrees east. However, a recent study (Phillips and Majkowski, 2011) shows that the faults typically dip less than 50 degrees along the northern part of the SNFFS near Bishop. More recent work (Shagam 2011) near Independence shows that faults there dip 29-34 degrees east. My hypothesis is that faults farther south between Independence and Lone Pine dip shallowly (~30 degrees). I will test this hypothesis by mapping and measuring fault orientations of the SNFFS at George and Bairs Creeks using hand-held GPS and differential GPS devices at three fault exposures. The selected faults will have sufficient elevation exposure to allow for three-point analysis to determine their orientations in detail. Deliverables of this study will include a detailed map of a section of the SNFFS, an analysis of fault outcrop exposures to determine fault dip, and a report discussing the data and conclusions found in this study.

References

Le, K., Lee, J., Owen, L.A., Finkel, R. 2007. Late Quarternary slip rates along the Sierra Nevada frontal fault zone, California: Slip partitioning across the western margin of the eastern California shear zone-Basin and Range, Geological Society of America Bulletin, vol. 119, p 240-256

Phillips, F. M., Majkowski, L., 2011. The role of low-angle normal faulting in active tectonics of the northern Owens Valley, California, Lithosphere, v. 3, p. 22-36.

Shagam, Greg. Analysis of the Sierra Nevada Frontal Fault orientation in the vicinity of Lone Pine and Independence, California. B.S. Thesis, California State University, Fullerton, 2012.

Origin of Debris Flow Deposits on Starvation Canyon Fan,Death Valley, California

Department of Geological Sciences, California State University, Fullerton

Kelly Shaw
Advisor: **Jeffery Knott**

Abstract

Hunt and Mabey (1966, Fig. 48, pg. 67) mapped three debris fan lobes on the Starvation Canyon alluvial fan in western Death Valley, California, with an estimated volume of 8 to 25 million cubic yards. These deposits were mapped on top of the fan indicating that they were younger than the ~70 ka Qg2 alluvial fan. Tertiary volcanic rocks were mapped at the mouth of Starvation Canyon in the Panamint Range by Hunt and Mabey, with faulted Cambrian and Precambrian metasedimentary rocks to the west. The Precambrian Sterling Quartzite and Johnny Formations are composed of distinctive brown quartzite and purple shale. Granite at Hanaupah Canyon is 8.6 km upstream from the piedmont.

Field observations show that the northern lobe of the Qg2 deposits of the Starvation Canyon fan and consists of grusified granite boulders at the surface. The debris flow deposits, which generally line a wash channel, are composed of 1-6 m diameter, varnished, but relatively unweathered, granite boulders with rare metasedimentarty boulders (<2%). The wash channel is incised through the Qg2 deposits with debris flow boulders overtopping the channel margins and resting atop the Qg2 deposits at the distal end. The southern lobe of the Starvation Canyon fan is composed of Qg3 gravels with an intervening active channel (Qg4). Both the younger and active channels are lined with large granite boulders (up to ~9 m diameter) that are unweathered and have very little, if any, varnish. Based on my field observations, Hunt and Mabey were correct in their conclusion that younger, unweathered granite boulders from debris flows are found atop the Qg2 gravel. However, my observations are that the debris flow boulders are limited to the channel and channel margins. I infer that these deposits traveled along the incised channels and did not overtop the Qg2 gravels, aside from the area at the distal end of the fan. I infer that the flow volumes were significantly lower volume than previously thought and, based on the dominance (>98%) and size (up to ~9 m) of granite boulders these flows would have traveled ≤ 16 km before being deposited on the Starvation Canyon fan, with the origination point >8 km from the apex of the fan. My interpretation is that these debris flows were emplaced in multiple events rather than one large event. This is based on the differences in weathering of the boulders as well as the disconnected nature of the debris flow deposits. The active channel (Qg4) is comprised of similar debris flow deposits as are found in the older channels. This would indicate that the same processes continue to repeat, that this process has been ongoing for ~70 ka and that debris flow travel may be ≤ 16 km.

References

Hunt, C. R., Mabey, D. R., Stratigraphy and Structure, Death Valley, California, 1996

Landslide Hazards in the Pai Valley, Northern Thailand

Department of Geological Sciences, California State University, Fullerton

Garcia, Dylan J., Shaw, K., Rhodes, B.P., Chantraprasert, S.

Abstract

In this study we look at past mass wasting events in the Pai Valley of Northern Thailand to better understand the size and scope of previous slide events. Thailand is located in southeast Asia, sharing borders with Burma, Laos, Cambodia, Malaysia, The Andaman Sea and The Gulf of Thailand. Pai Valley is in the northernmost region of Thailand, close to the Burmese border. Research on past slides in this area is significant because of the population living on these slides. The inhabitants of Pai had an estimated population of about 2,284 as of 2006 (www.geonames.org). Residents built their homes on the suspected slide area and are dependent on the agriculture grown on the slides. A future slide event could be catastrophic for the valley, including major loss of life. It is hypothesized that in the past there was a massive debris flow that originated west of the valley. Our research shows the suspected slides are "Torrent Slides", or "Long Slides". These are channelized type slides discharging water and debris composed of bedrock, unconsolidated sediment or organic material and can be triggered by changes in ground water conditions or the geometry of the slope. Given the tropical setting of the area, it is presumed these slides were triggered by excessive rains or unseasonably wet periods. Through satellite imagery and on the ground mapping we determined at least two slide events and possibly a third. The mapping process involved driving and looking for locations with boulders versus bedded areas. When we determined the end of the boulder section and beginning of bedding, we plotted the location on our map. We continued until we had the boundaries of the slide determined. Clast counts were performed at a number of locations within the slide area. At each location an average of 50 clasts were counted, classified and measured. Sizes ranged from sand grain to 16 meters. The main lithologies of the clasts were granites and sandstones, this is expected if the slides originated west of the valley. There was no clear difference in weathering from one location to the next which leads us to believe these events occurred in a relatively short time frame. Future research could determine the precise origins of the slides and cosmogenic dating could be helpful in determining how old the slides are.

Petrographic and geochemical analysis of the Summit Gabbro and associated granitoids of the Kern Plateau, Southern Sierra Nevada, California

Department of Geological Sciences, California State University, Fullerton

Elizabeth White
Advisor: **Dr. Diane Clemens-Knott**

Abstract

The Sierra Nevada Batholith (SNB) formed from multiple arc magmatism events that occurred as a result of the collision between the North American and Farallon plates during the Mesozoic Era (251-65 million years ago). Arc batholiths, such as the SNB, are important to understand because they record episodes of continental crustal growth. Calculating the amount of material, in this case magma, transferred from the mantle to the crust is dependent on the composition of the mantle wedge located beneath the continent. The focus of this research is to understand the origin and formation of the SNB by geochemically characterizing the Summit Gabbro and suite from the Kern Plateau located in the southern Sierra Nevada Mountains, California. Gabbros are crucial to studies on mantle-wedge composition since they have experienced a minimal amount of chemical-evolution and represent a proxy for the chemical composition of the mantle source below the batholith. By definition gabbros are mafic, or silica-poor, and contain large quantities of magnesium- and iron-rich minerals, such as olivine. The amount of chemical evolution that was experienced by a magma can be quantified using the magnesium number (Mg#), which is calculated using the formula Mg # = [100 Mg/Mg + Fe2+]. In this formula, Mg and Fe2+ are atomic proportions, which are measured using scanning electron microscopes (SEM) and electron microprobe analyzers (EMPA). Higher Mg#'s indicate more primitive magmas—i.e., magmas that have not evolved much from their original, mantle-derived composition. Therefore, gabbros with higher Mg#'s are ideal for identifying mantle-derived magmas that have not experienced much differentiation, and the geochemical compositions of these high-Mg# rocks can be used to determine whether the mantle source rock was either "enriched" or "depleted". The methodology to complete this object includes: (1) study thin sections of each sample using a petrographic microscope to look for alteration that may have changed rock chemistry and identify minerals for SEM analysis; (2) analyze the rock chemistry of each sample using the x-ray fluorescence spectrometer (XRF) in order to identify the most mafic samples; and (3) analyze the compositions of minerals to determine their Mg# and further estimate the composition of the magma from which they crystallized. In order to complete this last task, I will use two instruments: the scanning electron microscope (SEM) and an electron probe microanalyzer (EPMA). A desired potential outcome of this study is to locate gabbros with Mg#'s between Fo72-Fo65. The highest Mg# recorded to date from the eastern SNB is Fo50 (Gevedon, 2013), a value significantly lower than what would be predicted for a pristine mantle melt. Thus, geochemical characterization of the Summit Gabbro may provide new information about the type of mantle underlying the eastern Sierra Nevada arc and is one of the first steps to determining how the western margin of the North American continent was modified and expanded by Mesozoic arc magmatism.

References

Gevedon, Michelle, 2013, Paired oxygen and hafnium isotopic analysis on zircon from gabbros: identifying potential Mesozoic mantle heterogeneity of the Sierra Nevada Arc: MS thesis. California State University, Fullerton; California. Print.

Permutation Test in Microarray Data Analysis

Department of Mathematics, California State University, Fullerton

Mirna Dominguez
Advisor: **Gülhan Bourget**

Abstract

Microarray technology has become one of the most powerful tool to simultaneously study thousands of genes at once. Since microarray data involves high dimensional data (i.e., number of genes are comparably higher than number of replicates), making inferences using current statistical methods are insufficient. Despite the inability of the F-test to draw valid conclusions in Microarray data, it is one of the common statistical methods that is used. One of the assumptions of the classical F-test is that groups (genes) are supposed to be independent. However, this assumption is violated in microarray data because gene-gene interactions are possible. In this paper, we suggest performing permutation test to explore if the p-value obtained from the F-test can be improved. We consider various magnitudes of correlation among genes from no correlation to strong correlations in a Monte Carlo study to compare the p-values of the F-test and the permutation test. Our findings show that the permutation test is preferred over the F-test.

Introduction

Microarray experiments study gene expressions of cells, organisms, or tissues. For example, in a cancer study healthy and diseased tissues of breast is compared to identify disease causing genes.

Microarray data has a high dimensional data structure which makes challenging todraw statistical inferences [1]. Various methods have been proposed to answer different kinds of questions. For example, clustering and classification are two common methods to identify groups of genes that share similar functions [2,3]. These methods search for similar genes, but they do not help to identify which genes are differentially expressed under different conditions. To find differentially expressed genes, we need to perform hypothesis tests of no difference in the means of gene expressions under different conditions. Fold change, linear models, as well as Bayesian methods [4–6] are some of the statistical tests; however because of computational difficulty, adopting these methods in microarray analysis have been slow in practice.

To test whether the difference in means is statistically significant, analysis of variance (ANOVA) method can be performed. If the populations from which the data were sampled violate assumption(s) of the ANOVA, then the results of the analysis may be incorrect or misleading. For example, if the assumption of independence of the observations is violated, then the ANOVA is not appropriate. The test statistics used in ANOVA is the classical F-test. Since gene-gene interactions can happen in nature, the F-test should not be considered to answer if there is a difference in means of genes under different conditions. Permutation test can be an alternative test to consider. Permutation test is particularly used when the distributions of the data is unknown, sample sizes are small, or outliers are present [7]. However, in this paper we suggest performing permutation test in lieu of the violation of the independency assumption in ANOVA. We run Monte Carlo Simulation studies by considering various magnitudes of correlations among genes to answer if the permutation test is preferable over the F-test.

In Section 2, we describe the data and outline the F-test and the permutation test, and in Section 3 we describe Monte Carlo Simulation study and present its findings. Finally, we draw conclusions in Section 4.

Methodology

A single multivariate observation is the collection of measurements on p different variables (genes) taken from the same trial (array). If n observations have been obtained, the entire data set can be represented in an $n \times p$ matrix

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}$$

The row vector $\mathbf{X}_j'$ represents the jth multivariate observation. The matrix $\mathbf{X}$ represents p genes each having n observations.

Now, consider a microarray experiment of n_1 and n_2 sample from population 1 and population 2, respectively. For example, population 1 can represent the disease group, while population 2 can represent the healthy group. Suppose that the expression levels of p genes are measured and matrix representation of the population 1 and 2 are defined as $\mathbf{X}$ and $\mathbf{Y}$, which is in the form of (1). Let X_{ij} be the expression level for gene j of sample i from population 1. The expression level vectors for sample i from population 1 can be expressed as $\mathbf{X}_i' = (X_{i1}, \ldots, X_{ip})$. The mean expression levels of gene j in population 1 is,

$$\overline{\mathbf{X}}_j = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{ij}.$$

The mean expression level vector for p genes for population 1 is given by $\overline{\mathbf{X}} = (\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_p)'$. Similarly, we can define these expressions for population 2. Now, we are ready to outline the hypotheses we are interested in.

Comparing Mean Vectors from Two Populations

Consider a random sample of n_1 and n_2 from populations 1 and 2. The observations on p variables can be arranged as follows:

$$\begin{aligned} &\text{Population 1:} \quad \mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_{n_1}' \\ &\text{Population 2:} \quad \mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_{n_2}'. \end{aligned}$$

We want to make inferences about the differences of the mean vectors of the populations. That is, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_i$ is the mean vector of population $i (= 1, 2)$. We want to answer the question of $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ or equivalently: Is $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$? We need to make some assumptions to provide answers to these questions. The assumptions are:

1. The sample $\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_{n1}'$ is a random sample of size n_1 from a p-variate population with mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$.
2. The sample $\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_{n2}'$ is a random sample of size n_2 from a p-variate population with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$.
3. $\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_{n1}'$ is independent of $\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_{n2}'$.

For large samples, these assumptions are enough to make an inference about $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. However, when the sample sizes n_1 and n_2 are small we need to have the following as- sumptions as well.

1. Both populations are multivariate normal.
2. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

The null (H_0) and alternative (H_1) hypotheses are:

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0 \text{ versus } H_a : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq 0$$

where $\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \ldots, \mu_{1p})'$ is the mean expression level of population 1, and $\boldsymbol{\mu}_2 = (\mu_{21}, \mu_{22}, \ldots, \mu_{2p})'$ is the mean expression level of population 2. That is, the null and alternative hypotheses can be rewritten as

$$\begin{aligned} H_0 &: (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22}, \ldots, \mu_{1p} - \mu_{2p})' = (0, 0, \ldots, 0)' \\ H_a &: (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22}, \ldots, \mu_{1p} - \mu_{2p})' \neq (0, 0, \ldots, 0)' \end{aligned}$$

***F*-Test**

The classical F-test compares the means of the columns of $\mathbf{X}$, and assumes that these columns are independent (univariate case). Here, we want to compare the differences of the p means of $\mathbf{X}$ and $\mathbf{Y}$. To adopt the data structure from multivariate case to univariate case, we consider the observations as the differences of the $\mathbf{X}$ and $\mathbf{Y}$. That is, we compute $\mathbf{X}_{ij} - \mathbf{Y}_{ij}$ and apply the univariate F-test on these observations. The statistic

$$F = \frac{\text{MST}}{\text{MSE}},$$

where MST is the mean square for treatments (genes) and MSE is the mean square for errors, follows F distribution with $p - 1$ and $p(n_1 + n_2 - 1)$ degrees of freedoms.

Permutation Test

Permutation test also called randomization or re-randomization test has been proposed for a long time, but it has become practical with the advance of high-speed computers. Permutation test is more useful when the distributions of the data are unknown, sample sizes are small, or outliers are present. The basic approach to permutation test that we used in the simulation follows:

- Decide a test statistic. In ANOVA, the test statistic is the F-test.
- Calculate the F-test for the data, called it F_{obs}.
- Repeat the following r times, where r is a number greater than 1000.
  - Shuffle the labels of the populations (genes).
  - Calculate the F-test for the shuffled data, call it F_s.
- Compute how many times F_s is greater than or equal to F_{obs}, call this number s. Then calculate the p-value as the ratio of s and r. That is, p-value=s/r.
- Decide if the -value can reject the null hypothesis.

Results

We generated two multivariate normal distributions by running Monte Carlo simulation studies in R software: MVN($\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_1$) and MVN($\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_2$), each with dimension p (genes).

The variance covariance matrices are defined as

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{pmatrix} \boldsymbol{\Sigma}_\rho & 0 & 0 & \dots & \dots & \dots \\ 0 & \boldsymbol{\Sigma}_{(-\rho)} & 0 & 0 & \dots & \vdots \\ 0 & 0 & \boldsymbol{\Sigma}_\rho & 0 & \dots & \vdots \\ \vdots & 0 & 0 & \boldsymbol{\Sigma}_{(-\rho)} & 0 & \vdots \\ \vdots & \vdots & \vdots & 0 & \ddots & \vdots \\ \dots & \dots & \dots & \dots & \dots & \ddots \end{pmatrix},$$

where

$$\boldsymbol{\Sigma}_\rho = \begin{pmatrix} 1 & \rho & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \dots & \dots & \rho^{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{n-2} & \dots & \vdots & \vdots & \rho \\ \rho^{n-1} & \rho^{n-2} & \dots & \rho & 1 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_{(-\rho)} = \begin{pmatrix} 1 & -\rho & \dots & (-\rho)^{n-2} & (-\rho)^{n-1} \\ -\rho & 1 & \dots & \dots & (-\rho)^{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (-\rho)^{n-2} & \dots & \vdots & \vdots & -\rho \\ (-\rho)^{n-1} & (-\rho)^{n-2} & \dots & -\rho & 1 \end{pmatrix}.$$

The matrix $\boldsymbol{\Sigma}_\rho$ and $\boldsymbol{\Sigma}_{(-\rho)}$ have dimensions $n \times n$, and the matrix $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ has dimension $p \times p$.

We considered sample sizes of $n_1 = n_2 = 10$ for $p = 50$ genes. We assume that there are 5 groups of 10 genes in each group, totaling of 100 genes. That is, we fixed $n = 10$ in $\boldsymbol{\Sigma}_\rho$ matrix. We assumed $\rho = 0, 0.1, 0.2, \dots, 0.9$ as various magnitudes of correlations, and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = (0.5,0.5,0,0,\dots,0)$. We run 1000 data sets to test the null hypothesis at the significance level of $\alpha = 0.05$. We followed the permutation test approach described in the previous section with $r = 1000$. We computed p-values, which is the probability of rejecting the null hypothesis when the null hypothesis is true, to draw conclusions about the F-and the permutation tests.

	ρ									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F-Test	0.040	0.053	0.054	0.056	0.073	0.083	0.094	0.106	0.125	0.163
Permutation Test	0.037	0.051	0.055	0.059	0.073	0.081	0.093	0.104	0123	0.165

Table 1. *The probability of rejections for F and Permutation tests at α = 0.05.*

We consider the test statistic is valid if its p-value smaller than the chosen significance level α or the p-value lies in the $(1 - \alpha)\%$ confidence interval

$$\alpha \pm 2\sqrt{\frac{\alpha(1-\alpha)}{s}}.$$

In our simulation with $\alpha = 0.05$ and $s = 10000$, the interval (2) becomes

$$0.05 \pm 2\sqrt{\frac{0.05 \times 0.95}{1000}} \iff (0.043, 0.057).$$

When $\rho = 0$, the assumption of the ANOVA is satisfied because populations are independent. Therefore, in this situation, F-test should be equal to the significance level. This observation is validated for tests from the table. To study the effect of the correlation among genes, we considered various degrees of magnitudes of correlations from small to high degrees. When the correlation is small to medium, we still suggest using F-test because type I error is well protected. However, when the correlations among genes are increasing F-test should be avoided to make any kinds of conclusion about the data. When correlation is medium to strong, permutation test is preferred.

Conclusion

Microarray experiments produce a large amount of gene expression data. The main goal is to apply appropriate statistical methods to determine which genes function differently under different conditions (i.e, normal lung cells versus diseased lung cells).

One of the most commonly used methods to test whether the difference in means is statistically significant is the ANOVA. However, since genes can be correlated among themselves, the use of ANOVA can lead to invalid conclusions of the data. In this paper, we investigated if the permutation test applied to ANOVA could improve the p-value of the classical F-test. Our findings show that we can perform F-test if the correlation among genes are small to moderate, but should be avoided if strong correlations are suspected. When correlations among genes are strong, permutation test is an alternative method. Overall, permutation test outperformed the F-test, hence we suggest to use permutation test in lieu of the F-test.

Acknowledgments

We thank the California State University, Fullerton (CSUF) for providing us with an intramural grant to work on this project. The author extend a special thanks to Dr. Gu ̈lhan Bourget for mentoring and allowing the opportunities to grow as undergraduate researcher. The authors also thank Dr. Bourget for her valuable comments and editions that helped improved the quality of this paper.

References

Mehta T, Tanik M, B AD (2004) Towards sound epistemological foundations of statistical methods for high-dimensional biology. Nature Genetics 36: 943-947.

Alon U, Barkai N, Notterman D, Gish K, Ybarra S, et al. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences of the United States of America 96: 6745-6750.

Brazma A, Vilo J (2000) Gene expression data analysis. FEBS Letters 480: 17-24.

Baldi P, Long A (2001) A bayesian framework for the analysis of microarray expres- sion data: regularized t -test and statistical inferences of gene changes. Bioinformat- ics 17: 509-519.

Wang S, Ethier S (2004) A generalized likelihood ratio test to identify differentially expressed genes from microarray data. Bioinformatics 20: 100-104.

Wettenhall J, Smyth G (2004) limmagui: a graphical user interface for linear mod- eling of microarray data. Bioinformatics 20: 3705-3706.

Good P (2000) Permutation Tests. Springer.

Mathematica and Clifford Projective Spaces

Department of Mathematics, California State University, Fullerton

Andrew Halsaver
Advisor: **Dr. Alfonso Agnew**

Abstract

Computing canonical equivalence classes (EC) in Clifford projective space (CPS) can be difficult, especially in high dimension. We created our own package in Mathematica to perform such computations. By passing in certain parameters to our user-defined Mathematica function, EquivalenceClass, we can analyze canonical representatives of EC in CPS. The examples in this paper involve $\mathcal{C}\ell_{2,0}$ explicity, but our Mathematica package works with Clifford algebras such as $\mathcal{C}\ell_{3,0}$ and $\mathcal{C}\ell_{1,1}$. Other Clifford algebras will be available in our package in the future.

Introduction

What is Clifford Algebra?

Clifford Algebra is an associative algebra equipped with a quadratic form. For example, consider the quadratic space $\mathbb{R}^{p,q}$. It is an n dimensional real vector space where $n = p+q$, and has a non-degenerate symmetric scalar product which induces the quadratic form:

$$\mathbf{x} \cdot \mathbf{x} = x_1^2 + ... + x_p^2 - x_{p+1}^2 - ... - x_{p+q}^2 \qquad [1,\ \text{p. } 205].$$

Furthermore, Clifford algebra is generally non-commutative. For example the Clifford algebra, $\mathcal{C}\ell_{2,0}$, has $e_1e_2 = -e_1e_2$, where $\{e_1e_2\}$ is a canonical basis of the algebra [1, p.188-189].

What is a projective space?

A projective space, denoted as $\mathbb{FP}^n$, is the set of lines through the origin of the vector space, $\mathbb{F}^{n+1}$. A simple example would be the real projective space, $\mathbb{RP}^1$. In this case, each element of $\mathbb{R}^2 - \{\vec{0}\}$ is mapped to its respective equivalence class (i.e. a line). Topologically, the set of EC, $\mathbb{RP}^1$ represents a circle since any two points on the same line in $\mathbb{R}^2$ belong to the same EC [2, p.2].

Methodology

Constructing CPS in Mathematica

It is discussed in Agnew-Childress' paper the standard approach to constructing projective n–space over a field. Further they discuss how to construct a projective space where a ring of $k \times k$ matrices, $\mathbb{M}$, is used for scalars. They note that since $\mathbb{M}$ is typically not a field, we are working with modules over rings instead of vector spaces [2, p.2]. Their discussion on constrcuing matrix projective spaces serves as the cornerstone for our Mathematica package.

The construction of a CPS requires that we take a module, $\mathcal{C}\ell_{p,q}^{n+1}$, Clifford algebra. Next, we impose an equivalence relation (denoted by $\sim$) on $\mathcal{C}\ell_{p,q}^{n+1} - \{\vec{0}\}$. The definition of $\sim$ is,

Definition 1.
$\vec{u} \sim \vec{v}$ if and only if there exists a $\lambda \in \mathcal{C}\ell_{p,q}^*$ such that $\vec{u} = \vec{v}\lambda$.

In this definition, $\mathcal{C}\ell_{p,q}^*$ denotes the invertible elements of $\mathcal{C}\ell_{p,q}$, and $\mathcal{C}\ell_{p,q}\mathbb{P}^n$ is the projective space defined as the set of EC [2, p.2].

Recall that Agnew-Childress discuss the use of matrices as scalars. As such we can make use of matrix isomorphims which relate to Clifford algebra. So, using $\mathcal{C}\ell_{p,q}^{n+1}$ to build $\mathcal{C}\ell_{p,q}\mathbb{P}^n$ we start by programming these matrix isomorphisms. An example of a matrix isomorphism would be,

$$\mathcal{C}\ell_{2,0} \simeq \text{Mat}\,(2, \mathbb{R}) \quad [1,\ \text{p.}205],$$

where Mat(2,$\mathbb{R}$) is short-hand for a 2-dimensional matrix having real entries. Other isomorphisms could be programmed using helpful properties, such as,

$$\mathcal{C}\ell_{p+1,q+1} \simeq \text{Mat}\,(2, \mathcal{C}\ell_{p,q}) \quad [1,\ \text{p.}214].$$

In a general case, given the CPS dimension (n), we build a column of $k \times k$ matrices whose entries belong to $\mathcal{C}\ell_{p,q}$. Exclusively for this discussion, we build a $2(\mathrm{n}+1) \times 2$ matrix, and consider if the columns are linearly independent, or linearly dependent (LI or LD)[a].

Now, there are several canonical forms which a matrix can take in order for its columns to be LI or LD. However, we can significantly reduce the number of cases by using matrix reductions. Moreover, matrix reduction is a result of applying elementary matrices from linear algebra, so it is a familiar process[b] [3, p.172].

In general, $\mathcal{C}\ell_{p,q}$ is non-commutative, so we have to distinguish between right and left scalar actions (i.e. multiply a matrix by a scaling matrix with said scalar on the left, or right, of the matrix itself) [2, p.2]. The use of row reduction on a matrix is equivalent to the action of using left scalar action on a $2\times2(n+1)$ matrix. Similarly, for column reduction, one must use right scalar action on a $2(n+1)\times2$ matrix. This is very convenient, since Mathematica can execute row and column reduction quickly. Furthermore, not only can we use matrix reductions to yield equivalence classes, we can also determine what scaling matrix is necessary to perform the reduction process.

Results

We can take advantage of matrix representatives of Clifford elements in order to determine what scaling matrix is required to reduce a matrix appropriately. To do this, we look to the definition of elementary operations on matrices from linear algebra [3, p.172]. We will work with right scalar action, but note that left scalar action is analogous and can be done by transposing matrices [2, p.2].

Consider a matrix with LD columns, and let's consider working with the projective line. That is let's suppose we have $u \in \mathcal{C}\ell^2_{2,0}$:

$$u_j = \frac{1}{2}\left[(x_k + \lambda x_{k+1}) + (x_k - \lambda x_{k+1})e_1 + (\lambda x_k + x_{k+1})e_2 + (\lambda x_k - x_{k+1})e_{12}\right] \quad (1)$$

where λ is a fixed real number, $k = 1,3$, and $j = (k+1)/2$. Also, assume for this particular case that x_1 is non-zero. As a result, we have the matrix representative

$$\begin{pmatrix} x_1 & \lambda x_1 \\ x_2 & \lambda x_2 \\ x_3 & \lambda x_3 \\ x_4 & \lambda x_4 \end{pmatrix}.$$

Using column reduction on this matrix yields

$$\begin{pmatrix} 1 & 0 \\ x_2/x_1 & 0 \\ x_3/x_1 & 0 \\ x_4/x_1 & 0 \end{pmatrix}.$$

Expressing this resulting matrix as its Clifford algebra representative, we find the EC to be

$$\begin{bmatrix} \frac{1}{2}(1+e_1) + \frac{x_2}{2x_1}(e_2 - e_{12}) \\ \\ \frac{x_3}{2x_1}(1+e_1) + \frac{x_4}{2x_1}(e_2 - e_{12}) \end{bmatrix}.$$

Now, consider Λ^* such that

$$\Lambda^* = \begin{pmatrix} 1/x_1 & -\lambda \\ 0 & 1 \end{pmatrix}, \quad x_1 \neq 0.$$

This matrix has a Clifford representative that is invertible. The Clifford representative is,

$$v = \frac{1}{2}\left[\left(\frac{1}{x_1}+1\right) + \left(\frac{1}{x_1}-1\right)e_1 - \lambda\left(e_2 + e_{12}\right)\right], \quad (2)$$

and has inverse,

$$v^{-1} = \frac{\overline{v}}{v\overline{v}} = \frac{x_1}{8}\left[\left(\frac{1}{x_1}+1\right) + \left(1-\frac{1}{x_1}\right)e_1 + \lambda\left(e_2 + e_{12}\right)\right], \text{ provided } x_1 \neq 0.$$

Looking back at equation (1), u · v (using right-scalar action for k = 1) the first component of the EC representative reduces to,

$$\frac{1}{2}(1+e_1) + \frac{x_2}{2x_1}(e_2 - e_{12}).$$

a. Recall the definition of linear dependence. As stated in Goode-Annin's text, "A finite nonempty set of vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_k\}$ in a vector space V is said to be linearly dependent if there exist scalars $c_1, c_2, ..., c_k$, not all zero, suchthat $c_1\vec{v}_1 + c_2\vec{v}_2 + ... + c_k\vec{v}_k = \vec{0}$." Furthermore, a set of vectors which is not linearly dependent is called linearly independent [3, p.269].

b. A matrix that is found by using a single elementary row operation on the identity matrix is, by definition, an elementary matrix [3, p.172].

For k = 3, the second component of the EC representative reduces to,

$$\frac{x_3}{2x_1}(1+e_1)+\frac{x_4}{2x_1}(e_2-e_{12}).$$

As required, we have the equivalence class

$$u\in\mathcal{C}\ell^2_{2,0}: u_j=\frac{1}{2}\left[(x_k+\lambda x_{k+1})+(x_k-\lambda x_{k+1})e_1+\right.$$

$$\left.(\lambda x_k+x_{k+1})e_2+(\lambda x_k-x_{k+1})e_{12}\right]\longmapsto$$

$$\begin{bmatrix}\frac{1}{2}(1+e_1)+\frac{x_2}{2x_1}(e_2-e_{12})\\ \frac{x_3}{2x_1}(1+e_1)+\frac{x_4}{2x_1}(e_2-e_{12})\end{bmatrix}, \text{ where } k=1,3 \text{ and } j=\frac{k+1}{2}.$$

We are inclined to think that equation (2) can be applied to other canonical matrix representatives of $\mathcal{C}\ell^{n+1}_{2,0}$. We do in fact find that this is true. We acknowledge this with the following theorem.

Theorem 1.
Consider $u\in\mathcal{C}\ell^{n+1}_{2,0}$: such that $u\simeq M\in\mathbb{R}^{2(n+1)\times n}$ where M has L.D. columns. A priori, we have

$$u_j=\frac{1}{2}\left[(x_k+\lambda x_{k+1})+(x_k-\lambda x_{k+1})e_1+(\lambda x_k+x_{k+1})e_2+(\lambda x_k-x_{k+1})e_{12}\right],$$

where $j=\frac{k+1}{2}$, λ is a fixed real number, and $k=1,\ 3,\ 5,\ \ldots,\ n{+}1$. The canonical equivalence class on u is determined by $v\in\mathcal{C}\ell^*_{2,0}$:

$$v=\frac{1}{2}\left[\left(\frac{1}{x_i}+1\right)+\left(\frac{1}{x_i}-1\right)e_1-\lambda\left(e_2+e_{12}\right)\right],$$
$$\text{whenever } x_i\neq 0, \text{ and } i=1,2,3,\ldots,n{+}1$$

Proof:
Let $u\in\mathcal{C}\ell^{n+1}_{2,0}$. The matrix representative of u is a $2(n+1)\times 2$ matrix with linearly dependent columns written as,

$$U=\begin{pmatrix}x_1 & \lambda x_1\\ x_2 & \lambda x_2\\ \ldots & \ldots\\ x_k & \lambda x_k\\ \ldots & \ldots\\ x_{2(n+1)} & \lambda x_{2(n+1)}\end{pmatrix}.$$

Let $\Lambda^*=\begin{pmatrix}1/x_i & -\lambda\\ 0 & 1\end{pmatrix}$, assuming $x_i\neq 0$.

Multiplying, we see that $U\cdot\Lambda^*$ yields

$$\begin{pmatrix}x_1 & \lambda x_1\\ x_2 & \lambda x_2\\ \ldots & \ldots\\ x_k & \lambda x_k\\ \ldots & \ldots\\ x_{2(n+1)} & \lambda x_{2(n+1)}\end{pmatrix}\begin{pmatrix}1/x_i & -\lambda\\ 0 & 1\end{pmatrix}=\begin{pmatrix}x_1/x_i & 0\\ x_2/x_i & 0\\ x_3/x_i & 0\\ x_4/x_i & 0\\ \ldots & \ldots\\ x_k/x_i & 0\\ \ldots & \ldots\\ x_{2(n+1)}/x_i & 0\end{pmatrix}.$$

This matrix corresponds to $w\in\mathcal{C}\ell^{n+1}_{2,0}$:

$$w=\begin{pmatrix}\frac{x_1}{2x_i}+\frac{e_1x_1}{2x_i}+\frac{e_2x_2}{2x_i}-\frac{e_{12}x_2}{2x_i}\\ \frac{x_3}{2x_i}+\frac{e_1x_3}{2x_i}+\frac{e_2x_4}{2x_i}-\frac{e_{12}x_4}{2x_i}\\ \ldots\\ \frac{x_k}{2x_i}+\frac{e_1x_k}{2x_i}+\frac{e_2x_{k+1}}{2x_i}-\frac{e_{12}x_{k+1}}{2x_i}\\ \ldots\\ \frac{x_{2n+1}}{2x_i}+\frac{e_1x_{2n+1}}{2x_i}+\frac{e_2x_{2(n+1)}}{2x_i}-\frac{e_{12}x_{2(n+1)}}{2x_i}\end{pmatrix}. \quad (3)$$

Furthermore, we see that this result is equivalent to right scalar multiplication of equation 1 by

$$v = \frac{1}{2}\left[\left(\frac{1}{x_i}+1\right)+\left(\frac{1}{x_i}-1\right)e_1-\lambda\left(e_2+e_{12}\right)\right]$$

which yields:

$$u_j \cdot v = \frac{x_k}{2x_i}+\frac{e_1x_k}{2x_i}+\frac{e_2x_{k+1}}{2x_i}-\frac{e_{12}x_{k+1}}{2x_i}.$$

These are precisely the components of (3), where
$k = 1, 3, 5, ..., 2n+1,\ \ j = \dfrac{k+1}{2},\ \ i \in \mathbb{N}.$

Since v is invertible, and since $v \simeq \Lambda^*$, it follows that the equivalence class for u is determined by v, and x_i is chosen to be the first nonzero element in the first column of U as determined by the definition of column-echelon matrices [3, p.147][c]. ✳

Examples
These operations are done with the function EquivalenceClass. Below are two examples of this function:

- Projective point example using $\mathcal{C}\ell_{2,0}$ coefficients:
  The first argument is an ordered list representing the coefficients $(u_0,\ u_1,\ u_2,\ u_{12})$ to the Clifford algebra having the form

$$u = u_0 + u_1e_1 + u_2e_2 + u_{12}e_{12}.$$

The last three arguments in EquivalenceClass represent p, q, and n in $\mathcal{C}\ell^{n+1}_{p,q}$.

$\mathrm{u1} = \{1,\ 1,\ \text{-}1,\ \text{-}1\};$
$\mathrm{eqc1} =$ EquivalenceClass[{u1}, 2, 0, 0]

This yields the equivalence class,

$$1 + e_1 - e_2 - e_{12} \ \longmapsto\ \left[\frac{1}{2}+\frac{e_1}{2}\right].$$

One might wonder why the first argument is passed in as a list of a list. We have set up the coding so that we can handle examples of CPS having higher dimensions which involves more than one Clifford element. We express this in the next example.

- Projective line example of $\mathcal{C}\ell_{2,0}$:
  $\mathrm{u2} = \{2,\ 1,\ 0,\ 1/2\};$
  eqc2 = EquivalenceClass[{u1, u2}, 2, 0, 1]
  The equivalence class here is,

$$\begin{pmatrix} 1+e_1-e_2-e_{12} \\ 2+e_1+\frac{1}{2}e_{12} \end{pmatrix} \longmapsto \begin{bmatrix} \frac{1}{2}+\frac{e_1}{2} \\ 1-e_1+\frac{15e_2}{4}+\frac{41e_{12}}{4} \end{bmatrix}.$$

Now, consider $u \in \mathcal{C}\ell^2_{2,0}$ with a matrix representative with linearly independent columns. In particular consider these six matrices:

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ 0 & 0 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ 0 & 0 \\ 0 & 0 \\ x_2 & y_2 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ x_1 & y_1 \\ 0 & 0 \\ x_2 & y_2 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$$

where $x_1y_2 - x_2y_1 \neq 0$. Using column reduction on each of these matrices yields these matrices respectively:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a & b \\ c & d \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ a & b \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ a & b \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{where } a = \frac{x_3\,y_2 - x_2\,y_3}{x_1\,y_2 - x_2\,y_1},\ b = \frac{x_1\,y_3 - x_3\,y_1}{x_1\,y_2 - x_2\,y_1},$$

$$c = \frac{x_4\,y_2 - x_2\,y_4}{x_1\,y_2 - x_2\,y_1},\ \text{and } d = \frac{x_1\,y_4 - x_4\,y_1}{x_1\,y_2 - x_2\,y_1}$$

c. This definition in Goode-Annin's text actually defines this as "row-echelon matri- ces". Even though we cite this definition as "column-echelon matrices", we note that the definitions are analogous where matrices are transposed to one another. The definition states, "An $m \times n$ matrix is called a reduced row-echelon matrix if it satisfies the following conditions: (1) It is a row-echelon matrix. (2) Any column that contains a leading 1 has zeros everywhere else." [3, p.147]

We find that the scaling matrix required to yield such reduced matrices is

$$\begin{pmatrix} \frac{y_2}{x_1\,y_2 - x_2\,y_1} & \frac{-y_1}{x_1\,y_2 - x_2\,y_1} \\ \frac{-x_2}{x_1\,y_2 - x_2\,y_1} & \frac{x_1}{x_1\,y_2 - x_2\,y_1} \end{pmatrix} \quad (4)$$

which is the inverse of the matrix

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$$

Now, matrix (4), as its Clifford representative, is Λ^*:

$$\Lambda^* = \frac{1}{2}\left[\frac{y_2 + x_1}{x_1y_2 - x_2y_1} + \frac{y_2 - x_1}{x_1y_2 - x_2y_1}e_1 - \frac{y_1 + x_2}{x_1y_2 - x_2y_1}e_2 - \frac{y_1 - x_2}{x_1y_2 - x_2y_1}e_{12}\right] \quad (5)$$

The inverse of equation (5) is

$$\frac{2(x_1y_2 - x_2y_1)}{(x_1 + y_2) - (x_1 - y_2)e_1 - (x_2 + y_1)e_2 + (x_2 - y_1)e_{12}}$$

It follows that (5) is invertible because the denominator of its inverse is nonzero. The only way the denominator would be zero is when $x_1 = x_2 = y_1 = y_2 = 0$ and this cannot happen per definition 1.

By this example, we may have found the canonical matrix representative for the module $\mathcal{C}\ell^2_{2,0}$. Moreover, we believe this applies for $\mathcal{C}\ell^{n+1}_{2,0}$. By this, we propose the following conjecture.

Conjecture 1.
Let $u \in \mathcal{C}\ell^{n+1}_{2,0}$ be defined as follows:

$$u_j = \frac{1}{2}\left[(x_k + y_{k+1}) + (x_k - y_{k+1})e_1 + (x_{k+1} + y_k)e_2 + (x_{k+1} - y_k)e_{12}\right],$$

$$\text{where } j = \frac{k+1}{2} \text{ and } k = 1, 3, 5, ..., 2n+1.$$

$$\text{Let } i < t \text{ such that } x_iy_t - x_ty_i \neq 0.$$

Also, assume that the matrix representative for u can be reduced to a matrix with its i^{th} element in the first column and t^{th} element in the second column as the leading coefficients (pivots) per the definition of a column reduced matrix (refer to footnote c. on page 6) [3, p.147]. Then, the equivalence class for u is determined by the scalar, $v \in \mathcal{C}\ell^*_{2,0}$:

$$v = \frac{1}{2}\left[\frac{y_t + x_i}{x_iy_t - x_ty_i} + \frac{y_t - x_i}{x_iy_t - x_ty_i}e_1 - \frac{y_i + x_t}{x_iy_t - x_ty_i}e_2 - \frac{y_i - x_t}{x_iy_t - x_ty_i}e_{12}\right]$$

Going about proving this conjecture can be done, but may be quite messy to do by hand. Consider the case of $\mathcal{C}\ell^2_{2,0}$. For this, we could perform the same procedure done to prove theorem 1 by exhaustion where we use the same method for each of the six matrices we considered on the previous page for $\mathcal{C}\ell^2_{2,0}$. However, as you may have already guessed, we wouldn't want to do that for $\mathcal{C}\ell^{n+1}_{2,0}$ as that would take too long. We will indeed look for a simpler approach to proving this conjecture.

Conclusion

Recall that from definition 1, we sought to establish a canonical form of an invertible scalar which would ultimately lead to the required equivalence class for a Clifford module. Using isomorphic properties whereby we map the Clifford module to a $2(n+1)\times 2$ matrix (in the case of $\mathcal{C}\ell^{n+1}_{2,0}$), we used matrix reductions to find the equivalence class. Realizing that matrix reduction methods from lin- ear algebra are achieved via elementary operations, it became a matter of using this definition from linear algebra step by step.

Mathematica can use matrix reductions quickly which is beneficial for calculating canonical equivalence classes in Clifford projective space in high dimension. Our Mathematica package implemented the use of the matrix representatives for a given Clifford module, and operated via matrix reduction to achieve the associated equivalence class.

We wish to continue our work with this Mathematica package to make it better so that it can handle any type of Clifford module. In particular, we feel that theorem 1 and conjecture 1 could be applied directly to $\mathcal{C}\ell_{3,0}$ as well as other Clifford algebras such that $\mathcal{C}\ell_{p,q} \simeq \text{Mat}(2, \mathbb{R})$.

Acknowledgments

To my mentor, Dr. Agnew, for his guidance and inspiration.

References

P. Lounesto, "Clifford Algebras and Spinors Second Edition," Cambridge University Press (2001).

A. F. Agnew and S. P. Childress, "Matrix Projective Spaces and Twistor- like Incidence Structures," Journal of Mathematical Physics 50, 122503 (2009).

S. W. Goode and S. A. Annin, "Differential Equations and Linear Algebra Third Edition," Pearson Education, Inc. (2007).

Biomimetic Pattern Recognition in Cancer Detection

Department of Mathematics, California State University, Fullerton

Leonila Lagunes
Advisor: Charles H. Lee

Abstract

Cancer treatments have been shown to be more effective if the cancer is detected and treated at an early stage. Current cancer detection methods include imaging as well as tissue and blood-sample testing. These methods are expensive and invasive for patients, thus scientists have been driven to develop new alternatives to detect cancer. Biomimetic Pattern Recognition (BPR) is a technique that constructs a hyper-dimensional (HD) geometric body by mimicking a biological system and uses it for classification. BPR is derived from the Principle of Homology-Continuity, which assumes elements of the same class are biologically evolved and continuously connected. In other words, between any two elements of the same class, there is a gradual connection. These connecting branches form HD line segments or hyper-surfaces. The resulting topological structure, known as a biomimetic structure, mimics a biological class. In recent years, BPR has been successfully used in voice, facial, and iris recognition software. In this project we developed new BPR algorithms and classification schemes to detect specific cancers using DNA microarray data. We investigated the performance of the proposed BPR methods on data for bladder, colon, leukemia, and liver cancers. Results indicate that the proposed BPR has an increase in recognition rate when compared to previous techniques. BPR has shown to be a promising approach for cancer detection using DNA microarray data.

Introduction

Cancer treatments have been shown to be more effective if detected and treated at an early stage. Current cancer detection methods include imagining and blood-sample testing. Cancer imaging encompasses various techniques including traditional X-Rays, X-Ray-based computed tomography, Magnetic Resonance Imaging, Positron Emission Tomography, ultrasound scans, and endoscopy [1]. Current detection methods can be expensive and invasive, driving scientists to develop alternative methods for detection, such as pattern recognition. Pattern recognition techniques such as Support Vector Machine, Discriminatory Analysis, etc. have been used in cancer detection in the past. In this paper, we consider and develop new Biomimetic Pattern Recognition techniques.

DNA Microarray Data

Diagnosis and treatment of cancer can be improved by characterizing gene expression levels in healthy and cancerous tissue. Gene expression levels can be studied through microarray technology. Microarray technology allows researchers to measure and monitor the expression levels of thousands of genes simultaneously for a given organism [1]. DNA microarray data can be used to determine which genes are expressed at different levels between cancer-free cells and cancer-containing cells [2]. Biologists gather DNA from both cancerous and healthy cells for comparison and is tagged with red fluorescence for cancer and green for normal. DNA fragments then bind to their complements in a microarray chip as a part of a process called *hybridization* (Figure 1). A red spot indicates that the gene is highly expressed in a cancer cell and minimally in a healthy cell [2]. Green signifies that the gene is minimally expressed in a cancer cell and highly in a healthy cell. Yellow fluorescence shows that a gene is almost equally expressed in both cells. A black spot indicates that the gene is inactive in both cell [3]. A laser then scans the microarray and determines the expression levels of each gene according to the intensity of the color and is given a numeric value. Each sample is defined as a sequence of numerical values of gene expression levels. In recent years, DNA microarray technology has provided a promising tool to determine the diagnosis and prognosis of different cancer types [4-7].

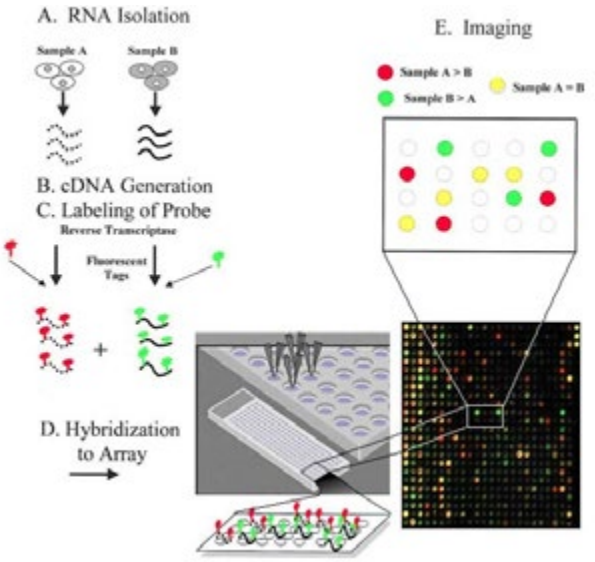


Figure 1. *DNA microarrays process from reference [3]. (A) DNA from a cell is extracted. Each DNA segment has a corresponding "spot" on the microarray chip. (B) When comparing gene expression levels, DNA from each cell is labeled with different fluorescent tags and hybridized on the microarray chip.*

Biomimetic Pattern Recognition

Biomimetic Pattern Recognition (BPR) is a technique constructing a hyper-dimensional (HD) geometric body to mimic a biological system for classification. BPR was first introduced by Shoujeu Wang in 2002 in Beijing, China and was derived from the Principle of Homology -Continuity (PHC) [8]. PHC assumes that the difference between elements of the same class is gradually changed. In other words, there is a gradual connection between any two elements that belong to the same class. These connecting branches can be HD line segments or hyper-surfaces and the resulting topo- logical structure forms a "biological" organism, which can be used for classification. One special characteristic of BPR is that it requires only a small number of samples as opposed to traditional pattern recognition algorithms. In recent years, BPR has been used successfully in voice recognition [9], iris recognition [10], and facial recognition [11]. BPR methods include different constructions as well as different classification techniques.

In this paper, the focuses are to develop two new techniques for developing BPR algorithms and apply them to DNA microarray data for cancer de- tection. We aim to build HD topological formations (skeleton-like structures) and pattern recognition schemes. We propose a new approach to the PHC, where elements of the same class are topologically assembled as nodes in a HD space and are con- nected by means of nearest neighbor.

Methodology

Data Sets

The BPR technique introduced in this paper is a general one, and can therefore be applied to any data set in a specified format. An applicable data set should be a n_S –by– m_G matrix, where n_S is the number of samples and m_G is the number of objects, such as genes, images, or frame data. However, for our purposes, we apply this technique to DNA microarray data.

BPR Algorithm with Hyper-Dimensional Line Segments

Training Process

By providing a new approach to the PHC, we connect nodes from the same class in a HD space by means of nearest neighbor. In order to build the HD topological formations, it is important to understand points and line segments in HD space. Let $\vec{x}$ be an element in $\mathbb{R}^n$ and $\overline{x_1x_2}$ be a line segment in $\mathbb{R}^n$ as well. The minimum distance, D, from $\vec{x}$ to a $\overline{x_1x_2}$ is determined based on whether its projection is inside the line segment. Let $\vec{u} = \frac{\vec{x_2}-\vec{x_1}}{\|\vec{x_2}-\vec{x_1}\|}$ be a unit vector going from $\vec{x_1}$ to $\vec{x_2}$ and $q = \langle \vec{x} - \vec{x_1}, \vec{u} \rangle$ be the projection of $\vec{x}$ onto the line segment. Note that $\|\cdot\|$ and $\langle \cdot \rangle$ denote the usual Euclidean norm and the inner product in $\mathbb{R}^n$, respectively. It can be shown that

$$D = \begin{cases} \|\vec{x} - \vec{x}_1\| & q < 0 \\ \sqrt{\|\vec{x}_2 - \vec{x}_1\|^2 + q^2} & 0 \le q \le \|\vec{x}_2 - \vec{x}_1\| \\ \|\vec{x} - \vec{x}_2\| & q > \|\vec{x}_2 - \vec{x}_1\| \end{cases} \tag{1}$$

Let S be the set of M elements of the training set and U be an empty set. Without loss of generality, let $\vec{x_1}$ and $\vec{x_2}$ be the two closest elements in S. Remove $\vec{x_1}$ and $\vec{x_2}$ from S and add them to U so that $U = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix}$. Then, we select the next element $\vec{x}_3$ in S so that its distance to the line segments in U is minimal; currently U simply contains a line segment connecting $\vec{x_1}$ to $\vec{x_2}$. Again, we remove $\vec{x}_3$ from S and add it to U in the following fashion:

$$U = \begin{pmatrix} \vec{x}_1 & \vec{x}_3 \\ \vec{x}_2 & \vec{x}_3^* \end{pmatrix} \tag{2}$$

where $\vec{x}_3^*$ is determined based on the proposed algorithms:

I.a. Nodal Connection: Connect $\vec{x}_3$ to the closest node of the line segment $\vec{x_1}$ to $\vec{x_2}$. In this case, $\vec{x}_3^*$ is either $\vec{x_1}$ or $\vec{x_2}$.

$$\vec{x}_3* = \begin{cases} \vec{x}_1 & \text{if} \quad \|\vec{x}_3 - \vec{x}_1\| < \|\vec{x}_3 - \vec{x}_2\| \\ \vec{x}_2 & \text{if} \quad \|\vec{x}_3 - \vec{x}_2\| \leq \|\vec{x}_3 - \vec{x}_1\| \end{cases} \qquad (3)$$

II.a. Segment Connection: Connect $\vec{x}_3$ to the closest element of $\overline{x_1 x_2}$. In this case, $\vec{x}_3^*$ could be $\vec{x_1}$, $\vec{x_2}$ or a new element, $\vec{x_t}$ (not from the original training set) on the segment from $\vec{x_1}$ to $\vec{x_2}$.

- If the projection of $\vec{x}_3^*$ lies outside $\overline{x_1 x_2}$, then $\vec{x}_3^*$ is defined as in Equation 3. Figure 2 shows how the projection of $\vec{x}_3$ can lie inside (A) or outside (B)

- If the projection of $\vec{x}_3$ lies inside $\overline{x_1 x_2}$, then

$$\vec{x}_3* = \vec{x}_1 + (\vec{x}_3 - \vec{x}_1) \bullet (\vec{u}) * \vec{u} \qquad (4)$$

where $\vec{u} = \frac{\vec{x_2} - \vec{x_1}}{\|\vec{x_2} - \vec{x_1}\|}$.

Continue in this fashion until S has been exhausted. At the end of the algorithm, the set U will contain $(M - 1)$ segments

$$U = \begin{pmatrix} \vec{x}_1 & \vec{x}_3 & \cdots & \vec{x}_M \\ \vec{x}_2 & \vec{x}_3^* & \cdots & \vec{x}_M^* \end{pmatrix} \qquad (5)$$

with at least one node of the segment being an element of the training set. Notice that while both algorithms use the same name set, a different structure is constructed. Figure 3 shows an example of the development of and the contrast between both proposed algorithms in $\mathbb{R}^2$. The Segment Connection algorithm provides a more compact structure than that from the Nodal Connection algorithm. Namely, the sum of all minimum distances in U is smaller for the Segment Connection algorithm. The algorithms are performed on each training class, hence, they produce two "biological organisms," one structure for the Training Normal class and one for the Training Cancer class.

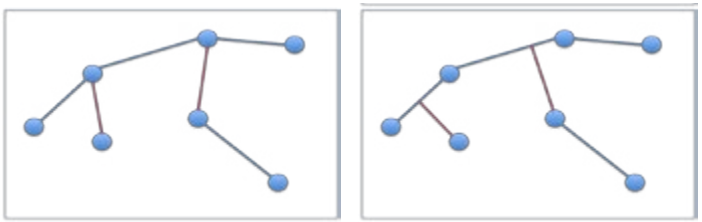


Figure 3. *Development of the skeleton-like structures using two assembling algorithms. In the Nodal Connection (left), the next point is selected based on its minimal distance to the current structure. In the Segment Connection (right) the next point is the closest point on the segment.*

We also defined two different extension pos- sibilities, where U can be composed of nodes which can connect to more than one node from S ("Multi- Lateral"), or U can contain nodes which can only connect to a one node ("Sequential").

1. **Multi-lateral:** The algorithm proceeds such that each line segment can connect to any previously constructed line as shown in Figure 4.

2. **Sequential:** The algorithm proceeds such that each line can only connect to a node that does not belong to U as shown in Figure 4.

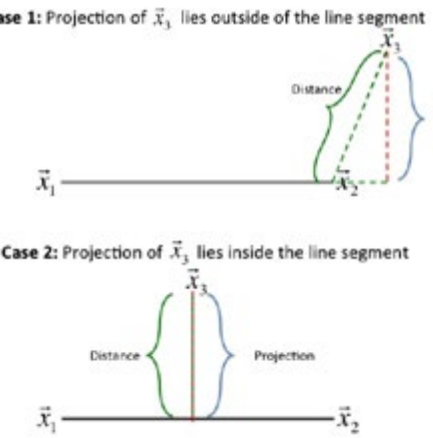


Figure 2. shows how the projection of a point $\vec{x}_3$ can lie outside (Left) or inside (Right) a line segment, namely $\overline{x_1 x_2}$ (Top) Case 1 shows the projecting landing outside the line segment. (Bottom) Case 2 shows the projection landing inside the line segment. In both cases, the projection is a red dashed line and the distance from the point to the line segment as a green dashed line.

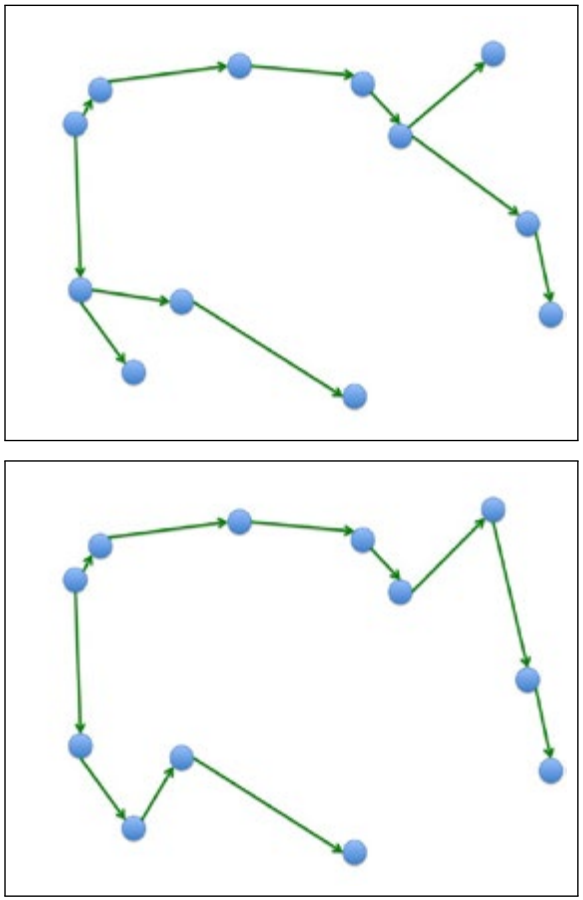


Figure 4. *Shows the different possible extensions. (Top) The multilateral extension shows how some nodes can connect to more than two nodes. (Bottom) The sequential extension shows how each segment can connect only two nodes.*

In Figure 4, we see how the "Sequential" Extension grows outward while the "Multilateral" can expand from a single node into different directions. The resulting skeleton-like structures, one for the cancer training set U_C and one for the normal training set U_N, are used for the classification of an arbitrary node from the test set, T_S. Two classification techniques are introduced. Accuracy for the algorithm is calculated based on the TruePositive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values with Equation (6) We also determine sensitivity and specificity in Equations (7) and (8) respectively.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP}. \quad (8)$$

Accuracy is used to calculate the percent of correctly classified samples. Sensitivity shows the ability of the algorithm to correctly identify a sample as cancerous, while specificity gives the ability of the algorithm to correctly identify a sample as cancer-free. Ideally, sensitivity should be close to 100% since, clinically, we do not want to identify a cancerous sample as cancer-free. An optimal method will reach close to 100% for all three of the above.

Classification Process:

Two structures are developed from the Training algorithm one for the cancer training set (U_C) and one for the normal training set (U_N). The resulting structures provide a basis for classification of an arbitrary node from the test set, T_S. We introduce a classification technique which we call "BPR Proximity" where an arbitrary node is classified as part of a class depending on its location relative to each "skeleton". If the distance from the node to the structure of class A is closer than the distance from the node to the structure of class B, then the node is classified as part of class A. Figure 5, depicts a visual representation of the **BPR** Proximity method. Mathematically, one can write the classification rule as follows:

$$\text{Class}(\vec{x}) = \begin{cases} \text{Cancer} & \text{if } \|U_C - \vec{x}\| \leq \|U_N - \vec{x}\| \\ \text{Normal} & \text{if } \|U_N - \vec{x}\| \leq \|U_C - \vec{x}\| \end{cases} \quad (9)$$

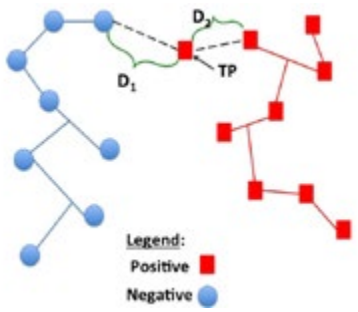


Figure 5. *The "BPR Proximy classification method. The True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values are determined based on the node's location relative to each derived "skeleton." Red nodes indicate cancer samples, and blue squares signify normal samples. The line segment structure represents the skeleton-like formation forthe cancer training set.*

For a fixed number of genes considered, a fixed holdout percentage (percentage of the total samples used for training), and a cancer type, the proposed BPR algorithms are run numerous times, each with different randomly picked training and testing sets. The average accuracy A_n^H (cancer) is recorded. Due to a large number of parametric variations (number of genes, holdout percentages, cancer types), a unified metric is needed to assess the performance of the proposed algorithms. A geometric mean,

$$G_n^H = \sqrt[Ntypes]{\prod_{Cancer=1}^{Ntypes} A_n^H(Cancer)} \qquad (10)$$

is calculated the for all considered cancer types, where n is number of genes considered and H is the holdout percentage. We use the geometric mean since the average accuracy is given in percentages across different parameters. A different form of generalized mean, like the arithmetic mean, cannot account for percent changes from one metric to another. We also define the algorithm performance as:

$$P_n = \sqrt[NH]{\prod_{i=1}^{NH} G_n^{H_i}} \qquad (11)$$

where $\{H_i\}_{i=1}^{NH}$ is a set of holdout percentages. We also use the overall performance for each cancer type, O_p, defined as:

$$O_{\text{Cancer}} = \sqrt[3]{\text{Sensitivity} * \text{Accuracy} * \text{Specificity}} \qquad (12)$$

for each cancer type. To then determine the performance for each method in order to identify the optimal technique, we use the Overall Performance across all cancer types, defined as,

$$O_{\text{All}} = \sqrt[4]{O(\text{Bldr}) * O(\text{Col}) * O(\text{Leuk}) * O(\text{Liv})} \qquad (13)$$

where, Bldr is Bladder, Col in Colon, Leuk is Leukemai, and Liv is Liver cancer. Equations 11 - 13 help determine the optimal technique.

Results

In this paper, the proposed BPR methods (two biomimetic construction algorithms and two classification methods) were applied to four different cancer types (bladder, colon, leukemia, liver), generating four data sets. Each data set contained caner-free (normal) and cancerous (cancer) samples. For bladder cancer, the data set contained 125 samples, of which 103 were cancerous and 22 were normal. Data for 6688 genes was provided by reference [15]. For colon cancer, the data set contained 62 samples, of which 40 were cancerous and 22 were normal. Data for 2000 genes was provided from reference [16]. For leukemia, the data set contained 73 samples, of which 48 were Acute Myeloid Leukemia (AML) and 25 were Acute Lymphoblastic Leukemia (ALL). We considered the AML samples to be cancerous and the ALL samples to be normal. Data for 7129 genes was taken from reference [17]. For liver cancer, the data set contained 181 samples, of which 105 were cancerous and 76 were normal. Data for 5520 genes was provided by reference [18]. Table 1 summarizes the microarray data sets used in this study.

Several metrics have been proposed for as- sessing the accuracy of the BPR algorithms. Accuracies are calculated based on the a verage of 100 runs. We ran 100 times in order to obtain objective view of the algorithm performance with different metrics since Training and Testing sets are randomly selected. Accuracy is calculated based on the remaining test set with Equation 6. Below are results obtained as well as how optimal conditions are determined and highest accuracy attainable for each cancer type. Table I summarizes the microarray data sets used in this study. Data for liver and bladder has been provided by genome- www5.stanford.edu in reference [12] and [13] respectively. Colon and leukemia data came from reference [14] and [15], respectively. Limitations of our study are discussed in the Conclusion section of the paper.

In order to develop the HD skeleton-like structure, we considered the gene expression level of each sample as a single node in the space, of either cancer-containing or cancer-free cells. The Cancer Training Set, S_C, and Normal Training Set, S_N, were randomly selected from the total number of samples, based on a Holdout value (percentage of the total samples used for training alone). Holdout percentages used in this study included 33%, 50%, and 80%. The remaining samples composed the Testing sets. Given S_C, and S_N, we defined the following two techniques when choosing the training and testing sets.

I. **Differential Mean:** foreachgene,g,wedefine the differential mean, D(g), as

$$D(g) = |\overline{S}_C(g) - \overline{S}_N(g)|. \qquad (14)$$

We then sorted the differential mean values for all the genes considered in descending order and chose the highest differential mean to construct a HD biomimetic structure.

II. **Cross Validation:** for each gene, *g*, we define the cross validation, crossvalind, as described in MatLab 2011a Bioinformatics toolbox. This function generates randomly selecting a train- ing set and test set.

For each simulation, the training sets contained a random selection of samples from all of the raw data amassed. Figure 6 shows the average gene expression levels for each type of cancer in our study.

After testing each algorithm, we determined the overall performance with Equations 10 and 12. In Figure 7 we see the overall performance across all cancers. Specifically, we see which of the four methods consistently reaches the highest performance score. In each plot, we see the dark blue bar, representing the Multilateral Segment Connection method, almost consistently score higher than the other methods. It is important to also note that between the two plots in each figure, the methods in the Cross validation plot score higher than the ones in the Differential Mean plot.

The average accuracy for each cancer type using the mentioned metrics is compared to previous algorithms and experiments. We limited comparison to studies using the same data sets of DNA microarray samples. Table 2 summarizes the overall performance of the proposed BPR and performance from other articles using the same DNA microarray data. In Peterson (2004), the accuracy is calculated using a single dominant mode of the Principal of Orthogonal Decomposition (POD) method. Testing was done separately with either cancer or normal set [7]. In Abbasi (2007), an improved POD classification method was introduced, where accuracy is determined from a combination of both cancer and normal sets [8]. Lee uses a multi-nodal POD along with Support Vector Machine, Self-Organized Map, and Neural Networks to determine the accuracy for each cancer type [9]. The last column shows the attained accuracy when using the proposed Biomimetic Pattern Recognition method for each cancer type.

From Table 2 we see the highest accuracy occurs at Cross Validation and Multilateral. Except in the case for liver cancer, where the highest accuracy occurs with Differential Mean and Sequential. From Table 2 we see that the proposed BPR reaches higher accuracies than previous studies. Except in the case for liver cancer. In Table 2, the reported accuracy is that with the following parameters: Cross validation as training set selection, Multilateral as extension method, and Segment Connection in construction.

Conclusion

In this paper we proposed a new BPR algorithm which employs a different approach to the PHC where elements of the same class are connected according to nearest neighbor. This approach allows for the development of two different biomimetic structure constructions (Nodal and Segment). The proposed methods were applied to bladder, colon, leukemia, and liver cancer data. We considered different number of genes to test highest recognition rate on each cancer type.

Results from Figure 7 indicate that the combination of the Segment Connection construction and Multilateral extension yield higher accuracies than the other combinations examined. Table 2 suggests that the given combination of schemes with a 33% holdout value give a higher accuracy for each cancer type. We also determined that experiments shows the new BPR algorithm has high recognition rate when compared to previous techniques, as seen by Table 3. It is important to note that the proposed BPR has higher accuracies than the previous studies for bladder, colon, and leukemia cancers. However, it is believed that liver cancer is better detected with the Differential Mean and Sequential implementations. This could largely be due to the fact that liver is the largest data set, as described in Table 1, and thus a sequential method suffices. We also note that colon cancer has the lowest accuracy. We believe this could be due to the small number of genes available. However, Biomimetic Pattern Recognition has shown to be a promising tool for cancer detection using DNA microarray data. Due to constrains, we also expanded our BPR algorithm into Planar structures. For further information, please refer to the thesis in the Department of Mathematics, California State University, Fullerton.

Acknowledgements

The author would like to thank Drs. Amy- beth Cohen and Laura Arce for their support, mentoring, and biology consultation. Graphics and computations were generated with MatLab2011a. The author would like to acknowledge Dai Nguyen for help in the MatLab 2011a coding. This work was supported by a Minority Access to Research Careers grant to from the National Institutes of Health (2T34GM008612-17) and the CSUF Mc- Nair Program.

Data Sets				
Cancer Type	**Bladder**	**Colon**	**Leukemia**	**Liver**
No. of Genes	6699	2000	7129	5520
No. of Cancer Samples	103 Primary Tumor	40 Primary Tumor	48 Acute Myeloid Leukemia	105 Primary Tumor
No. of Normal Samples	22 Healthy Tissue	22 Healthy Tissue	25 Acute Lymphoblastic Leukemia	76 Healthy Tissue

Table 1: *Data sets used in studies.*

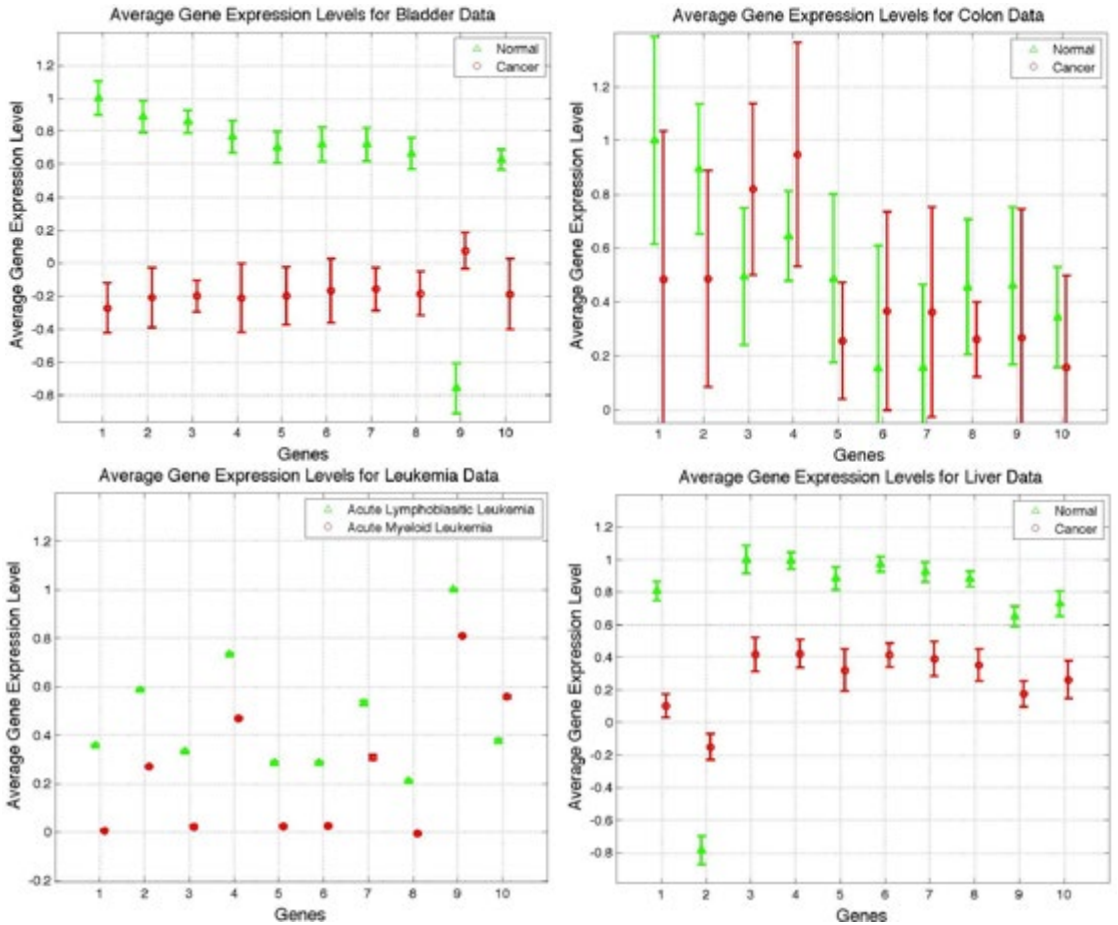


Figure 6. *Graphs represent the average gene expression level for genes with highest differential mean. Includes Bladder, Colon, Leukemia, and Liver average differential gene expression.*

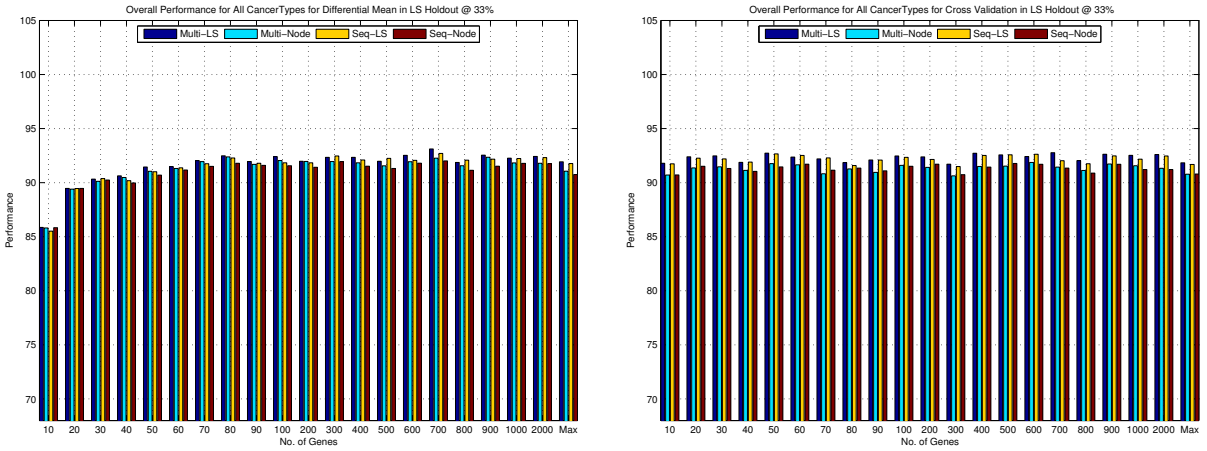


Figure 7. *Performance of Line Segment method across all cancer types. (Left) The performance across all cancers using Differential mean as the training set selection. (Right) The performance using Cross validation. For both plots, the x-axis represents the number of genes considered when implementing the technique and the y-axis shows the performance using Equation 12 and for all four methods. All methods were completed with 33% holdout 100 times.*

Overall Accuracy for each Cancer Type					
		Cancer Type			
Studies		**Bladder**	**Colon**	**Leukemia**	**Liver**
POD Cancer & Normal Separately Peterson (2004)		60.00%	N/A	N/A	75.00%
POD Cancer & Normal Together Abbasi (2007)		64.52%	N/A	N/A	82.30%
Multi-Modal POD Lee (2010)		N/A	65.35%	97.30%	96.43%
Cross Validation	Multilateral	98.95%	82.29%	94.09%	96.64%
	Sequential	98.61%	82.04%	94.03%	96.36%
Differential Mean	Multilateral	99.89%	81.59%	94.07%	95.95%
	Sequential	99.88%	81.24%	93.87%	97.00%

Table 2: *Overall Accuracy for each Method*

References

Hoopes. L. *Genetic Diagnosis: DNA Microarrays and Cancer.* (Nature Education, Scitable by Nature Education). 2008

Babu. M. *Microarray Data Analysis*,(Encyclopedia of Genetics, Computational Genomics, Proteomics and Bioinformatics, Horizon Press, UK). 2004

Vessey. K. *Use of Microarrays to Investigate Plant Response to Stress.* (GNC Training Workshop). 2008

Rhodes. D.R., Yu. J. *Large-scale meta-analysis of cancer microarray data identifies common transcrip- tional profiles of neoplastic transformation and progression*. Proceedings of the Natural Academy of Sciences of the United States of America. 2004

Peterson. D., Lee. C.H. *A DNA-based Pattern Recognition Technique for Cancer Detection*. Proceedings of the 26th Annual International Conference of the IEEE EMBS. 2004.

Abbasi. N., Lee. C.H. *Feature Extraction Techniques on DNA Microarray Data for Cancer Detection*. Proceedings of World Congress on Bioengineering. 200

Lee. C.B., Lee. C. H. *Extended Principal of Orthogonal Decomposition Method for Cancer Detection.* Proceeding of the International Journal of Bioscience, Biochemistry and Bioinformatics. 2010

Wang. S., Zhao. X. *Biomimetic Pattern Recognition Theory and Its Applications.* Chinese Journal of Electronics. Vol 3. No. 3. 2004

Qin. H., Wang. S., Sun. H. *Biomimetic Pattern Recognition for Speaker-Independent Speech Recognition*, Proceedings of the International Conference on Neural Networks and Brain. Vol. 2. (pgs 1290 - 1294). 2005

Zhai. Y., Zeng. J., Gan . J., Xu .Y. *A study of Biomimetic Pattern Recognition based Iris Recognition Method*, Proceedings of the International Sympo- sium on Information Processing. Vol. 2. (pgs. 71-74) 2009

Wang. Z., Mo. H., Lu. H., Wang. S. *A method of Biomimetic Pattern Recognition for Face Recognition*, Proceeding of the International Joint Conference on Neural Networks). Vol. 3. (pgs. 2216 - 2221). 2003

Chen, X., et. al., *Variation in Gene Expression Pat- terns in Human Gastric Cancer*s, Mol Biol Cell. 2003 Aug; 14(8): 3208-15. Epub 2003

Alon. U., et al. *Broad Patterns of Gene Expression Revealed by Cluster- ing Analysis of Cancer and Nor- mal Colon Tissues Probed by Oligonucleotide Arrays.* Proc. National Academic Science.

T. R. Golub, D. K. Slonim, et al. *Molecular classifi- cation of cancer: class discovery and class prediction by gene expression monitoring.* Science. 1999, 286: 531-537.

Chen, X. et al. *Gene expression patterns in human liver cancers.* Molecular Biology of the Cell. 2002.

A Statistical Analysis of Orange County K9-12 Mathematics Achievement Data

Department of Mathematics, California State University, Fullerton

Nathan Robertson, Soeun Park, Susan Deeb
Advisor: **Dr. Sam Behseta, Dr. David Pagni**

Abstract

In this work, we consider the problem of identification of those statistically significant variables that have an effect on the performance of K9-11 students in mathematics. We address this problem by constructing a number of statistical linear models with mixed-effects. Additionally, in order to identify the top performers, we employ three types of clustering techniques, namely, the method of *Kmeans*, *Hclust*, and *Mclust*. We compare the precision of these methods through measuring their misspecification rates via extensive simulation studies. Finally, we employ a Bayesian missing value imputation methodology in order to estimate some of the missing values associated with our data set.

Introduction

This data set was generated through an NSF funded project titled *Teachers Assisting Students to Excel in Learning Mathematics Phase II (TASEL-M2)*, led by professor David Pagni from the mathematics department at CSUF. As part of this extended study, California State Testing (CST) scores in mathematics were collected from a group of nearly 3000 K6-12 students whose names and personal information were removed to protect their privacy. An ultimate goal of this study is to examine students' socio-economic and demographic variables to view their effect on mathematical achievement.

Missing Values

In this blind data set, we had access to a number of variables such as students' math scaled scores from CST, mathematics courses taken, grade level progression, and mathematics raw scores for students in grade 6 to 11 from 2004-2008. Due to a steady flow of student migration in and out of the program at various time points, we encountered a considerable number of nonexisting values that may be interpreted as missing at random. It is those missing values that we aim to estimate.

The problem of missing values is a well-studied subject in statistical sciences (Little, 2011). Missing data problems often arise in the context of collecting educational data mainly because many students withhold personal and other types of information. Also, students get to take a variety of math courses, not necessarily in a sequence and thus, a gap will appear in a row associated with their records. In other words, our matrix of observations will be sparse due to missing observations per individuals.

In this work, we shall particularly concentrate on applying Bayesian methods for missing value estimation. This is beneficial for multiple reasons: 1 - we contribute to the sparse literature in the field, 2 - by applying model checking techniques, we will verify the quality of our missing value estimations, and 3 – we would potentially provide a more complete picture of the underlying relationship between the variables of interest.

Modeling

The original data set is a combination of surveys collected at different time periods, "waves" in order to measure students' and teachers' performance throughout the different school years. Our models are based on data from wave 3 of the survey which was conducted in the spring of 2005.

We employ a two-stage modeling scheme in order to first, identify the significance of teachers' effects, and second, quantify their strengths on the performance of freshmen and sophomore high school students in Orange County, California. The first stage borrows from the properties of the so-called "fixed-effect" linear modeling and is merely used to demonstrate whether teachers' performance would play any role on students' mathematical achievement. The second stage is built around the idea of utilizing multi-level regression modeling in order to measure the magnitude and statistical significance of teachers' effect on students' California Standards Test (CST) math scores.

Clustering

From the larger set of those students who were enrolled in either of the two studied high schools from the TASEL-M2 project, we selected a subset containing students enrolled in ninth grade in 2004 or 2005 totaling to 1070 students. We considered the following mathematical achievement variables: the students' standardized test scores from grades 9, 10, and 11, their highest CAHSEE score, and their overall math GPA. We implemented three clustering techniques in order to address two objectives: (a) identify which clustering technique attains the lowest misspecification rate, followed by (b) applying that method on student mathematical achievement data in order to study the trajectory of students' performances, as well as facilitate further comparisons between achievement variables.

Methodology

Missing Values

Suppose that the data are represented as an n x p matrix - a matrix with rows i = 1,..., n, and columns j = 1,..., p. Let $M_i = (M_1,\ldots,M_p)^T$ indicate a vector of zeros and ones such that $M_{i,j} = 1$ whenever the (i,j) cell is missing, and $M_{i,j}$ = 0, otherwise. Also, let $Y_{obs,i,j}$ be the observed value in row i and column j. Without loss of generality, we can assume a missing at random mechanism (our intent has been to demonstrate that in an extremely large data set such as the one we are working with, it is nearly impossible to seek out deterministic patterns for missing values, and thus, we treated the problem as one of missing at random). However, as noted in Little (2011), the mechanism has little effect on the Bayesian paradigm explained below. Indeed, Little recommends ignoring the missing data mechanism when it is justified. Note that the full data set will be the aggregate of $Y = (Y_{obs},\ Y_{missing})$, where $Y_{missing}$ are the missing data.

Now, suppose that $f(Y|\theta)$ is the distribution of all data, given all associated unknown parameters θ. Also, let $f(M|Y,\psi)$ represent the distribution of M given the full data set indexed by the unknown parameters ψ.

If there were no missing values, using Bayes theorem, one could write the posterior distribution of all parameters as

$$f(\theta,\psi|Y,M) = C \text{ x } \pi(\theta,\psi) \text{ x } L(\theta,\psi,|Y)$$

where, $f(\theta,\psi|Y,M)$ is the posterior distribution, C is a constant not dependent on ψ and θ, $\pi(\theta,\psi)$ is the full prior distribution on the parameter space, and $L(\theta,\psi,|Y)$ is the complete-data likelihood which can be written as

$$L(\theta,\psi|Y)=f(Y|\theta)f(M|Y,\psi).$$

However, with incomplete data, the full posterior distribution becomes

$$f_{full}(\theta,\psi|Y_{obs},M) \propto \pi(\theta,\psi)L(\theta,\psi,|Y_{obs},M)$$

where $L(\theta,\psi,|Y_{obs},M)$ is the likelihood due to observed values. Now, the observed likelihood is obtained by integrating all missing values out of the full likelihood:

$$L(\theta,\psi,|Y_{obs},M) = \int f(Y_{obs},Y_{missing}|\theta)f(M|Y_{obs},Y_{missing},\psi)dY_{missing}.$$

As in Rubin (1976), under the missing at random mechanism, one can assume

$$f(M|Y_{obs},\ Y_{missing},\psi) = f(M|Y_{obs},\psi)$$

for all missing values, and

$$\pi(\theta,\psi) = \pi(\theta)\ \pi(\psi).$$

Specifically, as per Hoff (2009), consider the case of data modelled via a multivariate normal distribution indexed with the mean parameter θ, and the covariance Σ. Subsequently, the sampling probability of data for student i can be given as

$$P(m_i,\{y_{i,j}:m_{i,j}=1\}|\theta,\Sigma)$$
$$= f(m_i)\ f(\{y_{i,j}:m_{i,j}=1\}|\theta,\Sigma) = f(m_i)\int(f(y_{i,1},\ldots,y_{i,p}|\theta,\Sigma)\prod_{y_{i,j}:m_{i,j}=0}dy_{i,j}).$$

The main consequence of this is that we can integrate out the missing data to obtain the marginal probability of observed data. This would allow us to utilize the simple Gibbs sampling algorithm to update the posterior distribution of θ and Σ along with implementing the Bayesian multiple imputation scheme.

In iteration (r+1) of the algorithm,

1) Sample $\theta^{(r+1)}$ from $f(\theta|Y_{obs},Y^{(r)}_{miss},\Sigma^{(r)})$;
2) Sample $\Sigma^{(r+1)}$ from $f(\Sigma|Y_{obs},Y^{(r)}_{miss},\theta^{(r+1)})$;
3) Sample $Y^{(r+1)}_{miss}$ from $f(Y_{miss}|Y_{obs},\theta^{(r+1)},\Sigma^{(r+1)})$

In the case of normal distributions, the marginal posterior distributions of θ and Σ as signified in steps (1) and (2) have closed forms when conjugate priors are used. That is, we consider

a multivariate normal prior on θ and an inverse-Wishart on Σ. It is also straightforward to obtain the marginal distribution of $(Y_{miss}|Y_{obs},\theta,\Sigma)$ in step (3) through $\prod_{i=1}^{n} f(Y_{i,miss}|Y_{i,obs},\theta,\Sigma)$, which would be proportional to the conditional distribution of the missing values given the observed values, the mean and the covariance parameters. The main advantage of this procedure is that in estimating missing values, the method is borrowing strength from the observed values.

Su et al. (2011) have developed a multiple imputation package in R called *mi*. In addition to implementing the Bayesian missing value imputation method, this program provides the users with sensitivity analyses associated with the missing value predictions to the effect that the efficiency of the predictions can be assessed. We aim to compare the various outputs of this software with the results obtained.

Modeling

Stage 1:

After removing students with missing information, we ended up with 1242 students who, in turn, were nested within 45 teachers. We consider three different measures of math achievement, taken as the response variable for our linear model. These measurements are students' mathematics GPA, CST scores and CAHSEE scores. In this stage, we would like to see whether there exists a teachers' effect. In other words, we would like to investigate the possibility of statistical significance of teachers' performances on their students' mathematics learning. To achieve that goal we construct a simple mixed-effects model with a nested structure in terms of student-teacher relationship. We can express each of the three models as

$$y_i \sim N\left(\alpha_{j[i]} + X\beta,\ \ \sigma_y^2\right),\ \ for\ i = 1,\ \ldots,\ 1242, and\ j = 1, \ldots, 45,$$

where y_i is student *i*'s math achievement (GPA, CST score, or CAHSEE score), X is the design matrix which consists of information on student's gender, ethnicity, English language proficiency, and the level of Math they are in, β is a vector of coefficients that corresponds to the predictors in X and is fixed for all teachers, and $\alpha_{j[i]}$ is the intercept for student *i* associated with teacher *j*. According to Gelman and Hill (2007), in usual regression, α_j would come from the classical least squares estimation for each teacher but in this model, α_j is assumed to be random so that

$$\alpha_j \sim \mathrm{N}(\mu_\alpha,\ \sigma_\alpha^2),\ \ for\ j = 1, \ldots,\ 45,$$

where μ_α is the mean level of all students, and σ_α^2 is the uncertainty for the j^{th} intercept . In multilevel modeling, α_js are partially pooled toward μ_α. The model pools teachers with fewer students toward the mean level more than teachers with a higher number of students.

Stage 2:

Taking the nested structure of this data set into account, we fitted a two level varying-intercept, varying-slope Mixed-Effects model including a teacher level predictor (see page 280 in Gelman and Hill, 2007). This model has a single student-level predictor *x* (the indicator for whether the student was Asian or Hispanic), and a single teacher level predictor μ (the average math GPA of students associated with each teacher). Thus,

$$y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]}x_i,\ \ \sigma_y^2\right),\quad for\ i = 1, \ldots,\ 1242,\ \text{and}\ j = 1,\ \ldots,\ 45,$$

and

$$\begin{pmatrix}\alpha_j \\ \beta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\gamma_0^\alpha + \gamma_1^\alpha \mu_j \\ \gamma_0^\beta + \gamma_1^\beta \mu_j\end{pmatrix},\ \begin{pmatrix}\sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2\end{pmatrix}\right),\ \ for\ j = 1,\ \ldots,\ 45,$$

where α_j and β_j vary by teacher and ρ is the between teacher's correlation parameter.

We can express this model by substituting the formulas for α_j and β_j into the equation for y_i:

$$y_i = \left[\gamma_0^\alpha + \gamma_1^\alpha \mu_{j[i]} + \eta_{j[i]}^\alpha\right] + \left[\gamma_0^\beta + \gamma_1^\beta \mu_{j[i]} + \eta_{j[i]}^\beta\right] x_i + \varepsilon_i.$$

Here $\gamma_1^\alpha \mu_{j[i]}$ and $\gamma_1^\beta \mu_{j[i]}$represent teacher *j*'s random effects on the intercept and slope of student *i*'s CST scores, respectively. The error term ε_i is iid $N(0, \sigma_\alpha^2)$ and is assumed to be independent of α_j and β_j. The model assumes that teachers' effects differ randomly across students with different ethnicities but that CST scores increase linearly with teachers' averaged GPA at the same fixed rate for all teachers.

An efficient way to estimate α_j, β_j, *and* σ_α^2 would be to use the likelihood function of y and maximize it with respect to each of these parameters. However, maximizing the likelihood function in the presence of repeated measures is known to produce biased estimations. To elaborate, this is a multilevel regression model in which students have repeated measurements and therefore the

regression coefficients are expected to carry high correlation between them and hence are not independent. Therefore, we consider a technique known as Restricted Maximum Likelihood Estimation (RMLE) for the purpose of parameter estimation in this setting (1975 DR. Cox. Partial Likelihood, Biometrika). RMLE is based on the idea of synthesizing the joint likelihood function for α_j, β_j, *and* σ_α^2 through forming conditional inferences based on the parameters that are not being estimated. Due to the lack of independence in this case, we consider a recursive estimation technique that would facilitate estimating a quasi-likelihood function (as opposed to the original likelihood function), which is an adjusted version of the original likelihood function that accommodates for bias.

To apply RMLE, we used the R package and we relied on codes described in "Data Analysis Using Regression and Multilevel/Hierarchical Models" by Andrew Gelman and Jennifer Hill.

Clustering

Kmeans Method

Suppose we have a data set consisting of n observations of a random D-dimensional variable x. Let the D-dimensional set of vectors μ_k for which $k=1,...,K$, be a prototype associated with the k^{th} cluster. We aim to assign all data points to clusters as well as a set of vectors $\{\mu_k\}$ such that we minimize the sum of squares of the distances of each data point to its closest vector μ_k (Bishop 2006).

For each x_n we assign a set of binary indicator variables $r_{nk} \in \{0,1\}$ where $k=1,...,K$ and $r_{nj} \in \{0, 1\}$, where $k=1,...,K$, such that x_n is assigned o or 1 dependent upon which cluster it is assigned to. We next define a function known as a distortion measure by

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n - \mu_k\|^2$$

which represents the sum of the squares of the distances of each data point to its assigned vector μ_k. We wish to minimize J by assigning values to r_{nk} for each x_n.

The *Kmeans* clustering method allows for us to choose the number of clusters that it will output. Since we generated three different distributions, we ran *Kmeans* with three clusters. We made these simulations based on a grand assumption that each cluster is uniquely represented by a multivariate normal distribution with distinct mean and covariance structures.

Hierarchical clustering method

This method clusters based on the distance between data points. It starts by clustering each data point as its own cluster. The distance between each cluster is then measured and the two points closest together are joined into one cluster. The process is repeated until the number of clusters is reduced to the desired amount. The algorithm used in each step is

$$D(r,s) = \max\{d(i,j)\},$$

where $D(r,s)$ represents the dissimilarity between nodes r and s, and $d(i,j)$ represents the distance between i and j where object i is in cluster r and object j is in cluster s.

We used the *Hclust* function in R to cluster the data hierarchically. The *Hclust* command requires that we select the number of clusters we prefer. Since we generated three different multivariate normal distributions, we chose three clusters. That is, the process of hierarchical clustering will repeat until there are only three clusters.

Model Based Clustering Method

Model based clustering computes an approximate maximum for the classification likelihood:

$$L_{CL}(\theta_1, \ldots, \theta_G; l_1, \ldots, l_n|y) = \prod_{i=1}^{n} f_{t_i}(y_i|\theta_{t_i}),$$

where l_i are labels classifying each observation. For example, $l_i = k$ if y_i belongs to the k^{th} component. Model based clustering merges pairs of clusters according to the greatest increase in the classification likelihood among all possible pairs. The process begins by treating each data point as its own cluster. When the probability model in the classification likelihood is multivariate normal with equal-volume spherical covariance λI, the selection criterion is the sum-of-squares criterion.

Mclust is a function that combines model-based hierarchical clustering, expectation maximization (EM) for Gaussian mixture models, and Bayesian Information Criterion (BIC) approximation (Fraley and Raftery 2002).

Most commonly, f_k is the multivariate normal (Gaussian) density $\emptyset_k$, parameterized by its mean μ_k covariance matrix Σ_k,

$$\phi_k\left(y_i\middle|\mu_k, \sum_k\right) = \frac{exp\left\{-\frac{1}{2}(y_i - \mu_k)^T \sum_k^{-1}(y_i - \mu_k)\right\}}{\sqrt{\det(2\pi \sum_k)}}.$$

Data generated by mixtures of multivariate normal densities are characterized by groups or clusters centered at the means μ_k, with increased density for points nearer the mean.

$$2\log(D|M_k) \approx 2\log p\left(D\middle|\hat{\theta}_k, M_k\right) - Y_k \log n = BIC_k \ ,$$

in which, D represents data and M_k depicts the k^{th} possible model to be considered. In this setting, the corresponding surfaces of densities are ellipsoidal. Multivariate normality facilitates clustering of data by changing shape, value, and orientation of each density (Fraley, Raftery, 1997). Another advantage of model-based clustering is the determination of optimal number of clusters which is achieved by criteria such as the Bayesian Information Criterion (BIC). The algorithm calculates the BIC for varying number of clusters, and selects the number of clusters for which the BIC is minimized.

Misspecification Rate: Comparing the Three Methods via Simulation

We started by generating random data in R from three separate multivariate normal distributions with equally spread means. Subsequently, we used *Kmeans*, *Hclust*, and *Mclust* on the simulated data. We then calculated what percentage of points from each clustering method was misspecified from the true scenarios. We then repeated this process while increasing the number of distributions being sampled as well as increasing the dimensionality of each underlying distribution.

Time Dependent Classification

We then applied the clustering methods on variables from the data set. Variables Math GPA and CST scores were paired on a yearly basis for analysis. The students were classified into different groups based on the clustering in order to observe their mathematical achievement relative to other students throughout their high school years.

The count of students with the highest mathematics achieving cluster is represented by $y_{(i_1,i_2,i_3)}$ in which: $(i_1, i_2, i_3) \in \{1,0\}$ where i_1, i_2, i_3 determine whether or not the students were in the highest scoring cluster.

This way $i_1 = 1$ specifies the student was clustered in the highest scoring cluster in 9th grade while $i_1 = 1$ specifies that the student was not placed in the highest scoring cluster in 9th grade. Similarly, i_2 refers to the clustering placement of student in grade 10 and i_3 refers to the clustering placement in grade 11.

Results

Missing Value

We implemented the multiple imputation approach of the previous sections on the achievement data set containing the scaled math scores of 3474 observations acquired between the years 2005-2008 from the TASEL-M2 project assembled by AnneMarie Conley from UC Irvine. To assess the quality of the imputation technique, we concentrated on the longitudinal data associated with math scaled scores in the same period. We employed the Gibbs algorithm of section 2 with a normal model having normal conjugate priors.

In Figure 1, we include a missing pattern graph. The graph is comprised of four blocks (rows), each showing the portion of available observed data in black, as opposed to the portion of missing data in white.

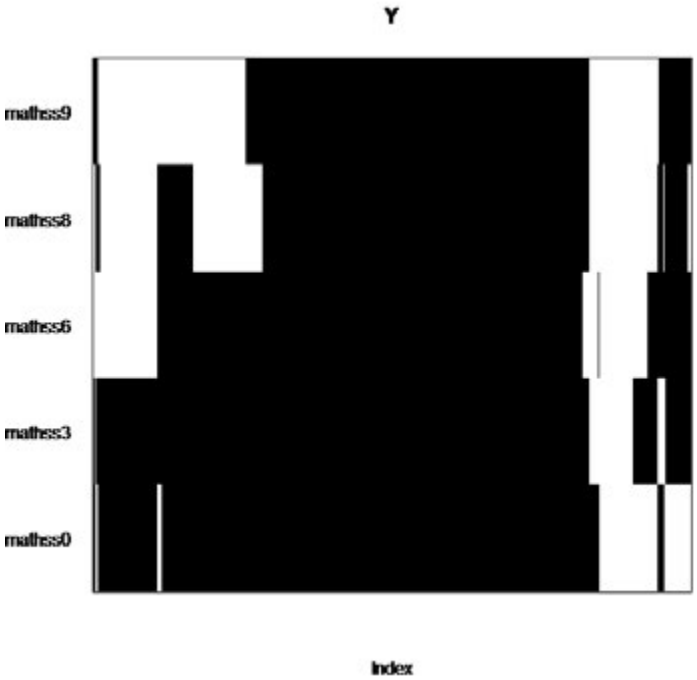


Figure 1. *MP plot used with the mi package for R. The black parts of the plot are the available data while the white are the missing data.*

The scaled math scores of 2008 are on the top block followed by those of 2007, 2006, 2005. It is apparent from Figure 1 that a considerable portion of the response variable is missing over the period of study.

In Figure 2, we show the posterior distribution of scaled math scores obtained from the Gibbs sampling scheme of section 2.

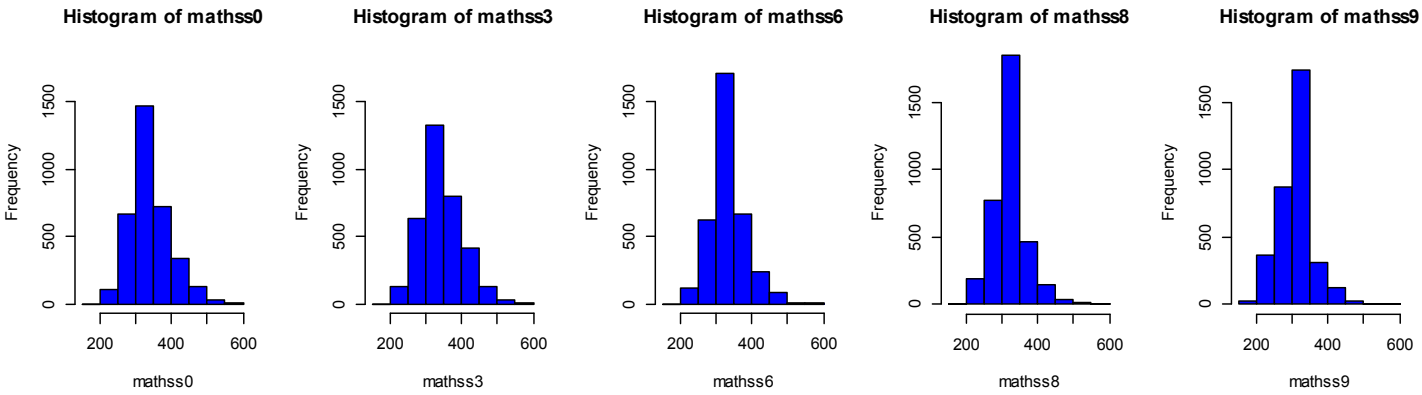


Figure 2. *Histograms of the filled out data sets generated from the Gibbs Sampler method.*

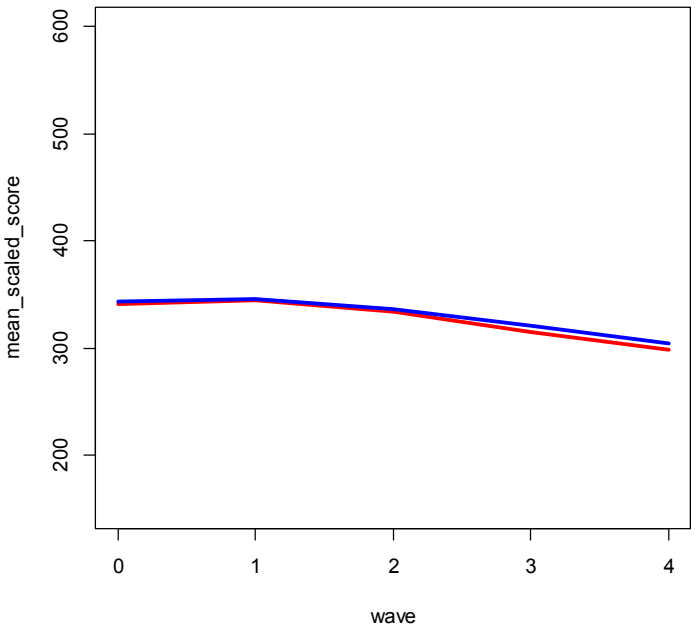


Figure 3. *A plot comparing the mean values generated by the mi package in R and the Gibbs sampler method. mi is in red while the Gibbs method is in blue.*

It is quite important to note that these distributions are realized through borrowing strength from the information of the full data (observed and missing) as manifested by the hierarchical algorithm of section 2.1. Finally, in Figure 3, we demonstrate the mean of each of the posterior distributions of the previous graph.

The decline in the mean scores is not surprising, especially considering that we followed students progressively from grade 6 to grade 11 meaning that, in each wave of the study, the students would have to deal with a more challenging set of mathematical problems. The Bayesian methodology applied in this work is quite robust and is computationally inexpensive. This approach is considerably more reliable than ad-hoc methods such as data deletion and estimating missing values with averages.

Modeling

As seen in Figure 4, teachers' performances play a major role in students' mathematics achievement.

We note that for some teachers there is a significant overlap between the three types of intervals of interest, while this pattern is easily distorted for some other teachers. Similarly, we could not establish a general pattern of preference for one measure of achievement in comparison to the others. That is, for a given teacher we can have students performing better in CST as opposed to GPA and CAHSEE, whereas for another teacher students performances for CAHSEE is shown to be superior.

Figure 5 demonstrates that Asian students universally scored higher on the CST than Hispanic students.

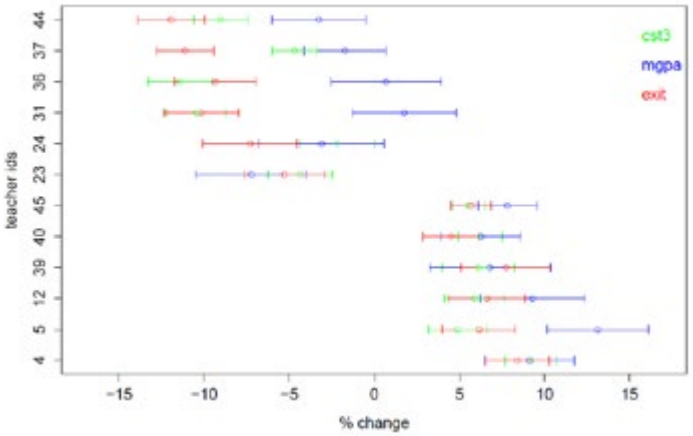


Figure 4. *A plot of 12 selected teachers (6 from the high performance group and 6 from the low performance group). The x-axis demonstrates the percent change from the overall mean. The y-axis represents teacher id. Different colors show a specific measure of achievement. The intervals are obtained by adding/subtracting one standard deviation from the parameter estimate in the mix-effect model. Clearly teachers for whom there is a significant overlap between the intervals have demonstrated higher consistency. Also intervals on the positive side of the x-axis can be perceived as effective mathematics learning.*

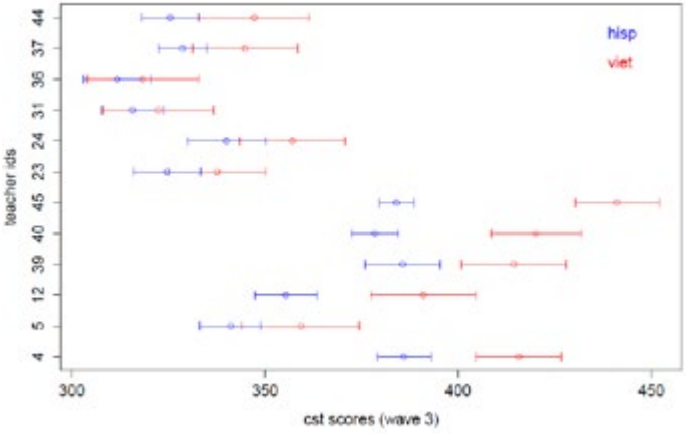


Figure 5. *A plot of the same12 selected teachers. The x-axis demonstrates CST scores taken from wave 3 of the study. The y-axis represents teacher id. Red shows the scores for Vietnamese students while blue show the scores for Hispanic students. The intervals are obtained by adding/subtracting one standard deviation from the parameter estimate in the mix-effect model.*

It is also worth noting that in most cases, the intervals did not overlap indicating a significant difference in scores for a given teacher.

Figure 6, takes gender into account and hints that there may be more significant differences between the CST scores for Asian males and females compared to those of Hispanic males and females.

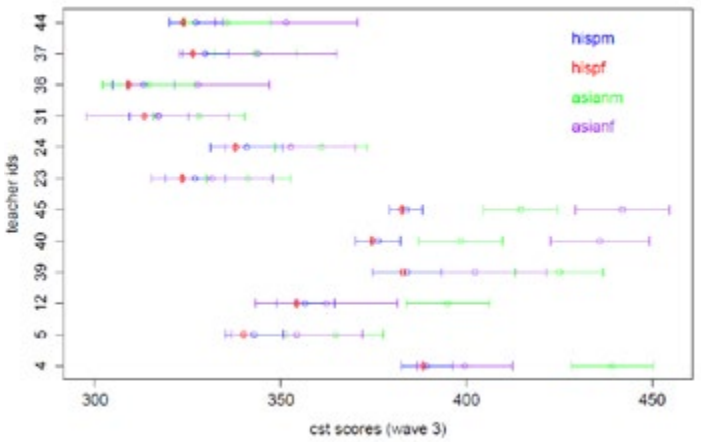


Figure 6. *A plot of the same 12 selected teachers. The x-axis demonstrates CST scores taken from wave 3 of the study. The y-axis represents teacher id. Different colors are used to show pairings of ethnicity and gender. Intervals are obtained by adding/subtracting one standard deviation from the parameter estimate in the mixed-effect model.*

Mathematical achievement does not have a unique definition. As such, we see that different measures yield different results indicating that the entire picture cannot be seen by looking at one dimension of mathematical achievement. Teachers' effect on mathematical achievement cannot be ignored.

Clustering

We notice from Table 1 that the *Kmeans* method yields a significant number of misclassified data points in the control case.

	Scenario I Centers far apart			Scenario II Centers closer together			Scenario III Centers closest		
Method	3 D	6 D	10 D	3 D	6 D	10 D	3 D	6 D	10 D
K-means	0	24.85	14.76	.067	24.95	14.88	26.03	19.22	10.19
Hclust	0	0	0	.033	0	0	38.77	42.57	26.47
Mclust	0	0	0	0	0	0	27.63	19.8	10.08

Table 1. *Misclassification rates, in percentages, for each tested scenario pertaining to 3, 6, and 10-dimensions.*

It is worth noting that sometimes the *Kmeans* method would misclassify entire distributions by generating artificially large or small number of clusters. *Kmeans* performs better in high dimensions as well as higher number of clusters that are close together. Additionally, under scenario III (see Figure 7) where underlying distributions are taken to be very close, we found *Hclust* consistently performing worse than the other two methods.

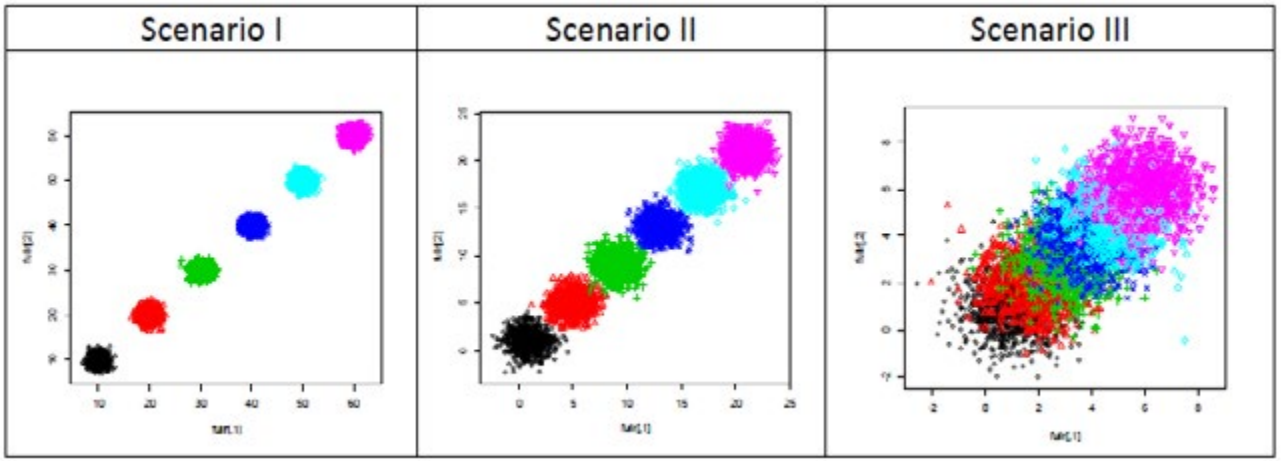


Figure 7. *Scenarios generated using Mclust for distributions of different means of 6-dimensional Gaussians. Scenario I is realized under the assumptions that the means of the distributions are 10 units apart. Scenario II is the case when the means are 4 units apart and, finally, in scenario III, the means are taken to be 1 unit apart. In the figure, each color represents its own cluster. Clearly, as the means become closer to each other, the clusters are also closer together allowing for more misclassification. In scenario I, the distributions are clustered perfectly while in scenario III, the clusters overlap.*

In general, *Mclust* performs quite well. However, when the predetermined number of clusters grows, the computational complexity of *Mclust* increases significantly.

A solid majority of the students in our data set were consistently classified in the lower achieving group throughout the three years of the study. There is a noticeable decrease in performance between students classified as being in the highest cluster, compared to those students who were not. We can observe from Table 2 that groups with smaller counts pertain to higher precision.

					GPA			Mathss averaged		
Trajectory			Count	Freq.	Mean	Median	Standard deviation	Mean	Median	Standard deviation
1	1	1	20	.019	3.774	3.917	.316	444.9	440.2	25.567
1	1	0	18	.017	3.488	3.5	.308	410.5	406.2	20.710
1	0	1	4	.004	2.688	2.688	.452	434.7	438.7	24.160
0	1	1	7	.007	3.446	3.5	.512	422.6	426.3	9.409
1	0	0	32	.030	2.463	2.464	.546	375.2	376.2	22.660
0	1	0	41	.038	3.328	3.375	.327	374.3	373.3	18.755
0	0	1	1	.001	2.75	2.75	NA	387.3	387.3	NA
0	0	0	947	.885	1.806	1.75	.738	306.1	302.7	32.890

Table 2. *Time Dependent Clustering; showing the count, frequency, and the mean, median, standard deviation regarding Math GPA and CST scores (mathss9, mathss10, mathss11) averaged for grades 9-11 for each trajectory.*

Consequently, data summaries in those groups can be used for a more precise estimation of the parameters of the underlying population from which this data is extracted.

Conclusion

In this work, we considered three approaches for the analysis of TASEL-M2 data. Even though these methods, namely, missing data imputation, modeling, and clustering are addressing different aspects of measuring the effect of explanatory variables on the response, they share the common objective of identifying those students whose mathematical performance was significantly higher (lower) than others. In this process, we encountered a number of technical and computational challenges, mainly stemmed from the fact that the underlying data set had an extremely complex structure. This was the main motivation for our simulation studies. In the future, we plan to tackle the same set of problems from a more theoretical perspective.

Acknowledgments

We would like to sincerely thank: the Mathematics Department Summer Research Program for providing us with the opportunity and funding for this research project; Dr. Sam Behseta for his professional guidance, sparking our interest in research; Dr. David Pagni for providing insight regarding the data set; Dr. AnneMarie Conley and Nayssan Safavian for their enthusiastic direction on the project. We would also like to thank Selene Black, Antouneo Kassab, Jimmy Kwon, and Calvin Pham for their contributions to the modeling portion of this project as well as Reina Galvez for her contributions to the clustering part of the project.

References

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York, NY.

Fraley, C., Murphy, T.B., Raftery, A.E., Scrucca, L. (2012). Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Seattle, WA.

Laird NM and Ware JH. (1982). Random-Effects Models for Longitudinal Data. *Biometrics.* Vol 38, 963—974.

Little RJA (2011) Calibrated Bayes, for statistics in general, and missing data in particular (with discussion). *Statistical Science.* Vol 26, No 2, 162-174.

Little RJA, Rubin DB (1987). *Statistical Analysis with missing data.* Wiley, Hoboken, N.J., 1st edition.

Little RJA, Rubin DB (2002). *Statistical Analysis with missing data.* Wiley, Hoboken, N.J., 2nd edition.

Lockwood JR, McCaffrey DF, Mariano LT, Setodji C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational And Behavioral Statistics.* Vol 32. 125-150.

Pinheiro J C and Bates DM.(2000). *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

Rubin DB. (1976). Inference and missing data (with discussion). *Biometrika* Vol 63, 581-592.

Rubin DB (1978). Multiple imputations in sample surveys. In *Proceedings of Survey Research Section* 20-34, American Statistical Association, Alexandria, VA.

Rubin DB (1987). *Multiple imputation for nonresponse in surveys.* Wiley, New York.

Sanders W, Saxton A, and Horn B. (1997). The Tenessee value-added system: a quantitative outcones-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluational measure?* (137-162). Crownin Press, Thousands Oaks, CA.

Spiegelhalter D, Thomas A, and Best N. (1997). WinBUGS: Bayesian inference using Gibbs Sampling (Technical Report). Biostatistics Unit. Cambridge, UK:

Su Y, Gelman A, Hill J, Yajima M. (2011) Multiple imputation with diagnostics (mi) in R: opening windows into the black box. *Journal of statistical software*. 45(2), 1-31.

Tanner MA, Wong WH. (1987) The calculation of posterior distributions by data augmentation (with discussion). *Journal of American Statistical Association.* Vol 82, 528-550.

Scattered Light Measurements for Advanced LIGO's Output-Mode-Cleaner Mirrors

Department of Physical Sciences, California State University, Fullerton

Adrian Avila Alvarez

Abstract

Abstract: The Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO), with sites in Livingston, LA and Hanford, WA, and its international partners, Virgo, GEO600, and KAGRA, are being built to detect gravitational waves, a phenomenon theorized by Einstein in his Theory of General Relativity (GR). By direct prediction of GR, gravitational waves are ripples in space-time that propagate at the speed of light and are created by violent astrophysical processes. The gravitational-wave detectors are all based on the Michelson interferometer, which has an input laser, a beam splitter, and two perpendicular arms with mirrors at each end. However, their configurations have significantly more complexity to augment the sensitivity. Higher order spatial modes can create 'junk light' that decreases the shot-noise limited sensitivity of the detectors. To combat this, each LIGO detector has an Output Mode Cleaner (OMC) at its detection 'dark port.' Scattered light from the OMC mirrors can reduce the shot-noise limited sensitivity of the instruments, and add noise via stray and counter-propagating light. Thus, it is important that the light scattering from the OMC mirrors in Advanced LIGO be minimal. This paper will describe measurements of the scattered light from sample Advanced LIGO OMC mirrors.

Authors

Adrian Avila Alvarez

Adrian is currently a junior physics major at CSUF and a member of both the Society of Advancing Chicanos and Native Americans in Science (SACNAS) and the Gravitational-Wave Physics and Astronomy Center (GWPAC). As part of GWPAC, Adrian conducts research under the supervision of Dr. Joshua Smith. As part of his research, Adrian measures how much light is being scattered on an optic's surface in attempt to contribute towards the detection of gravitational waves, which are a thousand times smaller than the diameter of a proton. Adrian has had the opportunity to present his research findings at numerous meetings, symposiums, and conferences. In October of 2013, he presented his research at the SACNAS National Conference in San Antonio, TX. Adrian plans to continue his gravitational-wave research while pursuing a PhD program. With a passion for working with equations, gravitational-wave research, and physics, expect to read much more about Mr. Adrian Avila-Alvarez.

Neha Ansari

Neha Ansari is a third year biochemistry major minoring in print journalism. Ansari currently serves as the Chemistry and Biochemistry Section editor for Dimensions. At Cal State Fullerton, Ansari conducts research in Dr. Peter de Lijser's organic chemistry lab where she studies the use of a carbonyl group as a nucleophile in photosensitized electron transfer reactions of oxime ethers. In the summer of 2013, Ansari was selected as an Amgen Scholar at the University of Washington, where she investigated the development of micelles in self-immolative polymers for the purpose of drug delivery in Dr. AJ Boydston's organic chemistry lab. Ansari is a member of the ASI Board of Directors representing the College of Natural Sciences and Mathematics, the University Honors Program, and is a President's Scholar.

Aaron Justin Case

"Aaron Justin Case is a geology enthusiast who is soon to graduate this summer with a Bachelor of Science degree Geology. He enjoys and participates in meetings and conferences to expand his knowledge and love for geology. His participation in theses conferences is indicative of his passion to learn. His research includes studies of geochemical correlation of basalts in northern Deep Springs Valley, California. Methods to correlate the basalts include X-Ray Fluorescence Spectroscopy (XRF). His research has been published in the GSA Abstracts with Programs Vol. 46, No. 5 and presented at the at the 2014 GSA Joint Rocky Mountain/Cordilleran Section Meeting. It is titled Geochemical Correlation of Basalts in Northern Deep Springs Valley, California, by X-Ray Fluorescence Spectroscopy (XRF). His research has made him proficient with the use of X-ray Fluorescence Spectroscopy, lab work, and geochemical understanding of basalt as well as associating rocks. He actively leads and arranges geology club hikes, areas includes corundum explorations in the San Gabriel Mountains and Salton Sea trips. Aaron is quite the personality and strives to succeed in his field."

Jennifer Castillo

Jennifer Castillo is hoping to graduate this spring with a Bachelor's degree in Chemistry with a minor in Criminal Justice. She joined Dr. Rogers' lab to research a method to synthesize anhydrous cyclopropanol. This method involved Schlenk techniques, vacuum filtration, extraction, rotary evaporation, 1H-NMR, and infrared spectroscopy. She hopes to have a career in forensic science. The experience she gained using the different instrumentation at CSUF and in Dr. Rogers' lab has been an extremely rewarding and beneficial to her career hopes.

Katherina Chua

Katherina Chua is a senior undergraduate majoring in Biochemistry. She started her undergraduate research experience in fall of 2012 and has participated in multiple research programs on campus and off campus, including Research Careers Preparatory Program in the fall of 2011, Minority Health and Health Disparities International Research Training Program in Buenos Aires, Argentina during the summer of 2012, and UCSF Summer Research Training Program during the summer of 2013. Katherina is currently conducting research under the guidance of Dr. Michael Bridges as a Minority Access to Research Careers (MARC) scholar studying the oligomeric structure of the intrinsically disordered protein stathmin and measuring interspin distances using electron paramagnetic resonance spectroscopy. Through this research, she is able to verify that stathmin exists as an oligomer in solution via site-directed spin labeling technique. In her multitude of undergraduate research experience, Katherina has decided to continue her pursuit of a research career and has applied to PhD graduate schools that focus on combining pharmaceutical sciences and biophysics. After graduating this May, she will be entering these PhD programs next fall.

Ashley Chui

Ashley Chui is a fourth-year undergraduate majoring in Biochemistry. She began her undergraduate experience in the spring of 2012 through the Research Careers Preparatory Program under the guidance of Dr. Michael Bridges studying the characterization of aggregates of the intrinsically disordered protein, stathmin. She was able to exhibit the reversibility of the aggregation of stathmin in its native environment as well as show that stathmin demonstrates a concentration-dependent behavior. Currently as a HHMI Two-Year Scholar, her project has evolved into the verification of the oligomerization of stathmin through a number of techniques, one being site-directed spin labeling electron paramagnetic resonance. Through Ashley's experience in the Bridges lab, she has developed a strong interest in biophysics and would like to continue conducting research in biophysical chemistry for graduate school, which she hopes to attend in the fall of 2015.

Kim Conway

I am a graduating senior double-majoring in Biology and Health Science. I am currently working in Dr. Paul Stapp's Vertebrate Ecology and Conservation Biology laboratory studying the prevalence of bot fly infestation in thirteen-lined ground squirrels in northern Colorado, while also attempting to identify the species of bot fly through molecular genetics techniques. I am a BURST (Biology Undergraduate Research Scholar Training) scholar at CSUF, and have been working as a veterinary assistant for 4 ½ years. I will be attending the UC Davis School of Veterinary Medicine next fall to obtain a DVM degree with a focus on wildlife medicine.

Jeremy Cordova

Jeremy Cordova is a graduating Geology student interested in structural geology and how it relates to geologic hazards and mineral deposits. For the last two years he has been working extensively in the field at Los Penasquitos Lagoon, in San Diego County, analyzing a potential paleotsunami deposit. As a member of multiple geological organizations Jeremy plans to become highly involved in his field. Graduating in the Fall of 2014 he will remain in the West working in geologic hazard assessment or the mineral industry, while pursuing a Master's Degree in Structural Geology.

D'Lisa O'hara Creager

D'lisa O'hara Creager is an undergraduate student that will be finishing up this summer with a Bachelor of Science degree in Geology. She has been working with Dr. Brady Rhodes on her senior thesis project studying paleotsunami deposits in Seal Beach, CA. Dr. Rhodes also advised her in completing a hydrogeochemical study done in Chiang Mai, Thailand during her participation in the Environmental Science Research in Thailand (ESRT) program in the summer of 2013. O'hara has contributed to the California State University, Fullerton Geology club for the past year and served as its historian. She enjoys attending conferences and has given a number of poster presentations on her research. In her spare time she enjoys playing volleyball, which she spent three years coaching at Sunny Hills High School in Fullerton, CA. O'hara plans on earning her Master of Science degree in Geology and is currently deciding on a graduate program.

Susan Deeb

Susan Deeb is a fourth year undergraduate student hoping to graduate Fall 2014 with a Bachelor of Arts degree in Mathematics as well as a minor in Physics. For the past year, she has been working on statistical analyses of TASEL-M2 data. TASEL-M2 is an NSF funded project, led by Dr. David Pagni, conducted on Orange County schools over several years. The data collected includes demographic information and mathematics scores of over 3000 K6-12 students. Susan has been working with a group of undergraduate students, under the supervision of Dr. Sam Behseta, to analyze this portion of the data. Together, they have presented their results to a group of researchers at University of California, Irvine. They have also had the opportunity to present in a small conference to mathematicians at Chapman University, as well as at SCCUR 2013 and most recently, the CSUF Research Competition, in which Susan was the lead presenter. They have recently been recognized for their research activities by the College of Natural Sciences and Mathematics at CSUF.

Mirna Dominguez

Mirna Dominguez is currently in her fourth year and graduating spring 2015 with a Bachelors in Mathematics. The most fascinating part about research for her was the one-to-one learning interaction. Under the supervision of Dr. Bourget statistical applications such data arrays were analyze through simulations; Gene-gene interactions are analyzed through comparison of methods. She is very excited to have her first publication and desires to do further research work.

Tiffany Duong

My name is Tiffany Duong and I am an undergraduate student getting my bachelors of science degree. My research is on water conservation in hopes that one day my research will be of use in the fight against California's current water crisis. In the future after I graduate from California State University, Fullerton, I plan on pursuing my career in health sciences by getting my medical degree.

Dylan Garcia

Dylan Garcia is an undergraduate student working towards a B.S. in Geology. He has been working with Dr. Matthew Kirby in the Paleoclimatology and Paleotsunami Laboratory as his lab assistant and thesis student. He has been learning geochemical and sedimentological lab techniques, as well as training other students. Aside from his paleotsunami research, he has also participated in the Environmental Science Research in Thailand program as a Louis Stokes Alliance for Minority Participation student. Here, he worked with Dr. Brady Rhodes in Pai, Northern Thailand on a landslide hazards research project. Dylan is also a volunteer mentor for America on Track, a non-profit organization dedicated to helping children of prisoners. He has jump-started an organization on campus where he recruits mentors for AOT and gives STEM lectures to the mentees and mentors – he is interested in making science accessible to others through science communication. He frequently attends conferences, gives research presentations, and is currently deciding on a graduate program. "Whenever we are curious about something, we inquire about it, and we look for an answer. What we don't realize is that we are conducting research – we are being scientists – we are exploring the new and undiscovered. Question Everything." djrg99@csu.fullerton.edu

Andrew Halsaver

Andrew is a first year graduate student in the Applied Math program at CSUF, and he graduated in the Fall of 2013 at CSUF with a Cum Laude degree in Applied Math. He began his research with Dr. Agnew in the summer of 2013 where he worked on using Mathematica with Clifford Projective Spaces. Andrew's work can be used to help Dr. Agnew's research advance regarding Clifford Projective Spaces. Also, Andrew's work may be useful to other areas of study such as theoretical physics. He hopes to continue his work with Dr. Agnew while studying for his Masters degree at CSUF.

Taylor Kennedy

Taylor Kennedy is a California native from Huntington Beach. His love for geology was discovered when working toward his Associate of Arts degree at Orange Coast College. Taylor is finishing his thesis and will graduate from California State University Fullerton with a Geol Sci degree this fall. Taylor also works for his teacher Professor Laton and pursues an active life performing martial arts and suffers through mudruns.

Estefania Larrosa

Estefania Larrosa is a sophomore at Santiago Canyon College. She started working at the Bridges Lab last summer as a part of the STEM^2 CSUF Summer Research Experience Internship. After that summer, Dr. Bridges was kind enough to let her stay for an additional semester and work on this kinetics project. Larrosa's major is Biochemistry, and after she finishes her bachelor's degree she plans to apply to medical school.

Calvin Lung

Calvin Lung is an alumnus of CSU, Fullerton and has his BS in biology, with a concentration in molecular biology and biotechnology. In May 2013, Calvin received the CSUF MSTI Scholarship for Future Teachers, a scholarship given to those pursing a career in education. Working under Dr. Joel K. Abraham, Calvin's research explored the effects of diet optimization has on worm cast quality and quantity. Eisenia fetida, a type of worm, can break down organic matter, such as food wastes, and turn them into a nutrient-rich soil amendment that plants can easily uptake. Calvin found that food sorting had no measurable impact on cast production or nutrient content. Calvin is currently taking a year off of school by working before pursing a PhD in biology.

Jesus Manuel Mejia

Jesus Manuel Mejia Ochoa is in his senior year as an undergraduate student majoring in Biochemistry. He started his research experience in fall 2013 by joining Dr. Michael Bridges' laboratory. Jesus is currently conducting research under the guidance of Dr. Bridges which focuses on the intrinsically disordered protein stathmin, its oligomerization, and measuring interspin distances using electron paramagnetic resonance spectroscopy. Through this research he was able to measure intra and inter-molecular distances in stathmin and further corroborate the existence of stathmin as an oligomer in solution through the use of double mutants and site directed spin labeling. After graduating, Jesus plans on working on a research team or obtaining a teaching position, before continuing his education in a graduate program.

Miriam Morua

My name is Miriam Morua and as an undergraduate at CSUF, I joined the urban agriculture community based research experience (U-ACRE) program with Dr. Abraham where I was able to develop my own independent research project with Dr. Jochen Schenk. My research focused in water management strategies where I compared the efficiency of different irrigation systems such as surface and subsurface irrigation for growing pepper plants in the Fullerton Arboretum. I was able to communicate my research in conferences such as the 2012 HACU student track conference in Washington D.C and the 2013 Latinos in Agriculture in El Paso, TX. Along with travel grants, I was also awarded a scholarship from the Rose Society of Saddleback Mountain Research Assistant Award in Plant Sciences in 2013. This year, I was accepted to the Watershed Management Internship (WMI) program that collaborates with the USDA and California State Universities to promote and expose students to career opportunities with government agencies while performing important research in conservation. Being accepted to the WMI program will help me reach my goal to obtain a master's degree in Biology because it will directly fund my research.

Garrett Mottle

Garrett Mottle is a fifth year senior graduating this summer pursuing a Bachelor of Science degree in Geological Science. He has spent the past couple years doing undergraduate research in the eastern Sierra Nevada mountains studying normal fault orientations near Lone Pine, CA. Garrett has attended field trips such as those with the Pacific Section SEPM (Society for Sedimentary Geology). He attends professional meetings such as those with the South Coast Geological Society (SCGS). As of late, he will be attending Geology Field Camp in Dillon, MT, and then look to pursue a career as a professional geologist and/or continue his studies in geology and obtain a Master of Science Degree in Geological Science.

Cristy Rice

Cristy Rice is hoping to graduate in spring 2015 with a B.S. in Biology and an emphasis in Marine Biology. She transferred here and also became a Southern California Ecosystems Research Program (SCERP) scholar in the Summer of 2012. Under the direction of Dr. Kristy Forsgren, she has worked on California pipefishes to establish a faster method of identification and determined species via genetic analysis. She has presented her work at numerous symposiums and conferences and won a research grant and best poster award from the Southern California Academy of Science annual meeting. She hopes to continue on into a graduate program after she graduates and is very interested in ecology and systematics among other topics.

Nathan Robertson

Nathan Robertson is a fourth-year undergraduate student graduating this spring with a Bachelor of Arts in Mathematics and a concentration in Statistics. Nathan has spent the past two years researching mathematics achievement with Dr. Behseta. Nathan hopes to continue his studies by pursuing graduate school in Statistics.

Kelly Shaw

Kelly Shaw is an undergraduate student set to graduate in December 2014 with a Bachelor of Science Degree in Geology. Kelly has done geologic research studying mass wasting events in Thailand through the Environmental Science Research in Thailand (ESRT) program at CSUF. She has been dedicated to the Geology Department serving first as the club treasurer and currently as the club president. She is also a member of Phi Beta Delta Honor Society on campus at CSUF. Kelly regularly attends professional meetings and lectures. In the past Kelly has served her community by being actively involved with the Boy Scouts of America and the Parent Teacher Association (PTA) in the Brea-Olinda School District where she held various office positions including PTA President at Brea Jr High. Kelly plans to apply to the CSUF graduate program in the spring of 2015 to work on a Master of Science Degree in Geology.

Jennifer Spencer

Working under the mentorship of Dr. Melanie Sacco, my research involves studying disease responses of the 14-3-3 gene family in *Nicotiana benthamiana* and Solanum lycopersicum. Through this research, we hope to substantiate a role for 14-3-3 genes in effector-triggered immunity in plants. I have conducted this research while participating as a Minority Access to Research Careers (MARC) scholar and Ronald E. McNair scholar. After my undergraduate degree, I will obtain a Ph.D. involving infectious disease biology, virology and immunology. I aspire to work for the Center for Disease Control and Prevention or the World Health Organization to research and help mitigate detrimental diseases.

Elizabeth White

Elizabeth White is a senior pursuing a Bachelor of Science degree in Geology. She is set to graduate this summer after attending field camp and investigating the regional geology of Montana at the University of Montana-Western. She also will be traveling to Thailand for the Earth Science in Thailand (ESIT) class offered as a joint program between CSUF and Chiang Mai University. This fall, Elizabeth will begin her master's degree with a focus in geochemistry. When she is not immersed in her coursework, Elizabeth can be found in the University Learning Center (ULC) where she works as a geology tutor.

Editors

Neha Ansari

Neha Ansari is a third year biochemistry major minoring in print journalism. Ansari currently serves as the Chemistry and Biochemistry Section editor for Dimensions. At Cal State Fullerton, Ansari conducts research in Dr. Peter de Lijser's organic chemistry lab where she studies the use of a carbonyl group as a nucleophile in photosensitized electron transfer reactions of oxime ethers. In the summer of 2013, Ansari was selected as an Amgen Scholar at the University of Washington, where she investigated the development of micelles in self-immolative polymers for the purpose of drug delivery in Dr. AJ Boydston's organic chemistry lab. Ansari is a member of the ASI Board of Directors representing the College of Natural Sciences and Mathematics, the University Honors Program, and is a President's Scholar.

Chris Baker

Chris Baker (Editor-in-Chief) is a graduating Geology student. He is currently working with Dr. Phil Armstrong to determine the rates of exhumation of the southern Alaskan Mountains. Chris plans to further his education and obtain a graduate degree with emphasis in hydrogeology. Currently Chris works as a geologist for The Source Group, is involved with multiple professional societies, and serves as an officer in the South Coast Geological Society.

Kali Prasun Chowdhury

Kali Prasun Chowdhury (Duke) graduated with high honors from University of California, Riverside in two and a half years with a degree in Business Economics. Since then, he has worked in consulting for five years before embarking on his graduate career. His endeavors have led him to complete a Master of Arts in Economics, from University of California, Santa Barbara and he is in his last semester of finishing up his Master of Science in Statistics, here at California State University, Fullerton. With a curiosity and a passion for research on some of the most fundamental questions of the Standard Social Science Model, Duke's ultimate goal is to attain his doctorate in Statistics. He wishes to apply it to the current models of human decision making, to improve predictive accuracy and inferential ability, over and above what is currently observed in such settings.

Kim Conway

I am a graduating senior double-majoring in Biology and Health Science. I am currently working in Dr. Paul Stapp's Vertebrate Ecology and Conservation Biology laboratory studying the prevalence of bot fly infestation in thirteen-lined ground squirrels in northern Colorado, while also attempting to identify the species of bot fly through molecular genetics techniques. I am a BURST (Biology Undergraduate Research Scholar Training) scholar at CSUF, and have been working as a veterinary assistant for 4 ½ years. I will be attending the UC Davis School of Veterinary Medicine next fall to obtain a DVM degree with a focus on wildlife medicine.

D'Lisa O'hara Creager

D'lisa O'hara Creager is an undergraduate student that will be finishing up this summer with a Bachelor of Science degree in Geology. She has been working with Dr. Brady Rhodes on her senior thesis project studying paleotsunami deposits in Seal Beach, CA. Dr. Rhodes also advised her in completing a hydrogeochemical study done in Chiang Mai, Thailand during her participation in the Environmental Science Research in Thailand (ESRT) program in the summer of 2013. O'hara has contributed to the California State University, Fullerton Geology club for the past year and served as its historian. She enjoys attending conferences and has given a number of poster presentations on her research. In her spare time she enjoys playing volleyball, which she spent three years coaching at Sunny Hills High School in Fullerton, CA. O'hara plans on earning her Master of Science degree in Geology and is currently deciding on a graduate program.

Niv Ginat

Niv Ginat is a graduating graphic design student from Los Angeles, California. Motivated by his love for design and hunger for knowledge, Niv draws his experience from a wide range of projects that he has worked on since embarking on his creative journey. In 2013, he was awarded the Jerry Samuelson Scholarship for his commitment and artistic potential as a graphic designer. He hopes to one day be able to carry his work with him as he travels around the world.

Carose Le

Carose Le is a graphic designer from Seattle, Washington. She is expected to graduate Fall 2014 with a Bachelor of Fine Arts degree in Graphic Design. She serves as a graphic designer for the Spring 2014 issue of CSUF's literary magazine of the College of Communications, TUSK. Carose has previously worked as a designer for City of Fullerton Parks & Recreation and CSUF Mihaylo College of Business & Economics. On behalf of College of the Arts and the Jerry Samuelson Scholarship Foundation, Carose was selected as a Jerry Samuelson Scholar. The scholarship recipient is selected by faculty and awarded to a visual arts student who has demonstrated talent and artistic potential. She enjoys using bright and vibrant colors when compiling her use of typography, photography and graphics.

Phillipe Diego Rodriguez

Phillipe is currently a second year physics major at CSUF and part both the Honors Program and the Dan Black Physics Business Program. With an interest in both physics and business, Phillipe plans on a career in technology based new venture start ups. Phillipe is a member of many physics, business, and honors organizations on campus. In addition, he volunteers with a youth soccer organization as a coach, referee, and mentor. Phillipe is currently working with Dr. Jocelyn Read on crust interactions of neutron star binaries during inspirals and mergers as part of his Senior Honors Project. Phillipe plans to attain his Masters in Business Administration while giving back to the organizations that helped him reach his goals.