Journal of Undergraduate Research

University of Rochester

Volume Six
Issue Two
Spring 2008

The *Journal of Undergraduate Research* (*jur*) is dedicated to providing the student body with intellectual perspectives from various academic disciplines. *jur* serves as a forum for the presentation of original research, thereby encouraging the pursuit of significant scholarly endeavors.

Acknowledgements

Faculty Advisor
Priscilla Anchinclose, Ph.D.

The College
Richard Feldman, Dean

Professorial Review Board
Amit Basu, Ph.D.
Udo Fehn, Ph.D
John Jaenike, Ph.D.
Paul Pearson, Ph.D.
Robert Poteda, Ph.D.
Dawn Pregraves, Ph.D.
Jacob J. Schlesinger, M.D.

From the Editors

Our journal exists to encourage University of Rochester undergraduates to participate in research and to exhibit some of the greatest contributions Rochester students make to research in their respective fields. But what exactly is research, and why do we feel so strongly that college students should be engaged in it? Of course, many undergraduates at the University of Rochester work in laboratories with professors of biology, chemistry, physics, engineering, and other sciences, and often this work involves serious research elements. But research comes in many forms. On one hand, there are biologists who observe life in a Petri dish and chemists who examine different reactions. On the other hand, there are geologists who perform field work in a remote mountain range and historians who evaluate old, forgotten letters and newspaper clippings in a library. The Journal of Undergraduate Research has featured articles from all of the natural sciences, and also from many disciplines of the humanities and the social sciences, showcasing the diverse and broad range of research that Rochester undergraduates complete.

And why do we feel it is so important for undergraduates to involve themselves in research? There are two compelling reasons for this. First, any original research adds at least a small piece of information to the global pool of scientific knowledge. Additionally, it is an invaluable experience for those who intend to go into a research-oriented career. This is reason enough for many students to participate in research, but doing so has other benefits as well; engaging in research is a remarkable opportunity for a student to experience both academic and personal growth. There are few better ways to nourish one's ability to reason than through a sustained research effort. Research as part of a group can provide excellent exposure to "real world"–style teamwork. And unlike the usual coursework, a student engaged in original research is seeking out knowledge that is entirely new; such an experience is a chance to explore one's creativity, to push one's intellectual boundaries, and to begin to truly understand oneself.

Sincerely,

Andrew Niles, 2008 — Harshika Satyarthi, 2009

Editors-in-Chief

Perspectives on Research

jur interviews Thomas Hahn, Ph.D.

Thomas Hahn is a Professor of English at the University of Rochester.

jur: Did you do any research as an undergraduate and did that help you in graduate school?

Hahn: Although I took some graduate courses while I was an undergraduate, I don't think that I did any research; I didn't work with a faculty member on a research project. This contrasts with what is often the case here at the University of Rochester.

jur: What about during graduate school? When did you really start doing research?

Hahn: As a first-year graduate student, I began with a "methods" course, which itself requires some research. I also wound up being a research assistant in several different capacities for different faculty members while I was a graduate student, so I participated in their research as I also began to do my own.

jur: You focus on medieval literature and pop culture; how did you first get interested in these fields?

Hahn: I had an education that oriented me toward the ancient world and the Middle Ages; I studied Latin and Greek when I was in high school, and I think that carried over to some degree in terms of what I became entranced with in college. When we think about the Middle Ages, we have this tiny, elite segment of the population, maybe one or two percent of people, who were literate and who produced books, and who sponsored and read the books. But in fact, vast numbers of people heard all kinds of stories, so my interest in popular culture is, to some degree, an interest in the origins of modern popular culture in the Middle Ages in stories like that of Robin Hood, for example, as well as the ways in which those medieval stories have continued to be popular in the centuries since the Middle Ages.

jur: In what ways has the influence of medieval literature manifested itself into popular culture? You mentioned Robin Hood...

Hahn: My main interest is in Robin Hood, but the Arthurian materials, which are pretty much everywhere in our culture, would be another instance of this. But any set of stories about knighthood, chivalry, and so on would certainly draw upon medieval origins in some distanced way. So for example, our sense of what it means to be a lady—and of things like chivalry and so on—is traceable back to the Middle Ages, even though they are somewhat mutated in modern times.

jur: Now in the humanities, or more specifically in the UR English Department, what sort of research is being produced by undergraduates?

Hahn: I'm actually teaching a research seminar this fall. It's a course that will actually have a lab—a humanities lab—attached to it. It's a course about Robin Hood, and what it will ask students to do is a variety of research tasks that involve looking at not only the earliest surviving texts about Robin Hood, but also the audiences and the publishers who made these things popular and who consumed them in the first place. In addition, students will actually work on post-medieval, more recent material, including popular books on Robin Hood in the 19th century—children's books, novels, etc. And then, they will work on children's books and other kinds of popular formats for Robin Hood in the 20th century, including, of course, film. All students in the class will actually take on a series of research projects in which they will work with some authors' published books on Robin Hood, most of whom are not very well known; it requires a certain amount of archival work just to dig out who this person was, what else he or she wrote, who the illustrators were for this material, whether they did other work, etc. Eventually, all this material will actually be contributed to a digital archive on Robin Hood materials that we will be producing. So this will essentially take the material that individual undergraduate researchers have found and deliver it to a much wider public once the Robin Hood digital archive is actually online.

jur: Do you have any advice for anyone who is pursuing research?

Hahn: It seems to me that one of the things that is valuable

about a notion of a humanities lab or humanities research is that it confirms for undergraduates the pragmatic value of the skills that they learn in the classroom. I think that we tend to perhaps regard things that are learned in the classroom as appropriate only to the classroom, when in fact they have real-world value. That real-world value comes out very clearly with research and writing and editing and working with digitization. When all of those things come together, it seems to me that what undergraduates gain is not just some nugget of knowledge—as in particular for some research project of which they only have control because they are the ones who have done the work—but also a validation of the skills that they have developed over their course of time as undergraduates. And of course, these projects draw on skills that are inert and static, and they help change and hone those skills as one continues to do research.

jsr: What is it about the English Department here at the University of Rochester that makes it unique?

Hahn: We are one of the smallest English Departments in the country that offers a Ph.D. program. This means that we are very much in contact with undergraduate students on a daily basis in both traditional classroom situations and in terms of research. As a Ph.D.–granting research institution, we as a faculty are actively engaged in research. Additionally, as a department that mirrors what a small liberal arts college department might look like, we have this connection to the undergraduates, and I think that there is an inevitable carry-over of the research interest of a faculty like this to the undergraduates, both inside the classroom and through individual projects.

Innate Color Preference and Flower Switching in Heliconiinae Butterflies

Sarah Green, 2009

Advised by Allen Masters, Ph.D.

Council on International Educational Exchange

Heliconiinae butterflies have numerous opportunities to alter their foraging preferences during their unusually long nine-month lifespan. They feed on nectar from a wide range of flowering plants, such as the orange and yellow flowers of *Lantana camara* (Verbenaceae; Fig. 1A). They use their keen sense of vision to learn by color association. Past research has revealed that orange flowers are innately preferred to yellow in *Agraulis vanillae* (Heliconiinae) butterflies.[1] The ability to distinguish between different colors has evolved in these invertebrates and can also be seen in experiments in which innate color preferences subside when *Heliconius charitonius* (Heliconiinae) butterflies were presented with a reward for visiting less preferable colors.[2] Members of the *Heliconius* genus have also been shown to learn the locations of landmarks, host plants, and nectar sites.[3] Butterflies in other families have been shown to be capable of associative learning as well. In a study by Takeuchi et. al. (2006), nectar was placed on disks of different colors, and when given a choice, Papilionidae butterflies would preferentially fly to those same colors days later.[11] Butterflies have the ability to change both innate and learned behaviors based on flower color[1,4]; however, little is known about the comparative ability of different species to change innate behavior based on their surroundings.

The ecological goal of every species is to maximize fitness. Butterflies are faced with a dilemma in the case of nectar foraging: the degree to which they specialize on certain flowers has a direct effect on their fitness. Specialization and generalization have both positive and negative consequences. Specialization allows individuals to seek out certain food sources more easily and to feed more efficiently, but in doing so, they may be missing out on more beneficial food sources. In contrast, generalization lets individuals look for the most favorable flowers in a given environment, while at the same time increasing the likelihood that they will encounter harmful substances. Generalists are also able to find food more often than specialists, who encounter their food sources at a lower rate.

Thus, the willingness of a species to change its innate foraging pattern reflects evolutionary responses to maximize fitness. The genetically-based biases of butterflies can change based on experience. The innate color preferences shown by Masters (1990) and Swihart and Swihart (1970) can be overridden if a greater reward is offered by other flowers or if a toxic substance is added to the preferred flowers.[10, 4] This foraging flexibility is important because plants produce flowers at different times throughout the year. Therefore, if a butterfly is able to forage more generally, it can find flowering plants at all times. On the other hand, specialization decreases the threat of toxic plants and decreases competition because it creates small, non-overlapping niches.

This study investigated learning based on early exposure to different flowers in young adult heliconiinae butterflies of the species *Heliconius erato*, *Heliconius melpomene*, and *Heliconius sara*. First, butterflies were exposed to two different species of flowers: *Lantana camara*, which is orange and yellow (Figure 1A), and *Lantana groseverata*, which is white (Figure 1B). Based on past studies, the yellow flowers were expected to be preferred.[1, 10] Secondly, white flowers were introduced into a garden that previously contained only yellow and orange. It was anticipated that all three species would visit this addition with less frequency because of their specialization on the more familiar, innately preferred yellow flowers. Yellow and orange flowers were considered together because they were both present initially, although butterflies generally feed from the yellow flowers upon first landing (there was nectar added to both colors daily). If the preferences of the butterflies changed to a great degree, it would provide evidence that they are less specialized than anticipated.

Approach

Research was conducted at the Monteverde Butterfly Garden at 1340 meters in the premontane wet forest life zone in Monteverde, Costa Rica. Both gardens were enclosed greenhouses containing plants and other butterflies not being studied, and both contained orange and yellow flowers for the whole study. The experimental group was in Garden 1, which contained white flowers only for the second portion of the experiment. Garden 2 had white flowers in both phases of the experiment and was used as the control. Orange and yellow *L. camara* flowers growing outside the Monteverde Butterfly Garden were used as nectar sources. White *L. groseverata* plants were obtained from the premontane wet forest life zone of the San Luis EcoLodge and Research Station, at approximately 1105 meters.

Both species of the flowers were placed in cut plastic bottles hanging from the ceiling of the gardens. A sucrose solution was added daily to all the flowers in the study so that they all offered the same reward. When both were present, the inflorescences of the two *Lantana* species occupied the same

Figure 1: *Flowers of Lantana camara (left) are orange and yellow; orange flowers are older and generally nectar-free while yellow flowers are less than 24 hours old and are nectar-producing. In contrast, flowers of L. grossevenata (right) are entirely white.*

hanging bottles, so they were side by side and intermixed. Other flowers in Garden 1 included *Aristolochia grandiflora* (Aristolochiaceae), *Carludovica rotundifolia* (Cyclanthaceae), *Calathea crotalifera* (Marantaceae), *Stachytarpheta frantzii* (Verbenaceae), *Rondeletia amoena* (Rubiaceae), *Gurania* spp. (Cucurbitaceae), *Passiflora* spp. (Passifloraceae), and *Impatiens wallerana* (Balsaminaceae). Garden 2 contained *Asclepias curassavica* (Asclepiadaceae), *Stachytarpheta frantzii* (Verbenaceae), *Impatiens wallerana* (Balsaminaceae), *Rondeletia amoena* (Rubiaceae), *Gurania* spp. (Cucurbitaceae), *Passiflora* spp. (Passifloraceae), and *Heliconia latispatha* (Heliconiaceae). The second garden was smaller and contained fewer flowering plants at the time of the study, but both contained equal amounts of *Lantana* flowers.

Butterflies were obtained as pupae and released in the gardens as adults in equal numbers of individuals from each species. Observations started four days after release began, and butterflies continued to be released throughout the observation period. During the first phase, orange *L. camara* flowers were placed in bottles in both gardens, and white *L. grossevenata* flowers were only placed in Garden 2.

Observations took place in half-hour increments in each garden. The species and color of flower visited were recorded every time a butterfly landed on a *Lantana* flower and extended its proboscis. The duration of a feeding bout was considered to begin when an individual landed on a *Lantana* flower, to continue as the individual fed on this species, and to end when it visited a different species or returned to a leaf to perch. Six days after observations began, white flowers were also placed in Garden 1. The butterflies were examined for six subsequent days. Throughout the study, flowers of both colors were replaced when they were no longer living (about every five days). Any butterflies released after the start of the second phase were identified with a paint pen marking on the left wing, and their visits were not included in the results because they had no experience prior to the addition of white flowers in the garden.

Results

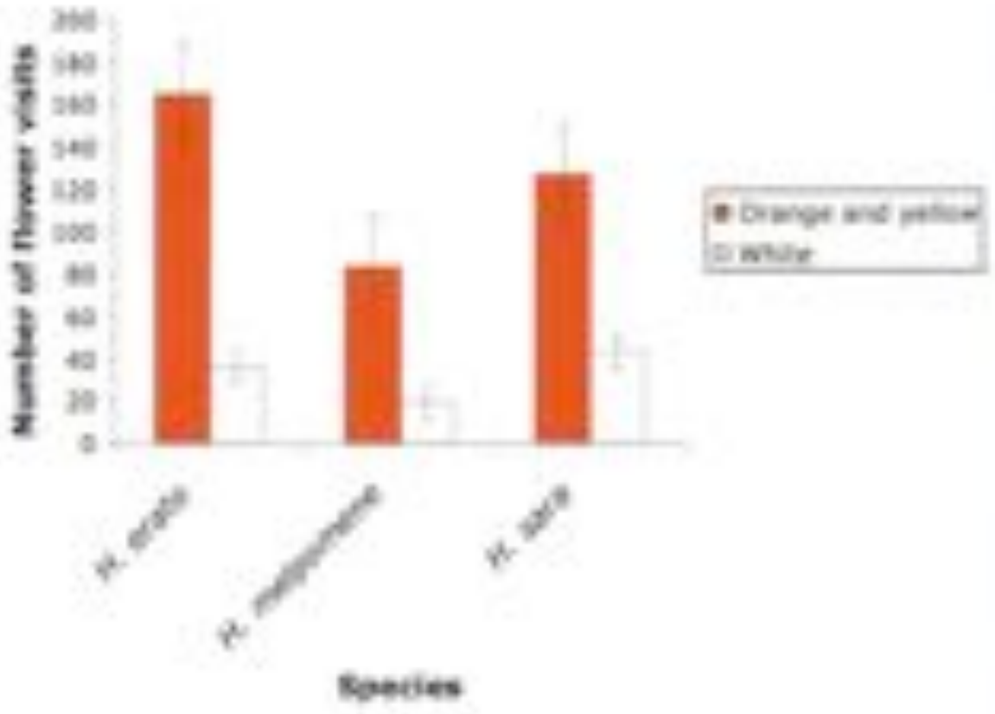
Innate Color Preferences

The data from Garden 2 (control) confirmed that all three species have similar color preferences when presented with the two species of *Lantana* (Figure 2). *H. sara* (74.7% ± 0.05), *H. melpomene* (81.6% ± 0.04), and *H. erato* (82.1% ± 0.06), all visited the orange and yellow flowers in the control garden with a higher frequency than in the experimental garden. While there was some individual variation, members of each species mostly visited the orange and yellow flowers with a higher frequency.

Sequential Specialization on One Color

Individuals tended to forage from a single color of flower. To test the significance of this observation, bouts consisting of more than four flowers were analyzed, and color specialization was defined to occur if more than 70% of the flowers visited

Figure 2: *Frequency of flower visitation in Garden 2 (control) by the species H. erato (N = 31), H. melpomene (N = 23) and H. sara (N = 28) to Lantana spp. Orange/yellow and white flowers were present since the time of release and constantly refilled with artificial nectar. Bars represent number of visits ± standard error.*



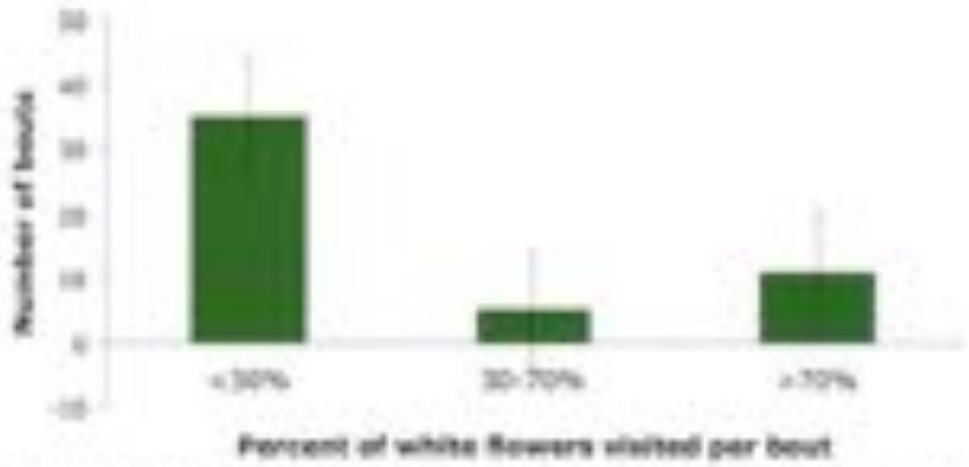


Figure 3: *Percent of white flowers visited for bouts in which butterflies fed on four or more flowers in both gardens (N = 51). Bouts comprised mostly of orange and yellow flowers occurred most often. However, white flower specialization was also present. When only those bouts containing more than 30% white flowers were considered, white flower specialization took place more often than with generalized bouts. Data represents number of bouts ± standard error.*

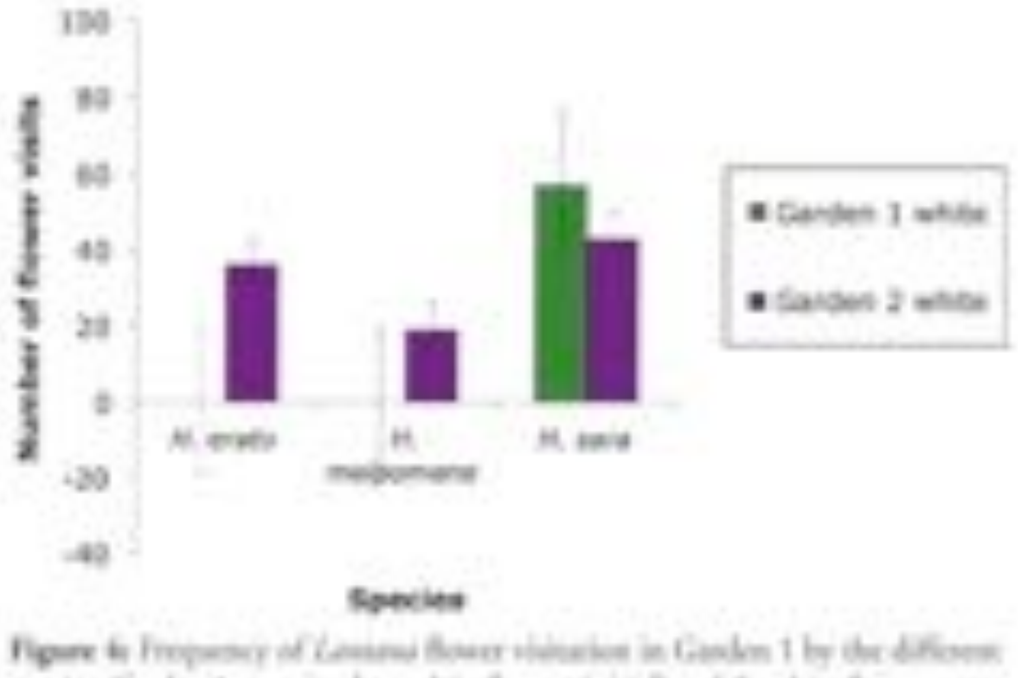


Figure 4: *Frequency of Lantana flower visitation in Garden 1 by the different species. Garden 1 contained no white flowers initially, while white flowers were present in Garden 2 throughout the entire study. Data represents number of visits ± standard error.*

in a single bout were of one color. Of the 51 bouts with at least four flowers, 35 specialized in orange and yellow, 11 specialized in white, and only five were unspecialized (Figure 3). Orange and yellow specialization occurred in a significant number of cases (χ^2 = 38.20, *df* = 2, P < 0.05), which is consistent with known preference for these flowers. When the orange and yellow specialization was discounted, butterflies were more likely to specialize in white flowers than a 30-70% mixture of the two strains (χ^2 = 11.44, df = 1, P < 0.05). The difference between the visits with a greater percentage of white flowers and the visits with mixed bouts suggests that these butterflies foraging for white flowers were doing so intentionally and not just landing on them by chance or because they were conveniently located. Due to the large population of butterflies present, it was assumed that each visit was from a different individual, and that specialization occurred throughout the population studied and was not the result of a few specializing individuals.

Effects of Early Experience

Despite the tendency to favor yellow and orange flowers, most of the species were affected by early exposure to the different flower strains (Figure 4). In Garden 1, no white flowers were available early in the lives of the butterflies, whereas Garden 2 contained white flowers throughout the lives of the individuals. While *H. erato* showed no change in preference regardless of early experience (χ^2 = 0.65, df = 1, P = 0.42), the data for this species is inadequate due to only three individuals observed foraging on *Lantana* plants after the white flowers were added. These butterflies were often observed foraging on other flowers instead. *H. melpomene* butterflies in Garden 1 favored orange flowers more after the white flowers were introduced (χ^2 = 5.02, df = 1, P < 0.05). In contrast, *H. sara* butterflies still visited orange flowers more often, but the frequency of their visits to white flowers increased in comparison to the control (χ^2 = 15.18, df = 1, P < 0.05). For this species in Garden 2, 39.0% of the flowers visited were white, but only 25.3% of the flowers visited were white in Garden 1.

Discussion

All three of the species studied preferred orange and yellow flowers over white (Figure 2). This reflects an innate preference for more colorful flowers, which supports the findings of Masters (1990) and Swihart and Swihart (1970). Butterflies with these preferences tend to find flowers with more nectar,[10] thus increasing their fitness. The color preferences that have evolved in these butterflies have allowed them to be somewhat specific foragers, focusing more energy into finding certain reliable flowers based on their color.

The analysis of the longer bouts revealed further specificity in foraging patterns. Most butterflies specifically visited orange and yellow flowers, while others sought those that were white (Figure 3). The white flowers would have been especially appealing to the butterflies foraging on them because there is less competition, as fewer butterflies take their nectar. After learning about this advantage of white flowers, individuals may have visited the white flowers even more frequently. In contrast, others who had not previously fed from white flowers were less likely to know of their advantages and seek them out. Therefore, bouts tended to be predominated by one color.

However, a degree of generality was also shown in many cases. *H. erato*, which has been categorized as a generalist species,[3] did not exhibit a change in color visitation frequency when the white flowers were introduced into Garden 1 (Figure 4). They may have a higher frequency of visitation to other plants because these butterflies have red stripes very similar in color to flowers such as *R. amoena*. Sexual selection may have caused coevolution of these butterflies to the similarly-colored flowers, making the red color of both attractive, whereas this did not occur with the *Lantana* flowers. It is possible that these butterflies are slightly general in their tendencies and began their adult lives by foraging on the orange and yellow flowers. By the time the white flowers were introduced, they had already finished their explorations and had found the other flowers in the garden more preferable.

The greater specificity in *H. melpomene* foraging explains why they visited orange and yellow flowers more often if this was the only strain they were exposed to initially (Figure 4). The timing of the addition of the white flowers coincided with the time that many of the *H. melpomene* butterflies began exploring the advantages of the *L. camara* flowers. Though all of the species are generalists to some degree, the extent of their generality differs. These butterflies perhaps waited a short time before exploring *L. camara*, and instead fed from other appealing flowers first, such as the red flowers, for the

same reasons they were visited by the *H. erato* butterflies. On the other hand, *H. erato* explored the orange and yellow flowers earlier. These two species may have evolved different approaches to finding food sources in order to avoid competition.

H. sara visited white flowers with a higher frequency if they had not been exposed to them initially (Figure 4). This generalist type of behavior is beneficial because it allows organisms to explore their environment in order to maximize resources. For heliconiinae butterflies, which have long lifespans, it is important to be able to forage for many months.[4] Therefore, they must be able to find plants flowering throughout the year, and it is necessary to drink nectar from plants that are not flowering when they emerge. Additionally, there is evidence that heliconiinaes maintain energy stores for long periods of time because they produce small numbers of eggs throughout their lifetime, instead of producing many when they first emerge.[1] Therefore, they are in need of a constant food supply, and a new resource may be appealing. If the plant is just beginning to flower, it will probably be available for a longer amount of time than a flower that has already been in bloom for an unknown amount of time.

If *H. sara* is better adapted to feed from *Lantana* flowers than from the other flowers in Garden 1, then niche partitioning may have been taking place more in this garden, whereas the different species may have been equal competitors in Garden 2 due to similar abilities to feed from all of the flowers. The generality and degree to which the different species are willing to forage on different kinds of flowers has evolved to be different so that they are not competing for the same flowers at the same times. It has been observed that *Asteraceae* spp. flowers are dominated by Heliconiinae butterflies in the field, which prevents other species from visiting the flowers (there is a high attraction of the butterflies to these flowers because of their orange color). However, under controlled conditions, this does not occur and other species are able to visit the flowers more. It was speculated that the low density of the flowers in nature allowed the *Heliconius* spp. butterflies to efficiently trapline, so not enough nectar was available for other butterflies.[5] It may be that the superior foraging by the *H. sara* butterflies in this study prevented the other species from visiting these flowers in Garden 1. In Garden 2, perhaps it was more difficult for the *H. sara* to trapline due to factors such as the flowers being concentrated in a smaller area, making traplining unnecessary. Therefore, the other two species could access the *Lantana* flowers more easily.

The differences in the reactions to early experiences are beneficial for these species because they have similar distributions. They are all foraging for nectar from similar plants in the same places, so fitness is increased if they are not foraging together (as this could cause a depletion of resources). Therefore, they are able to react to the same conditions in different ways, allowing them to avoid competition. The results from Garden 2 showed that all three species were willing to try to feed from white flowers despite their innate preference for orange and yellow. This small degree of generality is seen also in the temporal differences among the species in their desire to forage on the *Lantana* plants in Garden 1. *H. sara* visited the white flowers with higher frequency than the other butterflies, so this species exhibited the most generality, allowing individuals to reap the benefits of these flowers with less competition. It remains unknown whether or not this generality extends to other unfamiliar flowers. Future studies could explore the effects of different flowering species on foraging behavior or how foraging is affected by time of exposure prior to the introduction of novel flowers.

References

1. Boggs, C. L. 1997. Reproductive allocation from reserves and income in butterfly species with differing adult diets. Ecology 78: 181-191.
2. Brower, A. V. Z. 1996. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: a phylogenetic hypothesis from mitochondrial DNA sequences. Evolution 50: 195-221.
3. Cardona, M. Z. 2001. Patterns of pollen collection and flower visitation by *Heliconius* butterflies in southeastern Mexico. Journal of Tropical Ecology 17: 765-768.
4. DeVries, P. J. 1987. The Butterflies of Costa Rica and Their Natural History. Princeton University Press, Princeton, NJ, pp. 186-187.
5. Gilbert, L. E. and M. C. Singer. 1975. Butterfly Ecology. Annual Review of Ecology and Systematics 6: 365-397.
6. Masters, A. R. 1990. Dual role of pyrrolizidine alkaloids in nectar. Journal of Chemical Ecology 17: 195-203.
7. Schemske, D. W. 1983. *Lantana camara* (Greco Negrito, Lantana). In: Costa Rican Natural History. D. H. Janzen, ed. The University of Chicago Press, Chicago, IL, pp. 266-268.
8. Scott, J. A. 1986. The Butterflies of North America. Stanford University Press, Stanford, CA, pp. 82.
9. Swihart, C.-A. 1971. Colour discrimination by the butterfly *Heliconius charitonius*. Linn. Anim. Behav. 19: 156-164.
10. Swihart, C. A., and S. L. Swihart. 1970. Colour selection and learned feeding preferences in the butterfly *Heliconius charitonius*. Linn. Anim. Behav. 18: 60-64.
11. Takeuchi, Y., K. Arikawa, and M. Kinoshita. 2006. Color discrimination at the spatial resolution limit in a swallowtail butterfly, *Papilio xuthus*. The Journal of Experimental Biology 209: 2875-2879.
12. Weiss, M. R. 1997. Innate colour preferences and flexible colour learning in the pipevine swallowtail. Animal Behaviour 53: 1043-1052.

DM Domain Proteins and their Role in the Generation of Sexual Dimorphism in C. *elegans*

Victoria DiMarco, 2008

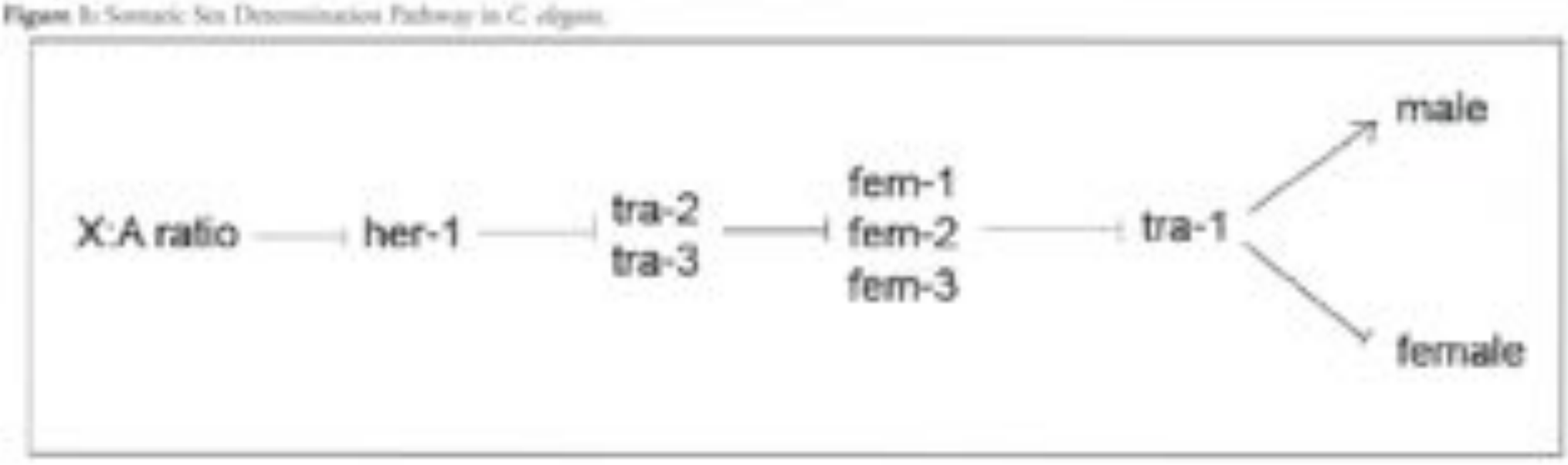
Advised by Douglas Portman, Ph.D.

Department of Aging and Developmental Biology, UR Medical Center

Differences between the sexes are not unique to humans and related mammals. Sexual dimorphism is present to some degree in most higher eukaryotes, although the mechanisms that lead to the development of these dimorphisms are just beginning to be understood. The nematode *C.elegans* has been especially amenable to studies in sexual differentiation. The wild type population normally contains only hermaphrodites, with males occurring rarely (0.15% of the time) through nondisjunction of the X chromosomes during meiosis.[1] The nervous system of the nematode has been particularly useful for characterizing specific sex differences. The adult hermaphrodite nervous system contains 302 neurons; males have 381. The complete neural wiring of the adult hermaphrodite has been reconstructed via electron micrograph, and the majority of wiring pathways in male-specific neurons have been reconstructed as well.[2] The different neurons in the male are implemented in male-specific mating behaviors. The sensory rays in the male tail are innervated by a series of male-specific sensory neurons and correlated motor and interneurons. Anatomical features specific to the male such as the spicules and hook, which are essential for proper male mating behavior, are innervated by male-specific neural pathways as well. The hermaphrodite, on the other hand, utilizes egg-laying neurons which are lacking in the male.[10] The anatomy of the worm is such that phenotypic deviations are often easy to score, making the animal useful for genetic studies. This, as well as its easy maintenance on agar plates and short generational turnover, as well as a completely sequenced and mapped genome, make the worm ideal for these studies.

The first example of a switch gene, the name conveniently given to master regulatory genes that lead to a complete transformation of sexual phenotype, was found in *D. melanogaster* in the mid-1940s. The gene, named *transformer (tra)*, was identified by chance because it lead to a number of males that were wild type for sex-linked traits that should only have existed in the female of the species in a genetic study.[3] Mutagenesis was carried out in the 1970s in *C. elegans*, and ten transformed mutations were isolated. Mapping of the alleles identified three genes: *tra-1, tra-2,* and *tra-3*. These genes were shown to be regulated by the X:A ratio of the animal and to result in total transformation of the sexual phenotype. Specifically, mutations in these genes granted male sexual phenotype to XX hermaphrodites.[4] Both *tra-2* and *tra-3* were shown to lie upstream of both *fem* genes (the activity of which is required for correct male development) and of *tra-1*, whose function is necessary for correct hermaphrodite sexual development. The *tra-1* gene was sequenced and shown to contain a number of zinc finger domains, implicating the protein in gene regulation as a transcription factor.[4] The SRY gene was later identified in *H. sapiens* through special cases of sex reversal and has been identified as a switch gene required for proper male development.[5] Identification of these genes in both vertebrates and invertebrates demonstrates a common evolutionary theme of an upstream master regulatory gene that responds to a primary signal (i.e. the X:A ratio in *C. elegans* or *D. melanogaster*) and that is essential for sex-specific development.

Figure 1: Somatic Sex Determination Pathway in *C. elegans*



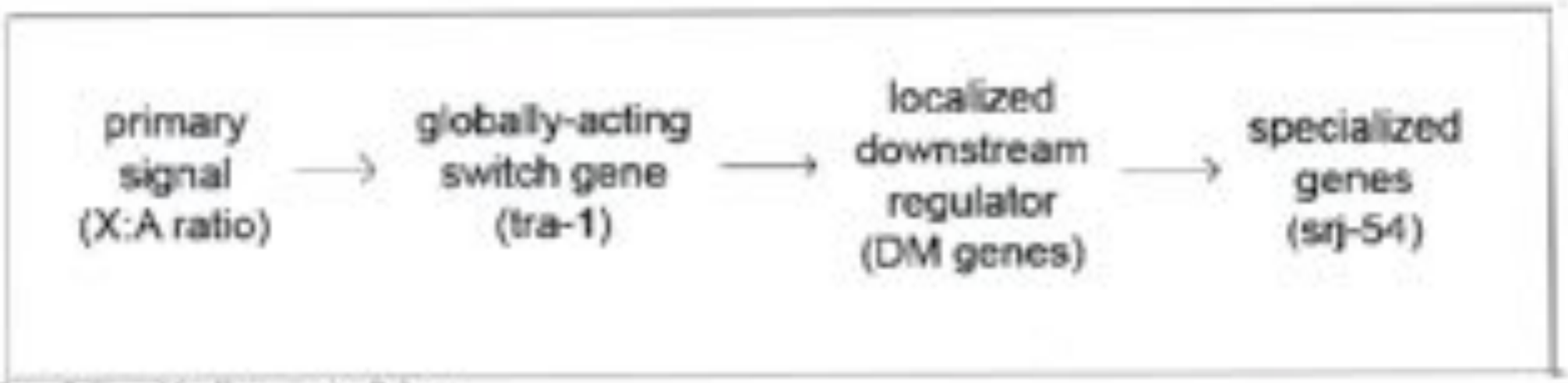


Figure 2: General Sex Determination Pathway

The above somatic sex determination pathway (Figure 1) is limited in that it does not describe anything between the global switch gene *tra-1* and the eventual sex phenotype. Downstream regulators in this model remain to be identified. While the earliest acting class of genes is the switch genes, which act globally, a second class of more localized sexual regulators most likely lay between the globally-acting regulators and those downstream genes that serve very specialized tissue-specific functions.[4] The first such gene to be identified was the *doublesex* (*dsx*) gene in *D. melanogaster*. This gene was found to control aspects of sexual development in the fly via two alternatively spliced isoforms. While the gene does not act so globally transform the sexual phenotype of the animal as would a switch gene, the DSX(F) and DSX(M) isoforms, which differ in their C-termini, direct aspects of sex-specific anatomical development.[5] While the female isoform appears to serve no function, the male isoform directs male-specific development. The gene was found to code for a DNA-binding motif known as the zinc finger. The downstream regulator *mab-3* had earlier been identified in *C. elegans* and shown to control localized sexual differentiation of the male tail.[6] Although the two genes had been identified separately, eventual sequence alignment of *dsx* and *mab-3* showed remarkable similarity in the DNA-binding motif, although DSX contained only one such domain and MAB-3 contained two. The two proteins were later shown to bind similar sites on the DNA.[4] This DNA-binding motif was then named the DM (Doublesex/MAB-3) domain.[5]

The DM domain was shown through site-directed mutagenesis to contain a series of cysteines and histidines essential for protein function and to have zinc-dependent DNA binding, making it likely that proteins containing this motif would be transcription factors. Studies of the side chains and basic regions showed that the motif binds to the DNA minor groove, unlike the more common zinc-dependent DNA binding domain that binds in the major groove of the helix.[4] Since both *dsx* and *mab-3* have sexually dimorphic functions, further search for genes encoding DM domains looked promising in identifying other genes implicated in sex differentiation. Searches in *C. elegans* turned up two additional DM domain genes: *mab-23* and *dmd-3*.

The first DM domain gene identified, *mab-3*, was found in a general screen for mutations that reduced the ability of males to mate. The *mab-3(e1240)* allele is most likely a null allele. This allele results in improper V-ray development in the male tail, and more specifically in the absence of six of the nine sensory rays. MAB-3::GFP reporters were expressed in the male tail, hindgut, intestine, and a head neuron.[4] Visual sexual dimorphism arises in the early to mid L3 stages, when cell divisions occur from the V5, V6, and T ray lineages, which form the precursor cells that will develop into the sensory rays. Mutation in *mab-3* results in improper differentiation of the V5 and V6 cells, eliminating rays derived from descendants of these cells.[7] Studies of *mab-3* with the gene *lin-32* were performed later, in order to further elucidate the sex determination pathway and identify interactions between regulators. Both *mab-3* and *lin-32* are necessary for sensory ray formation. The absence of *lin-32* results in a decreased number of rays in the male tail, with the number lost depending on the allele.[16] Genetic epistasis analysis and comparison of mutant phenotypes demonstrated that *mab-3* played a permissive role in V ray formation, while *lin-32* played a more integral role in determining whether rays developed at all.[7] Mutation in *mab-3* also resulted in the production of yolk proteins by the male, a normally hermaphrodite-specific event, as well as production of excess alae.[4] The MAB-3 protein has no known function in hermaphrodites, since *mab-3* hermaphrodites display no observed defects. The gene was shown to lie downstream of *tra-1*, placing it as a downstream regulator in the pathway of sexual development.[5]

The relationship between *mab-3* and *dsx* is characterized by their common DM domain. Both genes play a part in the differentiation of sex-specific neuroblasts in their respective organisms, as well as regulate the expression of yolk protein genes. To test whether the two proteins were interchangeable, the male isoform of DSX was introduced into *mab-3* mutants. Remarkably, the DSX(M) protein was capable of restoring V-ray formation to the *mab-3* mutant, implying that the DSX(M) and MAB-3 proteins function by similar means and can regulate similar genes in the nervous system.[5]

The *mab-23* gene was identified via a forward genetic screen for mutants defective in male nervous system patterning. The gene was found to encode for a DM domain, identifying it as a purported transcription factor. Mutation in *mab-23* results in improper male tail development, as does *mab-3*.[17] Each sensory ray in the male tail is derived from a single precursor cell. A MAB-23::GFP fusion protein was expressed in the ray lineages that give rise to rays 1 through 4, as well as 6. Through use of a *tax-2* reporter, which is expressed in those neurons that synthesize dopamine, it was found that mutation in *mab-23* resulted in improper synthesis of dopamine in several rays. It was also shown that the axons of several neurons projected incorrectly and failed to synapse with their targets. Defects in male turning behavior indicated defects in male-specific sex muscles, resulting in defective mating behavior. Defects in the development of the proctodeum result in ineffective sperm transfer, and therefore sterility. The MAB-23::GFP fusion protein was found to be expressed in both males and

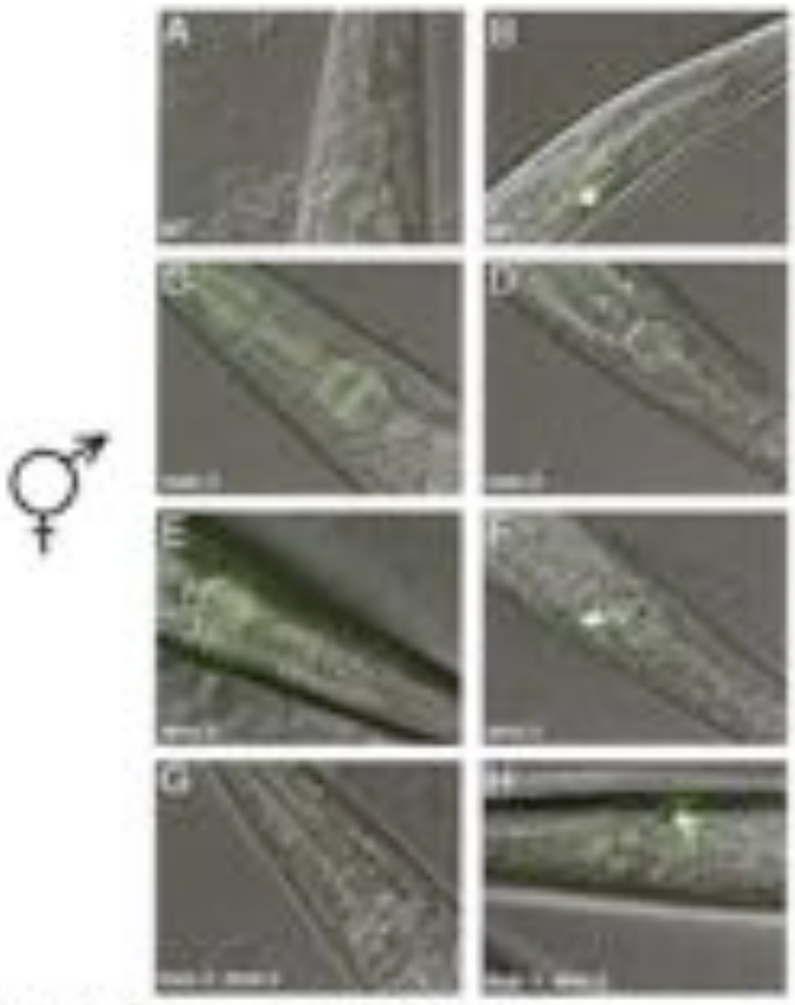


Figure 3: *fdf-3* expression in DM domain mutants.

hermaphrodites, although the mutation appears to have no deleterious role in the hermaphrodite, as demonstrated by the existence of a naturally-occurring population of animals homozygous for *mab-23* mutation.[10] There is some, although incomplete, overlap between those tissues affected by *mab-23* and those affected by *mab-3*, although studies show that they are independently regulated.[10]

A third DM domain transcription factor, *dmd-3*, was recently identified through a microarray screen for genes expressed in the male tail sensory rays. Mutation in this gene causes improper male tail retraction, resulting in maintenance of the tail tip and a male tail that resembles that of the hermaphrodite. While normally the tail tip cells fuse in the early L4 stage of the male, they fail to do so in the *dmd-3* mutant and persist in the tail spike. Reporters for the *dmd-3* gene are expressed in the rays, tail tip, hindgut, and gonad, which are all sexually dimorphic tissues. Expression in the hermaphrodite is limited to the anchor cell, showing predominantly sex-specific expression. Although dopamine synthesis is normally limited to the RnA neurons in rays 5, 7, and 9, *dmd-3* males express dopamine markers in all RnA neurons in all rays, a phenotype somewhat reminiscent of that of *mab-23*. Also similar to *mab-23* mutation, the axons of many RnA neurons fail to migrate correctly (D.A. Mason & D.S. Portman, pers. comm.). However, study on *dmd-3* is not as extensive as that on the other DM domain genes identified in *C. elegans* and further characterization of the defects remains to be performed.

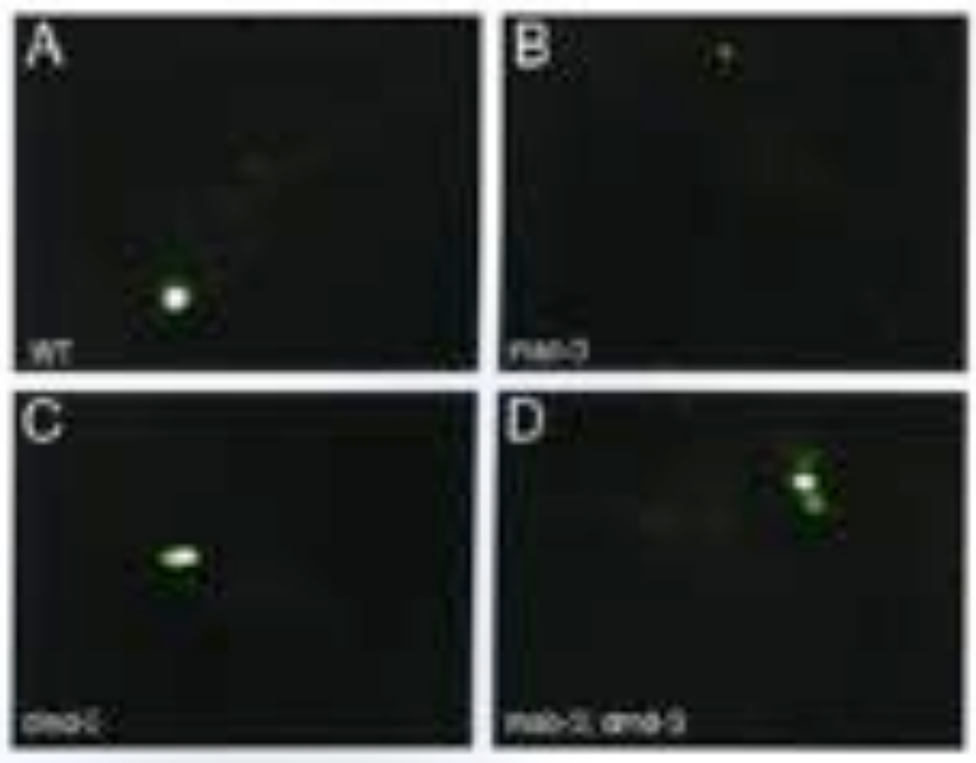
All three of the above DM domain genes in *C. elegans* are integral for proper male tail development, showing their importance in sexual differentiation and development. They each have overlapping roles, whether it be the involvement of *mab-3* and *mab-23* in male muscle and ray development, or the role of both *mab-23* and *dmd-3* in limiting dopamine synthesis to a specific subset of ray neurons and ensuring correct migration of neural axons. Interestingly, the *mab-3;dmd-3* double mutant male has a tail that appears almost identical to that of the hermaphrodite in that it is lacking all rays and maintains the pointed tail tip.

It is interesting to note that a search for homologous DM domains in the human genome have thus far turned up at least eight genes, among which are *dmrt1* and *dmrt2*. These genes are expressed in the differentiating gonad and are associated with human sex reversal, showing themselves to be integral in sexual development just as their invertebrate counterparts.[3] The genes were mapped to the distal end of chromosome 9. FISH performed on a group of XY female individuals confirmed that *dmrt1* and *dmrt2* were both hemizygous, having been lost in a deletion of the distal arm of chromosome 9. The fact that loss of these genes is required for sex reversal in humans implies that they play a role at least similar to that of other DM domain proteins in Nematoda and Arthropoda, serving an essential function in the sex determination pathway.[11]

A more global search for DM domain genes identified members in many other organisms, including the mouse, a variety of fish, and several other vertebrates. The range of DM domain genes makes it possible that this family of genes may represent the first example of conservation of sexual regulatory genes across phyla. For this reason alone, further analysis of the function of individual DM domain genes is necessary, and comparison of DM domain genes in diverse phyla may prove evolutionarily informative. It is at least likely that these DM domain genes may have played an ancient role in metazoan sexual development, some of which has been conserved in their current functions.[11]

The DM domain genes in *C. elegans* and *D. melanogaster* play the role of downstream regulators, controlling the development of more localized areas than those controlled by the more globally-acting *tra-1* and *dsx*, respectively. Lying downstream of these local regulators are any number of genes that play specified roles in development, and can be viewed as sitting at the distal end of the sexual development pathway (Figure 2). One example of such a gene is *srj-54*. The SRJ-54 protein is a seven-transmembrane serpentine receptor that is dimorphically expressed in the AIM interneuron of the head in the *C. elegans* male alone. The gene is one of a number of class J serpentine receptors, several of which share remarkable homology with *srj-54* and one of which has been examined below and found not to be sexually dimorphic in expression. Due to its sexually dimorphic expression pattern, the *srj-54*

Figure 4: GFP Images of *fdf-3* expression in DM domain mutants.



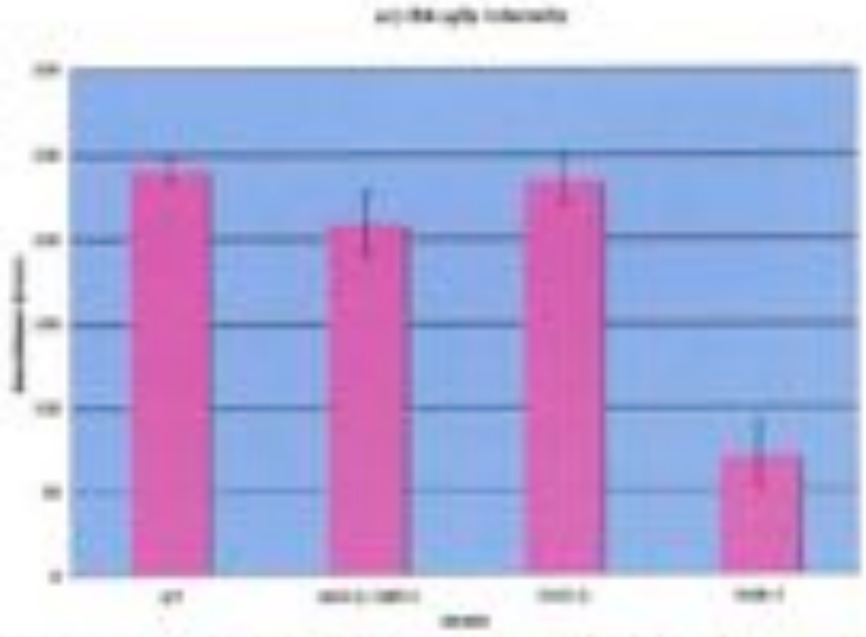


Figure 5: Quantification of GFP expression in DM domain mutants.

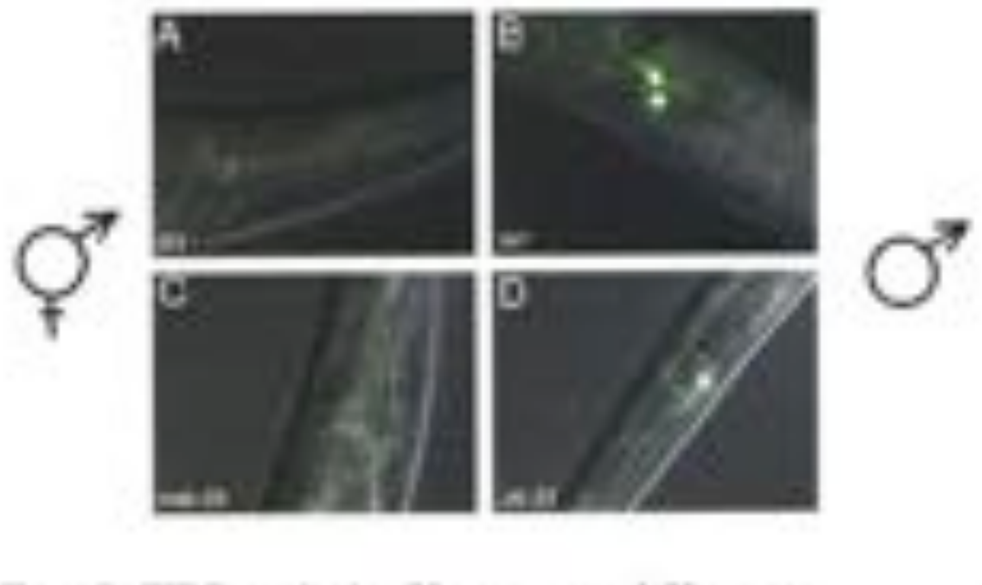


Figure 6: GFP Expression in wild type versus *mab-23* mutants.

gene is asserted to lie downstream of both *tra-1* and the DM domain proteins, although the role of the DM domain proteins in its expression has only recently been examined.

The work described below aims to identify upstream regulators of *srj-54*. A candidate gene approach was used, specifically looking at the DM domain genes *mab-3*, *mab-23*, and *dmd-3*. Expression of an integrated *srj-54* transgene fused to a GFP reporter was examined in single mutants of each DM domain gene, as well as a *mab-3;dmd-3* double mutant. Expression was found to be reduced in the *mab-3* mutant. Given this fact, a strain carrying the *mab-3* promoter driving expression of GFP, as well as *srj-54* driving mCherry expression, was examined in order to identify possible co-expression of *mab-3* and *srj-54* in the same neurons. The two genes were shown to have overlapping expression in a male head neuron. Knowing then that the genes were co-expressed, it was necessary to test the hypothesis that *mab-3* directly regulates *srj-54*. Possible *mab-3* binding sites were identified on the *srj-54* promoter, and deletions of these regions made via PCR. Given that *srj-54* is sex-specifically expressed, steps were taken to examine other *SRJ* family genes in a search for a similar expression pattern. A transgene carrying *srj-57* fused to a GFP reporter was not found to be dimorphic in expression. The assumption that *srj-54* lies downstream of *tra-1* was tested through construction of a *tra-1* XX pseudomale strain and examination of *srj-54* expression.

Materials & Methods

C. elegans strains and culture. Culture and genetic manipulation of *C. elegans* were performed as previously described.[14]

RNAi Feeding Assay. Using an anti-GFP RNAi construct obtained from another lab, *E. coli* cultures harboring the RNAi construct were spread on NGM plates. Six to eight adult hermaphrodites from the *Ex[mab-3p::gfp, srj-54::mCherry]* strain were placed on each of four RNAi feeding plates and left overnight. The parents were removed in the morning, and the eggs allowed to grow to adult. When the animals reached the young adult stage, they were examined for GFP and dsRed expression.

Quantification of fluorescent expression. For each of the strains carrying the *fdIs5* transgene, quantification of GFP expression levels was performed with the AxioVision program. This program was also used in conjunction with the Zeiss compound microscope to take the DIC and GFP images used in quantification, as well as the dsRed images used in the co-expression study. Using this program, a small rectangle of a given size 535.96 pixels squared was created. The rectangle was made of a small enough size such that it fit inside the cell body of the neuron in the wild type and was placed on the center of the cell body so as to obtain readings of optimal expression without interference from fading expression along the cell body edges. For each subsequent image, the rectangle was copied and pasted so as to maintain the same size. The rectangle was moved in each image and placed in the center of the cell body. In instances where it did not cover the cell body, efforts were taken to place the cell body in the center of the rectangle. In the vast majority of cases, the cell body was greater in size than the rectangle. Using the program, the mean density of GFP expression and the mean standard deviation of GFP expression were take for each of the twelve animals for the four strains examined. Data analysis was then performed with Microsoft Excel.

Results

In order to test whether the known DM domain genes in *C. elegans* had any regulatory effect on *srj-54*, strains were first constructed using an extrachromosomal array carrying a *srj-54::gfp* transgene. This work was performed the previous semester, and images were taken of animals. The data gathered using this extrachromosomal array implied that GFP expression was severely reduced in the *mab-3; dmd-5, him-5* double mutant, and slightly in the *mab-3; him-5* single mutant (data not shown). However, it was decided to reconstruct the strains using the newly-obtained integrated *srj-54::gfp* strain, called *fdIs5*. Using this integrated transgene, the following strains were constructed:

fdIs5; mab-3; him-5
fdIs5; dmd-3 him-5
fdIs5; mab-3; dmd-3 him-5

Images were then taken of twelve adult males from each strain that were representative of the general level of GFP expression seen in the strain population. Images were also taken of a hermaphrodite from each strain to demonstrate that GFP expression is absent. Images were taken of animals of the same age from each strain, grown under identical conditions. Figure 3 below shows GFP expression in the wild type hermaphrodite and male *fdIs5; him-5* strain, as well as in the two single mutant strains and the double mutant.

Figures 3A and 3B show the wild type level of expression,

which is clearly visible. Figures 3E and 3F show expression in the *fdf5; dmd-3 bnm-5* mutant, in which GFP expression is not visibly reduced. The same is true for the *fdf5; mab-3; dmd-3 bnm-5* double mutant, pictured in Figures 3G and 3H. Figure 3C and 3D show the *fdf5; mab-3; bnm-5* single mutant, in which GFP expression is severely reduced. Figure 4 pictures the GFP images of a characteristic male of each genotype.

It is clear that the *fdf5; mab-3; bnm-5* mutant (Figure 4B) displays severe reduction in GFP expression. It is interesting that this single mutant displays such drastic reduction in expression, while the *fdf5; mab-3; dmd-3, bnm-5* double mutant shows a level of expression indistinguishable from the wild type. Quantification of GFP expression was taken from twelve individuals of each genotype characteristic of the average level of expression in their respective strains (Figure 5).

Taking into account error, the expression levels of the wild type *fdf5; bnm-5* strain, *fdf5; dmd-3 bnm-5* and *fdf5; mab-3; dmd-3 bnm-5* mutants are not significantly different. However, it is abundantly clear that expression levels in the *fdf5; mab-3; bnm-5* mutant are drastically reduced, indicating that *mab-3* is required for normal levels of *srj-54* expression in males.

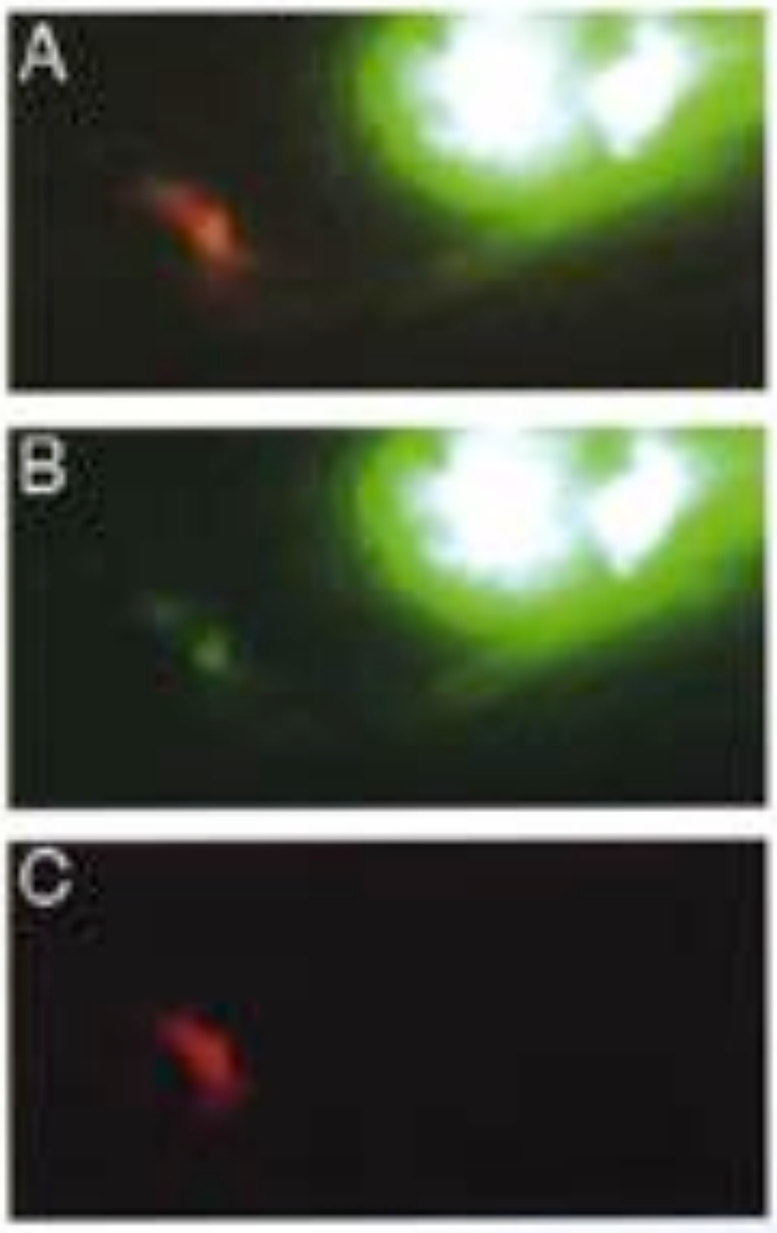
While the above strains were constructed and examined simultaneously, construction of an *fdf5; mab-23; bnm-5* strain proved more difficult. The first cross proved unsuccessful in finding recombinants and maintaining *mab-23* as homozygous in the strain. This may indicate that *fdf5* (and *fdf6*, as was later discovered), integrated inconveniently close to the *mab-23* locus, making recombination rare. The cross was then attempted using *fdf6*, a second integrated *srj-54::gfp* strain. This too proved unsuccessful. The first cross, using *fdf5*, was then performed again. However, fifty F2 progeny were picked, rather than the normal twenty to twenty-five. With this number of progeny, it was possible to isolate recombinants and maintain both *mab-23* and *fdf5* homozygous in the strain. This *fdf5; mab-23; bnm-5* strain was then examined in a fashion similar to that of the strains mentioned above. When GFP expression levels were examined in this mutant versus the wild type *fdf5; bnm-5* strain, it was found that there was no difference in expression (Figure 6).

Having seen that GFP expression levels were reduced in the *fdf5; mab-3; bnm-5* mutant, it was hypothesized that *mab-3* may regulate *srj-54*, whether directly or indirectly. If regulation is direct, the two genes should be co-expressed. Both were known to be expressed in a head neuron in the male, likely the AIM interneuron. A strain was obtained that possessed an extrachromosomal array carrying the *mab-3* promoter driving GFP expression, as well as the *srj-54* sequence driving mCherry expression. However when viewed, the GFP expression in the pharynx was so intense that it made it impossible to draw any conclusion about co-expression, since fringing from the pharynx obscured visualization of the nearby AIM neuron. It was decided that an anti-GFP RNAi feeding assay was likely to knock down GFP expression in the pharynx, where RNAi may easily reach, while leaving expression in the neuron relatively untouched. After a single generation, GFP expression had not been knocked down sufficiently to allow identification of co-expression. Adults from the generation that grew up on the RNAi feeding plates were placed on a second set of feeding plates in the same fashion as above, and allowed to lay eggs. The parents were removed, and the offspring examined once they had reached the young adult stage. This second generation had sufficient knock-down of GFP expression in the pharynx to allow visualization of the AIM neuron in many instances, although approximately half of the animals still had expression levels too high to allow analysis by eye. In several cases, it was abundantly clear that *mab-3* and *srj-54* are expressed in the same location. Figure 7A below shows an overlay of both GFP and dsRed expression in the same neuron. The cell body is prominent, with both reporters visibly overlapping. Figures 7B and 7C show individual GFP and dsRed images. The shape of the cell body is visible in both, and the unique shape of the cell common to both images. From these data, it is obvious that *mab-3* and *srj-54* are co-expressed in the AIM interneuron, suggesting that the regulatory relationship between *mab-3* and *srj-54* could be direct.

Knowing then that *mab-3* and *srj-54* are co-expressed, and that *srj-54* expression (in the form of *fdf5*) is severely reduced in the *fdf5; mab-3; bnm-5* mutant, it remained to be determined whether *mab-3* directly regulates *srj-54*. The sequence of the *srj-54* promoter was obtained (Figure 8), and regions that match the known *mab-3* consensus binding sequence TGTAACA were located. This search turned up four such sequences in the promoter (highlighted sequences in Figure 8).

None of these sequences matched the *mab-3* consensus binding sequence exactly, but they each shared a significant number of common base pairs with the consensus binding sequence. Should *mab-3* directly regulate *srj-54*, we would expect to see a reduction in *srj-54* expression if these consensus binding sequences were eliminated. As such, primers were designed so as to eliminate varying numbers of the suspected *mab-3* binding sequences in the *srj-54* promoter (underlined regions in Figure 8). The *fdf5* transgene, which has already

Figure 7: *srj-54* and *mab-3* reporter co-expression.



```
   1  taagcggtcc tcgattaatg tcggctactt ttttttaggt cagtgcacca
  51  ttctcctgc tctcagcccc ggtggcttat ttgttcacta caatttattt
 101  caacttctat caccaagtgg ccaacaattt ttgctcactt ttacttgcag
 151  ttcatggctc ttcatcgtca atattcctga tattaattca caaatcatat
 201  cggaaagtcg tatgggatct gctgaagttt gatatcttca cgagaaaaaa
 251  caatcaggtg tcaagttttg tcgtctcaag aaactcagtt gctatagatc
 301  gtcggaaaaa gatcgaaaat gattcgagct ggaaacgttt aatttctgta
 351  ttcaatcgaa ccattcgagt ttgatcttaa aacatcaaat tgcgaaaata
 401  taaatttcat aaattaaata aacaatttaa agcgaaaact gacactcgaa
 451  agtcagtctc acattctaca gtattcattc gaaacgttct tgtctagact
 501  gatttaactt ccggcccccc atgtttctgt tctcggaaaa aggaaattat
 551  tgagaagttc tgagagcatt acatttttg tgacagaagc caagcttccc
 601  tttttaattc gtatagttgc ttctggtagc tgtcttaacac ttttctgaa
 651  cattctgttc gcatcatgct tgacgat
```

Figure 8. *srj-54* promoter sequence.

been constructed and used in prior experiments, will be used as template DNA in the PCR reaction to make constructs with varying degrees of deletion in the *srj-54* promoter. These constructs will then be injected into wild type animals, and the levels of *srj-54* expression examined. This work is currently in progress.

Since it has been seen that *srj-54* is sexually dimorphically expressed, other genes were examined in order to identify other possible dimorphic expression. The *srj-54* gene is part of a larger family of class J serpentine receptors. Those *srj* genes with greatest homology to *srj-54* were found, and from those it was seen that *srj-57* had the greatest degree of homology. A strain was obtained that carried *srj-57* on an extrachromosomal array, as well as a *dpy* rescue. Since the extrachromosomal array is not maintained with 100% fidelity, it was necessary to using the *dpy* rescue to identify those worms carrying the extrachromosomal array. The strain did not include the *him-5* allele, and thus there were not enough males to allow analysis. Males were induced through heat shocking and subsequently crossed back to the strain to generate a strain that had the *him-5* allele. The resulting *dpy-5; him-5; Ex[srj-57::gfp]* strain was then examined for GFP expression.

It was found that GFP was expressed in an unidentified neuron, located in the head, in both males and hermaphrodites (Figure 9). Having ruled out sexually dimorphic expression in *srj-57*, it was decided not to examine further members of the *srj* family.

During the course of each of these experiments, it was assumed that *srj-54* lay downstream of the master switch gene *tra-1*. However, this hypothesis remained to be tested. In order to do so, it was necessary to create a *tra-1 -/-* strain and examine *srj-54* expression in XX pseudomales. The first step in the intended construction was to cross *fil5; him-5* to *pha-1* to generate an *fil5; pha-1* strain, thus removing the *him-5* allele. The *pha-1* strain contains a temperature sensitive allele that does not function at room temperature. The *pha-1* allele was used because of its extremely close location to *tra-1*. The *pha-1* allele serves as a balancer of *tra-1*: when performing the cross, wherever there is *pha-1* there will not be *tra-1*, making this temperature sensitive allele useful in selecting for progeny and identifying the desired genotypes. Due to mislabeling of the *pha-1* and *pha-1; him-5* strains, the *pha-1; him-5* strain was crossed instead. It was thus necessary to perform a second cross in order to remove the *him-5* allele and complete the *fil5; pha-1* strain. From this point, the strain will be crossed to the *tra-1* strain, and a *tra-1; fil5* strain constructed. Any males in this strain will be XX pseudomales. Should *srj-54* lay downstream of *tra-1*, we would expect GFP expression in these XX pseudomales, whereas that expression would not normally exist in XX hermaphrodites. The strain has yet to be constructed, and this work is currently in progress.

Discussion

srj-54 expression is reduced in mab-3 *mutants.*

Strains were constructed with *fil5*, the integrated *srj-54::gfp* transgene, and each of the DM domain genes. While GFP expression levels in the *fil5; dmd-3, him-5* single mutant and *fil5; mab-3; dmd-3, him-5* double mutant strains were not significantly different from that of the wild type *fil5; him-5* strain, expression levels in the *fil5; mab-3; him-5* strain were drastically reduced. It is interesting that expression is reduced in the *mab-3* mutant, implying that *mab-3* is necessary for proper *srj-54* expression, while expression levels in the *mab-3; dmd-3* double mutant are not reduced. This hints at a more intricate genetic model in which *dmd-3* may have a negative influence on *mab-3* expression. However, this data remains to be verified. In order to verify the expression levels of the double mutant, it would be useful to either perform anti-dmd-3 RNAi on the *fil5; mab-3; him-5* mutant and look for an increase in expression to that of the wild type, or to introduce an extrachromosomal array carrying *dmd-3* into the double mutant and look for a reduction in expression level to that of the *fil5; mab-3; him-5* mutant.

Since a delay occurred in the construction of the *fil5; mab-23; him-5* strain, it was not possible to construct the *fil5; mab-3; mab-23; dmd-3; him-5* triple mutant strain. It may be informative to construct and examine this strain in comparison with the other four strains already examined. A further reduction or augmentation of expression levels would give further insight into the interactions between DM domain proteins and *srj-54*.

srj-54 and mab-3 *are co-expressed.*

An examination of strains harboring an extrachromosomal array that carries *mab-3pro::gfp* and *srj-54wCherry* reporters proved difficult at first, due to intense pharyngeal GFP that blocked view of any nearby neurons. Since the AIM interneuron, in which *srj-54* is known to be expressed, is near the pharynx, it was necessary to take steps to knock down the pharyngeal GFP levels. Using an anti-GFP RNAi construct, pharyngeal GFP was sufficiently knocked down after two generations of growth on RNAi feeding plates. However, expression was knocked down only enough to visualize the neuron in a sizable, but statistically small, number of animals. It remains to be discovered why the pharyngeal GFP was not knocked down as drastically as was expected, but the reason may lie in the RNAi construct itself. This being said, after examination of a number of animals, it was clear that GFP and dsRed expression overlapped, implying that *mab-3* and *srj-54* have overlapping expression in the AIM interneuron. This suggests that *mab-3* is likely to function cell-autonomously in *srj-54* regulation.

The srj-54 promoter may be directly regulated by mab-3.

The sequence of the *srj-54* promoter was obtained, as well as the sequence of the *mab-3* consensus binding sequence. The promoter was searched for sequences matching the consensus binding sequence to a significant degree, turning up four such matches. Primers were designed such that varying numbers of these consensus binding sequences would be deleted. This will be done via PCR with the *fdf5* transgene as template DNA. GFP expression levels will then be examined. This procedure will give a general impression of whether *srj-54* is directly regulated by *mab-3*. However, should *srj-54* expression levels be knocked down, it should not be assumed that it is due to the absence of a direct regulator. The deletion of such large regions of the promoter runs the risk of eliminating other regulatory elements, some possibly belonging to other regulatory factors that have yet to be identified and which play a more integral role than does *mab-3*. Should expression be knocked down, site-directed mutagenesis or targeted deletion of the *mab-3* consensus binding sequences will be more reliable in determining whether direct regulation by *mab-3* is taking place.

srj-57 does not exhibit sexually dimorphic expression.

Knowing that *srj-54* is dimorphically expressed, *srj-57* was chosen based on its homology to *srj-54* as a likely candidate for similar sexually dimorphic expression. After inducing production of males in the *dpy-5; Ex[srj-57::gfp]* strain via heat-shocking and crossing these males back to the original strain, the resulting *dpy-5; him-5; Ex[srj-57::gfp]* was examined for GFP expression. It was found that both males and hermaphrodites expressed the GFP reporter in a neuron located in the head, thus ruling out sexually dimorphic expression.

srj-57 was chosen because it exhibited the highest degree of homology to *srj-54*. However, other genes in the SRJ class of serpentine receptors also demonstrate a high degree of homology to *srj-54*. Although *srj-57* failed to display any expression patterns of interest to this study, it is not without possibility that other genes in the family may exhibit such

Figure 9. *srj-57::gfp* expression.

patterns. In particular, *srj-53* and *srj-38* showed a high degree of homology to *srj-54* and may be worth examining in the future.

srj-54 is assumed to lie downstream of tra-1.

While *srj-54* has been assumed to lie downstream of *tra-1*, this assumption has not been tested. To do so, expression of *srj-54* in *tra-1* XX pseudomales must be examined. After crossing *fdf5; him-5* to *pha-1; him-5* and then eliminating the *him-5* allele from the strain, an *fdf5; pha-1* strain was completed. The *tra-1* must be crossed into the strain, and *pha-1* eliminated for it to be completed. This will leave an *fdf5; tra-1* strain consisting only of XX hermaphrodites and *tra-1* XX pseudomales. Should expression persist in these XX animals, it will be verified that *srj-54* lies downstream of *tra-1*.

These studies showed that *srj-54* may be directly regulated by the DM domain gene *mab-3*, while *dmd-3* may exert a negative influence on the activity of *mab-3*. It was also shown that the related *srj-57* gene did not display sexually dimorphic activity. These studies offer further insight into the role of DM domain proteins as downstream regulators in the sex differentiation pathway.

References

1. Hodgkin, J. & Brenner, S. Mutations causing transformation of sexual phenotype in the nematode Caenorhabditis elegans. Genetics 86, 275-287 (1977).
2. Hodgkin, J. Sex Determination and the Generation of Sexually Dimorphic Nervous Systems. Neuron 6, 177-185 (1991).
3. Raymond, C. S. & Zarkower, D. Evidence for evolutionary conservation of sex-determining genes. Nature 391, 691-695 (1998).
4. Yi, W. & Zarkower, D. Similarity of DNA binding and transcriptional regulation by Caenorhabditis elegans MAB-3 and Drosophila melanogaster DSX suggests conservation of sex determining mechanisms. Development 126, 873-881 (1999).
5. Sturtevant, A. H. A gene in Drosophila Melanogaster that transforms females into males. Genetics 30, 297-299 (1945).
6. Zarkower, D. & Hodgkin, J. Molecular Analysis of the C. elegans Sex-Determining Gene tra-1: A Gene Encoding Two Zinc Finger Proteins. Cell 70, 237-249 (1992).
7. Shen, M. M. & Hodgkin, J. mab-3, a gene required for sex-specific yolk protein expression and a male-specific lineage in C. elegans. Cell 54, 1019-1031 (1988).
8. Narendra, U. & Wynn, M. A. Sex-specific Gene Regulation: the doublesex DM motif is a bipartite DNA-binding domain. Journal of Biological Chemistry 277, 43463-43473 (2002).
9. Yi, W. & Zarkower, D. mab-3 is a direct tra-1 target gene regulating diverse aspects of C. elegans male sexual development and behavior. Development 127, 4469-4480 (2000).
10. Lints, R. & Emmons, S. Regulation of sex-specific differentiation and mating behavior in C. elegans by a new member of the DM domain transcription factor family. Genes & Development 16, 2390-2402 (2002).
11. Hodgkin, J. The remarkable ubiquity of DM domain factors as regulators of sexual phenotype: ancestry or aptitude? Genes & Development 16, 2322-2326 (2002).
12. Raymond, C. & Zarkower, D. A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. Human Molecular Genetics 8, 989-996 (1999).

13. Volff, J. & Zarkower, D. Evolutionary Dynamics of the DM Domain Gene Family in Metazoans. Journal of Molecular Evolution 57, S241-S249 (2003).
14. Sulston, J. & Hodgkin, J. The Nematode Caenorhabditis elegans (ed. Wood, W.W.) 587-606 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988).
15. Schafer, W. Egg-laying. WormBook. 2005.
16. Portman, D.S. & Emmons, S.W. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. Development 127, 5415-5426 (2000).

JUR

Explore your passions.

Contribute articles.

Join JUR.

sa.rochester.edu/jur

The Effects of Coal-Fired Power Plant Emissions and Fly Ash on a Regional Watershed

Amanda F. Carey and Zoë R. Harrold, 2008

Advised by Robert Poreda, Ph.D.

Department of Earth and Environmental Sciences

The Roaring Creek watershed, located near an anthracite coal field in central Pennsylvania, is an essential drinking water resource to the surrounding population. Located within the Conrad Weiser State Park, the watershed has been receiving an atmospheric deposition of fly ash (coal combustion residue, or CCR) as well as acidic mine discharge from abandoned mines, contaminating local groundwater through centuries of coal mining.[1-10] Past and present use of the land surrounding the watershed for coal mining, coal fired power production, and most recently, disposal of CCR, provides ample opportunity for trace metal contamination within the watershed. The surrounding area is heavily mined and provides a prime location for coal power production and disposal of fly ash in unreclaimed mines. There are 15 independent coal combustion power plants in the area, which increasingly produce nine million tons of fly ash each year.[10] Therefore, it is important that we understand the effects that coal fired power production and atmospheric CCR deposition will have on the quality of water and general watershed health in areas near coal fired power plants.

Our initial focus was to determine the trace and heavy metal inventories of water and sediments within the watershed and surrounding areas to evaluate if the concentrations are within the Environmental Protection Agency's (EPA) standards. Ion chromatography was used to analyze major element concentrations, while an inductively coupled plasma mass-spectrometer (ICP-MS) was used to analyze trace elements and major lead isotopes (^{206}Pb, ^{207}Pb and ^{208}Pb) for source characterization including CCR, gasoline, and smelter and natural background lead isotopic ratios. By determining these compositions and inferring the chemical processes that underlie the complex watershed system, we were able to predict potential water quality issues. We quantified the effects of fly ash and inorganic aerosols emitted from a nearby power plant and mine acid drainage from reclaimed mines by observing spatial variations in trace metal concentrations within surface sediments, snow and water samples. In addition, a sediment core was used to determine the variations of trace metal input and source through time.

Figure 1: Sampling map from the second sampling trip in 2006. Red flags denote sample locations, the blue dot denotes initial snow sample and the red 'Y' is the location of the coal-fired power plant (created on TopoQuad using field latitude and longitude values).

General Setting:

The Roaring Creek watershed is contained in a synclinal sandstone aquifer characterized by a low buffering capacity and is directly adjacent to the northern most outcrop of the central anthracitic coal region of Pennsylvania. The watershed contains three reservoirs and is the main water source for a 15 mile radius. The creek flows from east to west through the Conrad Weiser State Park with its origin approximately eight miles north-east of the Mount Carmel cogeneration power plant. The 44MW coal fired cogeneration power plant, which opened in 1990, is located less than two miles from the main reservoir.[9]

Acidic precipitation, prevalent throughout most of the northeastern U.S., is a significant factor in the water's pH. In addition, local mine acid drainage is concentrated near the creek's origin. Both factors undoubtedly contribute to the watershed's chemistry, resulting in high sulfate and nitrate concentrations near known mine acid drainage inputs and a mean pH of 4.96. Although some component of agricultural input cannot be discounted near the origin of Roaring Creek, the watershed is overall largely protected from other potential sources of contamination such as landfills, agriculture, and gasoline stations by regional flow divides.

Sampling and Analysis Methodology:

Several techniques were implemented in order to achieve a broad range of data regarding the major trace and heavy element inventories. Water, surface sediments, sediment cores, and snow samples were collected from within the watershed

on three occasions between 2006 and 2007. Figure 1 displays sample locations of the second sampling trip along with the location of the power plant and the initial snow sample taken in 2006.

Field parameters (i.e. temperature, pH, and conductivity) and longitude and latitude (using Global Positioning System technology) were taken on location. All water samples were collected in nitric acid washed HDPE (high-density polyethylene) bottles and filtered through a 0.45 µm filter. Snow samples were collected in zip-lock bags to be melted and filtered later. Two sediment cores were taken from opposite ends of the watershed in half-inch copper tubes and were driven approximately 20 cm into the reservoir sediment to observe long term changes in deposition. A bulk sample of CCR obtained from within the power plant was used to determine specific inorganic trace metals that are concentrated by the combustion process.

Watershed and core sediments were sieved at 150 microns and treated with an oxalic solution (ammonium oxalate and oxalic acid) buffered to a pH of 3 in order to extract the exchangeable and adsorbed trace metals and the breakdown products of oxyanion complexes.[7] Ten mL of oxalic solution was added to 0.5 grams of sediment, agitated for ten minutes, and placed in an ultrasonic bath for forty minutes. The samples were then centrifuged twice for twenty minutes to obtain a supernatant which was decanted for analysis.[7] The extraction process was carried out in the dark to prevent degradation of the oxalic solution and all instruments and containers were acid washed prior to use.

To prepare the sediment extracts for trace metal analysis on the ICP-MS, the samples were further diluted with 18 MΩ water and acidified with ultra pure nitric acid. Analyses of water and sediments were done on the X7 Thermo Electron ICP-MS at the University of Rochester according to the U.S. Environmental Protection Agency Method 2008.[8] Indium and bismuth were added to all sediment digestion samples as internal standards for instrument calibration.[7]

Measurements of major anions and cations were taken on a Dionex ion chromatograph (IC) according to standard procedures at the University of Rochester.[7] Grain compositions and images were acquired using a scanning electron microscope (SEM) equipped with EDAX capabilities and a particle-phase

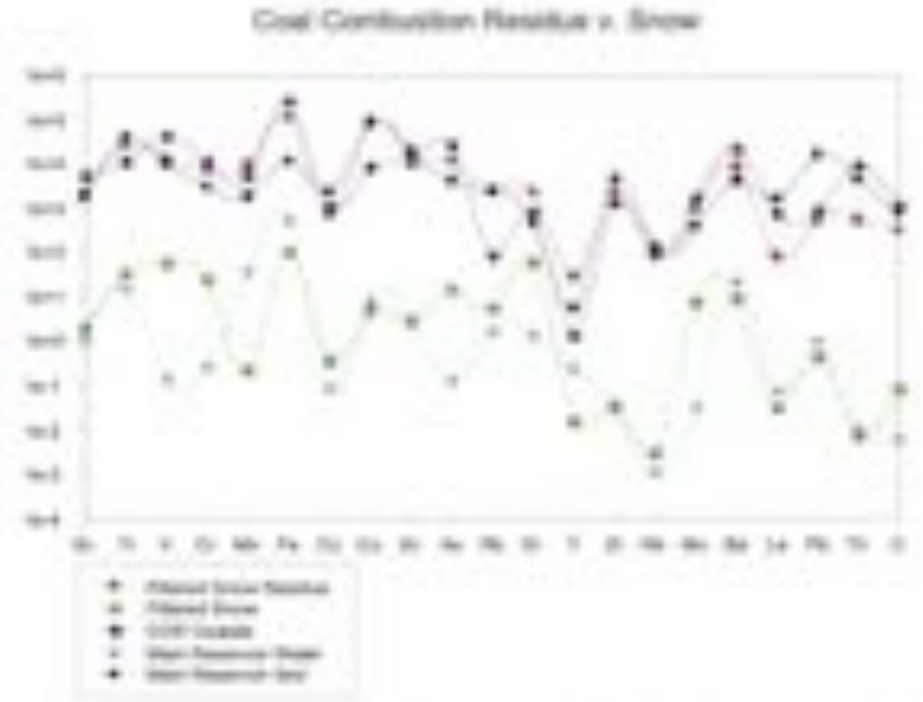


Figure 2: Concentrations of trace elements in filtered snow residue, filtered snow, oxalate extracted CCR, main reservoir water and sediments, representing mobile versus bulk silicate trace elements concentrations and a correlation with pure CCR.

analysis program.

Results and Discussion

Initial analysis compared the trace metal concentrations from an oxalic extraction of fly ash, filtered snow residue, and filtered snow precipitation as well as sediments and water samples from the main reservoir (Figure 2), representing concentrations of mobile trace elements versus those sequestered in bulk silicate samples. All samples show a strong elemental correlation with the exception of oxyanions (e.g. V, Cr, As, Mo, U). Specifically, the main reservoir water samples showed anomalously low concentrations of oxyanions when compared to filtered snow, sediment, and snow residue sample.

During sampling, snow samples taken in close proximity to the power plant had a distinct red coating, which is denoted "filtered snow residue" in Figure 2. The strong correlation between the filtered snow residue and coal combustion residue suggests that this coating is fly ash. SEM analysis of the filtered snow residue (Figure 3) showed a general Ca-Al silicate composition. In addition, various other grains such as pyrite and quartz were found. The residue also contained iron oxy-hydroxide clusters in obvious amorphous shapes that can be interpreted as a result of rapid cooling following the

Figure 3: SEM image of CCR.

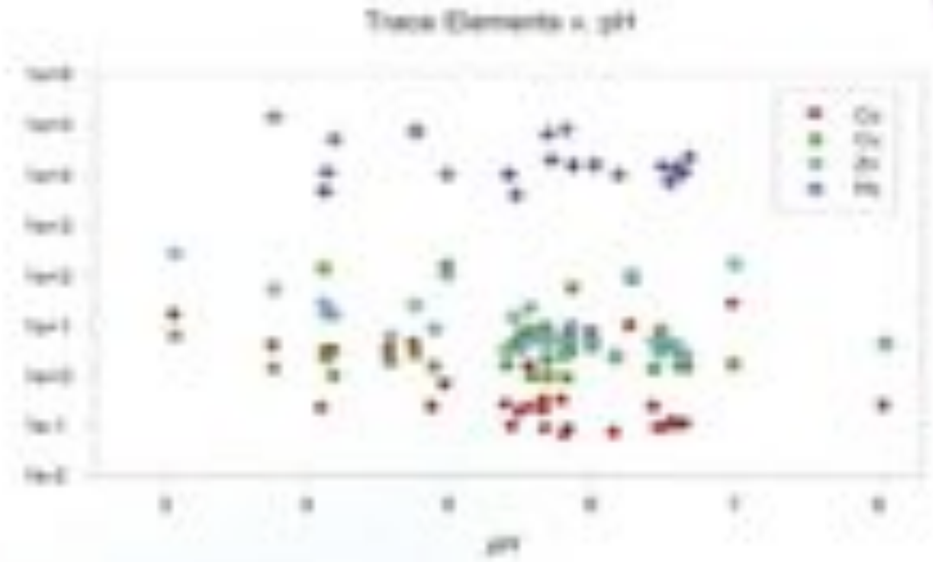


Figure 4: Concentration of trace elements versus water pH show a decrease of concentration with increasing pH.

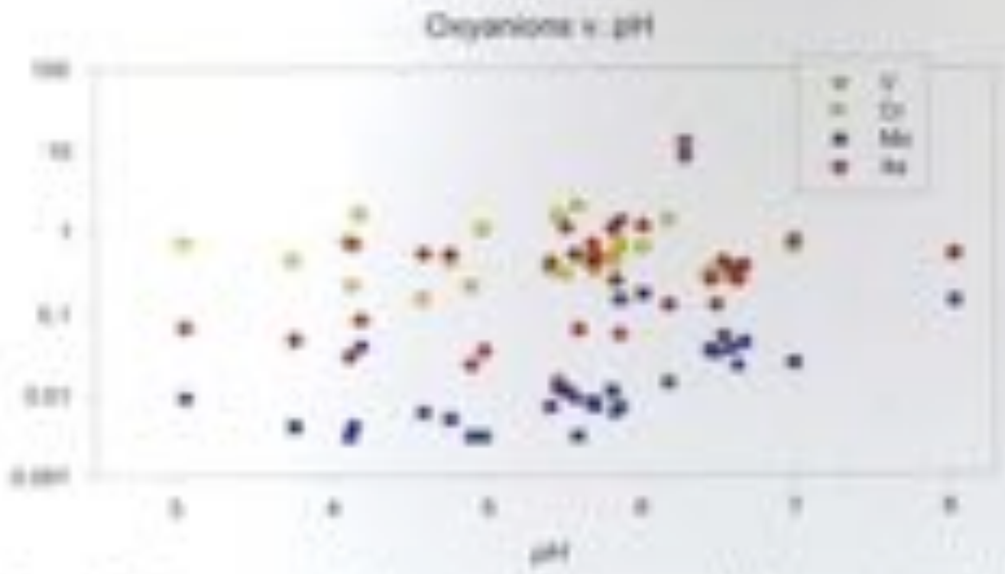


Figure 5: Concentration of oxyanions versus water pH show an increase of concentration with increasing pH.

Figure 6: Trace element concentrations in water versus distance from origin show decreasing concentrations until approximately 8 miles, where concentrations begin to show an increasing trend.

combustion. Analysis via an EDAX equipped SEM showed an average iron grain composition of 3% within a random distribution of grains.

The red tint seen on the snow sample can most likely be attributed to the presence of iron oxy-hydroxides. The high Ca-Al silicate background observed in the SEM results implies a unique and highly oxidizing environment as a result of the combustion of coal slag composed mostly of sandstone and limestone.

Several studies, including a study done in the Bengal Basin, have linked the release of arsenic and other cations to the dissolution of iron oxy-hydroxides (FeOOH) in reducing environments such as a shallow sub-surface environment.[3] This supports the data showing higher oxyanion concentrations found within the filtered snow samples relative to water from the main reservoir. We attribute this to the direct incorporation of iron oxy-hydroxides in the dry deposition on the snow, which was melted and filtered for analysis. FeOOH is known to be transported as a colloidal particle, and therefore would not be filtered out of melt water; hence, the higher concentrations with the release of additional adsorbed elements.

Analysis of anion and trace metal concentrations in snow precipitation in relation to distance from power plant yields no strong correlation. However, it was noted that higher oxyanion concentrations were measured in samples within close proximity to the power plant suggesting an anthropogenic dry deposition. It was also noted that samples obtained from similar distances from the power plant that were collected a week versus approximately 24 hours after a fresh snow appeared to have a larger amount of dry deposition, suggesting a time integrated dry deposition of CCR.

A strong correlation was observed between trace metal and major element and oxyanion concentrations when compared to the water samples' measured field pH (Figures 4 and 5). Major elements and trace metals show increasing concentrations with decreasing pH (increasing $[H^+]$), while oxyanion concentrations exhibit an expected inverse relationship. These results follow the expected K_d curves for the quantified elements.[4] Due to this observation, it can be inferred that the pH and individual K_d values are major factors in determining the aqueous concentrations of trace metals in the reservoir.

Trace metal concentrations in the water samples were compared to their location relative to the origin in an attempt to assess any spatial change in water chemistry (Figure 6). Results indicated a general decrease in trace metal concentrations along the flow path until approximately eight miles from the creek's origin, at which point the concentration begins to increase. This decrease in trace metal concentration along the flow path is expected due to the presence of a significant input of mine acid drainage near the origin of the creek and the knowledge that acidic mine drainage has an effect on trace metals such as Pb, Zn, Co, etc. The mine acid input decreases along the flow path, except for a small contribution near the western most reservoir. This data indicates a general dilution factor with increasing distance from origin from 0 to approximately eight miles along the flow path of Roaring Creek. At this eight mile point, the creek flows past the power plant and an area of known acid mine drainage. The source of increased concentrations remains unclear but may be related to acid mine drainage or an increased amount of dry deposition from the plant at this location. Similar comparisons done with water samples showed a slight correlation between major elements (e.g. Fe, Mn, Ba) versus distance from the power plant, yielding no definitive conclusions.

Raw sediment data showed no significant changes in concentration with respect to distance from the power plant. To correct for ambient geological/erosional input of other trace metals, normalizing to *Zr* was employed. Zirconium is a high field strength, refractory element that is not expected to be released in any significant proportion by the combustion process because of its refractory nature. This was confirmed by the data, as *Zr* was found at consistent concentrations within

Figure 7: Trace metal concentrations normalized to *Zr* versus distance from power plant show a general decrease with increased distance from the power plant in agreement with normalization to Fe.

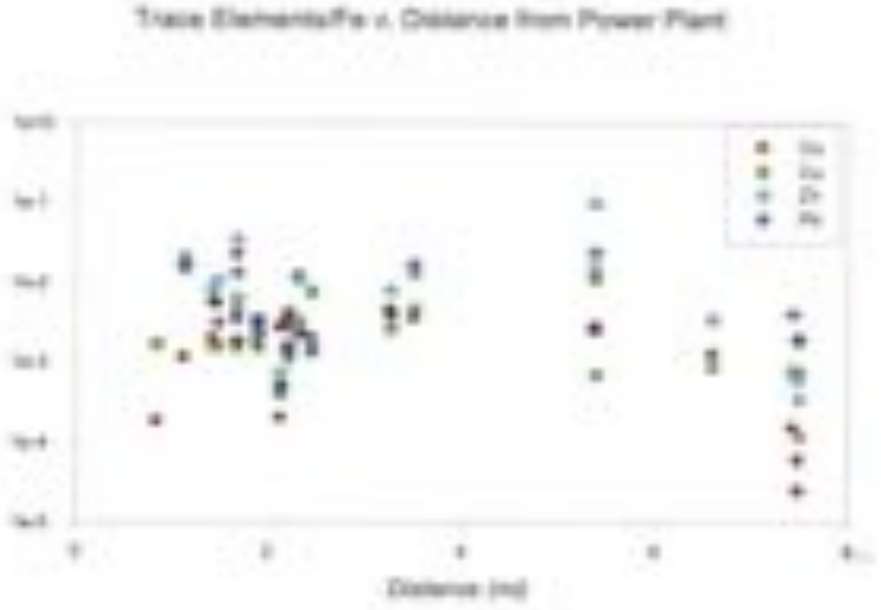


Figure 8: Trace metal concentrations normalized to Fe versus distance from power plant show a general decrease with increased distance from the power plant.

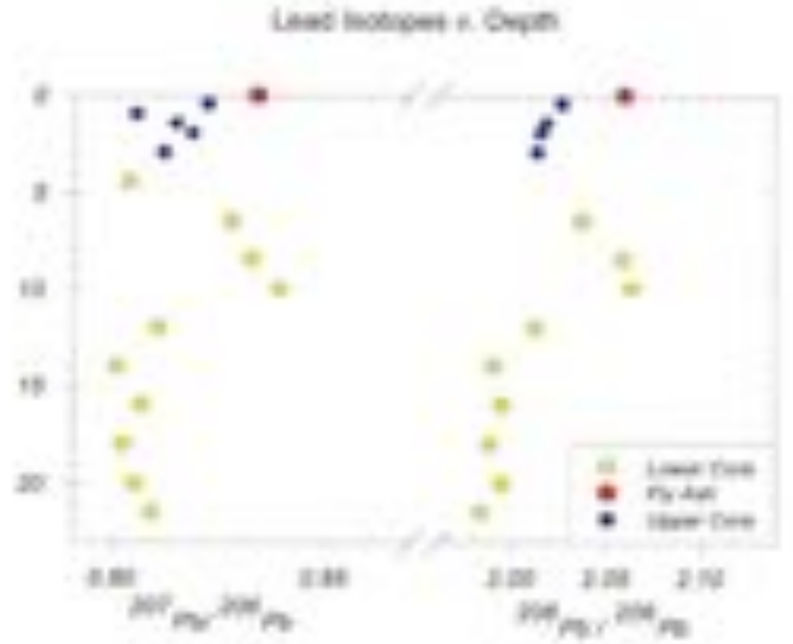


Figure 10: Lead isotopic ratios 207Pb /206Pb and 208Pb /206Pb versus depth within a core sample show younger sediments trending toward fly ash/gasoline lead.

the watershed and was in low concentrations in the fly ash. Normalization to *Zr* resulted in decreasing trends in trace metal, organics, and major element concentrations with increasing distance from the power plant. Data was also normalized to Fe to account for sample specific iron oxyhydroxide concentrations and their effects on ion adsorption. Similar trends were found in both the *Zr* and Fe normalizations (Figures 7 and 8).

Core samples were obtained to assess long term, temporal variations of measured trace metals in the region. The lower part of the core shows relatively consistent concentrations with a drastic input at approximately 4 cm depth. A decreasing concentration of Pb was found in younger sediments, an opposite trend from the trace metals. This decrease could potentially be explained by the mandated removal of tetra-ethyl lead from gasoline (Figure 9).

In an attempt to determine the source of the lead within the core, ^{207}Pb /^{206}Pb, and ^{208}Pb /^{206}Pb isotopic ratios were quantified (Figure 10). Deeper sediments showed relatively consistent lead ratios, similar to ratios seen in Mississippi valley lead, suggesting high concentrations most likely due to Pb smelting. In contrast, younger sediments trended toward ratios found in fly ash obtained directly from the power plant as well as gasoline lead ratios. A notable change in the cores' isotopic ratio occurs at approximately 4 cm, which correlates to the observed increase in trace metals from the same core. An additional change, occurring at approximately 10 cm may be attributed to large gasoline and/or fly ash contributions to the watershed sediments.

When lead isotopic ratios were plotted against each other, there was an obvious mixing component between smelter Pb and anthropogenic fly ash and gasoline Pb (Figure 11). This change in lead isotopic ratios suggests that fly ash is an increasingly dominant source of lead in the Roaring Creek watershed. In addition, temporal variations in total lead input into the sediment coupled with changes in lead isotopic ratios are expressed in Figure 12. Both graphs demonstrate that original lead sources have since moved toward a gasoline and fly ash lead composition. At the point where we believe our data corresponds with the removal of tetra-ethyl lead from gasoline, overall Pb concentrations decrease but the isotopic composition appears to trend towards the measured fly ash lead ratio. It is important to highlight that additional sources of lead in the youngest part of the core and sediment samples are unambiguously of fly ash composition, yet the overall concentration of lead in the core has decreased over time. Overall, lead isotopic data clearly suggests that the power plant is affecting the natural elemental inventories of the watershed.

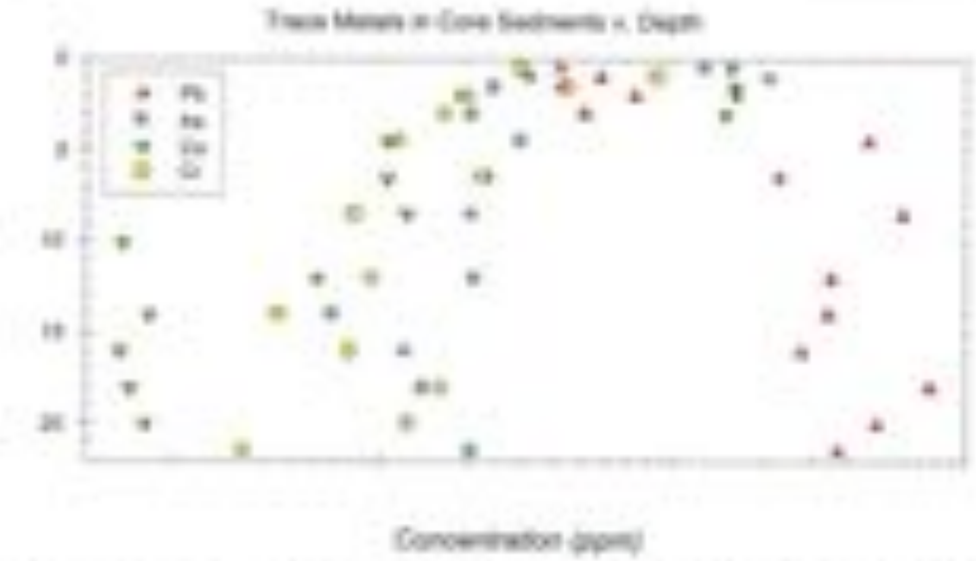


Figure 9: Trace metal concentrations in core samples versus depth. Older sediments display relatively consistent concentrations while younger sediments show increasing concentrations of trace metals. Lead displays a decreasing trend in younger sediments.

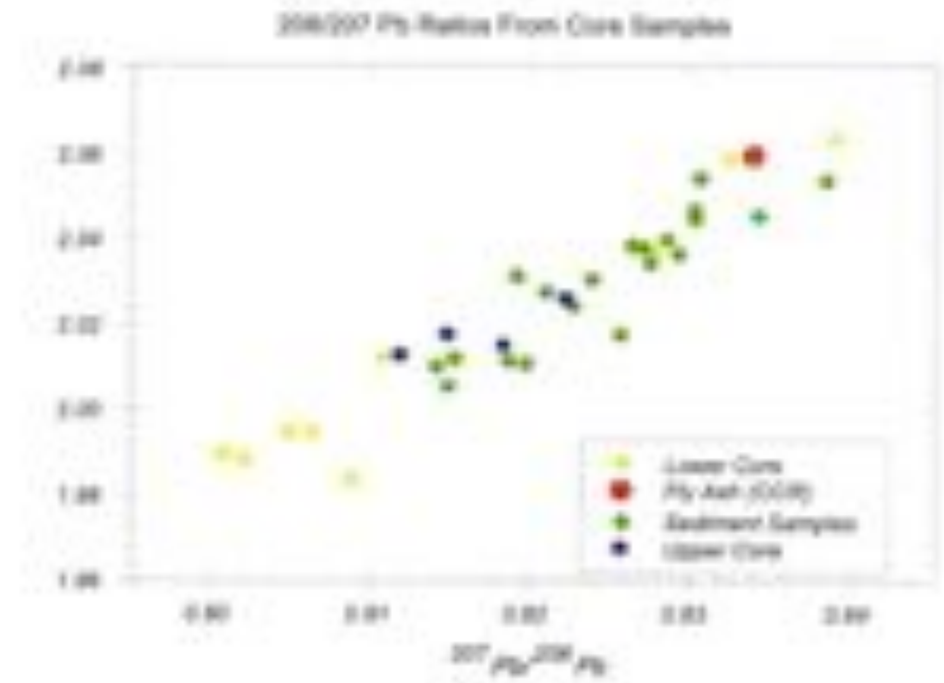


Figure 11: Lead isotopic ratios from core samples show mixing component between smelter Pb and fly ash/gasoline composition in younger sediments.

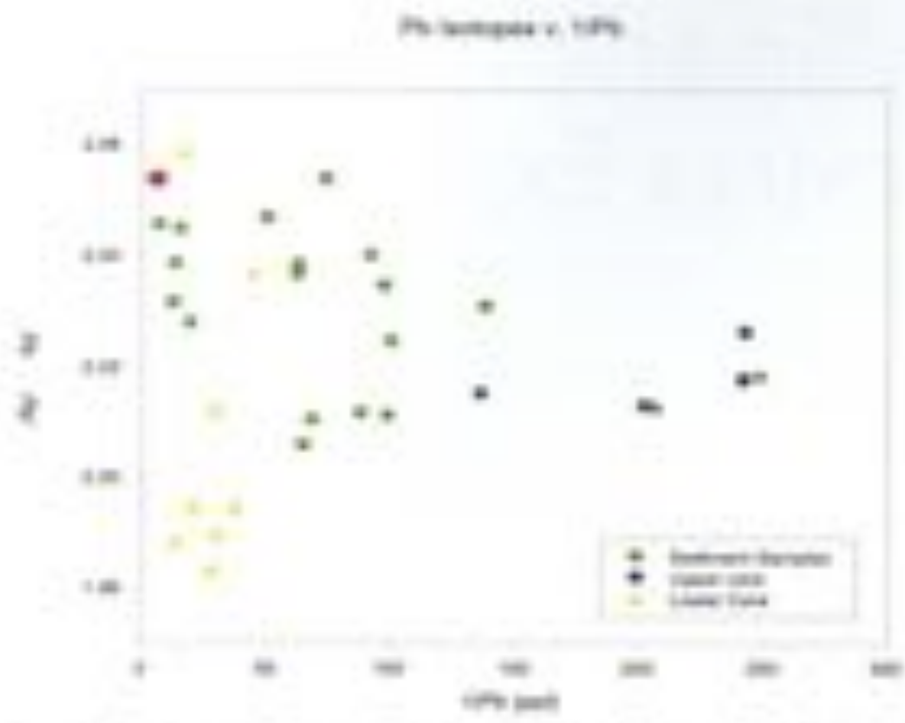


Figure 12: Lead isotopic ratios versus 1/Pb (ppt) shows changes in lead concentrations with related trending toward a fly ash/gasoline lead source. Red dot indicates fly ash sample.

Conclusion

Sediment core profiles, Pb isotopes, enrichment concentrations, and erosion corrected trace metal concentration data suggest that power plant emissions and fly ash are reaching the watershed in detectable amounts. However, this anthropogenic input does not appear to be impacting the drinking water quality beyond the limits set by the EPA. This particular power plant has proven to be unique due to its highly oxidizing combustion environment, leading to the presence of iron oxy-hydroxides, which may be a significant control of trace metal transport within the watershed. Furthermore, the degree of trace metal adsorption in this watershed is controlled by pH. This pH is a delicate balance of acidic inputs from mine drainage and precipitation as well as an alkaline contribution from the dry deposition of fly ash and the low inherent buffering capacity of the aquifer. The basic nature of fly ash may even prevent the release of trace metals into the reservoir. Ultimately, K_d and pH are the dominant controls of the trace metal concentrations within the watershed and currently mitigate any impact to water quality.

Acknowledgements

We would like to thank our advisor, Robert J. Poreda, Ph.D. for his guidance during this project; Tom Darrah for his help in the field and lab; Ashish Basu, Ph.D. and Arundhuti Ghatak for use of clean lab space, equipment, and prepared chemical reagents; Rachel Skellie for her assistance in the field and lab; Thomas, Ed and Sue Darrah for their assistance during field work, and Buddy the dog.

References

1. California University of Pennsylvania, Department of Earth Sciences. The Effects of Subsidence
2. Resulting from Underground Bituminous Coal Mining on Surface Structures and Features and on Water Resources: Second Act 54 Five-Year Report, for the Pennsylvania Department of Environmental Protection, February 6, 2005.
3. Dowling, C., Poreda, R., Basu, A., Peters, S., Geochemical Study of Arsenic Release Mechanisms in the Bengal Basin Groundwater. Water Resour. Res., 38, 2002.
4. Fehn, U., G. Snyder, and P.K. Egeberg, Dating of pore waters with 129I: relevance for the origin of marine gas hydrates. Science, 289, 2332-2335, 2000.
5. James, R.O., T.W. Healy., Adsorption of hydrolyzable metal ions at the oxide-water interface. J. Colloid Interface Sci., 40, pp. 42-52, 1972a.
6. Li, X., B.J. Coles, M.H. Ramsey, and I. Thornton, Sequential extraction of soils for multielement analysis by ICP-AES. Chem. Geol., 124, 109-123, 1995.
7. Long, S.E., and T.D. Martin, Method 200.8, Determination of trace elements in waters and wastes by inductively coupled plasma-mass spectrometry, in Methods for the Determination of Metals in Environmental Samples, pp. 83-122, U.S. Environ. Prot. Agency, Washington, D.C., 1991.
8. McKeague, J. A., Manual on soil sampling and methods of analysis, 2nd ed., Can. Soc. of Soil Sci., Ottawa, Ontario, 1978.
9. Stant, J., Evans, L., Gadinski, R., Norris, C., Impacts on Water Quality from Placement of Coal
10. Combustion Waste in Pennsylvania Coal Mines. Clean Air Task Force, 2007.
11. U.S. Environmental Protection Agency. "Air Toxics Website." Technology Transfer Network. 6 June 2007. <http://www.epa.gov/ttn/atw/>.

Production of Soluble Multimeric Dengue II NS1 Protein in the Baculovirus Expression System

Elizabeth Selleck, 2008

Advised by Robert Rose, Ph.D.

Department of Microbiology and Immunology, UR Medical Center

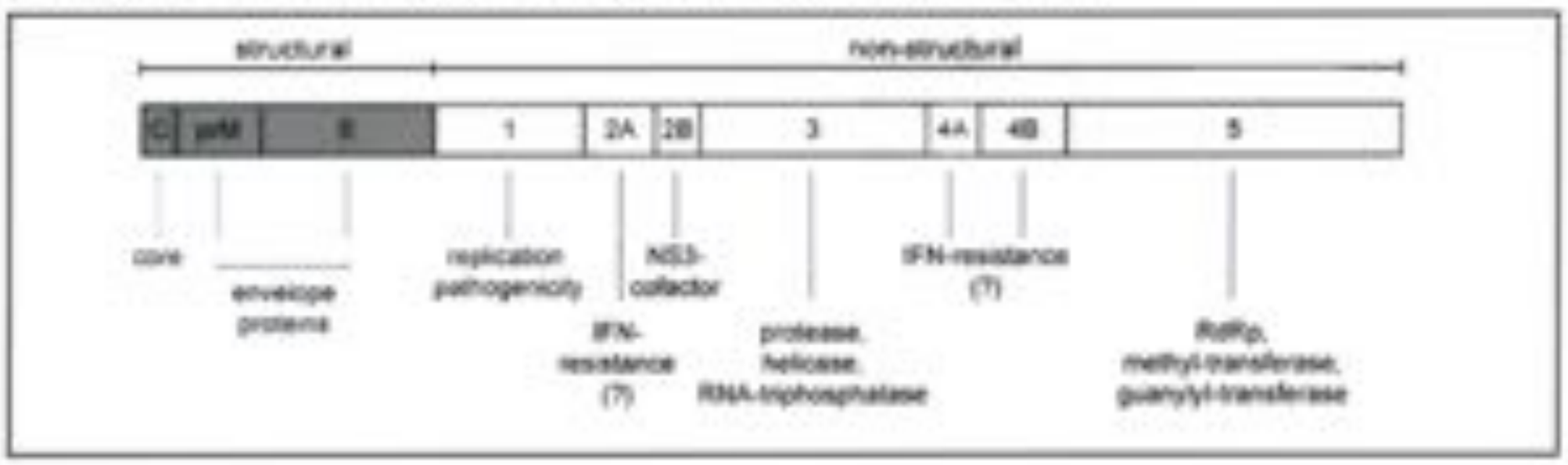
Flaviviruses are enveloped positive strand RNA viruses that cause significant disease burden among tropical human and animal populations. This genus contains nearly 80 viruses, many of which are human pathogens causing diseases that include fevers, encephalitis, and hemorrhagic fevers. Flaviviruses of global concern include Japanese encephalitis, yellow fever, and dengue viruses.[1] Dengue virus (DV) is of particular importance as a mosquito-borne virus with high morbidity and mortality rates, but it does not yet have an effective vaccine, as some other members of the genus do.[2] It is estimated that 50-100 million people living in tropical areas are infected with dengue virus each year, and infections can be fatal if not properly treated.[3,4] Infection with dengue virus can cause two syndromes: dengue hemorrhagic fever and dengue shock syndrome. These manifestations usually occur upon secondary infection and are most effectively treated with early, aggressive support care.[3] Several species of mosquito have been identified as transmission vectors including *Aedes aegypti* and *Aedes albopictus*, both of which have been shown to cause dengue epidemics.[5,6] Studies into various antigenic properties of dengue serotypes 1-4 are needed to decrease the number of affected individuals worldwide.

The 11kb dengue genome of dengue virus codes for three structural and seven nonstructural proteins[4] including the capsid (C), membrane (M), and envelope (E) proteins. The nonstructural proteins are known as NS1, NS2A, NS3, NS4A, NS4B, and NS5. The genome contains one open reading frame, which codes for a single polyprotein (Figure 1). After translation, the polyprotein is cleaved into 10 discrete products by a virally encoded protease. Structural proteins are located at the N-terminus of the polyprotein followed by the nonstructural proteins.[7,10] The virion particle is encapsidated by C protein, which is surrounded by a lipid bilayer with the envelope protein, a major surface component of the virus particle. The M protein is important for maturation of the virion. The nonstructural proteins have several functions in the cell. NS1 is of particular interest due to its largely unknown function in the virulence of dengue virus.

NS1 is a glycoprotein that is found in all dengue serotypes.[4,7] In its monomeric form, it has been found to be 45-48 kDa, although that value may vary according to the level of glycosylation present on the protein. Twelve cysteine residues are present and form disulfide bonds in the homodimer form. Dimerization has been shown to be a prerequisite for transport along the secretory pathway of infected cells.[8] This dimer is also associated with the membrane components of an infected cell.[7,9] A hexameric species, ~300 kDa, is secreted from infected mammalian cells and has been found in high levels, as much as 15 µg/ml, in the blood of infected humans.[7] The exact function of NS1 has not yet been verified, although it appears to be involved in viral RNA replication and viral multiplication due to its association with replicating RNA and its ability to increase DV1 production in infected cells.[1,4,9] Due to the fact that NS1 is found in high quantities in dengue infected patients, it has been implicated as a possible marker for viremia.[10]

NS1 is a possible vaccine candidate. Unlike the E protein, which has been shown to induce antibodies that may increase antibody dependent enhancement (ADE) and amplify severity

Figure 1: Dengue Genome with associated protein functions (Molecular Virology, University of Heidelberg)



of secondary infection. NS1 is not present in the virion, and therefore antibodies produced against it are unlikely to mediate ADE.[2] This humoral response to NS1 and protein E has been found in patients infected with dengue and other flaviviruses. It appears that the efficiency of NS1 as an immunogen relies on the protein being in an oligomeric form higher than a dimer; thus, it is necessary for protein produced by recombinant baculovirus infected insect cells to secrete NS1 in this higher molecular weight form if it is to be used in its native form for studies of the protein's vaccine capabilities as well as studies relating to how NS1 is involved in the virulence of the dengue virus.[3]

In this study, we found that recombinant NS1, when expressed in insect cells, is secreted in high molecular weight multimers, as demonstrated by western blot immunoassay. Cell free supernatant from recombinant pAc6067-N6His-DV2-NS1 baculovirus-infected *T. ni* cells was analyzed by SDS-PAGE. Western blot, and dengue NS1-specific and 6His epitope tag-specific monoclonal antibodies were used to confirm the presence of NS1. Supernatant was treated with 1%, 0.6%, and 0.4% β-mercaptoethanol, used to denature the protein and separate protein subunits, and monomers, dimers, and possible trimers were detected. We conclude that the baculovirus expression system is capable of expressing NS1 that is in turn capable of forming into higher molecular weight multimers. Thus, this may be a useful system for producing recombinant NS1 protein for studies of dengue virulence and possibly as a vaccine candidate for the prevention of dengue viral disease.

Methods

Recombinant NS1 Production

The recombinant baculovirus AcGP67-N6His-DV2-NS1 (kindly provided by Mallory Hamilton) was constructed briefly as follows:

C6/36 *Aedes albopictus* mosquito cells were cultured and infected with DV2 virus (16681) inoculum at a multiplicity of infection of 0.01. Cell free supernatant was collected on day 14. Typical viral titers range from $1x10^7$ to $1x10^8$ per mL of cell-free supernatant.

Total RNA from DV2 (16681) was extracted using Trizol reagent (Invitrogen, Carlsbad, CA). CDNA was produced from the RNA extraction by RT-PCR using the following primers:

pNS1 (BamHI) Forward:

5' GGC CGG ATC CGA TAG TGG TTG CGT TGT GAG CTG G 3'

pNS1 (XhoI) Reverse:

5' GGC CGA GCT CAG CTG TGA CCA AGG AGT TGA CC 3'

The ~1000 base pair product was cloned into pCR2.1-TOPO, the sequence verified, and used as a template to subclone the DVII NS1 gene into the baculovirus secretion vector pAC-GP67 (B.D. Pharmingen, Franklin Lakes, NJ). An N-terminal hexahistidine epitope tag was included to facilitate purification.

Log phase growth Sf9 cells (Invitrogen, Carlsbad, CA) were co-transfected with linearized BaculoGold DNA (B.D Pharmingen, Franklin Lakes, NJ) and pAc-GP67-N-6His-Den2 NS1 according to manufacture's instructions. Recombinant baculovirus titer was amplified through successive passages in fresh Sf9 cells. Expression of recombinant DV2 NS1 protein was verified by Western blot using antibodies against either the 6His epitope or an anti-DV2 NS1 polyclonal antibody raised in rabbit.

Expression of Recombinant DV2 NS1 in T. ni cells

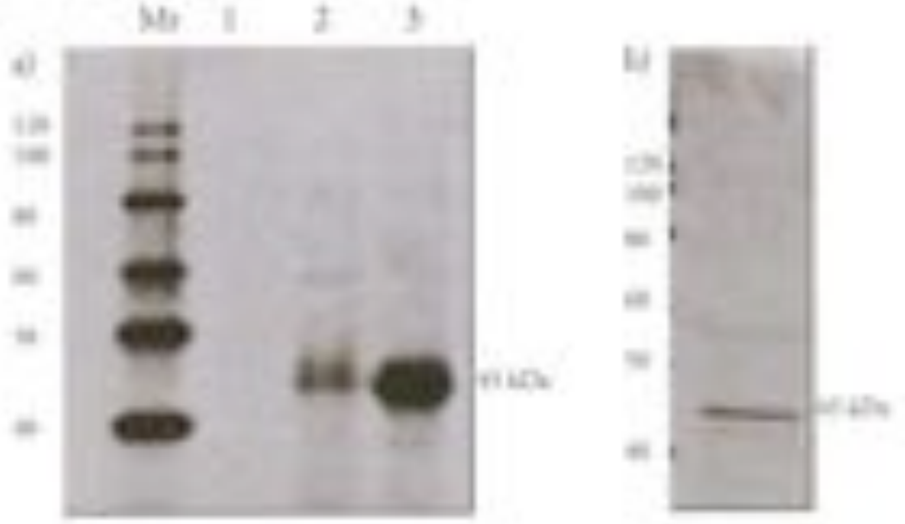
Log phase *T. ni* cells grown to $2.0x10^6$ cells/mL in a volume of 30 mL were infected at a dilution of 1:100 with recombinant pAcGP67-N-6His-DV2 NS1 baculovirus inoculum. This suspension was incubated for 48 hours at 28°C, shaking at 120 RPM. Cell-free supernatant was collected by centrifugation at 800 RPM for 10 minutes at 4°C. This supernatant was supplemented with EDTA-free protease inhibitors at 4°C.

Secreted NS1 was sequence verified, and presence in the cell-free supernatant was confirmed by SDS-PAGE Western blot; a 10% polyacrylamide gel with 4% polyacrylamide stacking gel was used. 60 µL of supernatant and 15 µL of 5xLaemmli Buffer with 5% beta- mercaptoethanol (14.3 M Omnipur), for a final concentration of 1%, were boiled at 100°C for 5 minutes before loading. The gel was electrophoresed at 150 volts for 30 minutes and then at 300 volts for 30 minutes in a gel box (Hoefer Scientific Instruments, San Francisco, California). The resolved protein was resolved onto a nitrocellulose membrane (Whatman, Dassel, Germany). The transblot was electrophoresed overnight at 25 volts in a transblot unit (BioRad). The nitrocellulose membrane was blocked in BSA immuno block solution (20 g/L BSA, 1xPBS, 0.2% Tween 20) for one hour, then incubated with polyclonal rabbit antibody RabD2 (provided by J. Schlesinger) at a dilution of 1:5000 in BSA immuno block solution or monoclonal mouse 6-His antibody at a dilution of 1:1000 in BSA immuno block solution for one hour. The RabD2 antibody was preabsorbed with 600 µL *T. ni* cell lysate and BSA-PBS without Tween. The membrane was washed 3 times for 5 minutes each in PBS wash buffer (1xPBS, 0.2% Tween 20). Secondary antibody of horseradish peroxidase conjugated-Donkey α rabbit or Sheep α mouse were added to the membrane at dilution of 1:20,000 or 1:15,000, respectively and incubated at room temperature for one hour. The membrane was again washed 3 times for 5 minutes each with wash buffer. The membrane was then

Figure 2: NS1 monomer presence in pAcGP67-6His-Den2-NS1 baculovirus infected *T. ni* whole cell fraction, insoluble cell fraction, and supernatant detected by Western blot with 6His epitope tag specific primary antibody. NS1 was probed with primary antibody at a dilution of 1:1000 and horseradish peroxidase-conjugated secondary antibody at a dilution of 1:15000.

(a) Lane 1: Uninfected *T. ni* cell supernatant. Lane 2: Whole cell fraction. Lane 3: Insoluble cell fraction. Cell fractions were treated with 1% BME. 200,000 cells were loaded in each lane. Mr is 7 µL of Magic Mark (Invitrogen, Carlsbad, CA) ladder.

(b) Cell free supernatant was collected from *T. ni* cells infected with recombinant pAcGP67-6His-Den2-NS1 baculovirus inoculum. 70 µL of supernatant was loaded.



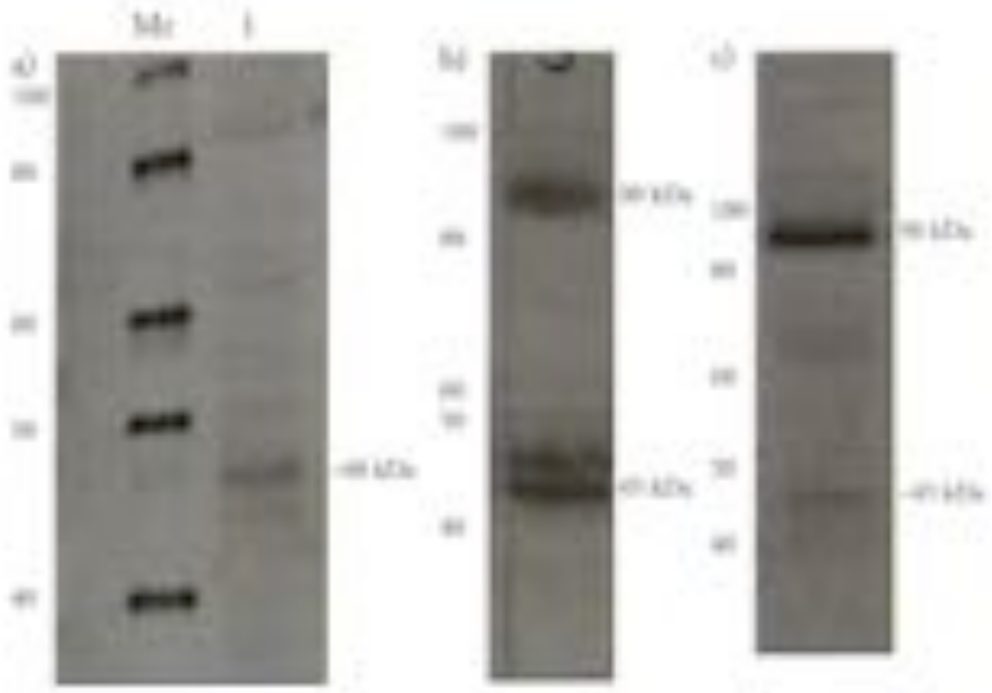


Figure 3: NS1 monomer and dimer in baculovirus infected *T. ni* supernatants with dilutions of BME detected by Western blot with Dengue NS1 specific RabD2 antibody at a dilution of 1:5000 and secondary antibody at 1:20000.
(a) Lane 1: 60 µL of supernatant from recombinant baculovirus infected T. ni cells treated with 1% BME. Mr is 7 µL of Magic Mark (Invitrogen).
(b) 60 µL of supernatant treated with 0.6% BME.
(c) 60 µL of supernatant treated with 0.4% BME.

incubated with 3 mL of each type of developing solution, 1 and 2 (GE Healthcare Little Chalfont, UK), for one minute and developed on Kodak film.

The supernatant was treated with dilutions of BME, and NS1 was detected by Western blot, as above. A whole cell fraction was collected from the cell pellet of the infected T. ni cells centrifuged at 800 RPM for 10 minutes at 4°C. Cell lysate buffer was added to the cell pellet at 1 mL/1x10^7 cells and mixed. An insoluble cell fraction was obtained by incubating the whole cell fraction for 15 minutes at room temperature, vortexing every 5 minutes. The lysed cells were centrifuged for 3 minutes at 2500 RPM. The supernatant was removed, and a volume of PBS was added to resuspend the pellet. Both fractions were prepared with a final concentration of 1xLaemmli Buffer and 1% BME. These samples were resolved on SDS-PAGE gels, transferred to nitrocellulose, and developed as above, with the exception of the dilution of 6His primary and Sheep α mouse secondary antibody, at 1:1500 and 1:10000 respectively.

Results

We sought to determine the complexity of NS1 structures present in the supernatant of T.ni cells infected with pAcGP67-NHis-DV2 NS1 recombinant baculovirus. First, the presence of NS1 in the supernatant was confirmed using whole cell and insoluble fractions of infected *T. ni* cells by Western blot with 6His epitope tag specific antibody. A ~45 kDa monomer was detected in the three fractions (Figure 2). NS1 was not seen in the uninfected cell supernatant, which indicates that NS1 is specific to dengue infected cell supernatant.

Infected cell supernatant was treated with dilutions of BME to determine the quaternary structures present. NS1 in supernatant treated with 1% BME was detected chiefly as a monomer at 45 kDa. Supernatant treated with 0.6% BME showed NS1 in both a 90 kDa dimer and 45kDa monomer with the dimer in greater amount. NS1 in the supernatant heated with 0.4% BME showed a 90 kDa dimer with a very small amount of 45 kDa monomer. A possible trimer was also detected in this dilution of BME; however, further study is required to confirm this. The monomer, dimer, and trimer were detected with dengue specific RabD2 antibody (Figure 3). The presence of higher molecular weight multimers of NS1 in the supernatant of recombinant baculovirus infected cells indicates this system expresses the protein in native multimeric forms. Variations in the molecular weight could be due to differing glycosylation of NS1.

The 45 kDa monomer of NS1 was detected in the supernatant of infected cells by Western blot with both dengue specific (RabD2) and 6His tag specific antibodies (Figure 4). A 90 kDa dimer of NS1 was detected at 0.6% BME with RabD2. However, the dimer was not detected by the 6His epitope specific antibody at the same dilution of BME. This indicates that in the homodimer structure, the amino terminus 6His epitope tag is masked.

Discussion

NS1 is seen in high levels in the blood of dengue infected patients. Studies showing its role in virus viability, such as this, have identified NS1 as an important contributor to disease as well as an indicator of infection.[1] NS1 is also a possible vaccine candidate against the dengue virus and perhaps other flaviviruses.[2] Thus, it is important to obtain the protein in its native form. Previous studies have shown that this native form includes dimers inside infected cells and secreted hexamers.[3,4] We have shown that recombinant baculovirus pAcGP67-NHis-DV2-NS1 infected cells are able to secrete NS1, which is detectable on a Western blot by dengue NS1 specific and 6His epitope tag specific antibodies. Also shown are monomer, dimer, and possible trimer forms of the protein detected in the supernatant after treatment with dilutions of BME with a dengue specific antibody. The dimer of NS1 was not clearly seen on the Western blot with 6His specific tag, which indicates that the 6His tag is masked in the dimer form; to use this tag for purification, it may need to be moved from the amino terminus to the carboxy terminus.

Future studies may focus on purification of secreted NS1. The purified protein may be used to study how it affects the virulence of dengue and other flaviviruses and how to possibly develop a vaccine. As such, it is important that the baculovirus expression system produces NS1 in a native form, which is able to form higher molecular weight monomers. The recombinant pAcGP67-NHis-DV2- NS1 is able to do this and is a candidate for producing NS1 for future studies.

Figure 4: NS1 monomer and possible trimer present in baculovirus infected T.ni cell supernatant detected by western blot with 6His epitope tag specific antibody at a dilution of 1:1500 and secondary antibody at 1:20000. No dimer is detectable.

References

1. Alcon-LePoder, S., Drouet, M.T., Roux, P., Frenkiel, M.P., Arborio, M., Durand-Schneider, A.M., Maurice, M., Le Blanc, I., Gruenberg, J., and Flamand, M., 2005. The secreted form of dengue virus nonstructural protein NS1 is endocytosed by hepatocytes and accumulates in late endosomes: implications for viral infectivity. Journal of Virology, Sept. 11403-11411.
2. Calvert, A.E., Huang, C.Y., Kinney, R.M., and Roehrig, J.T., 2006. Non-structural proteins of dengue 2 virus offer limited protection to interferon-deficient mice after dengue 2 virus challenge. Journal of General Virology. 87, 339-346.
3. Crooks, A.J., Lee, J.M., Easterbrook, L.M., Timofeev, A.V., and Stephenson, J.R., 1994. The NS1 protein of tick-borne encephalitis virus forms multimeric species upon secretion from the host cell. Journal of General Virology. 75, 3453-3460.
4. Flamand, M., Megret, F., Mathieu, M., Lepault, J., Rey, F.A., Deubel, V., 1999. Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in glycosylation-dependent fashion. Journal of Virology, July, 6104-6110.
5. Halstead, S.B., 2007. Dengue. Lancet. 370, 1644-1652.
6. Henchal, E.A., and Putnak, J.R., 1990. The Dengue Viruses. Clinical Microbiology Reviews, Oct., 376-396.
7. Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., and Straus, S.E., 2001. The Flaviviruses. Fields Virology 4th Edition, Chapter 32.
8. Lobigs, M., and Young, P.R., 1995. Maturation of the dengue-2 virus NS1 protein in insect cells: effects of downstream NS2A sequences on baculovirus-expressed gene constructs. Journal of General Virology,76, 979-984.
9. Mackenzie, J.M., Malcolm, K.J., and Young, P.R., 1996. Immunolocalization of dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. Virology, 220, 232-240.
10. Mackow, E., Makino, Y., Zhao, B.T., Zhang, Y.M., Markoff, L., Buckler-White, A., Guirakhoo, F., Chanock, R., and Lai, C.J., 1987. The nucleotide sequence of dengue type 4 virus: analysis of genes coding for nonstructural proteins. Virology, Aug, 217-228.
11. Price, D., Wilson, S., 2005. Dengue fever. Emedicine.com available at http://www.emedicine.com/emerg/topic124.htm.
12. Young, P.R., Hilditch, P.A., Bletchly, C., and Halloran, W., 2000. An antigen capture enzyme-linked immunosorbent assay reveals high levels of the dengue virus protein NS1 in the sera of infected patients. Journal of Clinical Microbiology. Mar., 1053-1057.

JUR

Explore your passions.

Contribute articles.

Join JUR.

sa.rochester.edu/jur

Evolution of Lake Basins in Northeast Tibet from Strontium Isotope Studies of Carbonates

Johanna Smith, 2008

Advised by Carmala Garzione, Ph.D.

Department of Earth and Environmental Sciences

The Tibetan Plateau is the Earth's largest geologic anomaly, composed of the highest mountain range on Earth with an average elevation of five km (16,400 feet). This vast high-elevation region sits north of the tectonic collision zone between Asia and the Indian subcontinent (Figure 2). The plateau consists of a series of small crustal blocks that were accreted to the southern margin of Asia.[1] India represents the most recent accretion event, colliding with Asia around 50 million years ago.[2] Although this region has a complex structural history related to each crustal block's accretion, the Indian collision and subsequent subduction under Tibet has caused the most recent deformation. This has raised some parts of the plateau from near sea level to their current elevations. Geologists continue to study this region to determine the processes that form broad, high, and uniformly elevated plateaus. One of the current issues of interest is how the margins of the plateau have evolved over the past 14 million years and what this tells us about surface uplift processes on the plateau.[3] This study is focused on the tectonic evolution of the northeastern margin of Tibet to explore the nature of deformation and basin formation over the past 12 million years.

This study employs strontium (Sr) isotope analysis of carbonates collected in sub-basins in northeast Tibet to determine whether sedimentary basins were formed in one large basin that was later segmented by the growth of large mountain ranges. Sedimentary deposits in this area indicate the presence of ancient lakes that existed roughly 7 to 12 million years ago (Ma). These paleolake deposits record information about the paleo-hydrology of the lake system and the sources of water and sediments to the lake basins. Lake deposits contain a Sr isotope fingerprint that is recorded in carbonate rocks formed throughout the lifespan of the lakes. This fingerprint is the ratio of the two major isotopes of Sr, ^{87}Sr, and ^{86}Sr. Sr is ideal for studying basin histories because the trace element Sr readily substitutes for calcium ions in carbonate rocks. This substitution occurs without any noticeable fractionation, or preference for one isotope over another. This means that the ratio of ^{87}Sr to ^{86}Sr in the water will be the same as the ratio of ^{87}Sr to ^{86}Sr in the precipitated carbonate.

If the source area feeding the lakes evolves over time, then the Sr isotope fingerprint of those source regions will also change. That change is recorded in the carbonate precipitated from the lake water. By comparing these fingerprints recorded in different paleolakes at the same time, we can better understand if the lakes occupied separate basins or if they were part of one large basin. If intervening mountain ranges rose in response to deformation from the collision of India and split a large basin into smaller sub-basins, then the Sr isotope fingerprint should also reflect this.

An application of this type was used for looking at the history of the Bonneville paleolake system whose modern analogue is now known as the Great Salt Lake.[4] During the Pleistocene (1.8 – 0.01 Ma), this was a vast basin that extended over 50,000 km, an area much larger than the present-day Great Salt Lake, which covers 4,400 km.[5] The study of its strontium isotope record helped to explain the evolution from

Figure 1: Author standing by the Yellow River, China in June 2006.

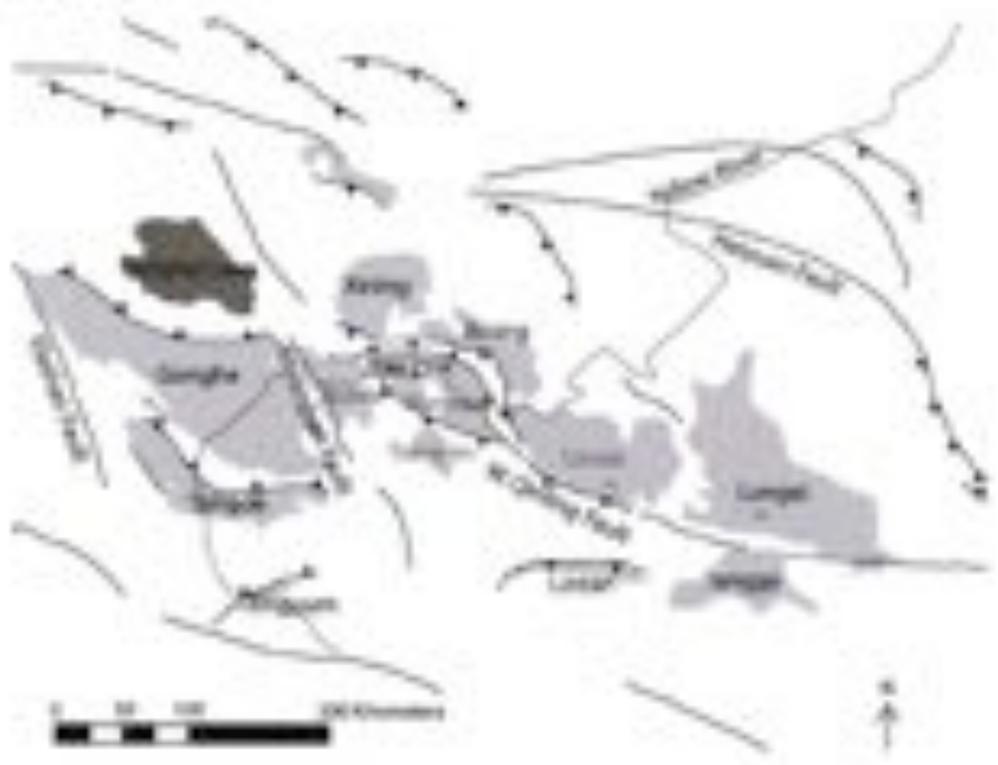
Figure 2: Map of China with study area depicted.

Lake Bonneville to the Great Salt Lake by interpreting changes in Sr isotopic compositions through time. Analysis of the Sr in carbonates from this lake showed an initial common Sr isotopic composition that diverged over time into two separate signatures. Hart *et al.* (2004) interpreted this pattern as a decreasing lake level which lowered to the height of an intra-basin ridge, splitting Bonneville into two separate lakes. One of these lakes eventually shallowed to nonexistence as the water level continued to drop. The other survives today as the Great Salt Lake. Similarly, our study hopes to show the evolution of paleolakes that existed in northeast Tibet in order to deduce whether or not they were initially linked. This will give us insight into the growth of mountain ranges on the northeast margin of Tibet that currently segment these basins.

Background Geology

Currently there are many small basins within northeast Tibet which are isolated by mountain ranges that rise 1 to 2 km above the basin floors. Several locations of similar ages have exposed sedimentary sequences containing lake deposits. These sequences have been dated using magnetostratigraphy, a technique based on periodic reversals of the magnetic poles recorded in the orientations of magnetic minerals in rocks.[5,6] By correlating the pattern of magnetic reversals in the sedimentary rock sequence with published precisely dated magnetic reversal patterns, the ages of the sedimentary rocks

Figure 3: Modern map of field area with basins in gray (lighter text denotes field sites).



can be inferred.[7]

The basins chosen for this study are Linxia ("lin-shah"), Xunhua ("shun-wah"), Tongren ("tong-wren"), and Guide ("gwe-dagh") (Figure 3). All of these basins have stratigraphic sections with lake deposits ranging from 7 to 12 Ma. If some of these basins were connected in the past as one large foreland basin, then we would expect the sub-basins to have the same isotopic signatures during this time period (Figure 4).

A foreland basin (Figure 5) forms in front of a growing mountain belt due to crustal thickening within the mountain belt. This thickened crust causes elastic deformation (bending) of the adjacent lithosphere.[8] The oldest documented fault motion in northeast Tibet comes from the West Qinling ("chin-ling") fault, which led to a linear mountain belt to the south of the Linxia and Xunhua basins.[9] Crustal thickening associated with this fault's motions might have been sufficient enough to cause the subsidence in the foreland.[7] If Linxia and Xunhua basins were part of a large integrated foreland basin, then their lake carbonate deposits would have produced uniform Sr isotopic compositions. As later ranges rose during further deformation in northeast Tibet, the Linxia and Xunhua basins would have been separated, and their Sr isotopic ratios would have diverged to reflect their present-day morphology (Figure 3).

Guide and Tongren basins, which are both south of the West Qinling fault, are piggy-back basins (Figure 5). Piggy-back basins form within zones of folding and faulting and are transported passively on active faults (such as the West Qinling fault) that are deforming beneath the piggy-back basins. If Guide and Tongren basins were unique piggy-back basins, then they should have unique isotopic compositions that reflect this separation.

Sr Isotopes

Sr is ideal for studying lake histories because the ratio of the isotopes is unique for each body of water and is clearly recorded in the lake's deposits of carbonates. Sr (atomic number 38) is a member of the alkaline earth metals, as is calcium. Although alkaline earth metals have some differences in electronegativity and atomic radius, they behave similarly due to their +2 oxidation, which makes them very reactive. Because of their reactivity, the alkaline metals precipitate from water as compounds such as $CaCO_3$ and $SrCO_3$ (carbonate rocks). Through hydrolysis and dissolution reactions, water and groundwater leach Sr from silicate minerals and dissolve

Figure 4: Map of hypothesized foreland and piggy-back basin model.

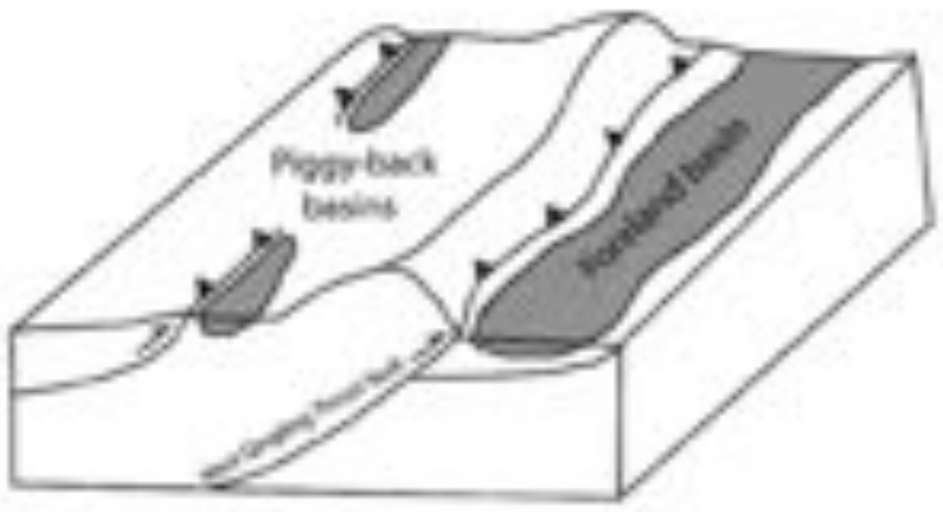


Figure 5: Foreland basin and piggy-back basins modeled in 3-D.

carbonate, which contains Sr as a trace element. This dissolved Sr is transported to lakes by rivers and groundwater.

The isotope ^{87}Sr is produced from the radioactive beta decay (loss of an electron) of rubidium (^{87}Rb), which leads to different ratios of the radiogenic ^{87}Sr to the stable ^{86}Sr isotope.[9] Source rocks in the drainage areas have varying ratios of radiogenic and stable Sr, and therefore the dissolved strontium derived from these regions reflects their different isotopic compositions. A lake will obtain its dissolved isotopic composition from the mixing of all the water flowing into it.

Methods

To analyze the $^{87}Sr/^{86}Sr$ recorded in carbonates, a small piece of the carbonate rock was separated. Because carbonates can be altered in the millions of years following original deposition, veins, diagenetic crystals, and weathered surfaces were avoided during sampling to remove the effects of alteration of the original strontium isotopic signature of the lake. The unaltered samples were dissolved in 0.5 M acetic acid solution. This extremely mild solvent was used so that the Sr from lake carbonate minerals could be dissolved out without leaching the Sr from other silicate minerals deposited in the lake.

Sr has a moderate solubility and a long residence (about 10^6 years) in freshwater and ocean basins compared to its mixing time (10^3 to 10^3 years). This long residence time allows for mixing of all the Sr sources to give a uniform signature for interconnected basins at one time. For example, in the oceans, Sr resides for approximately 4 to 5 million years but only takes 1000 years to mix. In lakes, these times are not as well defined due to varying rates of outflow. However, there is evidence from some of the basins in northeast Tibet indicating that the outflow was minimal, allowing the water to stay within the reservoir and mix thoroughly. Also, due to lower salinity, carbonates precipitate more slowly than in ocean water. Thus, we assume a similar relationship of long residence and rapid mixing time for these lakes.

Results and Discussion

Sixteen lake carbonate samples (five from Linxia basin, four from Tongren and Guide, and three from Xunhua) were analyzed for their $^{87}Sr/^{86}Sr$ composition. The samples were selected to span the approximately 5 million years of lake deposition. There are three main observations of the data (Figure 6), which can be interpreted to illustrate the connections between these basins. The most obvious is that the Guide basin deposits are very distinct, having a minimum separation of 0.003 $^{87}Sr/^{86}Sr$ from any of the other data points. This variation in the third decimal place indicates a significant difference in Sr isotopic composition, with a source of Sr for Guide that is either much older or much more felsic (silica rich). The second observation is that Tongren also has a distinct and increasingly more radiogenic signal. The oldest sample from Tongren has the lowest $^{87}Sr/^{86}Sr$ value, which suggests its waters sourced from either younger and/or more mafic (iron and magnesium rich) sources. Although Tongren lake deposits overlap the ranges of Linxia and Xunhua, the lake's slope shows a significantly different pattern that suggests changing sources over time. The last observation is that the Linxia and Xunhua samples all fall within the range of 0.710769 to 0.710990 and thus are within ±0.000121 of each other. They also share a very similar trend, especially in the older part of the record (11.5 to 8.5 Ma).

The interpretation of these observations is that the West Qinling fault was active by 12 Ma, separating Guide and Tongren from the basins on the north side of the fault. This is made clear from Guide's discrete Sr composition as well as Tongren's distinct trend. These basins' unique isotopic signatures and position to the south of the West Qinling fault

Figure 6: $^{87}Sr/^{86}Sr$ variations with time.

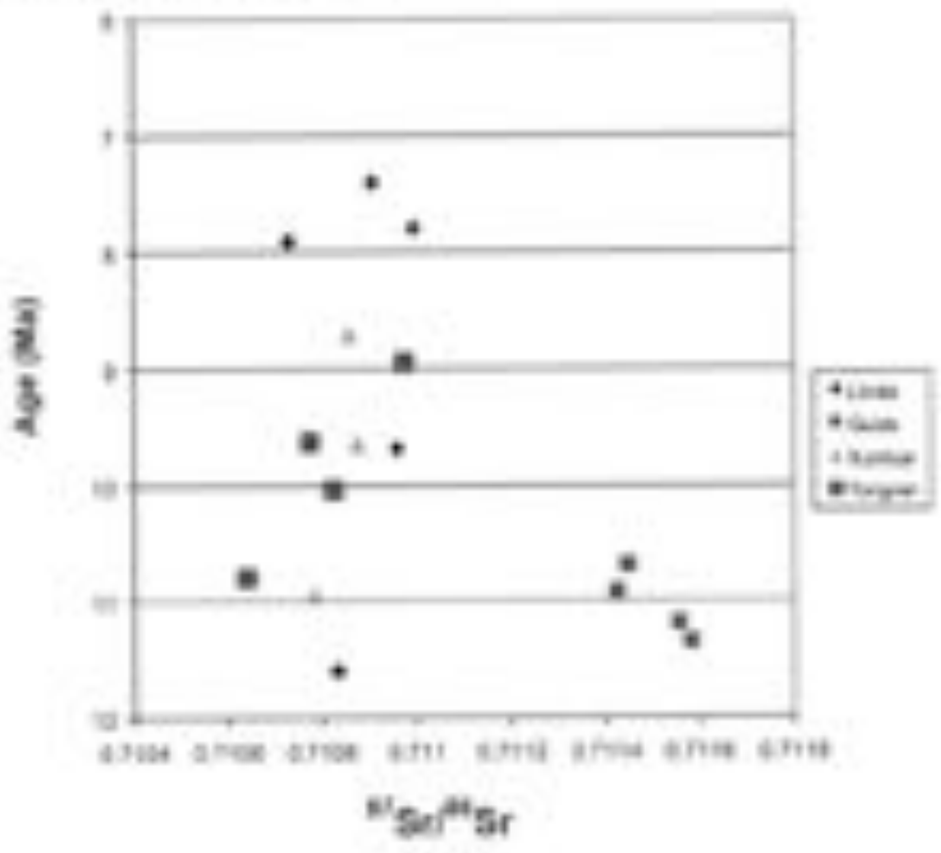


Figure 7: $^{87}Sr/^{86}Sr$ variations with time. Results of Linxia and Xunhua showing the invariant range of ±0.000027.

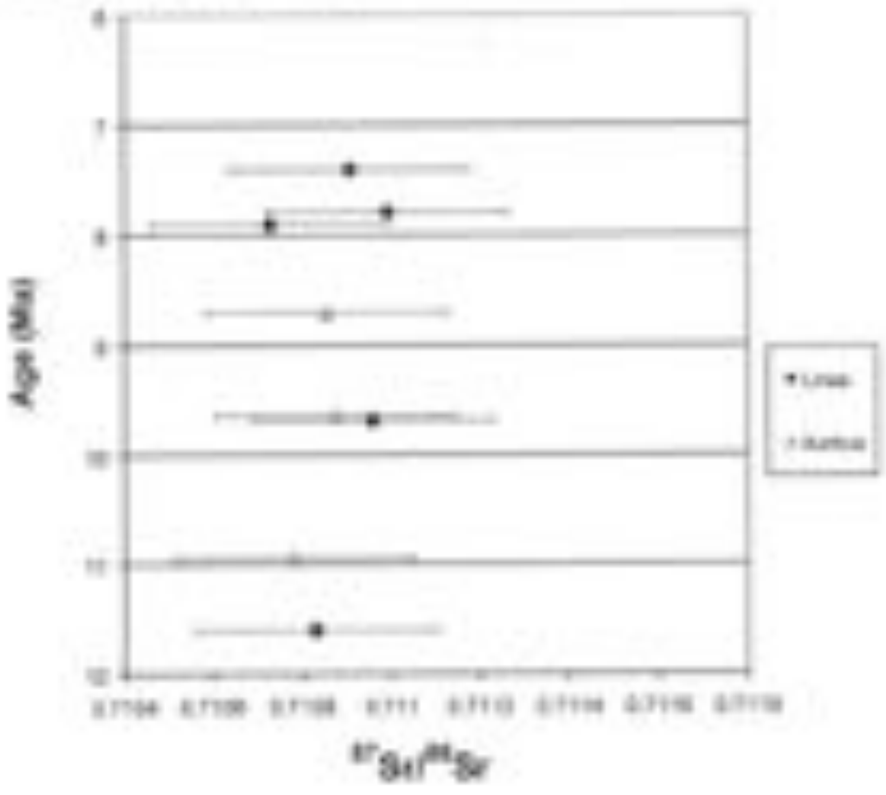


Figure 6. *Author standing at the top of the Guide deposits.*

define them as piggy-back basins.

During the comparable age range of 11.5 to 8.5 Ma, Linxia and Xunhua's similarity in range and trend does not preclude the possibility that they were part of the same basin. Carbonates that fall within a range of $^{87}Sr/^{86}Sr$ up to ±0.00027 are considered invariant (Hart *et al.* 2004), and therefore variability within this range is considered negligible. The variability in Linxia and Xunhua's record is less than half of this (Figure 7), and so it can be assumed that their signatures were derived from the same body of water. Furthermore, the proximity of these ancient lakes to the West Qinling fault is consistent with the idea that these basins were part of a large, foreland basin.

Sr isotopic evidence does not rule out a foreland basin setting for basins north of the West Qinling fault, but additional tests would be required to demonstrate the validity of this interpretation. Some uncertainty arises from the dating of the Xunhua record. Although there are tight magnetostratigraphic correlations for the Linxia and Guide basins, Xunhua's magnetostratigraphy is less well resolved. Therefore the ages could be off by several million years. If the Xunhua samples are younger than assumed for this study, their values would show overlap of the basins post-8.5 Ma. If they are older, then the records for Linxia and Xunhua may not overlap in time and thus would be incomparable. The integrated foreland basin hypothesis could be further tested with detailed mapping along the eastern and western flanks of the mountain range that separates the Xunhua and Linxia basins, as well as stable oxygen isotope ($\delta^{18}O$) studies of lake deposits to determine whether they reflect hydrologic connections between Xunhua and Linxia basins.

Conclusion

The study of strontium isotopes in northeast Tibet paleolakes has shown that the topography and geography changed over the last 12 million years and were different from the present configuration. Guide and Tongren basins show distinct Sr isotopic compositions from Xunhua and Linxia basins that are north of the West Qinling fault. The most likely explanation for this is that they were isolated basins. Linxia and Xunhua's records show striking similarity in the range of $^{87}Sr/^{86}Sr$, and therefore support the hypothesis that they were interconnected during the period of 11.5 to 8.5 Ma. The later records cannot be compared due to a lack of lake carbonates. Studies using stable oxygen isotopes ($\delta^{18}O$) and mapping along the boundary of the mountain range that separates Linxia and Xunhua basins may be able to further track the evolution and inevitable divergence of these basins to their present day separation.

The possibility that Linxia and Xunhua basins sourced a similar combination of rock types for Sr isotopes cannot be ruled out. Future work to clarify this issue would be to characterize both paleocurrents and the probable source areas for dissolved Sr to determine whether similar source areas can explain Sr isotopes in the basins through time. This would help differentiate similarities in their signatures that were caused by some similar sources in age and rock type from more precise similarities that indicate the basins were one and the same. Based on knowledge about the variability in Sr isotopes, the simplest interpretation remains that Linxia and Xunhua were part of a large, integrated foreland basin, and Guide and Tongren were separate, piggy-back basins. Their progression to the present day (Figures 3 and 4) has shown how the northeast Tibetan margin has evolved over the past 12 million years, and insinuates how this region will continue to change.

References

1. Yin and Harrison. "Cenozoic tectonic evolution of the Himalayan orogen as constrained by along-strike variation of structural geometry, exhumation history, and foreland sedimentation." Earth-Science Reviews. 59 (2000): 179-237.
2. Rowley, David. "Age of initiation of collision between India and Asia: a review of stratigraphic data." Earth and Planetary Science Letters. 145 (1996): 1-13.
3. Molnar, Peter and Joann M. Stock. "Slowing of India's Convergence with Eurasia at ~10 Ma and its Implications for Tibetan Mantle Dynamics." Tectonics. (In review).
4. Hart, William S. and Quade, Jay. "The $\delta^{18}O$ vs $^{87}Sr/^{86}Sr$ values of lacustrine carbonates and lake-level history of the Bonneville system." GSA Bulletin. 116 (2004): 1107-1119.
5. Fang et al. "Episodical subsidence by 29 Ma on the NE edge of Tibet from the magnetostratigraphy of Linxia Basin, China." Earth and planetary science letters. 210 (2003): 545-560.
6. Hough, Brian, C. Garzione, Z. Wang, R. Lease, D. Yuan, P. Zhang, D. Burbank, 2006. "Lithostratigraphic and isotopic correlations from the Xunhua and Linxia basins, NE Tibet: Implications for local climate variability." Eos Transactions AGU. 87(52). Fall Meeting Supplement. Abstract T33C-0532.
7. Butler, R.F. Paleomagnetism: Magnetic Domains to Geological Terranes. Boston: Blackwell, 1992.
8. Jordan, T. E. "Thrust loads and foreland basin evolution, Cretaceous, western United States." American Association of Petroleum Geologists Bulletin. 65 (1981): 2506-2520.
9. Clark, Marin. "Exhumation and plateau growth in northeastern Tibet." Eos Transactions AGU. 85 (2004): Abstract 0096-3941.
10. Wicander, John D. An Introduction to Isotopes and Metamorphic Petrology. New Jersey: Prentice Hall, 2000.

Counting Prime Paths in Fractals Built from Triangles

Elizabeth Munch, 2008

Advised by C. Douglas Haessing, Ph.D.

Department of Mathematics

Imagine running downtown with a set distance to go, but looking for an alternative path. Instead of turning back and taking the same road or merely repeating the same loop multiple times, how many other options are there? This situation could be modeled with graph theory, where each intersection is a vertex, and each road is an edge. The Ihara zeta function counts these special paths, which we call prime paths. Given a finite graph, G, the number of prime paths of a specific length k is finite. Finding the number of prime paths for large k can be immensely difficult. With the use of a generating function like the Ihara zeta function, where the coefficients are the solution to this counting problem, the original counting problem is easier to study by real, complex, and even p-adic analysis.

We examine the problem of counting prime paths in a family of fractals built out of triangles called the Sierpiński gasket. Since these figures are built from three-sided figures, we take the novel approach of examining our generating function using 3-adic analysis. The surprising and somewhat paradoxical result of this angle of examination is that as the fractals increase in size, the counting problem stabilizes to finite numbers using 3-adic analysis even though it is infinite using real analysis. As it turns out, this 3-adic viewpoint isolates counting the number of prime paths of length k through the three special intersection points on each fractal in the Sierpiński gasket, which is why the counting problem stabilizes using 3-adic analysis.

The Ihara zeta function was first introduced in the 1960s by Yasutaka Ihara. In its original form, it is an infinite product, given by

$$\zeta(u, G) = \prod(1 - u^{v(P)})^{-1}$$

where the product runs over the set of equivalence classes of prime paths in the graph. Bass proved that this function is in fact a rational polynomial. Given by the data obtainable from the graph including the adjacency matrix, the degree of each vertex and the sizes of the edge and vertex sets off the graph. This now gives us a simple way to count the number of prime paths of a large length k without needing to do it by hand. We also have the ability to compare graphs through their respective zeta functions. For example, Storm looked at the comparison of clique numbers, the counting of Hamiltonian cycles, and also determined if a graph is perfect or chordal.[1] Although for this, he utilized the edge based equivalent of the Ihara zeta function which was introduced by Stark and Terras.

The original Ihara zeta function is not the only way to count prime cycles. The function has a very useful property in that the logarithmic derivative,

$$u\frac{d}{du}\log\{\zeta(u, G)\},$$

can give us a different method of counting. For our purposes, this seems to be the more fruitful version of the equation to study in terms of the 3-adic convergence of the coefficients. However, it is possible that this study can be extended to the original function with the same consequences.

Here we will look at two aspects of the Ihara zeta function. We will look at the zeta function of a fractal, known as the Sierpiński gasket. We have written a procedure for Maple, a computer program for mathematics, which will create the n^{th} iteration of the gasket, included in the appendix. Although the Ihara zeta function is only defined for a finite graph, we will approximate it using a limiting approach of graphs and ultimately show that it converges 3-adically. This is not the first time a fractal graph has been studied from the point of view of the Ihara zeta function. However, where Guido *et al.* have normalized the zeta function so that it converges on the complex plane, we will look at the zeta function from the p-adic point of view, which will not necessitate any change of the function.

The Ihara zeta function of a graph

We must begin with some preliminary work in graph theory in order to define the function. First, we will assume that all graphs we are working with are undirected, leafless, loopless, finite (for the time being), and connected. Then, given G, a graph with vertex set $V(G)$ and edge set $E(G)$, a path $P=a_1a_2...a_n$ where a_j is an element of $E(G)$, has a **backtrack** if $a_{j+1}=a_j^{-1}$ for any j in $\{1,...,n-1\}$, as in Figure 1a. That is, it crosses the same edge twice in a row. The graph has a **tail** if $a_1=a_n^{-1}$, as in Figure 1b, where the ending edge and beginning edge are the same.

Now we define the prime paths which the Ihara zeta function counts. Let $C=a_1...a_n$ be a closed path in the graph G. We will notate repeating a path C f-times by C^f. Following path C and then path D is notated CD. Then C is a prime

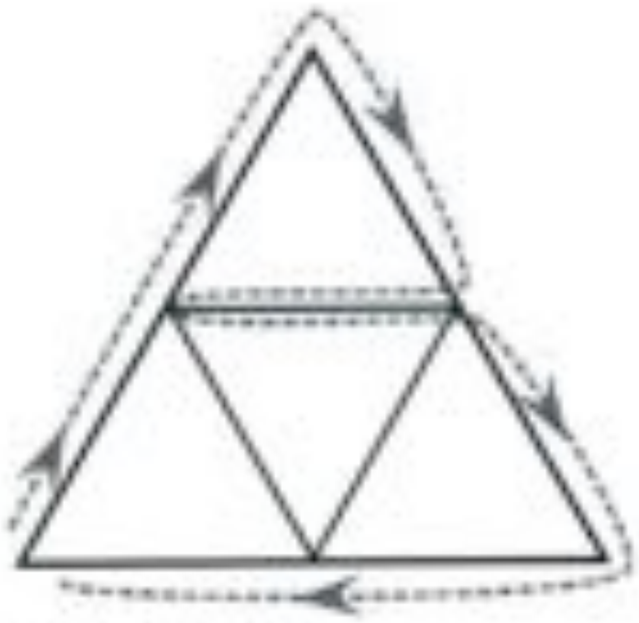


Figure 1a. A path with a backtrack

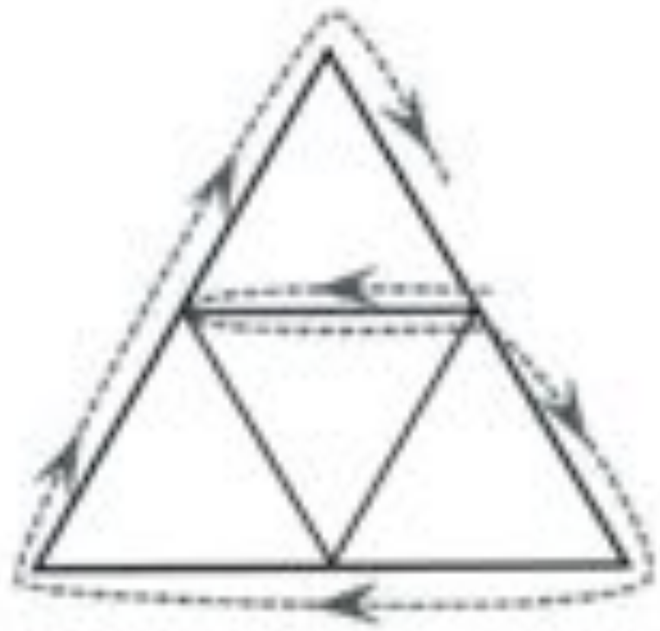


Figure 1b. A path with a tail

path, if it has no backtracks or tail and C is not equal to D^i for some closed path D. We can see by the definition that this still allows for repetitions of edges in G, so long as the path is not a repetition of one shorter closed path. As such, there are an infinite number of prime paths possible in any graph that has a closed path and which is not a cycle. We are also given an infinite number of options for new prime paths given two or more other prime paths; this occurs because, though D^i cannot be considered a prime path, CD is, as well as CD^2, and so forth.

We can also place an equivalence relation on this set of prime paths in the graph. Two closed paths are equivalent if they are the same path with a different starting point. Note however, that this means that the path $P=a_1a_2...a_{n-1}a_n$ is not equivalent to the same path traversed backwards, $P^{-1}=a_na_{n-1}...a_2a_1$. It follows that the equivalence class of C, denoted $[C]$, is the set $[C]=\{a_1a_2...a_{n-1}a_n, a_2...a_{n-1}a_na_1, ..., a_na_1...a_{n-1}\}$ which is the prime path started at each point in the path.

The Ihara Zeta Function is

$$\zeta(u,G)=\prod_{[P],\ \text{prime}}(1-u^{w(P)})^{-1} \tag{1}$$

with $w(P)$ as the number of vertices in $[P]$, and where the product runs over all prime equivalence classes. In this way we have an infinite product to work with. Bass proved that this infinite product is in fact a rational function. Let A be the adjacency matrix of G. Since our graphs have no loops or multiple edges, this is the $|V(G)|x|V(G)|$ matrix such that for $v_i,v_j\in V(G)$,

$$a_{ij}=\begin{cases}1, & \text{if } v_i \text{ is adjacent to } v_j\\ 0, & \text{else}\end{cases}$$

The degree of a vertex, notated $\deg(v_i)$, is the number of edges attached to that vertex. Let Q be the diagonal matrix such that entry $q_{ii}=\deg(v_i)-1$. Finally, let $r=1-|E(G)|+|V(G)|$ be the Euler characteristic. Then Bass proved that the Ihara zeta function satisfies

$$\zeta(u,G)^{-1}=(1-u^2)^{r-1}\det(I-Au+Qu^2). \tag{2}$$

We will use this exact equation as the Ihara zeta function for a self-similar graph, specifically the Sierpiński gasket.

In order to solve the counting problem presented, we have two options. We can use the coefficients of the Ihara zeta function, or we can use the coefficients of the logarithmic derivative of the Ihara zeta function. The method of counting intrinsically contained in each option is slightly different, but for our current purposes, the results will be easier to see in the logarithmic derivative.

Our first step is to understand precisely how the logarithmic derivative counts these prime cycles.

We begin by taking the log to obtain

$$\begin{aligned}\log\zeta(u,G)&=\log\left(\prod(1-u^{w(P)})^{-1}\right)\\&=-\sum_{[P]}\log(1-u^{w(P)})\end{aligned}$$

then the derivative and multiply by u to get

$$\begin{aligned}&u\frac{d}{du}\left[-\sum_{[P]}\log(1-u^{w(P)})\right]\\&=-u\sum_{[P]}\frac{1}{1-u^{w(P)}}(-w(P))u^{w(P)-1}\\&=u\sum_{[P]}\frac{(w(P))u^{w(P)-1}}{1-u^{w(P)}}\\&=\sum_{[P]}\frac{(w(P))u^{w(P)}}{1-u^{w(P)}}.\end{aligned}$$

Using the geometric series identity,

$$\sum_{n=0}^{\infty}u^n=\frac{1}{1-u},$$

for $|u|<1$, we have

$$\begin{aligned}&u\frac{d}{du}\log[\zeta(u,G)]\\&=\sum_{[P]}w(P)u^{w(P)}(1+u^{w(P)}+u^{2w(P)}+...)\\&=\sum_{[P]}w(P)(u^{w(P)}+u^{2w(P)}+u^{3w(P)}+...)\end{aligned}$$

If we let $M_k=\sum w(P)$, where the sum runs over all $[P]$ such that $w(P)$ divides k then we have

$$u\frac{d}{du}\log(\zeta(u,G))=\sum_{k=1}^{\infty}M_ku^k \tag{3}$$

Therefore, the coefficient of the term u^k will count the number

of prime paths with a number of vertices which divides k. This is extremely useful if we look at a prime k but will be more difficult to interpret for composite k.

The Sierpiński Gasket

Figure 2. The Sierpiński gasket

The Sierpiński gasket (Figure 2) is a standard, recognizable fractal. We will look at a graph theory analog of this fractal. It is constructed by creating three copies of the triangle C_3 and attaching them at the vertices. We will call this graph S_1. Then, we take S_1, create three copies of it, and attach them at three vertices which have degree 2, thus yielding S_2. This process can continue on to infinity, which is our fractal. Notice that except for exactly three vertices on each graph, every vertex has degree 4. We use this fact to create a program for Maple which creates each iteration of the gasket. This program has been included in the appendix.

We further define the concept of a fractal graph as defined by Guido *et al.*[2] A subgraph is a subset of vertices of a graph along with any edge which is adjacent to two vertices in the set of vertices. The ***frontier*** of a subgraph K of graph G, denoted by $F(K)$, is the set of vertices in $V(K)$ such that any vertex in the set is of distance 1 in graph G from some vertex in the set $V(G)$-$V(K)$. For an example, see Figure 3.

Consider the countably infinite graph G which has a bounded degree. We say that this graph is amenable if it has an amenable exhaustion. This means that it has a family of subgraphs $\{K_n\}$ such that $X = \bigcup K_n$ and

$$\lim_{n\to\infty} \frac{|F(K_n)|}{|V(K_n)|} = 0. \qquad (4)$$

Let us apply this concept to the Sierpiński gasket. In creating the gasket, we copied previous iterations and attached them at the points. Let K be the set of subgraphs which are isomorphic to some S_m, ordered by increasing vertex set size. By looking at any of these graphs, we see that $|F(K_n)|\le 3$. Since we have ordered them to increase as n goes to infinity, equation (4) holds.

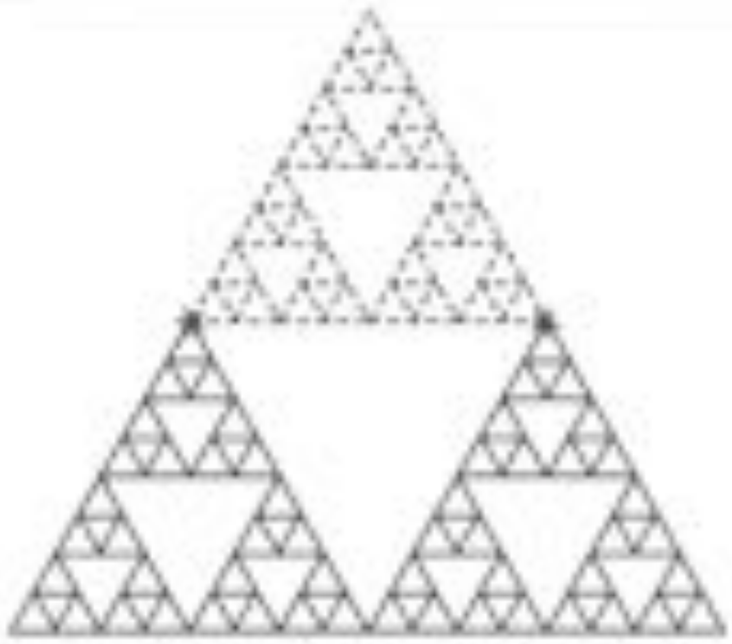


Figure 3. The vertices in the red stars are the frontier of the blue subgraph

Gaido *et al.* used these two concepts along with a set of isomorphisms within the graph to define the ***self-similar graph***. The idea is that K_{n+1} is created from a finite union of copies of K_n which have only the frontier in common. As we created the Sierpiński gasket, it is a self-similar graph.

Returning to the zeta function, we must look at the zeta function of the iterations to calculate the zeta function of the Sierpiński gasket. We begin by using Maple to calculate some examples of the zeta function. If we look at the Taylor approximation about 0 for the first through the fifth approximations of the gasket, we see that

$\zeta(u, S_1) = 1 + 8u^2 + 6u^3 + 6u^4 + 56u^5 + 72u^6 + 111u^7 + 432u^8 + O(u^{10})$

$\zeta(u, S_2) = 1 + 24u^2 + 18u^3 + 18u^4 + 374u^5 + 546u^6 + 963u^7 + 5690u^8 + O(u^{10})$

$\zeta(u, S_3) = 1 + 72u^2 + 54u^3 + 54u^4 + 2862u^5 + 4254u^6 + 6609u^7 + 87966u^8 + O(u^{10})$

$\zeta(u, S_4) = 1 + 216u^2 + 162u^3 + 162u^4 + 24150u^5 + 36114u^6 + 52093u^7 + 1894026u^8 + O(u^{10})$

$\zeta(u, S_5) = 1 + 648u^2 + 486u^3 + 486u^4 + 212430u^5 + 318318u^6 + 445017u^7 + 47223822u^8 + O(u^{10})$

However, for this paper we will mostly be dealing with the logarithmic derivative; we use Maple to calculate this as well. The first few functions are

$u\,(d/du)\log(\zeta(u,S_1)) = 24u^2 + 24u^3 + 30u^4 + 144u^5 + 168u^6 + 360u^7 + 1068u^8 + O(u^{10})$

$u\,(d/du)\log(\zeta(u,S_2)) = 72u^2 + 72u^3 + 90u^4 + 516u^5 + 798u^6 + 2952u^7 + 8982u^8 + O(u^{10})$

$u\,(d/du)\log(\zeta(u,S_3)) = 216u^2 + 216u^3 + 270u^4 + 1620u^5 + 2562u^6 + 10104u^7 + 30618u^8 + O(u^{10})$

$u\,(d/du)\log(\zeta(u,S_4)) = 648u^2 + 648u^3 + 810u^4 + 4952u^5 + 7854u^6 + 31560u^7 + 95526u^8 + O(u^{10})$

$u\,(d/du)\log(\zeta(u,S_5)) = 1944u^2 + 1944u^3 + 2430u^4 + 14868u^5 + 23730u^6 + 95928u^7 + 290250u^8 + O(u^{10})$

Recall that

$$u\frac{d}{du}\log\left(\zeta(u,G)\right) = \sum_{k=1}^{\infty} M_k u^k$$

where $M_k = \Sigma v(P)$ and the sum runs over all $[P]$ such that $v(P)$ divides k. Let $M_k^{(m)}$ be the k^{th} coefficient of $u\frac{d}{du}\zeta(u, S_m)$. If we define $|P_j^{(m)}|$ to be the number of prime paths of length j in S_m, then

$$M_k^{(m)} = \sum_{j|k} j\left|P_j^{(m)}\right|$$

First, let us look at paths of length 3. We can see that $M_3^{(m)}$ is the same as the number of equivalence classes of triangles in the graph S_m times 3. If we look at S_1, we can see that there are four triangles, and thus there are eight equivalence classes, and as a result, we have $M_3^{(1)} = 8 * 3 = 24$. In S_2, we have taken these eight equivalence classes and repeated them three times, since they appear in each copy of S_1. Therefore, $M_3^{(2)} = 3(M_3^{(1)}) = 3 * 24 = 72$. Likewise, $M_3^{(3)} = 3(M_3^{(2)}) = 216$. Because of the construction of the Sierpiński gasket, we will always have that $M_3^{(m)} = 3\,(M_3^{(m-1)}) = 8 * 3^m$.

$M_3^{(m)} = 8 * 3^m$ goes to infinity as m gets large, so we will

Figure 4. How one path multiplies 3 times in each subsequent iteration.

need to look elsewhere for convergence. We turn to the *p*-adic numbers, which are a different way of completing the rational numbers. A set of numbers is complete if every Cauchy sequence converges to a number which is also in the set. For example, we can complete the rational numbers with the irrational numbers to create the real numbers. This uses the normal absolute value as its metric. So long as *p* is a prime number, we can also complete the rational numbers to create the *p*-adic numbers. But for the case of the *p*-adic numbers, we will choose a different metric.

We start with the notion that any number can be represented as $\sum a_i p_i$ with a_i in $\{0,1,2,\ldots,p-1\}$. We can now define a new metric for these numbers. Given *p* prime, let $|x|_p = 1/p^n$ where $x=np^n$ and $\gcd(n,p)=1$. For example, if we are looking at the 3-adic numbers, let $x = 20/21 = (2^2 * 5)(3 * 7)$. Then $|x|_3 = 3$, $|x|_2 = 1/4$, $|x|_5 = 1/5$, and $|x|_7=7$.

We can also apply the principles of the geometric series to these numbers. In the real numbers,

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots$$

as long as $|x|<1$. We lift this notion to the *p*-adics. Here, the series is convergent as long as $|x|_p < 1$. With the new metric, these are convergent for numbers highly divisible by *p*. In the 3-adic numbers, $1 + 3 + 3^2 2 + \ldots = 1/(1-3) = -1/2$.

Now that we have defined the self-similar graph and the *p*-adic numbers, we can define the Ihara zeta function for the Sierpiński gasket graph. Our original goal was to change the function as little as possible, so we say simply that if

$$\zeta(u, S_\infty) = 1 + a_3^{(\infty)} u^3 + a_4^{(\infty)} u^4 + a_5^{(\infty)} u^5 \cdots$$

then

$$\zeta(u, S) = 1 + A_3 u^3 + A_4 u^4 + A_5 u^5 \cdots$$

where

$$A_k = \lim_{m\to\infty} a_k^{(m)}$$

in the 3-adic numbers. With this definition, we can also see that the coefficients of $u\frac{d}{du}\log(\zeta(u, S_m))$ will also be the limit of $u\frac{d}{du}\log(\zeta(u, S_m))$ as *m* goes to infinity. Our next step is to look at some of the coefficients of $u\frac{d}{du}\log(\zeta(u, S_m))$. In our previous example, we showed that $M_3^{(m)} = 8 * 3^m$. This then converges to 0 in the 3-adic numbers since

$$\lim_{m\to\infty} |8\cdot 3^m|_3 = \lim_{m\to\infty} \frac{1}{3^m} = 0$$

Figure 5. Prime paths of length 6 in S_3.

and thus the coefficient of u^3 in $u\frac{d}{du}\log(\zeta(u, S))$ is 0.

Similar to $k=3$, $k=4$ and $k=5$ are of the form $M_k^{(m)} = 3(M_k^{(m-1)})$, also going to zero. However, the case where $k=6$ is far more interesting. Because the minimum size of a prime path is 3,

$$M_6^{(m)} = 3\cdot\left|P_3^{(m)}\right| + 6\cdot\left|P_6^{(m)}\right|.$$

To calculate A_6, we first look at S_3. We already know that there are eight prime paths of length 3. Now we look at prime paths of length 6. These will either be made up of two triangles, or will be C_3, the cycle of length 6. Each of these paths can be seen in Figure 5. Remember that each path can be followed in two directions, so we have 20 equivalence classes. Then

$$M_6^{(1)} = 3\cdot 8 + 6\cdot(20) = 144$$

and we can return to the calculations from Maple to see that our counting is indeed correct.

We move on to S_4. As we saw previously, there are $8*3 = 24$ prime paths of length 3. As to paths of length 6, we have three copies of each of the 20 we had in S_3, two from the cycle C_3 in the large inner triangle, and then 12 completely new ones which were not counted in S_3. These are the prime paths which are comprised of two triangles where each are in a separate copy of S_1 which was used to build S_2 and go through the vertex which connects the two copies in S_2. Then we have

$$M_6^{(2)} = 3(3\cdot 8) + 6(3\cdot 20 + 2 + 4\cdot 3) = 516.$$

Next we look at S_5. Again, we have three times as many prime paths of length 3 as in S_4. There will also be three copies of each prime path of length 6 from S_4, but this time we only have 12 new paths to count from the paths which cross two copies of S_4 since the inner triangle is now large enough that a path of length 6 will not go around it. Thus we have

$$M_6^{(3)} = 3[3^2\cdot 8] + 6[3(74) + 12] = 1620.$$

We have reached an extremely important turning point. Now that we have a large enough *m*, no new paths will appear which encircle the center triangle. This means that of the prime paths of length 6, given $m \geq 3$, they can be categorized as either copies of paths found in S_{m-1}, or are one of exactly 12 new paths, no matter what *m* is. Thus

$$|P_6^{(m)}| = 3|P_6^{(m-1)}| + 12\,.$$

S_m will have three times the number of prime paths of length 6 in S_{m-1} plus the 12 additional ones created between the copies of S_{m-1}. Performing simple algebra, and noting that the number of prime paths of length 6 in S_{m-1} is three times the number of prime paths of length 6 in S_{m-2} plus 12, we calculate

$$M_6^{(m)} = 3M_6^{(m-1)} + 72$$

Thus as m gets very large, $M_6^{(m)}$ will converge 3-adically to -36.

$$\begin{aligned}
M_6^{(2)} &= 3M_6^{(1)} + 72 \\
M_6^{(3)} &= 3\left(3M_6^{(1)} + 72\right) + 72 \\
M_6^{(4)} &= 3\left(3\left(3M_6^{(1)} + 72\right) + 72\right) + 72 \\
M_6^{(m)} &= 3^{m-1}M_6^{(1)} + 3^{m-2}\cdot 72 + 3^{m-3}\cdot 72 + \cdots + 72 \\
&= 3^{m-1}M_6^{(1)} + (3^{m-2} + 3^{m-3} + \cdots + 1)\cdot 72
\end{aligned}$$

Recalling that $1 + 3 + 3^2 + \ldots = 1 / (1-3)$ in the 3-adics, and that $\lim_{m\to\infty} 3^{m-1}M_6^{(1)}$, we conclude that

$$\lim_{m\to\infty} M_6^{(m)} = -\frac{1}{2}144 = -72$$

Lemma 1. *Let L_k be the number of prime paths of length k in S_m which are contained in two copies of S_{m-1} used to create S_m. Given that k and m such that $k < \min\left(2^m + 6, 3\cdot 2^{m-1}\right)$, L_k is a constant which is independent of m.*

Proof. What needs to be shown is that L_k is not based on S_m for m large enough or, more specifically, that a prime path of length k in S_m cannot be contained in all three copies of S_{m-1}. Suppose we are looking at graph S_m and k, such that k is less than the minimum of 2^m+6 and $3*2^{m-1}$. There are two ways in which a prime path could be contained in all three copies. First, it could go around the large, inner triangle. This triangle has length $3(2^{m-1})$, since each side has length 2^{m-1}. The second is that it could be long enough to cross one side of the large inner triangle twice, with at least a triangle on each end contained in the other two copies. This can be seen in Figure 6. This path has length $2(2^{m-1}) + 2(3) = 2^m+6$. Thus, if a prime path has a length less than both of these, it can be contained in at most two copies.

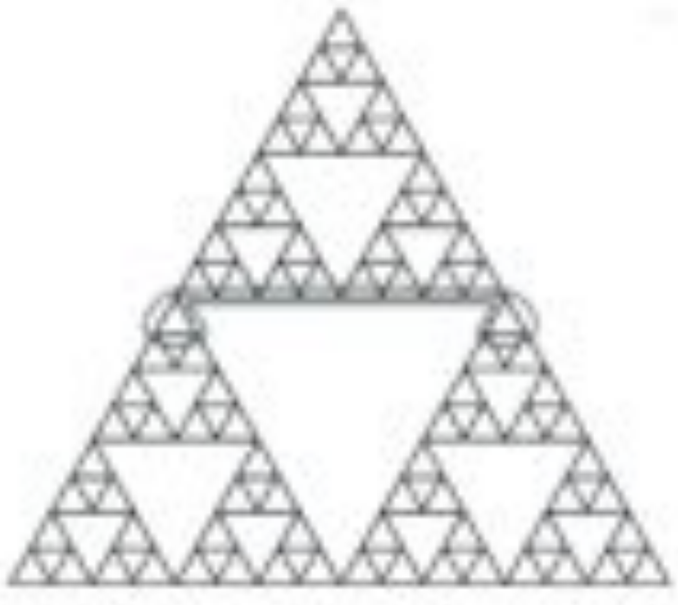


Figure 6. The shortest path in the Sierpinski gasket which has edges in all three copies of the previous iteration.

For a large enough m, each subsequent iteration will have a set number of new paths of length k which were not contained entirely in the previous iteration. Since locally, the area around the vertices where the copies meet will look exactly the same, there will be a set number of prime paths which go through that meeting vertex and are contained in both copies. Therefore, L_k is a constant which is not based on the graph S_m once m is large enough.

Lemma 2. *Let $|P_k^{(m)}| = |\{[P] : s([P]) = k, [P] \text{ in } S_m\}|$. If we fix* k, *then $\lim_{m\to\infty} |P_k^{(m)}|$ exists 3-adically and is equal to $-1/2$ L_k.*

Proof. The proof follows from Lemma 1. Note that once we reach m large enough, $|P_k^{(m)}| = 3|P_k^{(m-1)}| + L_k$ since any prime path in S_m can seen as either a copy of a prime path from S_{m-1} or is contained in two copies of S_{m-1}. Let μ be the largest integer such that $k \geq \min\{2^\mu + 6, 3\cdot(2^{\mu-1})\}$. Then

$$\begin{aligned}
|P_k^{(\mu+1)}| &= 3|P_k^{(\mu)}| + L_k \\
|P_k^{(\mu+2)}| &= 3^2|P_k^{(\mu)}| + 3L_k + L_k \\
|P_k^{(\mu+3)}| &= 3^3|P_k^{(\mu)}| + 3^2L_k + 3L_k + L_k \\
&\vdots \\
|P_k^{(\mu+n)}| &= 3^n|P_k^{(\mu)}| + L_k(3^{n-1} + 3^{n-2} + \cdots + 1)
\end{aligned}$$

Thus

$$\begin{aligned}
\lim_{n\to\infty} |P_k^{(m)}| &= \lim_{n\to\infty} |P_k^{(\mu+n)}| \\
&= \lim_{n\to\infty} 3^n|P_k^{(\mu)}| + L_k(3^{n-1} + 3^{n-2} + \cdots + 1) \\
&= \frac{-1}{2}L_k
\end{aligned}$$

So in the 3-adic numbers,

$$\lim_{m\to\infty} |P_k^{(m)}| = -\frac{L_k}{2}.$$

Theorem 1. *Let $\{n_1, n_2, \ldots, n_i\}$ be the divisors of* k, *where n_i = k, and let $M_k^{(m)}$ be the coefficient of u^k in the logarithmic derivative of the Ihara zeta function of the m^{th} iteration of the Sierpinski gasket. Then $\lim_{m\to\infty} M_k^{(m)}$ exists 3-adically and equals*

$$\lim_{m\to\infty} M_k^{(m)} = \sum_{i=1}^{i} n_i L_{n_i}.$$

Proof. From Lemma 1, we know that L_k is a constant for a large enough m. From Lemma 2, we know that $|P_k^{(m)}| = 3|P_k^{(m-1)}| + L_k$ for a large enough m. Recall that $M_k = \sum s(P)$, where the sum runs over all $[P]$ such that $s(P) \mid k$. We can expand the definition of $M_k^{(m)}$ to see that

$$M_k^{(m)} = \sum_{i=1}^{i} n_i |P_{n_i}^{(m)}|$$

Since we know that $\lim_{n\to\infty} |P_n^{(n)}| = -\frac{1}{2}L_n$, we know that 3-adically,

$$\lim_{n\to+\infty} M_k^{(n)} = \sum_{i=1}^{k} -\frac{n_i}{2} L_{n_i}$$

Now we can look at the Ihara zeta function itself, rather than at the logarithmic derivative. Consider the Ihara zeta function of S_n.

$$\zeta(u, S_m) = \prod_{[P]} \frac{1}{1-u^{v(P)}} = \prod_{[P]} (1 + u^{v(P)} + \cdots)$$

$$= \sum_{D\geq 0} u^{v(D)} = \sum_{k=0}^{\infty} N_k^{(n)} u^k$$

where the first sum is over all finite formal sums, $D = \Sigma\, n_i [P_i]$, with n_i nonnegative integers. Then if we define $v(D) = \Sigma\, n_i\, v(P_i)$, we have $N_k^{(n)} = |\{D \geq 0 \mid v(D) = k\}|$.

N_k counts all possible combinations of the form $n_1\, v(P_1) + n_2 v(P_2) + \ldots = k$, whether the P_i are connected or not. Rather, we do not count prime paths of length k, but the number of ways in which we can combine the lengths of the shorter paths to add up to k. It should also be noted that there are a finite number of prime paths of any length which are not comprised of a combination of shorter prime paths. Furthermore, paths of length 3, 4, or 5 cannot be created from a combination of shorter prime paths. The simplest example to look at in the Sierpiński gaskets are the $N_k^{(n)}$. As shown previously, there are eight equivalence classes $[P]$ of length 3 in S_n. Since for $k = 3$, we can only directly count equivalence classes, we have $N_3^{(1)} = 8$. Similarly, S_2 has 24 equivalence classes, and thus $N_3^{(2)} = 24$. What we see is that every subsequent S_n will have three times as many triangles as the previous, so $N_3^{(n)} = 8 \cdot 3^{n-1}$.

The cases where k=4 or 5 will be similar in that we are counting prime paths of length k, so let us look at the case where $k = 6$ in S_1. First, we can look at how many ways we can take these triangles and pair them up. These are any D of the form $D = 1 \vee [P_i] + 1 \_ [P_j]$ where $[P_i]$ and $[P_j]$ are each triangles in S_1. Since we have eight prime paths in S_1, we have $\binom{8}{2} = 28$ options of D. Next, we have the case where $D = 2\,[P_i]$, where there are eight combinations of a path with itself. Finally, we can count the Ds of the form $D = 1 \vee P_i$, thereby truly counting the prime paths of length 6. As shown previously, we have 20 prime paths of length 6 in S_1. Thus $N_6^{(1)} = 28 + 8 + 20 = 56$.

Let us look at the next gasket. We have 24 prime paths of length 3 and 74 prime paths of length 6 in S_2. We will have options of D of the form $D = 1 \_ [P_i] + 1 \_ [P_j 2]$, 24 options of $D = 2 \_ [P_i]$, and 74 options of $D = 1 \_ [P_i]$. Notice that $276 + 24 + 74 = 374 = N_6^{(2)}$.

Theorem 2. *Let $N_k^{(n)} = |\{D \geq 0 \mid v(D) = k\}|$. Then $\lim_{n\to\infty} N_k^{(n)}$ converges 3-adically.*

Proof. To begin, notice that

$$N_k^{(n+1)} = |P_k^{(n+1)}| + \left|\left\{D \geq 0 \;\middle|\; \begin{matrix} v(D)=k \text{ and } D \text{ consists} \\ \text{of more than one prime path} \end{matrix}\right\}\right|.$$

From Lemma 1, we know that $|P_k^{(n)}|$ converges 3-adically, so we should look at the other portion. We can expand this into

$$\left|\left\{D \geq 0 \;\middle|\; \begin{matrix} v(D)=k \text{ and } D \text{ consists} \\ \text{of more than} \end{matrix}\right\}\right|$$

$$= \sum_{i=1}^{k} \sum_{\substack{n_i=1 \\ i=1,2,\ldots,j}}^{k} \left|\left\{D \geq 0 \;\middle|\; \begin{matrix} D=n_1[P_1]+\cdots+n_j[P_j] \\ \text{with } v(D)=k \end{matrix}\right\}\right|$$

$$= \sum_{i=2}^{k} \sum_{\substack{n_i=1 \\ i=1,2,\ldots,j}}^{k} \sum_{\substack{n_i \geq 0 \\ \text{such that} \\ n_1+n_2+\cdots+n_j=k}}^{\infty} \left(|P_{n_1}^{(n+1)}| + |P_{n_2}^{(n+1)}| + \cdots + |P_{n_j}^{(n+1)}|\right)$$

As a result, we can see that $N_k^{(n)}$ is equal to a fixed number of $|P_k^{(n)}|$. Because from Lemma 2 we know that $\lim_{n\to\infty} |P_n^{(n)}|$ exists 3-adically, we know that $\lim_{n\to\infty} N_k^{(n)}$ exists 3-adically.

There are still many questions about the Ihara zeta function of the gasket. Most notably, we found an exact way to write the logarithmic derivative based on discernible data. But just how easy is it to write out?

Question 1. Is there a simple way to calculate L_n for any k? And, even more importantly, is there an easy way to write out the zeta function itself? We do not have a simple number like L_n from the logarithmic derivative with which to work. Once we have this, what else can we say about the function?

Question 2. To what do the coefficients of the Ihara zeta function of the Sierpiński gasket converge? What is the radius of convergence of the function? Is it 3-adic meromorphic? Is it a rational function?

Finally, this leads us to wonder to where this idea of p-adic analysis can be extended. One reason that the Sierpiński gasket lends itself so well to a study of prime paths is the method in which the gasket is created. Since there is only a vertex which connects two copies together, prime paths which are contained in two copies are necessary combinations of shorter prime paths. If we look at an example like the Sierpiński carpet, copies of the previous iteration are attached by far more vertices, and will thus be more complicated to study.

Question 3. Are there any other fractals whose Ihara zeta function converges p-adically? If so, is there a way to generalize what p will be necessary?

Appendix

Sierpiński Gasket Program for Maple

This program uses the property of the Sierpiński gasket that the degree of every vertex is 4 except for the three points. It creates three copies of the previous iteration, then connects the copies together by finding the points of each copy. It should be noted that the program requires the package which is far more useful than the Graph Theory package which comes with Maple 10.[1] This package is included with Maple 11.

```
Sierpc:=proc(b)
local G, G1, G2, H, m, n, C, L, p, q, v, w;
G:=CompleteGraph(3);

for w from 1 to b do

v:=nops(Vertices(G));
```

```
G:=RelabelVertices(G, [seq(i, i=1..v)]):
G1:=RelabelVertices(G, [seq(i, i=(v+1)..(2*v))]):
G2:=RelabelVertices(G, [seq(i, i=(2*v+1)..(3*v))]):
H:=DisjointUnion(G,G1,G2):
E:=0;
for n in Vertices(H) do
 for m in Vertices(H) do
    if Degree(H,m)=2 and Degree(H,n)=2
     and Distance(H,m,n)=infinity
     then H:=AddEdge(H, {m,n}):
        H:=Contract(H, {m,n}):
                if m<n then H:=AddVertex(H, n) else
H:=AddVertex(H,m) end if;
     end if;
 od;
od:
for p in Vertices(H) do
 for q in Vertices(H) do
  if Degree(H,p)=2 and Degree(H,q)=2 then
   if Distance(H,p,q)=3*2^(w-1) then
    H:=AddEdge(H,{p,q}) :
    H:=Contract(H,{p,q}):
  if p<q then H:=AddVertex(H, q) else H:=AddVertex(H,p)
end if;
     end if;
   end if;
  od:
 od:
 for p in Vertices(H) do
  if Degree(H,p)=0 then H:=DeleteVertex(H,p) end if;
 od:
 G:=H

 od;
 return(H);
 end:
```


Acknowledgement

The author would like to thank Prof. C. Douglas Haessig for all his help in the preparation of this paper.

References

1. J. Farr, M. Khatirinejad, S. Khodadad, M. Monagan, *A Graph Theory Package for Maple*, http://www.cecm.sfu.ca/CAG/papers/GT.paper.pdf.

2. D. Guido, T. Isola, M. Lapidus, *A trace on fractal graphs and the Ihara zeta function*, arXiv:math/0608060.

3. M. Horton, H. Stark, A. Terras, *What are zeta functions of graphs and what are they good for?*, Quantum graphs and their applications, 173-189, Contem. Math., 415, Amer. Math. Soc., Providence, RI, 2006.

4. C. Storm, *Some graph properties determined by edge zeta functions*, arXiv:0708.1933.

5. A. Terras, \emph{A Stroll Through the Garden of Graph Zeta Functions}, http://math.ucsd.edu/~aterras/newbook.pdf.

About the Journal

The *Journal of Undergraduate Research* (ISSN# 1547-9641), Volume 6, Issue 2, Spring 2008 was assembled on a Apple PowerMac G4 Quicksilver using Adobe InDesign CS. Graphics were produced using Adobe Photoshop 7.0 and Adobe Illustrator 10. Microsoft Office was used for text editing and review. Fonts used included Adobe Garamond, Futura, Times New Roman, and Symbol (for Greek characters). This 48-page journal was printed and perfectly bound by Mercury Print Productions Inc. of Rochester, NY.