piRNA Binding Characterization

January 10

2008

We aim to characterize the binding of piRNA to transposons by examining the difference between different families of both transposons and piRNAs.

MOL 455
Final Report

Adam Sanders
MOL 455
1/10/08

piRNA Binding Characterization

ABSTRACT

We aim to characterize the binding of piRNA to transposons by examining the difference between different families of both transposons and piRNAs.

INTRODUCTION

RNA interference (RNAi) plays a prominent role in biological events such as development, gene regulation, and defense against viruses and transposable elements[1]. RNAi is mediated by three groups of small RNAs including siRNAs, microRNAs, and piwi-associated small interfering RNAs (piRNA) (Meister & Tuschl, 2004). Differing from siRNA and microRNA in length and structure, very little is known about the action of piRNA and how they interfere with transposon amplification in the germline of organisms such as *Drosophila* or mouse. It is known is that the loss of piRNA machinery leads to the up-regulation of transposon elements harbored in the genome, and that these elements will duplicate themselves and reinsert their copies into the genome. The process of reinsertion generates double-strand chromosomal breaks, which can lead to the loss of chromosomal integrity (Chen, Pane, & Schupback, 2007) and arrest the developmental process of the embryo. The piRNA pathway is active in the germline from flies to humans, which is an indication of its early appearance in evolution. The human genome contains millions of remnant transposons, primarily evolved from viruses, which are considered to have been a driving force for evolution and species diversification. It is possible that co-evolution occurred between the transposon and piRNA sequences within a species, as well as between species. The answer is currently unknown.

With this in mind my project sets out to characterize the frequencies and locations of piRNA binding to transposons. The ultimate goal of this project is to better understand any interactions between transposons and piRNA. For a given transposon family across species, which piRNA sequences have evolved to silence them? For these transposons and piRNAs, do they operate against similar or different regions of the transposon? Finally I attempt to locate motifs found within areas of the transposon that are found to have a high preference for binding.

THE DATA SETS

There were two main data sets used in my portion of this project. The first of these was a total listing of 32,512 piRNAs that have currently been identified in *Drosophila*. The RNAs dataset was originally downloaded from the online supplement to a paper by Aravin et al on November 10th (Aravin, et al., 2006). This listing of piRNA elements was divided into three files, one for each type of protein family with which the piRNA have been known to interact. The protein families are aubergine (aub), piwi, and ago3. Each piRNA came with a sequence and a cloning frequency representing the number of times the piRNA was found in the *Drosophila melanogaster* genome. While this cloning frequency was not used in this analysis, it might serve as an interesting piece of future research into this subject. Notably absent from these files were the actual locations of binding to the genome. This information was later acquired by Tao Yue in his portion of this research project.

The transposable element data set originated from FlyBase, and was downloaded on the date mentioned previously. There were 1540 different elements broken up into 192 genomic regions. Given that the total number of nucleotides in these sequences was over 4 million, the average length of the elements in this data set was about 3000 nucleotides. In addition to the actual transposon sequence, this set contained the complete nucleotide sequence of every transposable element, its location in the genome, as well as the leading and trailing nucleotides surrounding that location. As with the piRNA data, the transposable elements were broken down into three specific types of transposable elements. The three types represented were long terminal repeat transposons (LTR), long interspersed transposon elements (LINE), and transposon inverted repeat (TIR). This transposon data and the piRNA data sets are summarized in Table 2 and Table 1 respectively. All original data files as well as those files used in later analysis can be found at www.princeton.edu/~acsander/school/mol455. In regard to Table 1 and Table 2 listed mentioned previously, the column labeled number represents the total count of either piRNAs or transposons. The length of those sequences is the number of nucleotides found in all the sequences of that type. These numbers serve as an important part of later analysis as they are used to normalize the binding frequencies of piRNAs.

Type	Number	Length
AGO3	10402	357808
AUB	8207	251756
PIWI	13903	202561
TOTAL	32512	812125

Table 1: piRNA Summary

Type	Number	Length
LTR	683	3090962
LINE	485	1014808
TIR	372	358275
TOTAL	1540	4464045

Table 2: Transposon Summary

BINDING DENSITY CHARACTERIZATION

The primary stage of this project was to determine whether or not there were significant differences in the binding frequency of any combination of piRNA and transposon. It would be very interesting to learn, for example, that piwi-associated RNA bound preferentially to LINE transposon elements. In regards to previous research in this area, we were unable to locate any research that had attempted to establish piRNA binding preference to a particular type of transposon. The first step in this process was to develop a master list of all possible interactions between piRNA and transposons. This master list, currently located on my website, can be queried for matches between any two data sets (e.g. PIWI RNA bound to a LINE Transposon as mentioned above) or even combinations of those sets (e.g. PIWI RNA bound to all types of transposons). In addition, the list makes it easy to identify all the piRNA that bound to any given transposon. For the purposes of this study, only direct text matches were considered. Matches that were found in regions upstream and downstream of the actual transposon or partial matches were not considered. One result of this direct text matching is that there should then be regions of every transposon with which no piRNA can interact. While it is not a perfect simulation, this feature did not appear to have any major effects on analysis. Before searching for interactions the upstream and downstream regions of the transposon sequence were removed from the data files. In order to accomplish the binding analysis, I developed a program in java, and later created several programs in bash and perl to aid analysis. The result of this interaction study was a total of just over ten-thousand interactions between all piRNA and transposon elements. An interesting way of visualizing this information is in Figure 1 below. This figure shows the location on the chromosome of every transposon broken down by type as well as every location that piRNA was found to bind to transposons. The y axis shows the different transposons while the x axis demonstrates position along those chromosomes in mega bases. The blue-colored dots are the location of transposons, where each shade corresponds to a different type, e.g. LINE or LTR. Likewise, the orange-colored dots represent the location of piRNA binding, where every shade represents a different type of piRNA, e.g. PIWI or AGO3.

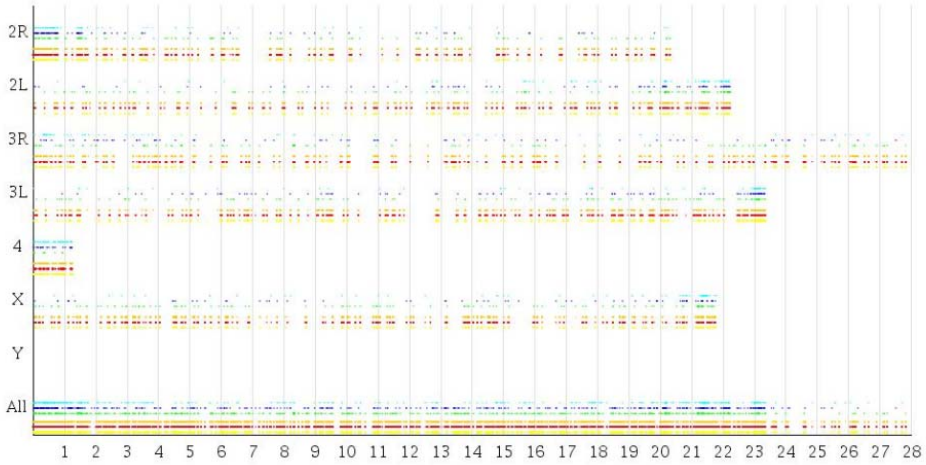

Figure 1: piRNAs and Transposons Mapped to the Drosophila melanogaster Genome

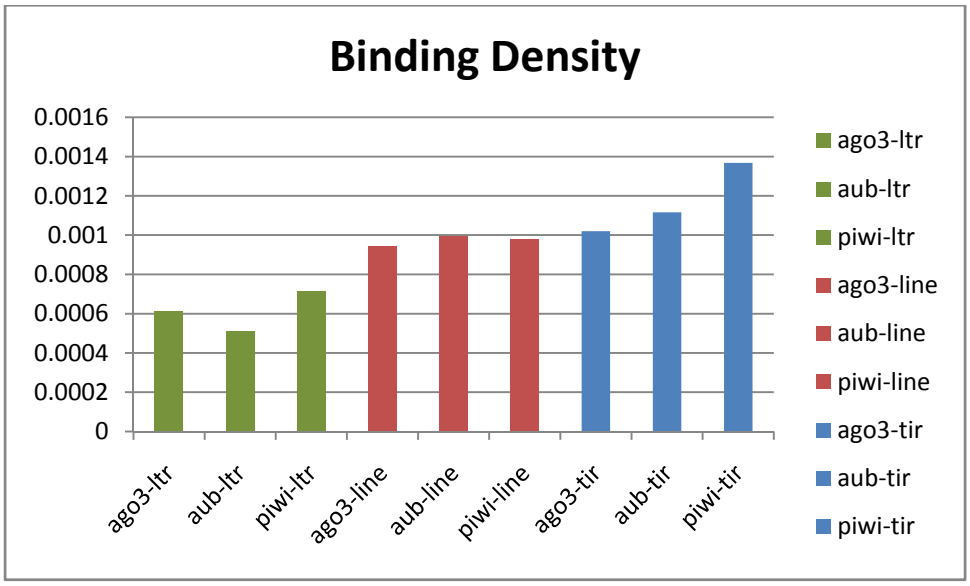

Figure 2: Binding density for all combinations of transposon and piRNA

Binding densities were used to determine if there was a higher than expected binding frequency for combination of piRNA and transposon. The density for any pairing is simply the number of interactions that were found between a given group of piRNAs and a corresponding group of transposons normalized for both the total number of piRNA sequences and the total length of all the transposon sequences queried. An important caveat is that rather than using the length of the piRNA sequence, the number of sequences was used to normalize these bindings. This was done because piRNA sequences vary in length very little, typically differing only by a few nucleotides. In contract to the variance in piRNA, transposable

elements show considerable discrepancies in length and include a range of about ten thousand nucleotides at their most disparate, though they regularly vary on the order of a thousand nucleotides.

$$Binding\ Density = \frac{\#\ of\ Interactions}{(Total\ number\ of\ piRNA)(Total\ length\ of\ transposon\ sequences)}$$

The results of the binding density analysis can be seen in Figure 2 above. This graph shows the binding densities of every single-piRNA to single-transposon match. That is, only one type of piRNA and one type of transposon element were compared, e.g. PIWI piRNA and LINE transposons. The different colors represent the different types of transposons, while each individual bar of those colors represents a specific type of piRNA. From this graph it becomes clear that there is little change found when varying the type of piRNA. The average binding density for PIWI and AGO3 when compared against all types of transposons did not experience any significant shifts. However, the type of transposable element does seem to affect piRNA interaction density. For example, interactions between every type of piRNA bound to LINE transposons occurred at a higher frequency than did any interaction where piRNA was bound to LTR transposons. Though significant variations were found in the binding frequency of different types of transposons to piRNA elements, this discovery is not particularly surprising. The hypothesis we proposed to explain this phenomenon was that the binding frequency variance found in different transposon types was a result of different amounts of sequence repetition unique to that particular type of transposon. While every transposon is expected to have some repeated sequences, the number of repeated sequences in any type of transposon varies. Our hypothesis given the data was then that TIR proteins have a greater number of repeated sequences than do LTR. The variations in the number of these repeat regions could result in artificially high binding frequencies when a single piRNA sequence is able to bind multiple times to a repeated region found in a given transposon. This explanation can also be shown from the perspective of piRNA. It is possible that piRNA elements have evolved in order to preferentially silence these repeated sequences. Given that sequence repeats are often found in transposons, this strategy would work well to quiet those elements. An additional benefit to this phenomenon is that fewer piRNA would have to be produced in order to accomplish the same goals, and the fewer piRNA that are produced, the less likely it is that those piRNA will result in unintended deleterious effects on other aspects of cell functioning.

Perhaps more interesting are the results of the genome-wide search for piRNA interactions accomplished by Tao Yue. Tao used a self-developed program similar to the one I had developed to work specifically with my data. His work also looked into locating direct text matches, and as a result our output was directly compatible with my output. When I worked out the binding frequencies of the entire genome using his data set and compared them to those I found within

transposons specifically I was surprised to learn that they were very similar. As piRNA are believed to be "hunters of genomic invaders" such as transposons (Hartig, Tomari, & Forstemann, 2007), it was expected that the binding density of these piRNAs within transposon regions would be far greater than the density with which they bound throughout the entire genome. In fact, it was discovered that binding to the entire genome actually occurred at a rate comparable to the highest rate found in my analysis. Initially, this seemed to us to be an impossible discovery. However, after a great deal of cogitation and an extensive programmatic bug search, it was realized that this finding is not a difficult one to explain. The dataset containing the transposons with which our group was working contains only a few transposable elements. These are the transposons found to be active by current research into the fly genome. However, the genome is filled with the wreckage of these transposable elements. Fragmented pieces and, very likely, entire copies of these transposons fill the genome. Having many matches of piRNA to the genome should really come as no surprise to those acquainted with transposable elements. After all, it was their ability to rapidly copy themselves to other locations within the genome that necessitated the evolution of the piRNA mechanism to begin with. Throughout vast stretches of time, many of these elements get broken and separated and eventually may not be recognizable as transposons. However, their sequences can remain in the genome at almost any location. In addition to this fact, the set of transposons with which I was working was far from the complete set known. I learned later that Xiaojing was working with a newer set of transposable elements that contained almost twice the number of elements as did my set. Figure 3 below shows the differences in binding densities for all piRNA bound to each type of transposon, as well as the genome and all of the transposons combined.

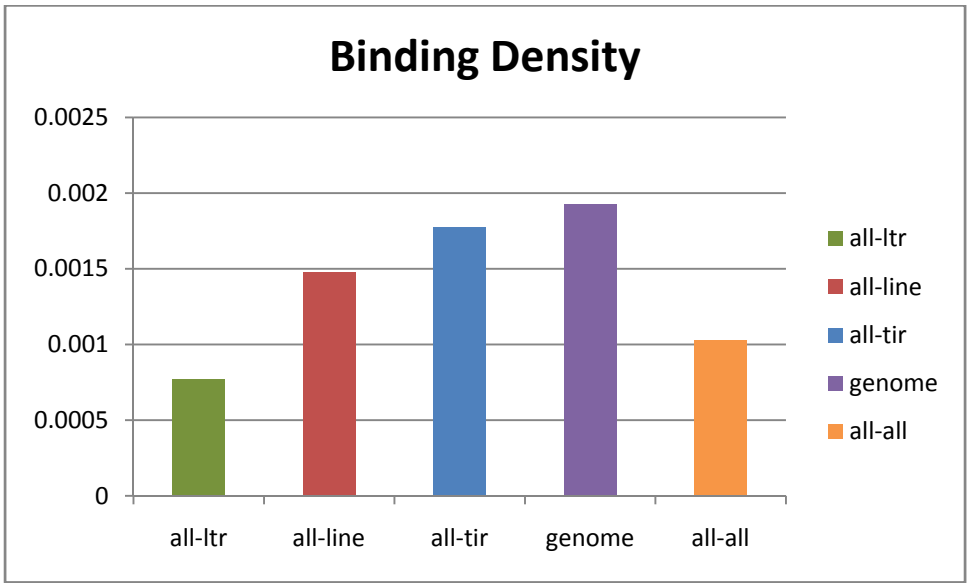

Figure 3: Binding density for all piRNAs against different types of transposon

In addition to running sense-strand piRNA matching, it was proposed that we investigate anti-sense piRNA interactions. The search for anti-sense piRNA interactions with the genome conducted by Tao turned up few differences from his initial sense results. He found the differences between sense and anti-sense piRNA matching to be surprisingly minimal. In spite of these results, I generated my own dataset of anti-sense piRNAs from the original. This anti-sense version set was capable of being developed and analyzed in exactly the same fashion as its normal-sense counterpart. The results of this aspect of the research were very similar to those discovered by Tao. The differences in frequency and density for the binding of sense piRNA verusus the binding of antisense piRNA were negligible. In comparison to the 4636 total interactions found when searching for sense matching, the anti-sense piRNAs bound 4826 times. This disparity represents a change of less than five percent when switching between sense and anti-sense versions of the piRNA. The results of this study in Figure 2 are shown below in Figure 4. In addition to demonstrating the small degree of variability seen between the total number of sense and antisense piRNAs bound to all transposons, this graph allows us to examine whether or not anti-sense matches preferentially to a particular transposon when compared with its sense alternative. One would expect, however, that such a preference would be unlikely given the results of the sense binding densities as well as the meager 5% difference in total number of bindings. Figure 4 supports this expectation as there is little variation between the number of sense and anti-sense interactions for any particular combination of files. Finally the reason for this change was discovered. From the literature, I discovered that AGO3 is actually the antisense version of PIWI and AUB (Brennecke, et al., 2007). The creation of my own data set was not useless, as the files with which I was working were not perfect inverses of one another. That is, the PIWI data set was not just a listing of the antisense versions of AGO3 piRNA.

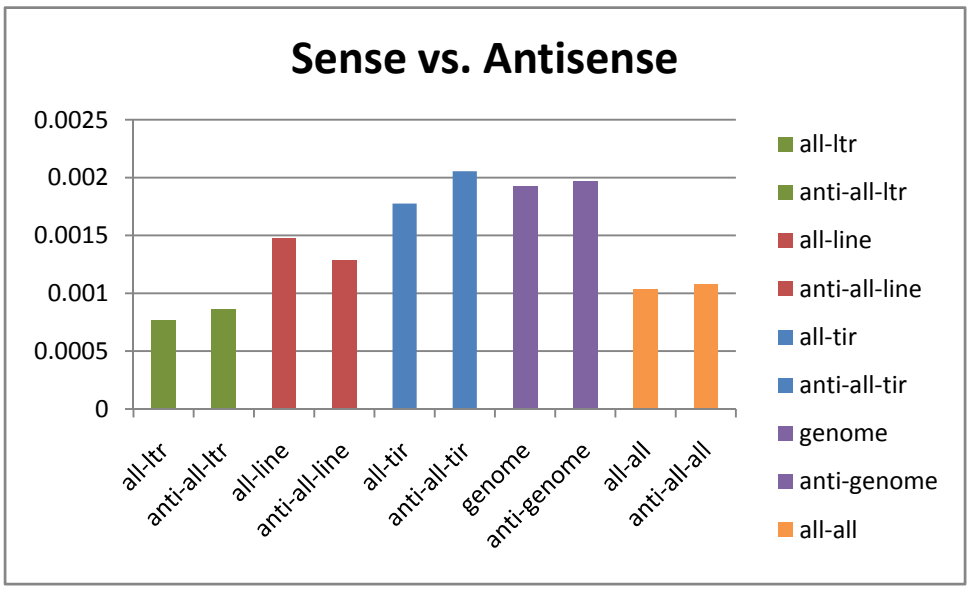

Figure 4: Binding of all sense and all antisense piRNAs

A final project related to piRNA binding density was an attempt to predict conservation of transposable elements across the different *Drosophila* genomes using binding density. The rationale behind such a search is that those transposons that are heavily bound are likely to have existed for a greater period of the evolution of that fly. That is, these elements should be more heavily bound because the different species of *Drosophila* have had more time to develop piRNA-based defenses against them. In order to test this hypothesis, it was necessary to narrow the list of transposons from the original 1540 transposable elements to a more reasonable number. Using the results from the binding density characterization mentioned previously, I developed a list of the most highly bound sequences, as well as a second list representing those transposons with absolutely no binding. Were these to be compared across genomes, it would be our hypothesis that those transposons that were highly bound, and by extension the piRNAs that bind to them, would be found in even distant relatives of *Drosophila melanogaster*, whereas those less bound would only be found in species very close to *Drosophila melanogaster* or even only within that organism itself. Figure 5 and Figure 6 below show all of the binding densities and frequencies of the 192 transposon regions. There is clearly a great degree of variation in both binding density and frequency of these transposon regions. Most importantly, in frequency there were regions that were bound by piRNA in a range between zero and over 2000 times.

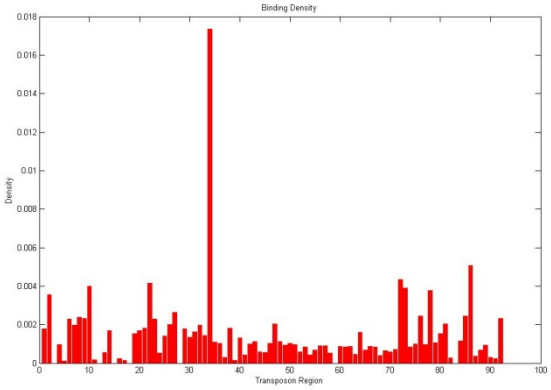

Figure 5: Binding density of all piRNA against every transposon

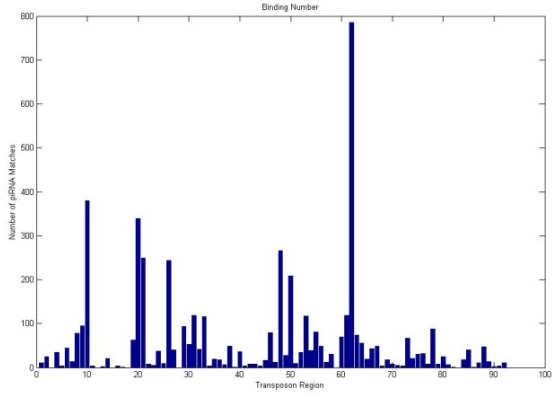

Figure 6: Binding frequency of all piRNA against every transposon

In order to make analysis simpler, the decision was made to work with ten of the highest and ten of the most highly-bound transposons, and ten of the least bound transposons. Since the majority of transposons that contained extremely high densities had sequences only a few hundred nucleotides long, it was decided to sort them first on the basis of their number of hits, and finally on their density. The results of the highest scoring transposons can be seen in Table 3 below. The lowest scorers were sorted in a similar manner and demonstrated surprisingly high sequence lengths for their very low binding

frequencies (all in this study were zero). Even in this small study it became obvious that certain regions were more fertile for piRNA binding than were others. In the high scoring table nine of the ten total transposons were either roo or F-element, while every single low scorer was from the Tirant family. While analysis was only done on individual transposons in order to save computation time, these results would likely generalize to every transposon within that region. In order to compare their genomic conservation among other *Drosophila* species, these transposons were put into BLAST.

Transposon	Hits	Length
roo_FBgnnnnnn_FBti0019681	24	2996
roo_FBgnnnnnn_FBti0019752	22	4451
F-element_FBgn0000652_FBti0019160	21	3612
F-element_FBgn0000652_FBti0020160	21	4694
F-element_FBgn0000652_FBti0019054	21	4697
roo_FBgnnnnnn_FBti0019558	21	4708
R1-element_FBgn0003908_FBti0019883	21	5364
roo_FBgnnnnnn_FBti0019023	21	8273
F-element_FBgn0000652_FBti0019398	20	4145
F-element_FBgn0000652_FBti0019382	20	4694

Table 3: Highest scoring transposons

Transposon	Hits	Length
Tirant_FBgnnnnnn_FBti0020010	0	8429
Tirant_FBgnnnnnn_FBti0019592	0	8431
Tirant_FBgnnnnnn_FBti0019495	0	8525
Tirant_FBgnnnnnn_FBti0019441	0	8526
Tirant_FBgnnnnnn_FBti0020028	0	8526
Tirant_FBgnnnnnn_FBti0018946	0	8526
Tirant_FBgnnnnnn_FBti0019543	0	8526
Tirant_FBgnnnnnn_FBti0018948	0	8532
Tirant_FBgnnnnnn_FBti0019313	0	8533
Tirant_FBgnnnnnn_FBti0019433	0	8536

Table 4: Lowest scoring transposons

This portion of my project is one I feel is ripe for further analysis. A similar project was completed by Xiaojing, however this project ran into some serious programmatic difficulties. The results from the BLAST searches for large numbers of transposons were difficult to interpret and organize. As a result, I decided to leave this portion of my project for future analysis. The generation of the list of highly bound transposons ended up being of great utility in the remainder of my project.

BINDING LOCATION CHARACTERIZATION

The second aspect of this project was to determine whether or not there was a preference for binding at a specific location within transposons. Through the use of the java program previously mentioned, I calculated the position along the transposon where the piRNA first bound. The results showed a fairly surprising trend. All piRNAs bound preferentially to the 3'end of the transposable elements. This effect was identified to the same degree in every single combination of piRNA and transposon element. As can be expected given the previous work in this area, the antisense versions of these same piRNA showed similar results. Both the sense and anti-sense graphs can be seen below in Figure 7 and Figure 8 below. It is clear from these graphs that there is a significant increase in binding over the last 20% of the transposon. In addition, a small increase was also found in binding at the 5' end of transposable elements. This increase was far less significant than was the

increase discovered at the 5' end and was not seen until later in the project. We were unable to locate studies that had investigated the same phenomena into which we were delving.

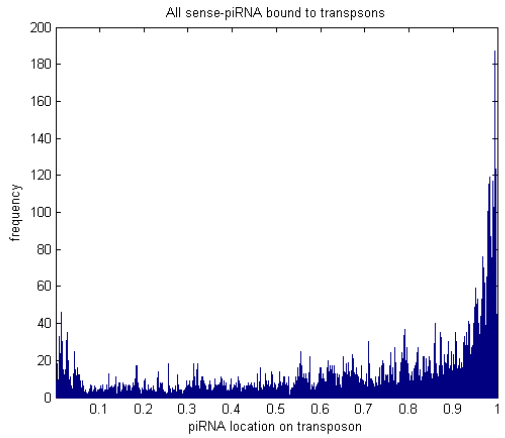

Figure 7: Binding location of all sense-piRNA bound to transposons

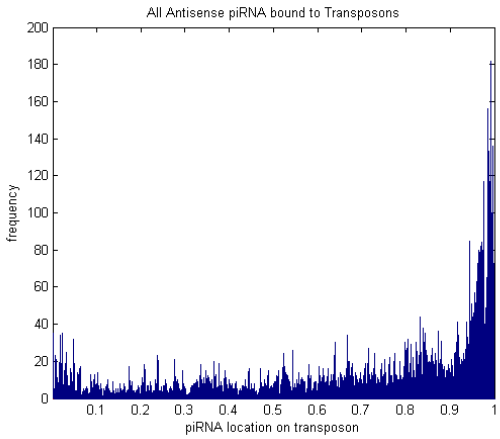

Figure 8: Binding location of all anti-sense piRNA bound to transposons

There were several initial hypotheses as to why this 3' preference might be the case. The very first of these was that there were piRNA that were specifically targeting the poly-A tail found in transposons and other proteins. Were this tail to be any length over the average sequence length of the piRNA, it would be relatively easy to develop a piRNA sequence (all A's or T'S) that was able to bind repeatedly to that transposon sequence. However, it was soon realized that such a piRNA, if it did exist, would have no transposon specificity. A close analysis of all the interactions discovered quickly showed this poly-A hypothesis not to be the case. In fact, not a single binding resulted from more than six A nucleotides in a row, and those that resulted from multiple A nucleotides were few and far between to begin with.

The poly-A hypothesis defeated, I turned to possible programmatic mistakes. It was entirely possible that some aspect of either the initial program that developed the master list, or those that were used in later analysis could result in skewed data. In fact it was mentioned previously that the direct text matching used in the original programs would undoubtedly result in zero binding at both the first 20-30 nucleotides as well as the final 20-30 nucleotides. However, this occurrence would actually lessen the 3' binding preference. The tremendous variability in the transposon size was also considered as a source of possible problems. It was thought that extremely small sequences would result in unusual data patterns. However, upon further consideration, these small sequences were relatively few. In calculating the average length of the top 5000 bindings (roughly half the total bindings) the average sequence length was just over 5000 nucleotides. This makes sense as more nucleotides result in more possibilities to bind. Running these sequences in the same analysis tool, the preference for the 3' end of transposon elements was still shown prominently. In a final attempt to ensure that no

programmatic difficulties were causing this binding phenomenon, I generated random perturbations of both the sense and antisense piRNAs and attempted to bind them to the genome. However, the randomization of these sequences resulted in a complete lack of interactions. Even though the incidence of any nucleotide was identical within the original piRNA, the result of randomizing their sequence location was not a single interaction.

One hypothesis that did warrant further investigation was that repeats within the transposons were responsible for binding toward the 3' end of those transposons. In order to falsify this hypothesis, one could simply show that the piRNA found binding to any given transposon were each unique. For this analysis, I again used the most highly bound sequences as a guide, though I also sampled other transposons to ensure that these highly bound transposons were not biasing results. The first step was to examine all of the piRNA that bound to any given transposon. For example, in the very most bound transposon, there were a total of 24 interactions and thus 24 possible piRNA that could have bound. Using the master list, which contained these sequences, the extraction of these sequences was relatively easy. The total number of piRNAs that bound was 22 as there were two piRNA that bound twice. This was surprising, as I had expected the repeated sequences to result in many more piRNAs that bound multiple times. As an aside, I also decided to examine the binding location within every sequence I tested, and the preference for the 3' end of transposons was evident even at the level of single transposon. Returning to the piRNA analysis, I took all the sequences that interacted with a certain transposon and ran them through motif finding algorithms. From this I was able to generate excellent motifs, often with nucleotide sequences around 20 nucleotides in length. For the topmost example, MEME created a 20 nucleotide motif with an e value of $7.05e^{-15}$. Figure 9 below shows the result for this element. The great majority of the piRNA that bound to this transposon were identical but shifted sequences. This analysis was completed on the rest of the top scoring list with similar results. Obviously, this analysis was impossible to complete on the low scoring transposons as they did not have any binding. The results of this piRNA motif analysis were that many of the piRNAs binding to any transposon are either very similar or simply shifted versions of one another. In addition, the 3' binding preference was visible even on the level of individual sequences.

Figure 9: MEME results for all piRNA interacting with roo_FBGnnnnnn_FBti0019681

While it was clear that the piRNA binding any individual transposon are often very similar, and show significant overlap, what was unclear was whether or not there are motifs conserved either within a single transposon (a repeat region) or across different transposons. In order to find the answers to these questions, further MEME analysis needed to be done. Additionally, since individual transposons showed a 3' preference for piRNA binding, I decided to extract different lengths of the 3' regions of transposons in an attempt to make sure any motifs in these regions themselves, or across different regions were identified. I created transposon datasets that included only the last 10, 20, and 30% of all nucleotides in every transposon, as well as the last 150, 300, 600, and 900 nucleotides. That is I tried both an identical fraction of nucleotides as well as an identical nucleotide number in the 3' region. In turning to the question of whether or not there are motifs conserved across different transposons, I ran multiple MEME analyses of randomly selected sequences from both the top-hit list as well as all of the transposons using the entire transposon sequence as input. The results of these efforts were far less significant than were those for the piRNA analysis. I was able to locate significant motifs only when running multiple transposons from the same region, e.g. two roo transposons shared significant sequence similarity. However, there were no recognizable motifs found when running any two different regions of transposon. I also attempted this same analysis using the extracted sequences mentioned previously. The results were no better than with the original sequences, and no motifs across transposons could be identified with reasonable certainty.

The final step of this process was to determine whether or not there were motifs repeated within every individual transposon. This hypothesis can also be tested through the use of motif-finding algorithms. In order to determine if there were repeating sequences within these transposons, I ran entire transposons in a MEME search looking for multiple instances of a motif within a single sequence. For this analysis, I selected both sequences at random, as well as tried every sequence in the high scoring list. Unlike the cross-transposon study mentioned previously, this search met with some decent results. I

was able to locate motifs often with p values around $1.3e^{-7}$. In addition, many of these motifs were found in the second half or even last quarter of the transposon. Figure 10 below is the result of one of these searches and demonstrates motifs occurring only in the last half of the transposon sequence. As portions of the previous research had suggested, the TIR elements that I tested did tend to have a higher number of motifs than did either LTR or LINE transposon elements. The particular sequence shown in Figure 10 below is for an LTR element that contains five repeats of such a motif. In addition it is also possible to see the overlap between these motif positions and the actual binding of piRNA shown in Figure 11 below.

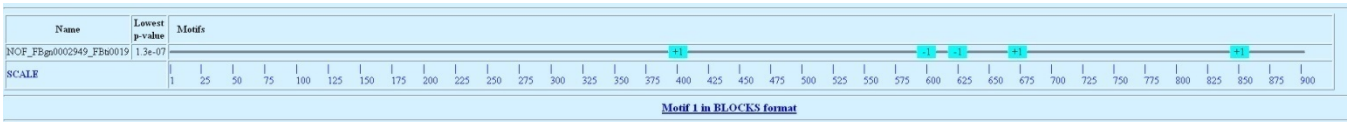

Figure 10: Results of a MEME query on a single transposon showing 3' motif preference in roo_FBGnnnnnn_FBti0019681

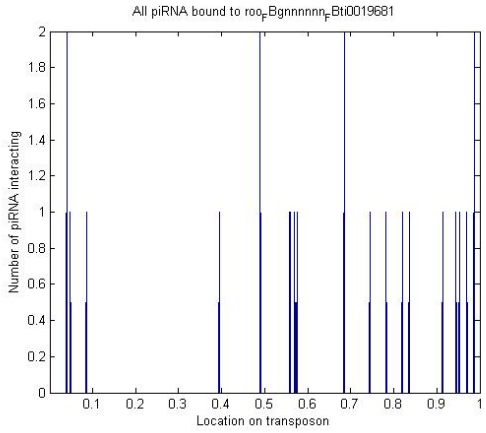

Figure 11: all piRNA bound to roo_FBGnnnnnn_FBti0019681 demonstrating 3' binding preference

DISUCSSION

From this information we can develop a model to describe the binding of piRNA to transposons. It appears as if there are great quantities of sequence-similar piRNAs that bind heavily to a few conserved areas within any transposon sequence. These conserved areas still tend to cluster towards the 3' end of the transposon, and are apparently the result of repeating sequences within those transposons. The repeated sequences, however, are not identical in many cases. Rather they are very similar, differing only by a few nucleotides over short stretches of 20-30 nucleotides. Finally I was unable to locate any significant motifs between unrelated transposons. These results seem to make sense in the schema of piRNAs as transposon-hunters. The evolution of piRNAs apparently target repeat-regions within transposons, as they are statistically more likely to occur. This evolution also resulted in the great majority of piRNAs developing similar and even overlapping sequences used to target these repeating transposons. The reason that there is such an increase in binding to the 3' end of

transposon sequences is that the sequences begin to repeat in these areas. These repeats can be discovered as motifs using motif-algorithms, and are heavily targeted by piRNA evolution. There are a great number of future directions for research in this area, many of which have already been addressed. Most importantly, it would be fascinating to learn which transposon elements are best conserved across different genomes.

CONTRIBUTIONS

- Adam Sanders
    - Everything mentioned in this paper (except that specifically stated otherwise)
    - Generation of master list of binding interactions
        - Analysis of these interactions using bash and perl scripting
        - Binding density characterization of
    - MEME analysis of piRNA and transposable elements
    - Creation of several randomized data sets for other individuals' projects
    - Some minor data parsing for other individuals' projects
- Tao Yue
    - Total genome analysis of piRNA binding
    - Research into the original locations of piRNA in the genome
- Xiaojing Huang
    - Very helpful scripts in MATLAB for displaying data in my project properly
    - BLAST searching
    - Intron hit searching using MATLAB and data generated by Tao
- Yanling Zhao
    - Actual idea to research piRNA
    - Much of the original direction of research
    - Very helpful discussions about the meaning of the results acquired in my project
    - Wrote background information
    - Majority of the research into piRNA

References

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., et al. (2006). A novel Class of small RNAs bind to MILI protein in mouse testes. *Nature* , 203-207.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* , 1089-1103.

Caspi, A., & Pachter, L. (2006). Identification of transposable elements using multiple alignments of related genomes. *Cold Spring Harbor Laboratory Press* , 260-270.

Chen, Y., Pane, A., & Schupback, T. (2007). *Current Biology* .

Girard, A., Sachidanandam, R., & Hannon, G. J. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* , 199-202.

Hartig, J. V., Tomari, Y., & Forstemann, K. (2007). piRNAs - the ancient hunters of genome invaders. *Cold Spring Harbor Laboratory Press* , 1707-1713.

Lakshmi, S. S., & Agrawal, S. (2007). piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Research Advance Access* , 1-5.

Lau, N. C. (2006). Characterization of the piRNA Complex from Rat Testes. *Science* , 363-367.

Meister, G., & Tuschl, T. (2004). *Nature* , 343-349.

Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., et al. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions int he Drosophila genome. *Genes and Development* , 2214-2222.

Vagin, V. V. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science* , 320-324.

Yin, H., & Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* , 304-308.