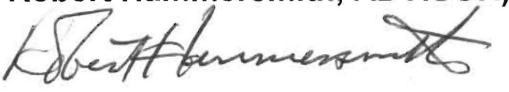
Analysis of Allelic Diversity of *Oxytricha trifallax* Utilizing the 81 Gene Locus

An Honors Thesis (Honors 499)

BY

Hunter Gibbons

(Dr. Robert Hammersmith, ADVISOR)



BALL STATE UNIVERSITY

MUNCIE, INDIANA

May 2014

Expected Date of Graduation

May 2014

ABSTRACT

Oxytricha trifallax is a ciliated protozoa located in the freshwater lakes and streams of Indiana. It contains a genetic organization including two separate nuclei, the micronucleus and macronucleus. These two nuclei separate the genetic functions of the cell between them. One highly characterized region of the genome for the *trifallax* is the 81 gene locus, but it has only been studied in two strains of this species. In order to broaden understanding of this region, the 81 locus was sequenced for 21 natural isolates to better understand the allelic variation in natural populations. All the strains were made up of homozygous or heterozygous combinations of the previously identified 3, A, and B alleles. There were only new variations of identified alleles previously identified in natural strains, and there were no identified natural triallelics. The lack of new alleles and level of homozygosity in their naturally isolated strains suggests that this species consists of isolated subpopulations with little interbreeding between these groups.

ACKNOWLEDGEMENTS

I would like to give a thank you to Dr. Robert Hammersmith for his mentoring and help on this project over years, as well as the help of Dr. Anne Blakey for all of her lab expertise. I would also like to thank Ball State University and the Honors College for the opportunity to work on this Project.

And a very special thank you to Paul Hickner and the sequencing core at Notre Dame University, who provide the expertise and guidance at every step to help us complete and build all of the sequences.

TABLE OF CONTENTS

TABLE OF FIGURES & TABLES

Page

FIGURES

TABLES

Introduction

Oxytricha trifallax

The *Oxytricha trifallax* is a ciliated protozoa from the class Spirotrichea (Prescott 1994). This single celled eukaryotic organism inhabits fresh water over a number of different regions. Like many other ciliated protists, it feeds on other microbial organisms and develops through a continuous life cycle *O.trifallax also* completes sexual reproduction through conjugation like many other ciliated protists. Altogether, there have been 29 different natural isolates of the *O. trifallax,* based off of morphological and molecular differences. Each of these strains are found in similar environmental locations, and can be found intermingled in different freshwater locals (Prescott 1994). In spite of the intermingling of all of these different strains, not all strains are capable of mating. Only some of the *O. trifallax* isolates can conjugate and form successful offspring, despite being viable in similar environments. This would suggest that inside of the overall population of the organism, there are divisive sub-populations that disallow mating among all strains.

The study of ciliated protozoan have resulted in a number of different discoveries over the past decades (Prescott 1994). The study of *tetrahymena thermophile* first introduced us to the concept of the telomere. This discovery eventually led to an entirely new understanding of chromosome architecture. Other forms of the *tetrahymena* also showed the functional work of non-coding RNA, and eventually led to our understanding of how RNA can function as an enzyme. Surveying *O. trifallax* for non-coding RNA genes resulted in identifying a full family of functional RNAs that were previously unidentified (Jung et al, 2011). These studies into epigenetic concepts as well as the use of non-coding RNAs are essential to expanding our understanding of how the genome functions as a system. The study of *O. trifallax* also provided a novel manipulation of DNA: the massive loss and rearrangement of DNA during conjugation. A DNA manipulation of this magnitude had never been seen before. All of these ciliated protozoa have certain things in common, and one of the most common parts of the organism is the presence of a dual organization of the genetic material (Prescott 1994).

Macronucleus and Micronucleus

The *Oxytricha trifallax* contains two nuclei at all times through its life cycle (Herrick 1994). These can be separately characterized as the micronucleus (MIC) and macronucleus (MAC). The MAC is the location used for all transcription and somatic processes for the organism (Prescott 1994), whereas

the MIC is transcriptionally inactive, but is used as the germline for reproduction (Herrick 1994). These two nuclei represent differences in function as well as form, as the MIC is diploid while the functional macronucleus is vastly polyploidy. The MAC can contain up to a 1000 copies of every single nanochromosome that is used (Seegmiller et al, 1997).

In conjugation, the sharing of the germline micronuclei results in the development of a new MIC and MAC, as a combination of both of the previous germlines emerge. (Seegmiller et al, 1997). This new "hybrid" MIC will cause the degradation of the current functioning MAC, and form a new MAC analogue built from the hybrid genome (Prescott 1994). This new MAC will be formed from the completion of a mitotic division as well as other processes. During the formation of this new MAC, there are a number of chromosome modifications that need to occur (Seegmiller et al, 1997). These can include, polytenization, fragmentation and rearrangement of chromsomes, as well as the addition of telomeres to some of the small chromosomes (Doak et al 2004). The addition of these telomeres can be as short as a hundred base pairs, and range all the way in size to 15 Kb. Each chromosome located in the mature MAC analogue contains an average of ~2400 bp. Altogether in the MAC, there are approximately 20,000 different minichromosomes amplified to about 1000/ploidy (Doak et al 2004). Most of these described minichromosomes contains approximately one gene once they have completely matured and completed MAC transformations. As research has expanded to better understand all the different parts of the MAC, we have learned that some of the larger nanochromosomes can have multiple genic regions located on them. One of the first nanochromosomes to be found with multiple functioning polypeptide products was the 81 gene locus. This chromosome had three separate products that are formed from it, as well as different alleles that can develop (Doak et al 2004).

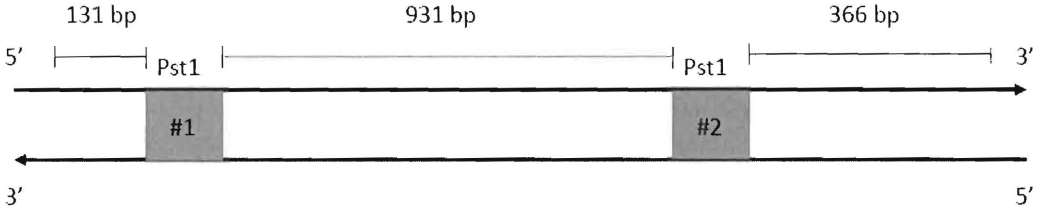
81 Gene locus

During the development of the MAC, after the mitotic division of the MIC when fragmentation occurs, a number of different sequences are separated from the final product chromosomes. These portions of the DNA that are separated out are called internal eliminated sequences (IES) (Segmiller 1996). These can be identified as a long IES (>4 Kb) or short IES (<4Kb). These IES's are cut out of the chromosomes and degraded away as they are unused. The sections of the DNA from the MIC that are retained are then reordered and ligated to create the final nanochromosomes. The order and sequence of the DNA in the MIC are not necessarily the same as that in the MAC. This means another level of rearrangement needs to occur (Seegmiller et al, 1996).

This study specifically looks at the 81 gene locus because of its development between the MIC and MAC. The 81 MAC chromosome family consists of three sizes of MAC chromosomes generated by alternate splicing events in the maturation of the MAC (Williams and Herrick 1991). These three

chromosomes all develop from different MIC processing events, and only chromosome 1 is a common region throughout all strains. Altogether, there are 3 subfamilies including 9 different chromosomes formed for this locus. This gene was studied according in the JRB310 and JRB510 strains, as these strains were available and readily mate with each other. The JRB310 strain had only one version of the locus, then characterized as 310, and now understood as the 3 allele. The JRB510 strain contained two different versions of the gene, referenced as 510A and 510B, or the A and B allele (Seegmiller et al, 1996). These three versions of the gene will be used as the reference as we compare this section of the genome for all of the other strains. These three alleles were distinguished according to their fragment sizes when cut with the restriction endonuclease Pst1 (New England Biolabs). Previous sequencing and restriction studies allowed for the development of the genetic map as shown in Figure 1.

Figure 1. *Genetic map outlining the 81 Gene Locus post amplification.* The map shows the two restriction endonuclease sites, as well as the fragments that the amplicon is broken into when the sites are cut. The A allele has a functioning #2 restriction site, the B allele has a functioning #1 site, and the 3 allele contains both.



These different versions of the gene have been characterized as alleles in the *O. trifallax* as they are believed to be diploid in nature in the micronucleus (Seegmiller et al, 1997). Yet, some of these ciliates have been identified as having all three versions of the gene, showing a possible triallelic/triploid condition. This condition has only been identified as isolates from cultured samples, and not from any samples whose DNA was extracted after being removed from natural habitats (Seegmiller et al, 1997). The possibility of a triallelic/triploid condition in a natural isolate could have implications for mating population structure as well as population dynamics, which would influence genomic studies of the organism.

Materials and Methods

Sample collection and DNA isolation

Samples of all 21 strains were provided by Robert Hammersmith's Lab at Ball State University. These samples were collected over a number of years from a variety of different locations throughout the state of Indiana. Indicator numbers were provided according to location and collection. The first three letters (TLA/JRB) indicate the location that the sample was taken from and which collection it was involved in. The numbers indicate the sample and isolate numbers. In order to keep these strains straight, this system has been standardized, so all JRB310 samples come from the same original isolate (Zoller et al 2012).

For DNA isolation, cultures of each individual strain were prepared. 600 to 800 of this ciliate culture was then collected in a microcentrifuge tube, and centrifuged for 3 minutes at top speed. After removing the supernatant from the resulting tube, a 5% chelex solution was added. This was then vortexed briefly, and incubated at 95°C. Once completed, it was vortexed and subsequently spun at top speed for 1 minute. The supernatant was then isolated and the resulting DNA was standardized to 50 ng/µL.

PCR Amplification

Amplification was carried out using the Carolina Biological PCR Reaction Bead system. Approximately 100 ng of isolated DNA from each strain was added to the beads, along with our forward (LCR1:ATAACAAATAAATCTCTACTTTAAC) and reverse primers (VHO':GCAATCAAGAAAGATGCCTAC), as previously described by William and Herrick, 1991. Each of these primers was brought to a final concentration of 0.5 µM each in every reaction tube.

The Amplification was completed using a standard hot top thermal cycler. The mixture was run for 35 cycles of 97°C for 15 seconds, 48° C for 1 minute, and 72°C for 1 minute. After this was completed, reaction mixes were stored at -20°C overnight before being used for sequencing studies.

Sequencing

Samples for all 21 strains being studied were sent to the sequencing corps at Notre Dame University and sequencing using the previous two primers mentioned, as well as two others designed specifically for the purpose of complete consensus on final sequences (SQF: GGTCTTGCTAAAGTAGCTGA) (SQR:TTGCACCAATCGTAAATGTT). In order to complete the reactions, DNA

concentrations were measured using the Nanodrop ND-2000 (Thermo Scientific), and 20 ng of DNA was loaded into each reaction. Sequencing was performed using BigDye® Terminator v3.1 chemistry (Applied Biosystems) and the 3730 DNA Analyzer (Applied Biosystems). This resulted in the production of four different sequences showing different sections of the amplicon

Once the samples were acquired, consensus samples were produced using two different programs. Sequence Scanner version 1.0 was used in order to produce trace files and show the confidence in reads. These sequences were then run through ClustalW Omega in order to create alignment files as well as produce consensus sequences for each sample. This program was also used to produce the percent Identity matrix as well as the phylogenetic tree.

In order to determine whether the strain was homozygous or heterozygous, the restriction endonuclease sites relating to studies involving the Pst1 enzyme were used for comparison. This allowed for the identification of representative alleles, as well as which variants occurred in each strain.

PCR Amplification

After PCR was completed, the sample concentrations were found using the nanodrop, and all came in with a concentration of at least 200 ng/µL. This indicated successful amplification of all samples, and were standardized to 10 ng/µL.

Sequencing

When sequencing was completed, the samples were analyzed using the alignment tool connected to ClustalW Omega. This software allowed for allelic diversity comparisons, the development of a phylogenetic tree, and a percent comparison matrix. Of the 21 strains used, three were unable to be sequenced as multiple attempts at reactions led to incomplete traces. These three included the two more distantly related samples, SHBA and JRA11, and JRB51. JRB51 showed indications of having a natural triallelic condition, but repeated attempts at sequencing failed. JRB37 was analyzed and portion of the consensus sequence was found, but a large portion of the sequence is still unknown. Despite this, the strain was still used for partial analyses.

-Allelic Diversity

The allelic diversity was identified from sequences by the presence or lack of a restriction site in two areas. The three previously identified alleles include the 3, A, and B alleles, as mentioned in the literature review. The two restriction sites used to separate all of the alleles are shown in alignment (Figure 2A and 2B). There was a third restriction location found in the 3/A heterozygotes, but only in this heterozygous state. Each strains allelic identifier is provided in Table 1.

Along with the basic identification of the alleles, there are more important points of note. The standard restriction site sequence for Pst1 is CTGCAG. The most common variation point was the 5^{th} base, as the adenine would convert to a thymine. This occurred at the first restriction site in the A allele, leaving only the second site intact. The B allele showed variation in both the 2^{nd} and 5^{th} base of the later restriction site in the sequence. This was only present in the A/B heterozygotes.

The A homozygous condition also showed two different variations. First, was a single base change in the first restriction site to CTGCTG, which was present in the JRA620 strain. The other three samples, the TLA samples, all showed a double mutation in the first restriction site resulting in a sequence of CAGCTG, yet there was no heterozygous condition and no variation in the second restriction site. This would indicate a variation on the A allele, what we

will from here on out indicate as the A* allele. The heterozygous A/B condition only showed the single mutation, suggesting a greater similarity to the normal A allele.

The third restriction site that was found during analysis was only found in the two samples that were the 3/A heterozygote. These two samples, the TLA25 and JRB615, are heterozygous in a position showing a variation in the 5th base. Every other sample shows a sequence of CTGCTG, while these two samples show the variation. This small change is shown in Figure 2C, but no other variation occurs in all of the other samples.

Figure 2. *Sequence alignment of the (A.) first restriction site, (B.) second restriction site, and (C.) the third novel restriction site of all of the usable sequences.* The strains with the correct sequence necessary for the restriction cut are highlighted, and the star underneath represents a consensus sequence.

(A.)

```
EMBOSS_TLA13      CCCATACCTCCAGCTGCTGTAGTATAAGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_TLA17      CCCATACCTCCAGCTGCTGTAGTATAAGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_TLA111     CCCATACCTCCAGCTGCTGTAGTATAAGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB211     CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB615     CCCATACCTCCTGCTGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB92      CCCATACCTCCAGCAGCTGTAGTATAAGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_TLA25      CCCATACCTCCTGCTGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB614     CCCATACCTCCTGCTGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB53      CCCATACCTCCAGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB510     CCCATACCTCCAGCAGCTGTAGTATAAGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRA52      CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRA620     CCCATACCTCCAGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB310     CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB63      CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB322     CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB27      CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
EMBOSS_JRB317     CCCATACCTCCTGCAGCTGTAGTATATGTAGCAACTAATTTTTTATTGAAGGCATCTGTG
                  ********** ** ********** **********************************
```

(B.)

```
EMBOSS_TLA13      ATACACTTACGCAGAATAGAATTTTCTTAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_TLA17      ATACACTTACGCAGAATAGAATTTTCTTAGACCACCACCTGCAGCATAAATTTCTTCCAT
EMBOSS_TLA111     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCATAAATTTCTTCCAT
EMBOSS_JRB211     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB615     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB92      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_TLA25      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB614     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB53      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCTGCGTAAATTTCTTCCAT
EMBOSS_JRB510     ACACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCTGCGTAAATTTCTTCCAT
EMBOSS_JRA52      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRA620     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB310     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB63      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB322     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB27      ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
EMBOSS_JRB317     ATACACTTACGCAGAATAGAATTTTCTGAGACCACCACCTGCAGCGTAAATTTCTTCCAT
                  * ************************* ************** ** **************
```

(C.)

Table 1. *Allelic categories including hetero- and homozygotes.* According to sequence alignment, each strain could be separated into one of four categories. Two remaining categories with no representative strains, which would be possible, just not present in our samples.

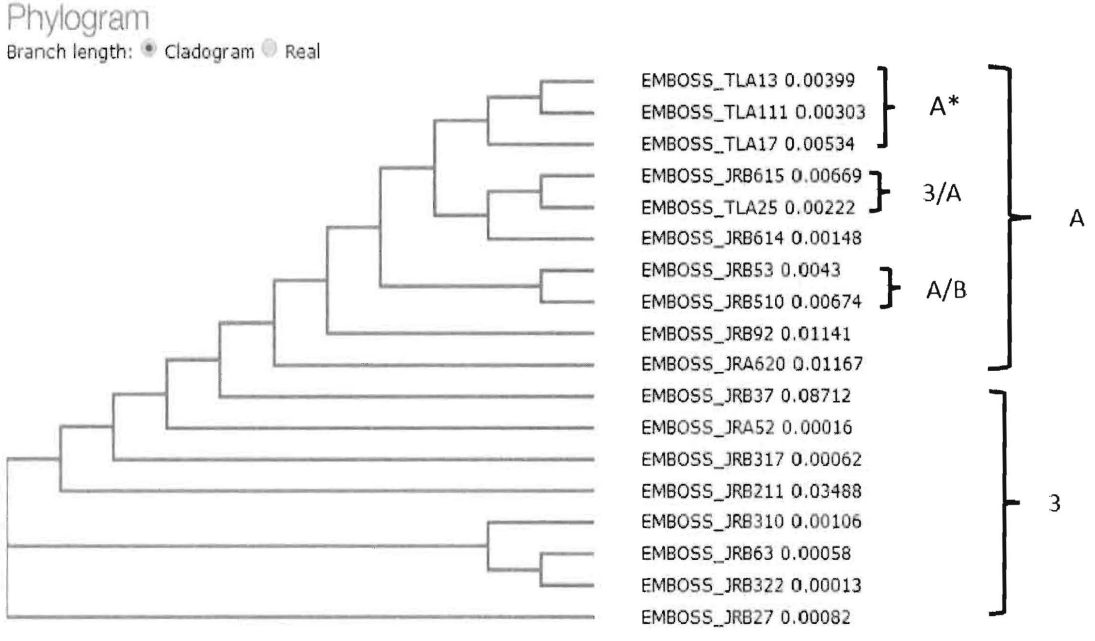
Alleles Present	*O. trifallax* Strains
3	JRB 310, JRB211, JRA52, JRB37, JRB317, JRB27, JRB63, JRB322, JRB92
A	JRA620, TLA13, TLA111, TLA17
B	None
3/A (Heterozygote)	TLA25, JRB615
3/B (Heterozygote)	None
A/B (Heterozygote)	JRB510, JRB53

-Phylogenetic Tree

The ClustalW Omega software allows for the collection of sequences into effective groups in order to allow for further analysis. Each sequence shows large sections of conserved sequence along with smaller areas of variation. This function of the software allows for immediate groupings and separation. This cladogram is located in Figure 3.

The tree prepared here shows three major separate original arcs beginning at the bottom left. These three arcs do not correlate to the three alleles previously identified, but the strains are generally broken up into the allele groups previously shown in figure 2.

Figure 3. *Cladogram showing each of the individual strains and how they separate.* The cladogram was created by the ClustalW Omega software by aligning all of the sequences together. Once this was completed, it sorted all of the strains into groups.



-Percent Identity Matrix

The ClustalW Omega software also was used to create a percent identity matrix that make a comparison of all sequences down to the tenth of a percent (Figure blah). This allows for a very exact comparison of each of the strains. This allowed for the discrimination of individual differences between even the strains that had shown to contain the same allele. Due to the size of the chart, it is provided on the next page.

Table 2. *Percent Identity Matrix showing relative comparisons between the conserved sequences once put in the alignment tool.* The percent identity matrix provides a more precise comparison between multiple sequences than the cladogram for increased understanding.

Percent Identity Matrix - created by Clustal2.1

1: EMBOSS_TLA13	100.00	99.07	98.89	93.97	96.70	96.43	97.14	96.80	96.73	96.69	95.33	95.80	87.77	95.19	95.26	95.33	95.39	95.56
2: EMBOSS_TLA17	99.07	100.00	98.89	93.69	96.88	96.43	97.30	96.96	96.48	96.77	95.07	95.55	87.07	95.25	95.31	95.39	95.23	95.11
3: EMBOSS_TLA111	99.30	98.89	100.00	93.97	96.70	96.43	97.14	96.80	96.73	96.69	95.33	95.80	87.77	95.19	95.26	95.33	95.39	95.56
4: EMBOSS_JRB211	93.77	93.69	93.97	100.00	93.64	95.00	94.16	94.33	95.27	94.89	96.44	95.03	86.05	96.36	96.29	96.36	96.36	96.36
5: EMBOSS_JRB615	96.87	96.88	96.70	93.64	100.00	96.96	99.11	99.11	96.89	96.95	95.76	96.10	87.28	95.92	96.01	96.09	95.91	96.01
6: EMBOSS_JRB92	96.32	96.43	96.43	95.00	96.96	100.00	97.14	97.47	96.80	97.03	97.03	97.23	89.32	97.25	97.32	97.40	97.24	97.25
7: EMBOSS_TLA25	97.23	97.30	97.14	94.16	99.11	97.14	100.00	99.49	96.97	97.11	96.21	96.38	88.46	96.46	96.55	96.63	96.45	96.46
8: EMBOSS_JRB614	96.96	96.96	96.80	94.33	99.11	97.47	99.49	100.00	97.13	97.28	96.46	96.37	88.74	96.71	96.80	96.88	96.70	96.71
9: EMBOSS_JRB53	96.43	96.48	96.73	95.27	96.89	96.80	96.97	97.13	100.00	98.90	97.14	97.48	89.11	96.90	96.81	96.73	97.05	97.23
10: EMBOSS_JRB510	96.51	96.77	96.69	94.89	96.95	97.03	97.11	97.28	98.90	100.00	96.60	96.86	89.02	96.60	96.52	96.60	96.60	96.60
11: EMBOSS_JRA52	95.43	95.07	95.33	96.44	95.76	97.03	96.21	96.46	97.14	96.60	100.00	97.73	91.07	99.78	99.56	99.63	99.63	99.78
12: EMBOSS_JRA620	95.73	95.55	95.80	95.03	96.10	97.23	96.38	96.37	97.48	96.86	97.73	100.00	90.04	97.65	97.40	97.48	97.47	97.65
13: EMBOSS_JRB37	86.90	87.07	87.77	86.05	87.28	89.32	88.46	88.74	89.11	89.02	91.07	90.04	100.00	90.66	90.23	90.26	90.57	90.96
14: EMBOSS_JRB310	95.29	95.25	95.19	96.36	95.92	97.25	96.46	96.71	96.90	96.60	99.78	97.65	90.66	100.00	99.79	99.79	99.72	99.58
15: EMBOSS_JRB63	95.21	95.31	95.26	96.29	96.01	97.32	96.55	96.80	96.81	96.52	99.56	97.40	90.23	99.79	100.00	99.93	99.79	99.64
16: EMBOSS_JRB322	95.29	95.39	95.33	96.36	96.09	97.40	96.63	96.88	96.73	96.60	99.63	97.48	90.26	99.79	99.93	100.00	99.86	99.65
17: EMBOSS_JRB27	95.13	95.23	95.39	96.36	95.91	97.24	96.45	96.70	97.05	96.60	99.63	97.47	90.57	99.72	99.79	99.86	100.00	99.86
18: EMBOSS_JRB317	95.29	95.11	95.56	96.36	96.01	97.25	96.46	96.71	97.23	96.60	99.78	97.65	90.96	99.58	99.64	99.65	99.86	100.00

Analysis

The evidence provided by three analyses showed that there was a clear, distinctive group of three separate alleles representing the 3, A, and B alleles as predicted by previous restriction studies (Seegmiller et al,1996, Seegmiller et al, 1997, William and Herrick 1991). These studies allowed for easy identification of the separate groups once sequencing had occurred. As seen in the separation of strains into their alleles, the 3 allele characterized in JRB310 is the most common allele present. But when compared to the other strains, there is an extremely high degree of similarity with over a 93% similarity to every strain. In fact all strains were over 93% similarity, excluding the single incomplete strain used, JRB37. Despite these degrees of difference, the sequencing would indicate that each of these strains of the organism have separated into small subpopulations allowing for selective sharing of genetic material.

In the comparison of different allele types, the only two strains showing the presence of the B allele were from the same water source and the same collection (JRB510 and JRB53). In the same way, the only forms in which we see the double mutation, which is the A* variation, of the first restriction site were all 3 from the same water source and the same collection (TLA13, 17, and 111). These small subpopulations would also indicate the high degree of homozygosity among the strains. As these subpopulations separate more effectively and only mate with other members of their subpopulation, single alleles will become more common and genetic variation will decrease because of the increased isolation.

Despite these separations, we do see consistencies throughout multiple different locations and collections. The TLA25 and JRB615 both showed the heterozygous 3/A condition despite originating from completely different locations. There was a presence of the 3 allele in every single collection and location. It may have been present in homozygous and heterozygous forms, but the 3 allele was shown to occur in all of the different collections.

The percent identity matrix and phylogenetic tree indicated the same separation according to alleles that was identified through the restriction site analysis. The strains grouped into one allele all show at least a 99% similarity according to the matrix. The cladogram also separates into the groups with the heterozygotes together and all of the individual homozygous allele groups are sorted together.

All three tests indicate the same result, and with all of these results we did not detect any natural triallelic conditions among the samples. The JRB51 which failed multiple times in sequencing looked like it may indicate a triallelic because of the way the sample was read. Despite the indications, there is no way we can truly indicate from the samples that this is true.

Future Directions

To further understand the genetic composition of this organism, multiple further analyses are necessary. First, a more comprehensive analysis into the changes that occur in the different allele groups should be done in order to identify possible location of SNPs, and all forms of variation in and among similar allele groups. Second, the strain JRB51 should be greater characterized using means different than basic sequencing. In order to identify if it truly is a triallelic, it could be identified more fully through a large comprehensive transformation study. Finally, the third possible restriction site in the 3/A heterozygote strains should be studied more closely using the sequencing data. This site could provide another tool for discriminating different groups from each other, and offer a better understanding of different allelotypes. Along with furthering the study of the Pst1 restrictions sites, expanding the restriction site analysis beyond Pst1 could provide a greater understanding of the alleles. This could provide a much better understanding of groupings inside of each previously known allele.

LITERATURE CITED

Prescott, D. M. (1994). The DNA of ciliated protozoa. *Microbiological reviews*,*58*(2), 233.

Jung, S., Swart, E. C., Minx, P. J., Magrini, V., Mardis, E. R., Landweber, L. F., & Eddy, S. R. (2011). Exploiting Oxytricha trifallax nanochromosomes to screen for non-coding RNA genes. *Nucleic acids research*, *39*(17), 7529-7547.

Herrick, G. (1994, February). Germline-soma relationships in ciliated protozoa: the inception and evolution of nuclear dimorphism in one-celled animals. In*seminars in DEVELOPMENTAL BIOLOGY* (Vol. 5, No. 1, pp. 3-12). Academic Press.

Seegmiller, A., Williams, K. R., & Herrick, G. (1997). Two two-gene macronuclear chromosomes of the hypotrichous ciliates Oxytricha fallax and O. trifallax generated by alternative processing of the 81 locus. *Developmental genetics*, *20*, 348-357.

Doak, T. G., Cavalcanti, A. R., Stover, N. A., Dunn, D. M., Weiss, R., Herrick, G., & Landweber, L. F. (2004). Sequencing the< i> Oxytricha trifallax</i> macronuclear genome: a pilot project. *Trends in Genetics*, *19*(11), 603-607.

Seegmiller, A., Williams, K. R., Hammersmith, R. L., Doak, T. G., Witherspoon, D., Messick, T., ... & Herrick, G. (1996). Internal eliminated sequences interrupting the Oxytricha 81 locus: allelic divergence, conservation, conversions, and possible transposon origins. *Molecular biology and evolution*,*13*(10), 1351-1362.

Williams, K. R., & Herrick, G. (1991). Expression of the gene encoded by a family of macronuclear chromosomes generated by alternative DNA processing in Oxytricha fallax. *Nucleic acids research*, *19*(17), 4717-4724.

Zoller, S. D., Hammersmith, R. L., Swart, E. C., Higgins, B. P., Doak, T. G., Herrick, G., & Landweber, L. F. (2012). Characterization and Taxonomic Validity of the Ciliate< i> Oxytricha trifallax</i>(Class Spirotrichea) Based on Multiple Gene Sequences: Limitations in Identifying Genera Solely by Morphology. *Protist*, *163*(4), 643-657.