Analysis of the Molecular Role of *COMT* in Bipolar Disorder

Sharon A. Lewis* and Cherie Ognibene

Chemistry Department, Langston University, Langston, Oklahoma 73050

Bipolar disorder is a chemical imbalance of neurotransmitters in the brain. This mental condition causes dramatic mood swings characterized by episodes of elation and high activity alternating with periods of low mood and low energy. Bipolar disorder affects approximately 5.7 million American adults, or about 2.6 percent of the U.S. population age 18 and older in a given year. As many as twenty percent of those with this disease who aren't treated commit suicide, according to National Institute of Mental Health. Some researchers think that alcohol abuse and psychiatric disorders might share vulnerability-enhancing genes. Many genes have been implicated in susceptibility to bipolar disorder. *COMT*, the gene symbol for catechol-O-methyltransferase, is involved in the breakdown of the catecholamne neurotransmitters. The enzyme introduces a methyl group to the catecholamine which is donated by S-adenyosyl methionine (SAM). The most common polymorphism in the membrane bound form of *COMT* is the Val/Met polymorphism resulting in low activity Met allele which increases susceptibility to bipolar disorder and may increase the chances for rapid cycling in these bipolar patients. This polymorphism is linked to psychiatric disorders related to the metabolism of catecholamine neurotransmitters like norepinephrine, dopamine, and epinephrine. The objective was to teach students to identify distinctive features in the protein sequences of *COMT* that show the variations in molecular genetics for multiple organisms and to include as many human sequences as possible to try to explain the phenotypic variations in populations. The research objective was to accumulate all of the information contained in the various databases on *COMT* to use bioinformatic visualization technology (multiple sequence alignments, phylogenetic trees, molecular interactions) for students to be able to understand the structure, polymorphisms, phenotypic variations between male and female, and population diversity for *COMT* as it pertains to bipolar disorder.

*Corresponding Author: Tel: (405) 466-3316; Fax: (405) 466-3638; E-mail: salewis@lunet.edu

Financial Support: This research was supported by the Oklahoma IdeA Network of Biomedical Research Excellence (OK-INBRE) National Institutes of Health Grant # P20RR016478-07; OK-INBRE/Oklahoma State Regents for Higher Education-Mini-Grant; Carnegie Mellon University, Pittsburgh Supercomputing Center (PSC) -National Institutes of Health Grant #5 T36 GM008789.

List of Abbreviations: BPD=bipolar disorder; *COMT*= Catechol-O-Methyltransferase; *S-COMT*=soluble form of *COMT*; *MB-COMT*=Membrane bound form of *COMT*; SNP=Single nucleotide polymorphisms; MSA=Multiple Sequence Alignments

Introduction

Bipolar disorder is a chemical imbalance of neurotransmitters in the brain. This mental condition causes dramatic mood swings characterized by episodes of elation and high activity alternating with periods of low mood and low energy. Sixty-six percent of people with bipolar disorder have at least 1 close relative who has it, too, or who has major depression [1]. According to the National Institute of Mental Health, as many as, twenty percent of untreated bipolar disorder patients commit suicide [1]. More than 12 million women and 6 million men in the US are affected by depressive illnesses in any given year, NIMH

estimates [1]. Fewer than half of people with depression seek treatment [1]. For at least 90% of those who have bipolar disorder the condition is recurrent.

Structure of *COMT*

One of the proteins implicated in susceptibility to bipolar disorder is catechol O-methyl-transferase. Catechol-O-Methyltransferase is the protein name for *COMT*. The O in the name stands for oxygen. This enzyme was first discovered by biochemist Julius Axelrod [2]. The enzyme commission number is 2.1.1.6. The crystal structure of catechol O-methyl-transferase is available in the Protein Database as Image ID: 1vid [3] (Figure 1).

Figure 1. Catechol-O-Methyltransferase protein

This image was resolved at a resolution of 2.00 angstrom and contains an alpha/beta fold with eight alpha helices and seven beta sheets. Alpha helices 4-8 and the parallel beta sheets 1-5 compose a Rossmann fold, which is also the nucleotide binding motif found in alcohol dehydrogenase. The C terminal has two anti-parallel beta sheets, which are beta 6-7 [4]. The active site includes the co-enzyme-binding motif and the catalytic site situated in the vicinity of the Mg^{2+} ion. The active site consists of residues from the amino terminal helices, through the nucleotide-binding fold to the loop region of the C-terminal beta strands (B6-B7) [4].

The *COMT* is involved in the breakdown of the catecholamine neurotransmitters. The enzyme introduces a methyl group to the catecholamine which is donated by S-adenyosyl methionine (SAM). This gene is involved in at least three metabolic pathways to include the following: tyrosine metabolism, estrogen metabolism, and dopamine metabolism (Figure 2) [2].

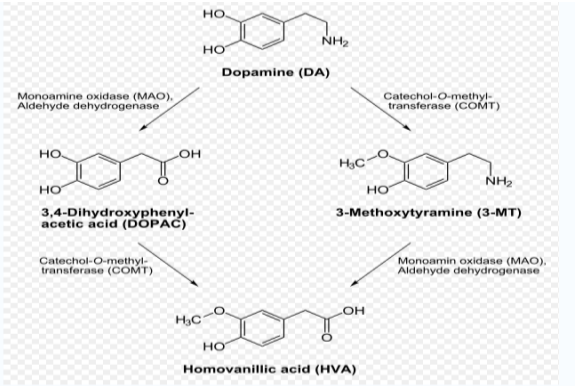

Figure 2. Dopamine Metabolism

The cytogenetic location for *COMT* is chromosome 22q11.21-q11.23. The molecular location is from base pairs 18,309,308 to 18,336,529. This gene has six exons. The amino acid sequence for *COMT* in humans and the secondary structures are displayed in Figure 3 [1]. The benefit of displaying the protein structure is to show that polymorphism 158 seems to be located in the loop region.

Polymorphisms in *COMT*

COMT exists in both a membrane bound form and a soluble form that are produced from alternative promotors on one chromosome 22. The membrane bound form (MB-*COMT*) contains an extra 50 AA at the N-terminus. The most well studied polymorphism in MB-*COMT* is a Val158Met polymorphism. This polymorphism is linked to psychiatric disorders because of the

function of *COMT* in the metabolism of catecholamine neurotransmitters. *COMT* acts on the prefrontal cortex to maintain appropriate levels of norepinephrine, dopamine, and epinephrine. The prefrontal cortex is involved with personality, planning, inhibition of behaviors, abstract thinking, emotion, and working (short-term) memory. To function efficiently, the prefrontal cortex requires signaling by neurotransmitters that *COMT* helps to maintain [5].

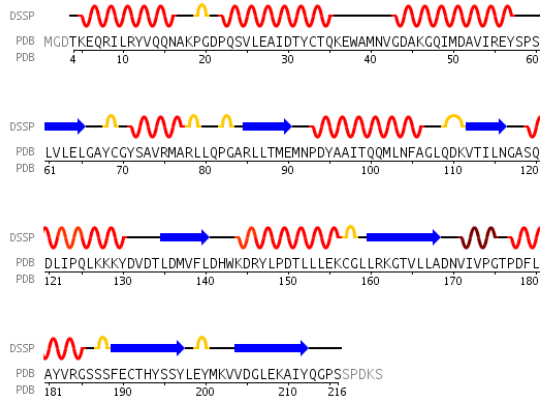

Figure 3. The Amino Acid Sequence of *COMT* and the Secondary Structures

The cytoplasmic soluble form (S-*COMT*) is the shorter form, which is produced by the liver, kidney, and blood. This form of the enzyme helps control the levels of certain hormones. The most well studied polymorphism in S-*COMT* is Val108Met polymorphism. The difference in activity between the two forms is a function of altered protein levels and is not the result of differences in the specific activities. MB-*COMT* has approximately a 10-fold greater affinity for dopamine and noradrenaline than S-*COMT* [5].

COMT Phenotypic Variations between Male and Female

Regarding its role as a regulator of catecholestrogens, researchers determined that the neurochemical and behavioral phenotype for *COMT* differs between male and female knockout mice and its enzyme activity in human brain is higher in women than men [6]. A separate group of researchers studying humans identified a significant association between bipolar disorder and the Val/Met polymorphism stating that it may increase the chances for rapid cycling in women. This research involved 217 unrelated men and women from the Ashenazi Jewish population all previously diagnosed with bipolar disorder. Using SNP genotyping, they calculated the odds ratio for allele and haplotype frequencies in bipolar disorder compared to schizophrenia patients [7]. The gene that encodes monoamine oxidase A (*MAOA)* and *COMT* met^{158} allele have been associated with obsessive compulsive disorder in men. This research involved 87 unrelated Caucasian men and women [6].

Homozygosity for **Val^{158}** predisposes women, not men, to schizophrenia, and predisposes these women to a sensation-seeking personality. Homozygosity for **Met^{158}** predisposes women to harm avoidance or neuroticism. Caucasian women also were found to have panic disorder with this allele, but Asian women experienced the panic disorder with the Val^{158} allele. The women experience panic disorder with the Met^{158} allele, but men experience obsessive-compulsive disorder. The protein and activity levels of *COMT* rise considerably in men in their thirties and fifties [8].

COMT Allelic Frequencies

Population diversity is an interesting area as it pertains to polymorphisms in genes implicated in susceptibility to bipolar disorder. Continuing to concentrate on *COMT*, the Val158Met polymorphism seen in Caucasians and African-Americans differs from the polymorphism in other populations. In Korean and Japanese

populations, the common polymorphism is Alanine to Serine [5]. Another group of researchers noted that in Caucasians (81 women, 53 men), pure African (11 women and 4 men), and Plains American Indians (149 women, 103 men), the low activity Met158 allele was associated with anxiety in women. In US Caucasians, the frequency of the Met158 allele was 0.42 and the frequency of the Val158 allele was 0.58. In Plains American Indians the Met158 allelic frequency was 0.26 and Val158 was 0.74. In various pure African populations and East Asians, the Met158 allelic frequency was 0.24. These researchers acknowledged that the Met158 allele in addition to being linked to bipolar disorder is also linked to alcoholism, increased alcohol intake in social drinkers, and obsessive compulsive disorder in men [9].

In addition to learning to use tools for bioinformatics, another objective was to teach students to identify distinctive features in the protein sequences of *COMT* that show the variations in molecular genetics for multiple organisms and to include as many human sequences as possible to try to explain the phenotypic variations in populations. After reading a few journal articles on the molecular genetics of *COMT*, we noticed that the articles were written covering several populations, but not the African American population. We agreed that differences will exist between the Yoruban (YRI) African population and the African American population since there are differences between Figure 4-the Yoruban Linkage Disequilibrium (LD) Plot for *COMT*, and Figure 5-the European (CEU) LD Plot for *COMT*, and Figure 6: Han Chinese (CHB) in Beijing, China LD Plot for *COMT.* When comparing the LD patterns corresponding to different populations we went to the International HapMap Project [10]. We selected HapMap Data Released 27 Phase II & III, Feb 2009, on NCBI B36 assembly, dbSNP b126. The Landmark Search query was *COMT*. Under Details, the Entrez gene selected was NM_000754. On the right hand side of the screen, under Reports & Analysis, we selected Annotate LD Plot, and selected the Configure button. We accepted the default parameters but specified the population and orientation.

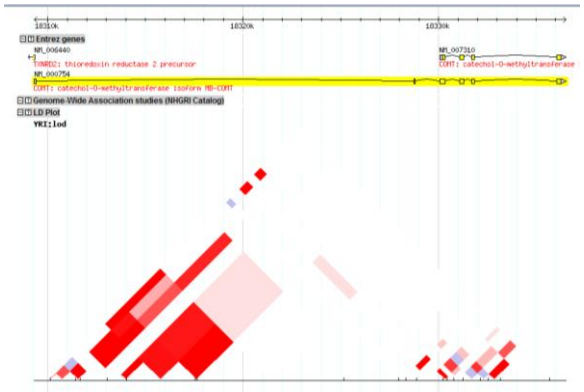

Figure 4. Yoruban Linkage Disequilibrium (LD) Plot for *COMT*

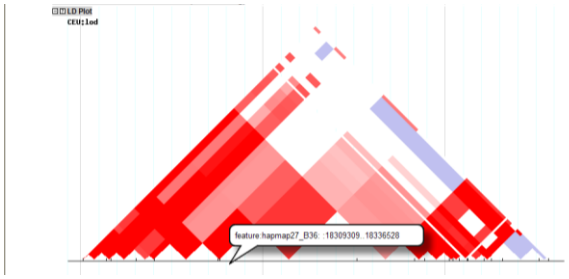

Figure 5. European (CEU) LD Plot for *COMT*

The students were able to obtain additional information regarding SNP genotyping in African-Americans from multiple websites to include the National Center for Biotechnology Information (NCBI) [11], Applied Biosystems Inc. [12], and the Pharmacogenetics and Pharmacogenomics Knowledge Base [13]. The ability to obtain the dbSNP (rs#) genotype query from NCBI Single Nucleotide Polymorphisms

was valuable in allowing the students to generate a list of rs#. At that website, they selected RS, gene, "COMT," next, African American Population with Genotypes. The result contained information for 71 individuals and 970 genotypes. Once the students selected "View Population" and "Display Results" the result was a report of the Genotype and Allele Frequency (Table 1). More specifically, the Genotype and Allele Frequency Report for four African American females revealed the single nucleotide polymorphism. Of the four individuals, the genotype was conclusive for only one of the African American females. More information is needed for African American populations.

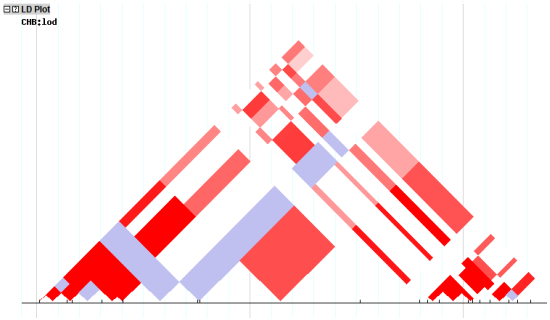

Figure 6. Han Chinese (CHB) in Beijing, China (LD) Plot for *COMT*

At the Applied Biosystems website, students were directed to select "Products," "Search for Taqman SNP Genotyping Assays," and identify a specific rs#, such as rs#165599. They were then directed to the National Center for Biotechnology Information, Single Nucleotide Polymorphism Database. In addition to viewing the SNPs listed for African-American populations, they viewed the allelic frequencies for different populations under the title Population Diversity (Table 2). For this report, only one African American population was noted having 120 samples with an adenine frequency of 0.367 and a cytosine frequency of 0.633.

Table 1. Genotype and Allele Frequency Report.

Individual Group	Sex	rs4680
African American	Female	T/T
African American	Female	
African American	Female	
African American	Female	N/N

Table 2. Population Diversity Reporting Allelic Frequency for *COMT* Locus.

ss#	Individual Group	No# of Individuals	Alleles	
Ss66862082	Global	180	A=.217	C=.783
	Global	52	A=.250	C=.750
	European	68	A=.191	C=.809
	African American	120	A=.367	C=.633
	Asian	90	A=.078	C=.922
	Sub-Saharan African	120	A=.425	C=.575
	Asian	24	A=.167	C=.833

Alternatively, the Pharmacogenetics and Pharmacogenomics Knowledge Base [13] allowed the students to obtain additional information on *COMT* SNPs. To use this search engine, they queried "*COMT*," variant genes, scroll down, dbSNP ID, rs#4680, to find the polymorphism that covered the Val/Met polymorphism (Table 3). Approximately 64.08% of the population had the G allele and 35.92% had the A allele.

rs4680 *[Homo sapiens]*

CCCAGCGGATGGTGGATTTCGCTGGC[A/G]TGAAGGACAAGGTGTGCATGCCTGA

The Pharmacogenetics and Pharmacogenomics Knowledge Base rs# also feed into the NCBI database. There were two reports for African Americans on Table 4. In the first report, 46

Table 3. PharmGKB *COMT* SNPs

Chromosome	SNP	Mutation	Location	Amino Acid Change	Occurrence	# Individuals	Method
22:18331271	Rs4680	G/A	Exon	Val/Met	64.08%/35.92%	426	PCR Taqman

samples were submitted having an adenine frequency of 0.326 and a guanine frequency of 0.674. The second report for African Americans involved 22 samples having a cytosine frequency of 0.636 and a thymine frequency of 0.364. This last population of African Americans with C/T alleles differed greatly from the other populations. The African-American population allelic frequencies differed from the Sub-Saharan African population. The one report for the Sub-Saharan African population had 120 samples having an adenine frequency of 0.292 and a guanine frequency of 0.708.

Table 4. Allelic Frequencies from PharmGKB to NCBI

Individual Group	# of Individuals	Alleles	Alleles
Global	164	A=.390	G=.610
European	48	A=.37500	G=.625
African American	46	A=.326	G=.674
Asian	48	A=.271	G=.729
Asian	184	A=.5	G=.500
European	120	A=.517	G=.483
Asian	90	A=.256	G=.744
Asian	88	A=.239	G=.761
Sub-Saharan African	120	A=.292	G=.708
Asian	74	A=.297	G=.689
Asian	90	A=.260	G=.740
Asian	92	A=.480	G=.520
African American	**22**	**C=.636**	**T=.364**

Materials and Methods

The research objective was to accumulate all of the information contained in the various databases on *COMT* to include its structure, polymorphisms, phenotypic variations between male and female, and population diversity as it pertains to allelic frequencies and use this information in visualization technology with such tools as multiple sequence alignment, phylogenetic trees, and gene ontology interactions. The PI established collaborations to teach high school students participating in the Research and Apprenticeship Program and the undergraduate students enrolled in the Introduction to Bioinformatics class and the Introduction to Chemical Research classes to use computers for analyzing molecular biology data. The students learned such techniques as retrieving sequences from the Universal Protein Resource (Uniprot) [14]; learning to use the web-based T-Coffee multiple sequence alignment (MSA) [15] and ClustalW [16]; producing multiple sequence alignments and identifying patterns using the Multiple EM for Motif Elicitation (MEME) [17] program on the web; learning to use Unix Operating System [18] to generate a T-Coffee and ClustalW MSA and MEME again for pattern identification and visualization of the pattern in the MSA using GeneDoc [19]; generating the phylogenetic gene tree and visualizing it using both TreeView [20] and FigTree [21]; generating a species tree using the National Center for Biotechnology Informatics' Taxonomy site; and finally learning gene ontology network (Systems Biology) using Cytoscape [22] available from the National Center for Integrative Biomedical Informatics.

After performing a brief literature review, database searching was necessary to obtain sequences for inclusion in the T-Coffee multiple

sequence alignment in Figure 7. These sequences were selected from Uniprot. Only six complete human sequences were found at this database and were consequently, included in this alignment. WinSCP [23] was the software providing integration between the Uniprot website and Unix. The UNIX system was accessed through Putty [24]. The host organization for Putty use and computer exercises for the T-Coffee multiple sequence alignments was the Carnegie Mellon University, Pittsburgh Supercomputing Center. GeneDoc [19] was the software tool used to view T-Coffee multiple sequence alignment for pattern identification and matching using MEME. The color coding on the multiple sequence alignment showed highly conserved patterns. TreeView [20] and Figtree [21] Software were used to generate the phylogenetic tree for these sequences.

Results and Discussion

In figure 7, the left hand column indicates the species of interest. Concentrate on Human 2-5 in the middle of the chart. Follow each of those sequences to the end of the line. Identify the last amino acid on the right hand side of the chart. For example, the last amino acid for Human 2 is a V (Valine) at position 162. The number 162 is on the right side of the chart. Now count amino acids over, from right to left, until you get to position 158. For example, Human 2 has V at position 162, K at position 161, D at position 160, K at position 159 and V at position 158. Also, Human 5 has V at position 158. But notice that Human 3 & 4 have M (Methonine) at 158. Remember that the most common single nucleotide polymorphism in Caucasians and African-Americans is a Valine to Methonine at position 158 in the longer membrane bound form of *COMT*. This is a clear example of how SNPs alter the genetic code created by the addition or deletion or substitution of a single nucleotide in a gene's long chain. This polymorphism was previously identified by other researchers. The students in this bioinformatics class generated the following multiple sequence alignments to visualize the polymorphism.

After the students grew accustomed to handling a few sequences in a multiple sequence alignment, they were encouraged to generate a MSA for a protein family. Using the methyl-transferase family, a MSA was generated with 96 sequences in Figure 8.

Table 5. Eight Motifs Identified by MEME.

MOTIFS ./meme.html (peptide)

MOTIF	WIDTH	BEST POSSIBLE MATCH
1	21	GPKRILEIGTCCGYSSICMAR
2	29	FDMIFIDADKSQYCKYFEWCLPLLRPGGV
3	21	PPDGHITTCERNPEHCQHARQ
4	34	RKMCSYNEWLMNHPNYTTTFIPIGDGMAISKKCE
5	21	NWQKAGVENQISCIEGPALET
6	15	IVVDNVLWHGRVADP
7	21	ELPRMQVSPEQGQFLCLLAKM
8	21	IEEMEQYAEENHVPIMDRETM

The MEME program was used for pattern identification [17]. This program is curator by the San Diego Supercomputer Center. MEME discovers patterns of amino acids that are distinctive of specific protein families and functions. The output is a very informative graphical display showing the order of identification and location of each of the motifs discovered by the MEME program on each of the sequences (Figure 9). The input was queried for a maximum of 8 motifs, with a minimum motif width of 6 and a maximum motif width of 50 (Table 5) for the sequence of the eight motifs. Again, the total number of sequences queried was 96.

Figure 7. *COMT* Multiple Sequence Alignment

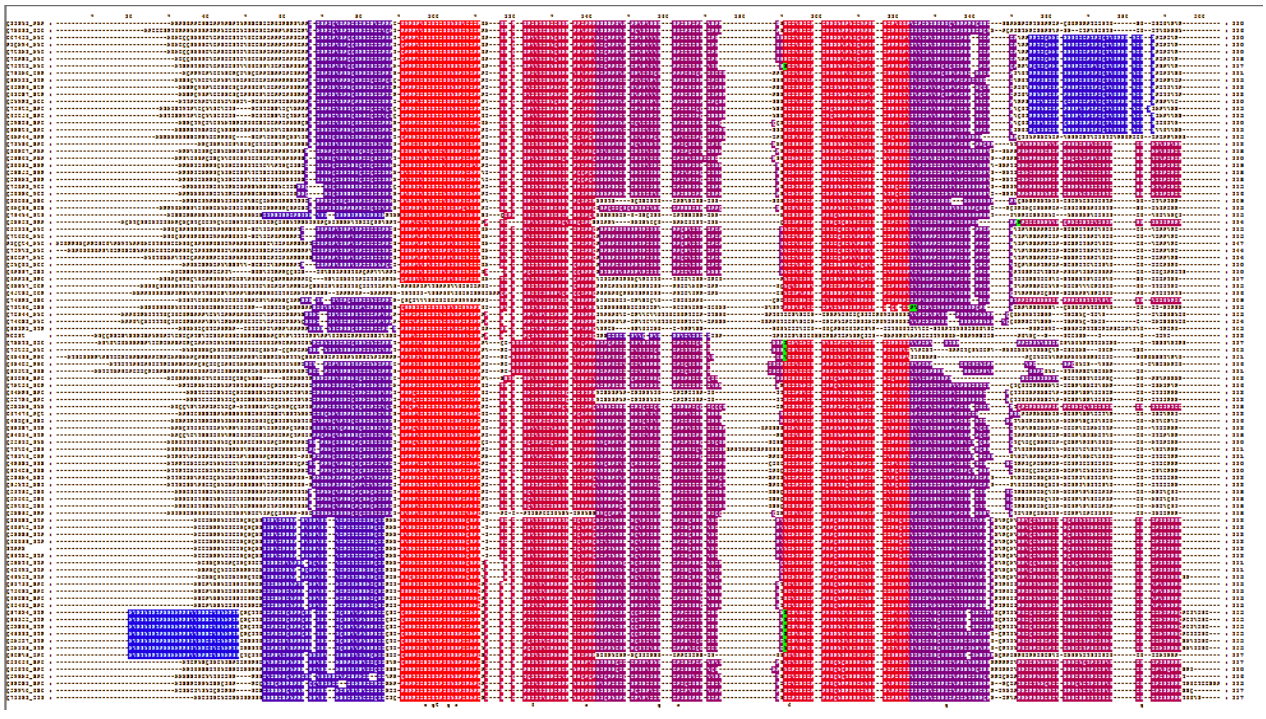

Figure 8. Methyltransferase Superfamily

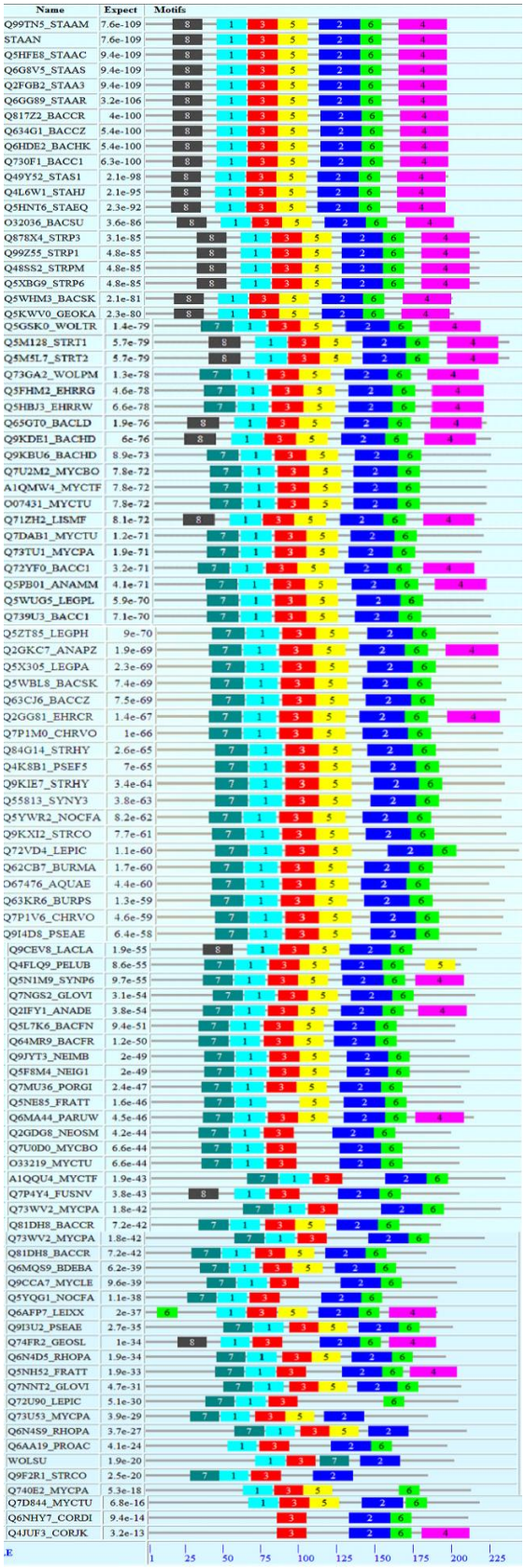

Figure 9. Eight Motifs Identified by MEME

The actual specifics for the MEME algorithm involved creating a Hidden Markov Model by selecting the Dirichlet mixture and Sibbald/Argos Voronoi sequence weighting algorithm; creating a PROFILE-SS script file using Henikoff method, BLOSUM60 matrix, and Gribskov's model for the open gap penalty of 10; extend gap penalty of 1; cutoff of 2.0; subalignments per pair of 5; minimum length of 50; and MEME, ZOOPs model.

Phylogenetic gene trees were constructed using TreeView [20] and Figtree [21] to consider the polymorphism in evolutionary context. Figure 10 shows a TreeView Phylogenetic tree of the organisms listed in the multiple sequence alignment from Figure 7.

The FigTree Phylogeny in Figures 11 and 12 were created using 96 sequences from the methyltransferase family. The goal was to produce a consensus tree using PHYLIP pairwise distances with Dayhoff PAM Matrix; create a neighbor-joining tree; and create a bootstrap consensus tree with 100 replicates of 10 data sets. This data was analyzed using the PHYLIP program consense.

We queried three databases to include the Pharmacogenetics and Pharmacogenomics Knowledge Base [13], Molecular Science Student Workbench [25], and the National Center for Biotechnology Information [11] to obtain a comprehensive list of genes implicated in susceptibility to bipolar disorder. After downloading Cytoscape [22] and JAVA runtime [26], the MIMI Plug-in was selected for the *COMT* query. Cytoscape gave a pictoral of the network of molecular interaction. In addition to showing the bioinformatics students to use the Cytoscape tool, this research attempted to combine these molecular interactions with the multiple sequence alignments and phylogenetic

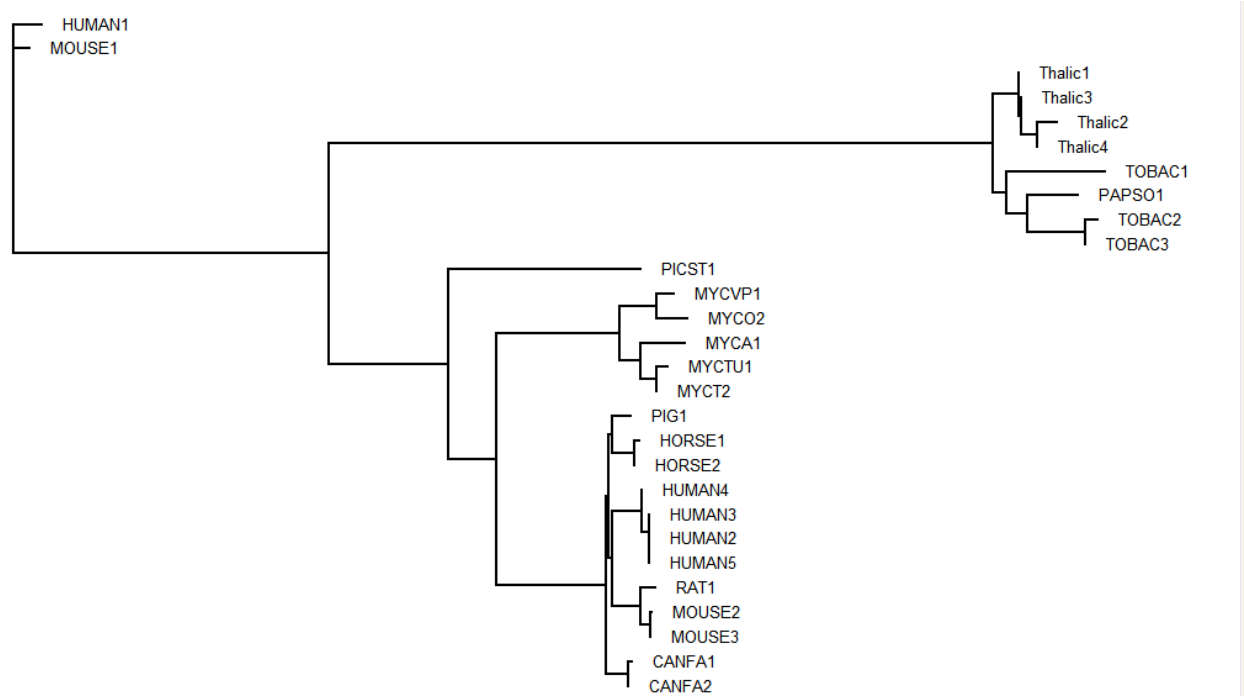

Figure 10. TreeView Phylogenetic Tree

trees. On Table 6 is a list of the bipolar query genes used to generate the Cytoscape Figures 13-17. Cytoscape identified 656 nearest neighbor genes using gene ontology.

Table 6. List of Genes Implicated in Bipolar Disorder

Gene Id	Symbol	Gene Id	Symbol
25120	AANAT	9248	GPR50
627	BDNF	9575	CLOCK
1312	COMT	22873	DZIP1
1812	DRD1	79796	ALG9
1814	DRD3	121278	TPH2
1815	DRD4	2904	GRIN2B
2099	ESR1	319100	TAAR6
2892	GRIA3	4684	NCAM1
2913	GRM3	7184	HSP90B1
2932	GSK3B	4552	MTRR
3360	HTR4	4524	MTHFR
3553	IL1B	57142	RTN4
3613	IMPA2	9145	SYNGR1
3628	INPP1	1848	DUSP6
4128	MAOA	23395	LARS2
6532	SLC6A4	23348	DOCK9
7494	XBP1	1393	CRHBP
8867	SYNJ1	468	ATF4

Teaching students using visualization technology bioinformatics tools allowed students to thoroughly enjoy learning bioinformatics because of the hands on aspect of using the computer. A T-Coffee multiple sequence alignment displayed the most common *COMT* polymorphism, Val158Met, and a phylogram displayed a phylogenetic tree. The addition of genotypes from African-American patients diagnosed with bipolar disorder will allow for a more thorough examination of population diversity. Future work includes identifying the number of nearest neighbor genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID) which uses KEGG; incorporating clinical data from UCSC Genome Browser; performing structural biological analysis of the motifs using Visual Molecular Dynamics (VMD); and studying the SNP allelic differences between African-American population as compared to the Yoruban and Caucasian populations.

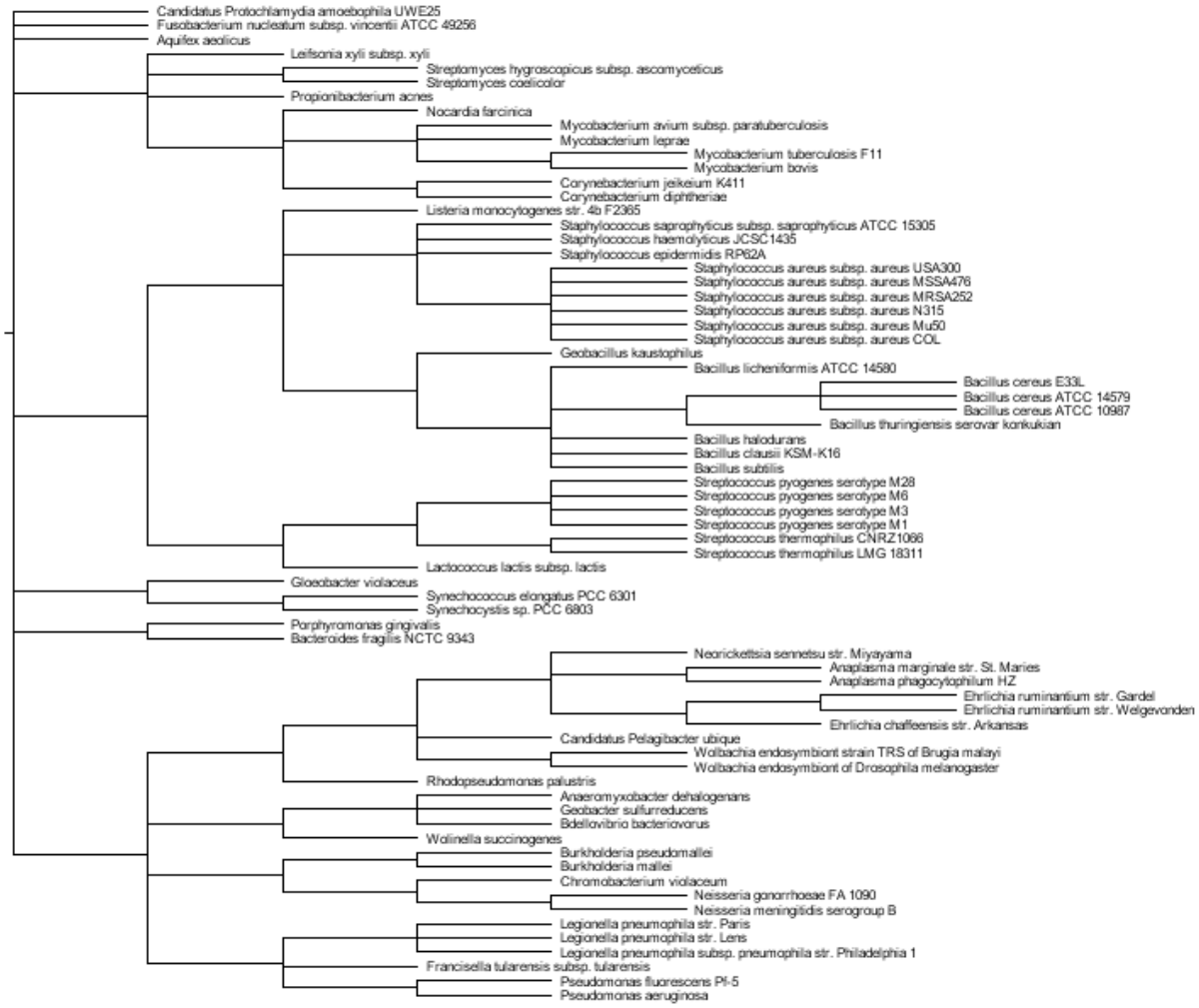

Figure 11. FigTree Phylogeny of Methyltransferase Superfamily

Figure 12. FigTree Phylogeny of Methyltransferase Superfamily

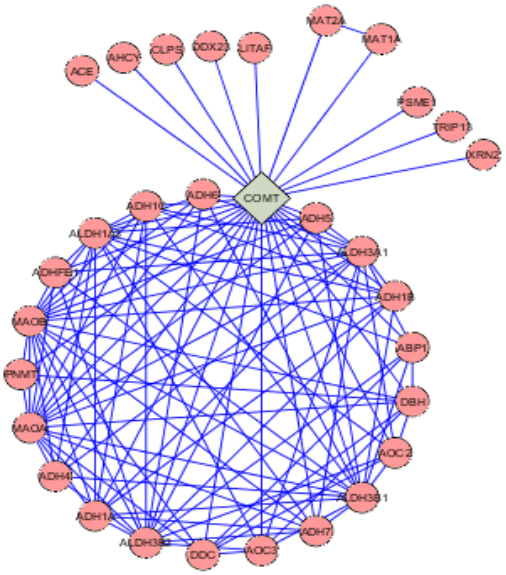

Figure 13. Cytoscape of *COMT*

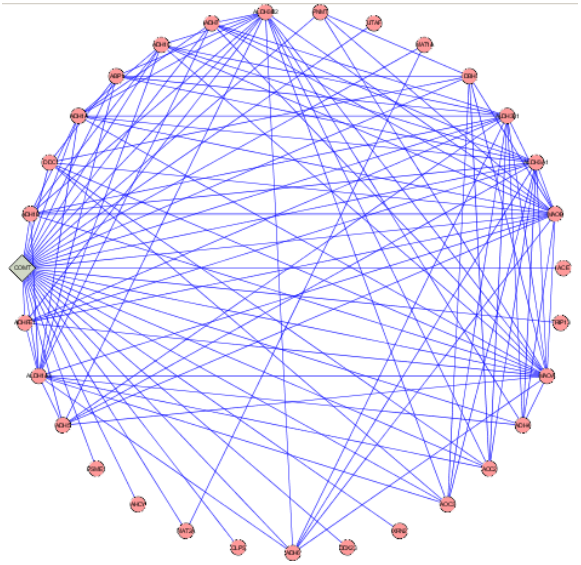

Figure 14. Circular Layout of *COMT* and Neighbor Genes.

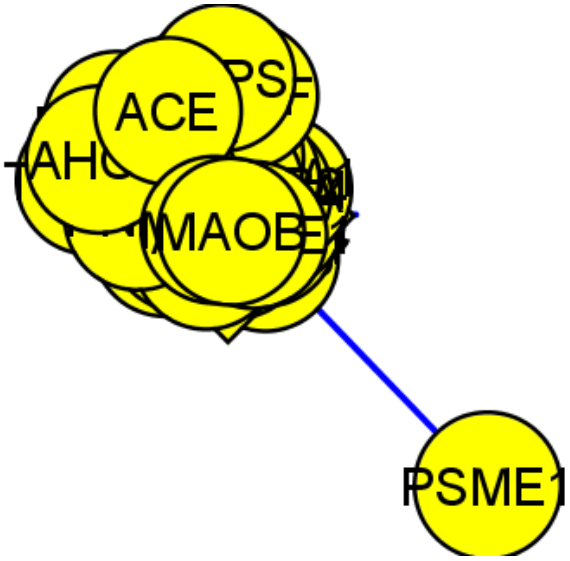

Figure 16. Edge Weighted Forced Directed-BioLayout of *COMT*

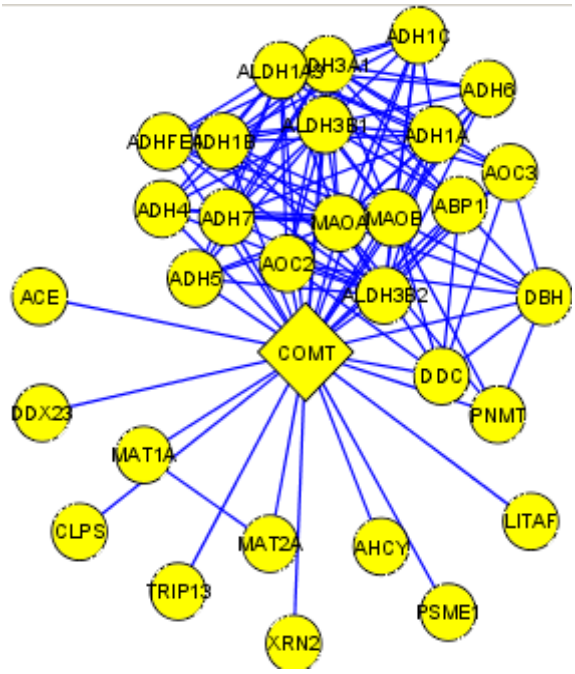

Figure 15. Force Directed Layout of *COMT*

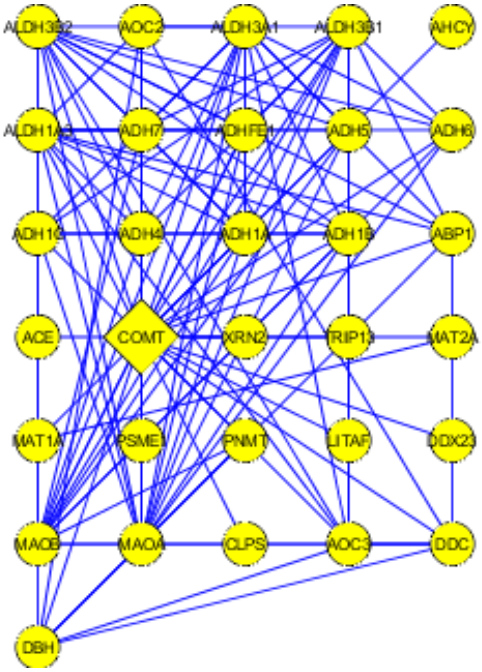

Figure 17. Grid Layout of *COMT* and Nearest Neighbor Genes

Acknowledgements

The author would like to thank Dr. JoAnn Haysbert, President of Langston University, for

her support in bringing the Introduction to Bioinformatics Class into fruition. Cherie Ognibene's persistence in generating the literature review and assistance with all aspects of this study was very valuable, and I wish her so much success. Without the financial support and the computer exercises supplied by Dr. Hugh B. Nicholas, Jr. and Dr. Alex Ropelewski's Minority Access to Research Careers (MARC) Grant funded by the National Institutes of Health, Grant #5 T36 GM008789, this *COMT* work nor the Bioinformatics Class would not have come to pass. The author wishes to thank Carnegie Mellon University and the Pittsburgh Supercomputing Center Staff for all of the wonderful accommodations and for an introduction to Dr. Jimmy Torres, our evaluator, and Dr. Ricardo Gonzalez Mendez, who allowed me to pick his brain regarding the availability of biomaterials at the National Institute of Health and the Pharmacogenetics and Pharmaco-genomics Knowledge Base. And finally, the author is indebted to Dr. Frank Waxman for his financial support through the OK-INBRE grant funded by the National Institutes of Health, Grant #P20RR016478-07. Without Franks help and sincere mentoring, my molecular biology/ bioinformatics laboratory at Langston University would not exist.

References

1. Rovner, S.L. 2004. "Depression". *CENEAR* 82(6): 33-40.
2. http://en.wikipedia.org/wiki/COMT
3. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. "The Protein Data Bank". *Nucleic Acids Research* 28: 235-242.
4. Vidgren, J., Svensson, L.A., Liljas, A. 1994. "Crystal structure of catechol O-methyltransferase". *Nature* 368: 354-358.
5. Williams, H.J., Owen, M.J., O'Donovan, M.C. 2007. "Is *COMT* a Susceptibility Gene for Schizophrenia," *Schizophrenia Bulletin* 33(3): 635-641.
6. Pooley, E.C., Fineberg, N., Harrison, P.J. 2007. "The met^{158} allele of COMT is associated with obsessive-compulsive disorder (OCD) in men: case-control study and meta-analysis". *Molecular Psychiatry* 12: 556-561.
7. Shifman, **S.**, Bronsein, M., Sternfeld, M., Pisant'e, M., Weizman, A., Reznik, I., Spivak, B., Grisaru, N., Karp., L., Schiffer, R., Kotler, M., Strous, R.D., Swartz-Vanetik, M., Knobler, H.Y., Shinar, E., Yakir, B., Zak, N.B., Darvasi, A. 2004. "*COMT:* A common susceptibility gene in bipolar disorder and schizophrenia*". American Journal of Medical Genetics, part B., (Neuropsychiatric Genetics)* 128B: 61-64.
8. Harrison, P.J., Tunbridge, E.M. 2007. "Catechol-O-Methyltransferase (COMT): A gene contributing to sex differences in brain function, and to sexual dimorphism in the predisposition to psychiatric disorders". *Neuropsychopharmacology* pp.1-9.
9. Enoch, M.A., Xu, K., Ferro, E., Harris, C., Goldman, D. 2003. "Genetic Origins of Anxiety in Women: a Role for a Functional Catechol-O-Methyltransferase Polymorphism," *Psychiatric Genetics* 13(1): 33-41.
10. The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs**.** *Nature* 449: 851-861.
11. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J. 2009. *Nucleic Acids Res*. 37(Database issue): D5-15. PMID: 18940862.
12. C_25746809_50, Applied Biosystems Inc., Foster City, CA 94404, (http://www.appliedbiosystems.com).
13. T.E. Klein, J.T. Chang, M.K. Cho, K.L. Easton, R. Fergerson, M. Hewett, Z. Lin, Y. Liu, S. Liu, D.E. Oliver, D.L. Rubin, F. Shafa, J.M. Stuart and R.B. Altman. 2001. "Integrating Genotype and Phenotype Information: An Overview of the PharmGKB Project". *The Pharmacogenomics Journal* 1: 167-170.
14. The Universal Protein Resource (UniProt), *Nucleic Acids Res.* 36:D190-D195 (2008), (http://www.uniprot.org).
15. Notredame, C., Higgins, D.G. Heringa, J. 2000. "T-Coffee: A novel method for multiple sequence alignments," *JMB*, 302: 205-217.
16. Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. "ClustalW: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap Penalties, and Weight Matrix Choice". *Nucleic Acids Research* 22: 4673-4680.
17. Bailey, T. L. and Gribskov, M. 1998. "Combining evidence using p-values: application to sequence homology searches". *Bioinformatics* 14: 48-54.
18. Dyson, P., Kelly-Bottle, S., Heilborn, Unix, Sybex Inc., 1999.
19. Nicholas, K.B., Nicholas H.B. Jr., Deerfield, D.W. II. 1997. "GeneDoc: Analysis and Visualization of Genetic Variation". EMBNEW. News 4: 14.
20. Page, R.D.M. 1996. "TREEVIEW: An application to display phylogenetic trees on personal computers," *Computer Applications in the Biosciences* 12: 357-358.
21. Drummond AJ, Rambaut A. 2003. BEAST version 1.3 [computer program] (http://evolve.zoo.ox.ac.uk/beast. Accessed 31 January 2006).

22. Shannon, P., Markiel, A., Owen, O. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks". *Genome Research* 13: .2498-2504.
23. http://winscp.net/eng/index.php.
24. Putty (http://www.chiark.greenend.org.uk/~sgtatham/putty).
25. G. Rendon et al. 2007. "Bridging Teaching, Learning, and Education in Grid-enabled Bioinformatics". Molecular Science Student Workbench. Teragrid07, Madison, WI.
26. Java weblink: http://java.sun.com.