

## Final Project

**Research Question:** Can the Ecdysozoa Theory be proved using modern phylogenetic analysis methods with the original sequences?

**Introduction:** There has been some difficulty classifying model organisms phylogenetically. It's classification is important because researchers want to figure out how closely *Drosophila* is related to humans as it is used as a model for some human diseases. There was a tension between the traditional hypothesis of Coelomata and the new, alternative hypothesis of Ecdysozoa. Ecdysozoa is a Superphylum that includes the Phylum Nematoda and Phylum Arthropoda. Before modern molecular phylogenetic methods, many of these phyla were based on the Coelomate hypothesis that theorized that these animals should be classified based on morphology of body cavities; for example no coelom, partial coelom, and true coelom ("Boundless Biology"). In the traditional view, fruit flies and humans were more closely related than either and nematodes. In the new theory, fruit flies and nematodes were more closely related than either and humans. "Ecdysozoa" refers to the fact that many animals in this clade shed their lightweight flexible exoskeleton, or cuticle, in a process called ecdysis (DefinedTerm).

I was able to find the original paper theorizing the Ecdysozoa theory published in 1997 by A.M.A. Aguinaldo, et. al. Based on their analysis of 18S rDNA from representative taxa, they found many moulting animals are closely related like Arthropods and Nematodes which includes *Drosophila Melanogaster* and *Caenorhabditis Elegans*. The researchers were careful about their selection for representative taxa because in the past, other researchers have used fast evolving sequences so they ensured slow evolving taxa instead. In Figure 1 of their paper, they show the difference in phylogenetic trees between using slow and fast evolving taxa. In the paper, the researchers' used a majority-rule consensus tree of four methods and their respective bootstrap values in their Fig. 2. One of those methods was maximum parsimony which I will to repeat with PAUP Version 4.0a (they used Version 3.1.1). They also used a maximum likelihood method since they have Jukes-Cantor distances as one of their methods, but RAxML was created in 2006 (Stamatikis). Based on my research question, I aim to align the sequences used in the original paper with modern methods like RAxML (which calculates maximum likelihood) and MrBayes (a Bayesian analysis created in 2001) (Wikipedia). Bayesian analysis incorporates a prior distribution as a parameter and uses a likelihood model which creates the posterior distribution (Stata). I predict this will provide a more likely tree because of this extra parameter.

**Methods:** First, I obtained the original paper that proposed the clade Ecdysozoa. At the bottom of this paper was a list of the sequences used for the researchers' analysis which included representative taxa. What was unfortunately missing were the sequences used for the outgroups, so I chose my representative taxa for the Cnidarian outgroup to be *Aurelia Aulita*. A list of the sequences used can be found below. Next, I organized all of these sequences into a text file and added it to my Github repository. When I loaded the file into Gitbash, I used Mafft to align my sequences.

### *Sequences*

Phylum	Genus	GenBank Accession Number
Rotifera	Brachionus	U49911
Annelida	Enchytraeus sp. (oligochaete)	U94948
Onychophora	Euperipatoides leukartii	U49910
Nematomorpha	Gordius sp.	U51005
Tardigrada	Macrobiotus sp.	U49912
Tardigrada	Milnesium tardigradum	U49909
Platyhelminthes	Stenostomum sp.	U95947
Annelida	Stylaria sp. (oligochaete)	U95946
Nematoda	Trichinella spiralis	U602311
Cnidaria	Aurelia aurita	AY039208.1

### *PAUP*

I converted my aligned sequence file to Nexus format with an online converter. I started PAUP, set Aurelia as the outgroup, and ran AllTrees first. Next, I ran Branch and Bound, followed by Heuristic Search. Lastly, I made a Consensus Tree to tell how the trees differ from each other, shown in Figure 1 below.

### *RAxML*

From there, I began running RAxML for this data with twenty maximum likelihood searches on twenty randomized stepwise addition parsimony trees. I used GTR, which stands for General Time Reversible, which assumes different rates of substitution for each pair of nucleotides as well as different frequencies for occurrence of nucleotides, and Gamma which is a model of rate heterogeneity (Choudhuri). I took the Newick notation from this tree and put it into FigTree, rooting the tree with my outgroup, Aurelia. The result can be seen below in Figure 2. In order to check this tree against a control, I also ran RAxML GTRGamma with Aurelia set as the outgroup. When I compared the two trees, they were the same. I generated 100 maximum likelihood trees, with distinct starting trees, and a tree with the best likelihood. With the replicates, I was able to draw bipartitions on the best maximum likelihood tree and look at the branch labels for the Bootstrap support values. I was also able to build a majority-rule tree and a strict consensus tree seen in Figures 3 and 4.

### *MrBayes*

I used the Nexus file in my MrBayes program, set the outgroup as Aurelia, and set the ngens (number of cycles of MCMC) to one million. I left the other commands the same as the tutorial (printfreq=100, samplefreq=100, nruns=2, nchains=4, savebrlens=yes, diagnfreq=1000, diagnstat=maxstddev.) After summarizing the samples (sump and sumt), the results of the program can be seen in Figure 5. I was able to produce unrooted trees of the Bayesian clade credibility values and the phylogram (branching tree that is an assumed estimate of phylogeny

with branch lengths proportional to amount of inferred evolutionary change) after MrBayes had stopped running as seen in Figures 6 and 7 (Deli).

## Results:

Figure 1 below shows the consensus tree for PAUP maximum parsimony analysis. This shows the tree for the most parsimonious tree that was conserved.

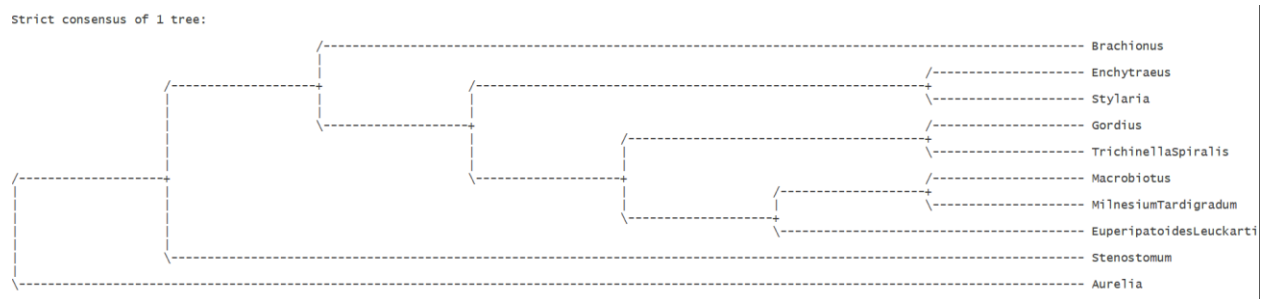


Fig. 1 PAUP Consensus Tree

Figure 2 below is the output for RAxML GTRGamma. It shows Aurelia (Cnidarian) as the outgroup as it is neither a protostome nor deuterostome. The inner branches show the other taxa as protostomes and within the Super Phylum Ecdysozoa. This tree shows Annelids and Rotifers with a MRCA (most recent common ancestor) and Tardigrades and Nematodes with a MRCA.

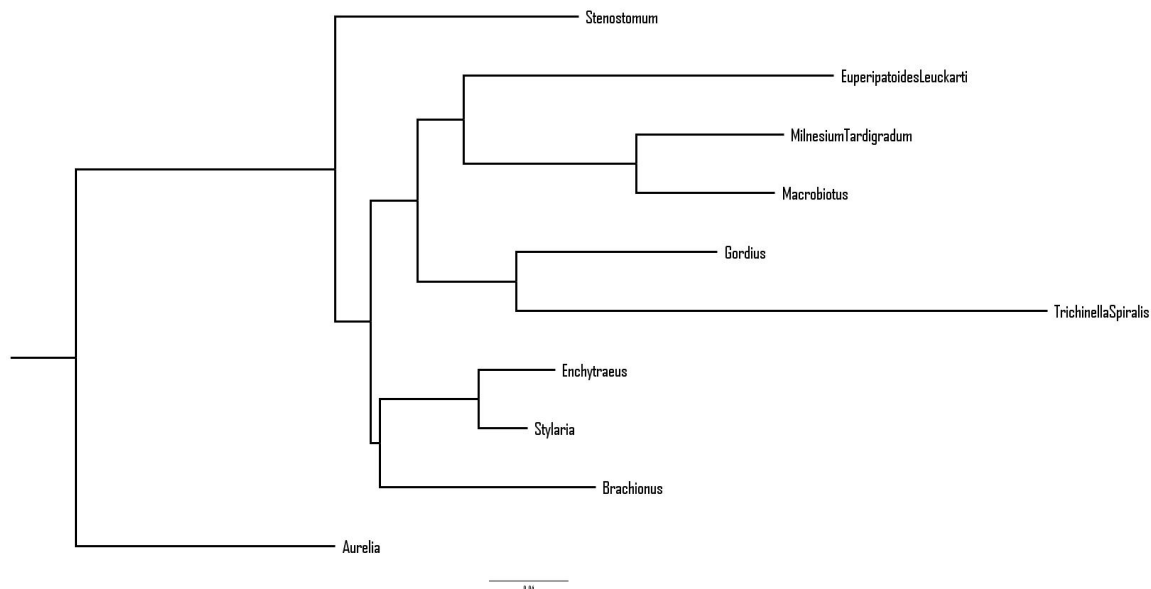


Fig. 2 GTR Gamma data for DNA

Figure 3 below is the Majority Rule consensus tree for RAxML one hundred replicates for bootstrap analysis. A Majority Rule tree shows the frequency of occurrence of clusters occurring in at least half of the trees and summarizes the bootstrap trees (“Consensus Trees”).

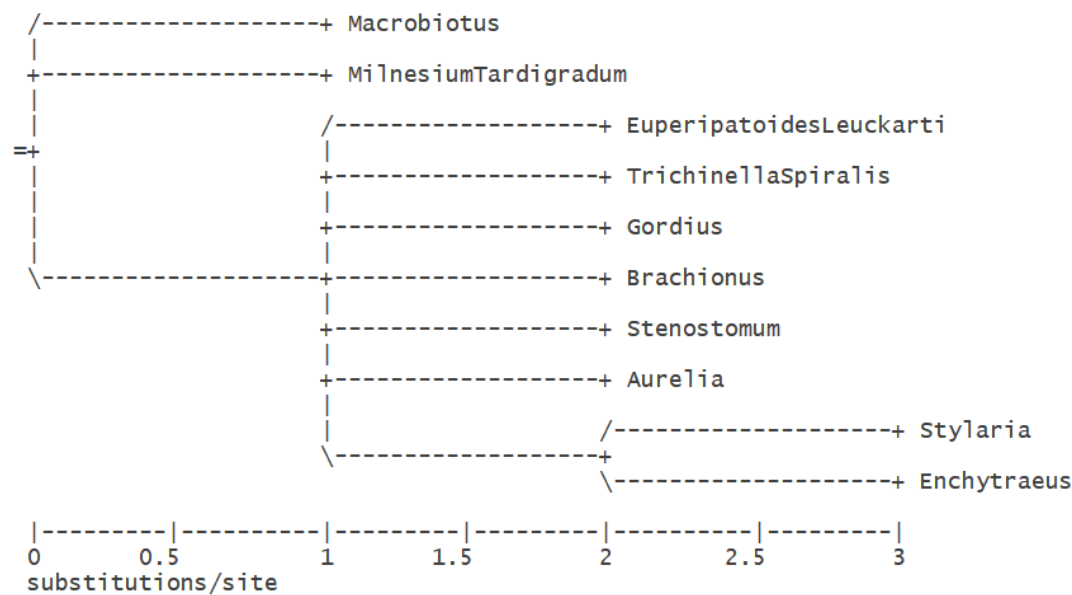


Fig. 3 Majority Rule Tree of Bootstrap Analysis

Figure 4 below is the Strict Consensus tree for RAxML one hundred replicates for bootstrap analysis. A Strict Consensus tree shows the frequency of occurrence of clusters occurring in all the trees and summarizes the bootstrap trees (Consensus Trees).

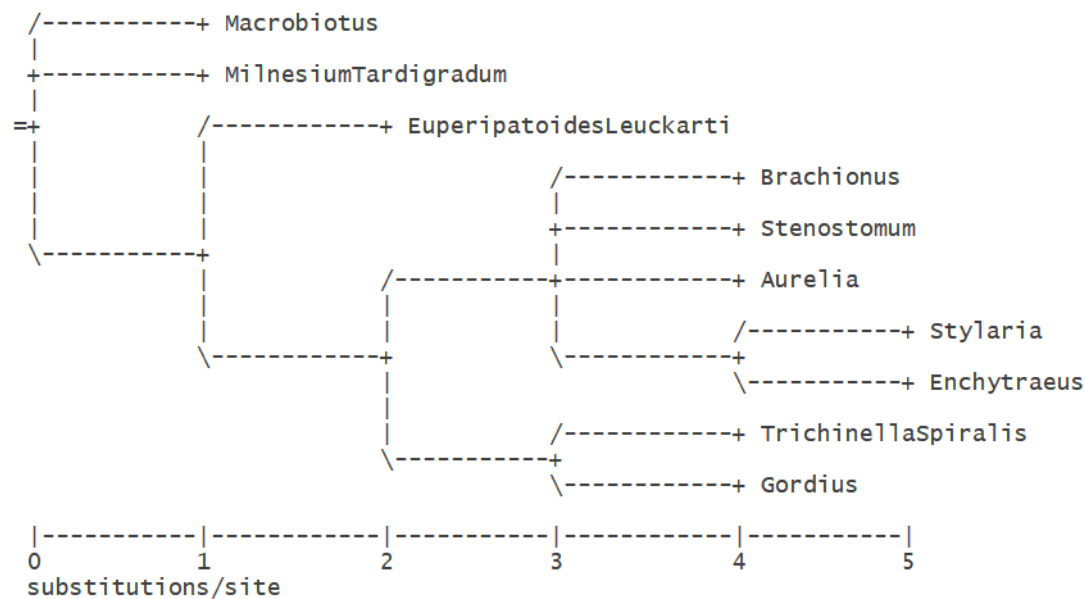


Fig. 4 Strict Consensus Tree of Bootstrap Analysis

Figure 5 below is the consensus tree for the MrBayes analysis. With Cnidarians as the outgroup, this tree shows that it is likely that these taxa evolved from an MRCA which shows that they all belong in a similar clade.

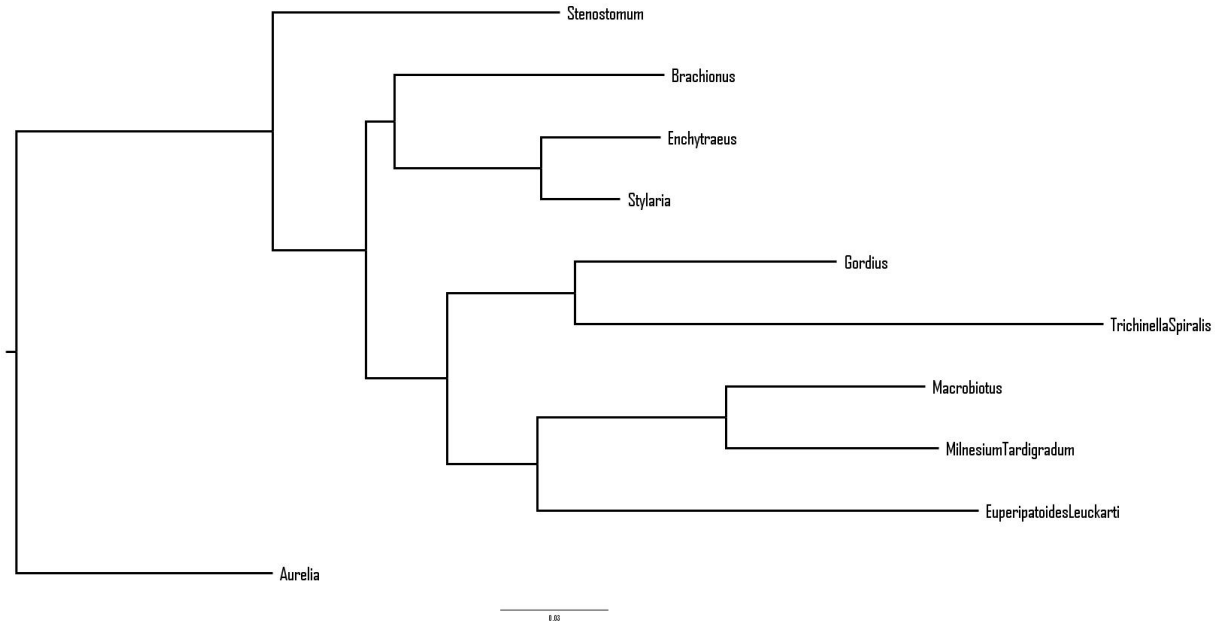


Fig. 5 MrBayes FigTree

Figure 6 below shows the clade credibility tree which is the tree with the highest score for maximum credibility. This score is found by evaluating each of the sampled posterior trees and giving it a score based on the fraction of times it appears in a set of sampled posterior trees (where the product of these scores is the tree's score) (Wikipedia).

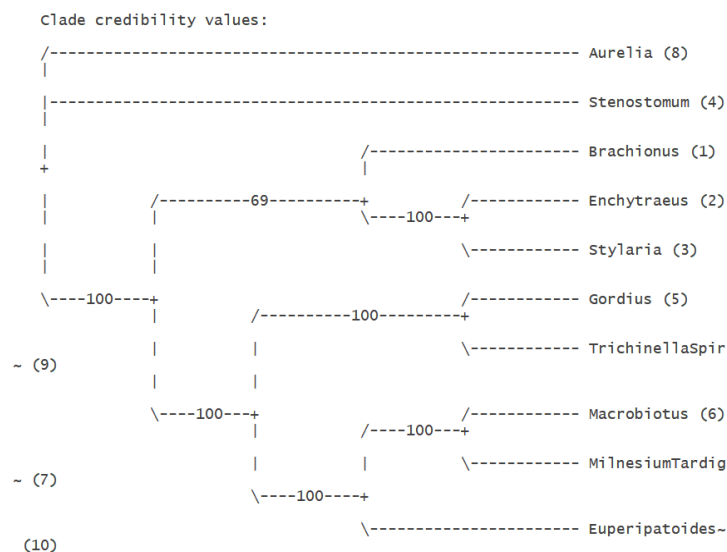


Fig. 6 MrBayes Clade Credibility Results

Figure 7 below shows the MrBayes Phylogram which has branch spans that are proportional to the amount of character change.

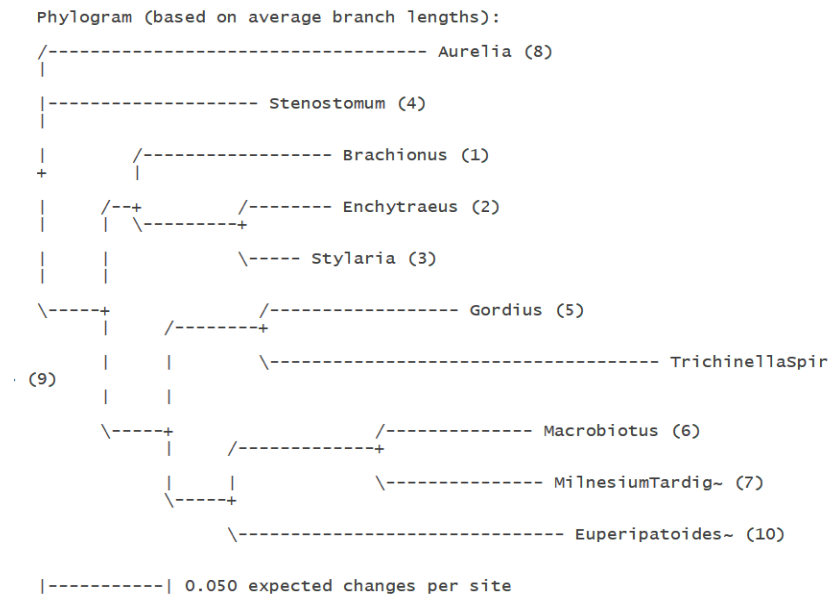


Fig. 7 MrBayes Phylogram

**Discussion:** My project was about testing the credibility of the Ecdysozoa Theory with newer methods. In rerunning the original testing done by researchers, I was able to set a control for my experimental methods. This control was maximum parsimony which I accomplished using PAUP. The version of PAUP the researchers used is different than the one I used. The Jukes-Cantor majority-rule consensus tree gave the researchers the highest bootstrap support overall (all above 71%.) In my tree, Rotifers share a common ancestor with nearly all other taxa, which is similar but not the same as the Figures 2, 3, & 4 in the research paper.

I then used RAxML to find the maximum likelihood trees. The result was seen in Figures 1, 2, and 3. These trees looked similar to the researcher's Figure 4 in that the taxa were descendent of an "Ecdysozoa" branch for an CA (common ancestor) from Protostomia. The tree seen in Figure 1 in my experiment is nearly identical to that seen in the researcher's Figure 4. In their Figure 4, they define the two main groups branching from Protostomia as Lophotrochozoa, which includes the annelids and rotifers, and Ecdysozoa, which includes arthropods (including *Drosophila*), tardigrades, onychophorans, nematomorphs, and nematodes. My Figure 1 has Lophotrochozoa defined with the Annelid taxa and Rotifer taxa branching from Ecdysozoa including those mentioned in their tree.

My MrBayes clade credibility values were 100% for nearly all branches. This is higher than the consensus trees made by the researchers and was likely more accurate due to the ability to use a million replicates. The consensus tree given by the researchers in their Figure 3 shows that Nematodes, Nematomorphs, and Arthropods in a separate branch (with 80% support within this

clade) from Platyhelminth, Rotifer, and Oligochaetes (average of with only 75% bootstrap supporting those clades being separate. My Figure 5 shows 100% bootstrap support for nearly all branches except for one which still has 69% support. Their main goal and conclusion was to provide evidence for a clade of arthropod-related moulting animals within the protostomes which is proven in both of our trees.

Their conclusion suggested that since *Drosophila melanogaster* and *C. elegans* are closely related, their mechanisms should not be expected in humans. Overall, they concluded that nematodes are not primitive metazoans but are within the protostome clade which supports the monophyletic protostome clade. My results showed the same result but organized a bit differently. My sequences were all seen in the protostome clade (besides the outgroup Cnidarian because it is a diploblast.) Based on this, I can conclude from my data that *Drosophila* and *C. elegans* are “more closely related” because they have a most recent common ancestor shared between them, not shared between either and humans (as they are deuterostomes.) The same is true for the proposed Ecdysozoa clade in my trees.

## References:

- Aguinaldo, Anna Maria A, et al. “Evidence for a Clade of Nematodes, Arthropods and Other Moulting Animals.” *ResearchGate*, ResearchGate GmbH, 29 May 1997, [www.researchgate.net/publication/14050829\\_Evidence\\_for\\_a\\_clade\\_of\\_nematodes\\_arthropods\\_and\\_other\\_moulting\\_animals](http://www.researchgate.net/publication/14050829_Evidence_for_a_clade_of_nematodes_arthropods_and_other_moulting_animals).
- “Boundless Biology.” *Lumen*, courses.lumenlearning.com/boundless-biology/chapter/superphylum-ecdysozoa/.
- “Chapter 4 Consensus Trees.” *COMPONENT Use's Guide*, pp. 1–3.
- Choudhuri, Soupratim. “Chapter 9 – Phylogenetic Analysis\*.” *Phylogenetic Analysis - Bioinformatics for Beginners - Chapter 9*, 2014, [www.sciencedirect.com/science/article/pii/B9780124104716000098](http://www.sciencedirect.com/science/article/pii/B9780124104716000098).
- “DefinedTerm: Ecdysozoa.” *Defined Term*, Coral Reef Information System Glossary, 2014, [definedterm.com/ecdysozoa](http://definedterm.com/ecdysozoa).

Deli, Temim. “ResearchGate.” *ResearchGate*, 2015,

[www.researchgate.net/post/What\\_is\\_the\\_difference\\_between\\_cladogram\\_dendrogram\\_and\\_phylogram](http://www.researchgate.net/post/What_is_the_difference_between_cladogram_dendrogram_and_phylogram).

“Maximum Clade Credibility Tree.” *Wikipedia*, Wikimedia Foundation, 27 June 2015,

[en.wikipedia.org/wiki/Maximum\\_clade\\_credibility\\_tree](http://en.wikipedia.org/wiki/Maximum_clade_credibility_tree).

Stamatikis, Alexandros. “RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses

with Thousands of Taxa and Mixed Models | Bioinformatics | Oxford Academic.” *OUP Academic*, Oxford University Press, 23 Aug. 2006,

[academic.oup.com/bioinformatics/article/22/21/2688/251208](http://academic.oup.com/bioinformatics/article/22/21/2688/251208).

“Stata: Data Analysis and Statistical Software.” What Is Bayesian Analysis? | Stata,

[www.stata.com/features/overview/bayesian-intro/](http://www.stata.com/features/overview/bayesian-intro/).