University of Colorado, Boulder

EBIO 4290 Phylogenetics Draft Paper

Gall Wasps: Evolution of Gall Chambers in Association to the Location of Gall

April 26, 2023

Sarthak Shukla



**Gall Wasps: Evolution of Gall Chambers in Association to the Location of Gall**

Sarthak Shukla

EBIO 4290, Spring 2022, University of Colorado Boulder

1. **Introduction**

There are varying examples of symbiosis between organisms in nature. However, it is always impressive when nature provides the means for one organism to manipulate another to do its bidding, at no benefit to itself. Gall wasps are one such example. These organisms are insects classified in the family of Cynipidae. Gall wasps are known to induce gall structures in a variety of plant species by injecting their eggs into plant tissue. Through this process, the plant will grow an external structure, a gall, that serves as a protective layer for the growing larvae. In addition to being a protective layer, the gall is also utilized as a source of nutrients for the larvae [3].

The complexity of gall structure can vary depending on the host-plant species as well as the gall-inducing species [3]. Gall structure can be classified as cryptic, distinct swelling or complex. Galls that appear inside stems or twigs are denoted as cryptic galls, which do not induce external structures on the plant. Distinct swelling structured galls are external structures that appear to be similar to the normal structure of the plant, whereas complex galls are external but present themselves as extremely different from the host-plant. Moreover, galls can be categorized as integral or detachable, depending on how they are attached to the host. Integral galls will be sturdy on the host-plant, while detachable galls are easily removed. Another important feature of galls is that they can contain a single larval chamber or be multi-chambered. Multi-Chambered galls are interesting to examine since they are composed of several single larval chambers that aggregate to form one, singular external gall. Lastly, galls may appear on several locations of the host-plant, which include: stems, twigs, runners, growing root tips, flowerheads, seed capsules, leaves, or buds. [1]

With the variety of gall structure, location, attachment, and chamber there are several opportunities for this species to show correlated discrete evolution among these characters. Two traits of interest in this study will be the number of gall chambers and the location of the gall on the host. Gall wasps specifically feed on a layer of plant tissue lining the inside of the gall, unlike most other phytophagous insects [1]. Additionally, they are a short-lived species, and therefore only feed during the larval stage of growth. Moreover, they specifically depend on nutritional sources of sucrose and carbohydrates to maintain metabolism and increase survival [2]. Thus, it seems that the location of the gall on the host plant would provide different nutritious values to the larvae, but also influence the structural complexity of the gall itself. In turn, it seems that reproductive organs such as flower buds, flowerheads, and seed capsules would provide a more nutritious layer of plant tissue for the larvae.

To explore the correlations between the number of gall chambers and location of the gall, this proposed study will estimate a bayesian phylogeny for the evolutionary history of gall wasps. Using methods of discrete correlation tests, such as Pagel tests, this study will investigate the relationship between these characters under the generated phylogeny. I hypothesize that galls present on the reproductive locations of a plant will be correlated with multi-chambered galls and on vegetative locations of the plant will present itself as single-chambered. Analysis of the correlation between these traits can help us understand more about the varying factors that lead to the development of galls.

1. **Materials and Methods**

**II.1 Data Information**

All of the data collected is provided by Ronquist et. al from the paper *Evolution of Gall Wasp-Host Plant Association*. This paper scored 175 morphological characters for 41 taxa in a Maximum Parsimony framework. Using this exemplar tree, Ronquist et al. created a meta tree for their trait mapping and estimations, which was limited to only the taxa that induce galls. Moreover, this paper provides character state information for the form, structure, position, chambers, and attachment traits of gall wasps [1].

**II.2 Phylogenetic Inference**

To generate a phylogenetic tree, the program MrBayes version 3.2.7a was used [4]. The dataset from the Ronquist et al. paper was a morphological character matrix scored for 175 characters. This matrix was in Nexus format, however it was not MrBayes compatible due to the scoring of the characters. For morphological data inputs, MrBayes can only support a total of nine character states that must be labeled with the numbers 0-9. The original morphological nexus data labeled the character states with the letters “a-p”, which thus prompted a necessary conversion. To accomplish this conversion, a python script (**can be seen in item x** of the appendix) was written that would take each morphological sequence and replace the most frequent character states with the numbers 0-9. The least frequent character states appeared at most 5 times over all sequences, and thus were replaced with a “-”. Following the conversion to MrBayes compatible Nexus file, the new morphology dataset was executed in MrBayes. The plan was to generate several high posterior probability trees, as well as a Baysesian consensus tree. Firstly, according to the example tree from Ronquist et al. there were 4 outgroup taxa labeled taxa who do not induce galls. Therefore, before generating trees the outgroup taxon was set to Ibalia rufipes. Figure 7 in the appendix displays all taxa in the phylogeny and their classification. The data was processed in an equal rates model where rates will be sampled from the gamma distribution. Then, MCMC was performed over 2 runs with each run having 100,000 generations, sampling frequency at 100, and discarding the first 25% of runs. After the MCMC walk had completed, the convergence of the runs was verified through the PSRF values [4]. Finally, the trees over the 2 runs were summarized using a measure of creating consensus of the posterior distribution using all clades compatible with each other. From this procedure, the highest posterior probabilities of all trees from the best log likelihood scored run were obtained, as well as consensus trees with strong posterior probability support for clades.

**II.3 Comparative Methods**

After the generation of our phylogenies, comparative methods are necessary to assess correlation between the number of gall chambers and position of the gall. To run Pagel tests [7] on the two traits, the software BayesTraitsV4 was used [6]. Primarily, two character matrices needed to be built. One character matrix is required for creating a mirror tree of trait mapping in Mesquite version 3.7.0 [5]. The second character matrix is required for BayesTraitsV4 analysis. Using the character state data in the table provided by Ronquist et al., a character matrix was constructed in Mesquite scoring values for the gall inducing taxa in the phylogeny. The scoring of the characters can be visualized in figure 5. Once the character matrix was built, a mirror tree window was created in Mesquite. The mirror tree window has parsimony traced character history for number of gall chambers on the left hand side and traced character history for location of the gall on the right hand side. The colors for the 4 states were coordinated in accordance to my hypothesis on both sides of the tree. Green was used to display single gall chambers and vegetative locations on the plant,while blue was used to display multiple gall chambers and reproductive locations on the plant [5]. The unscored taxa, which are either Inquilines (gall wasps that inject larvae into other existing galls) or non-gall inducing taxa, were left as gray since they should be unscored.

Using the character matrix that was developed in Mesquite, a new BayesTraitsV4 compatible character matrix was created using Visual Studio Code. To run the Pagel tests, two variations were performed. One variation was running the Omnibus test, Contingent Changes tests, and Temporal Order tests using the Baysesian consensus tree and the BayesTraits compatible character matrix. For all of the tests mentioned below, they were performed using a Maximum Likelihood framework and chi-squared values were computed for the test by taking log likelihood difference, which resulted in obtaining p-values from the VassarStats website [7]. The omnibus test was conducted by running the dependent test in BayesTraits and then the independent test, while setting no constraints. These tests were executed with the no constraints for the Omnibus test and setting constraints for Contingent Changes and Temporal Order tests. To understand the 8 rates of the dependent model and the constraints set, a graphic of the model can be seen in figure 4. The number of gall chambers was associated with the variable X (state 0 = single, state 1 = multiple) and the variable Y was associated with the location of the gall (state 0 = vegetative, state 1 = reproductive). Looking at the 8-rate model in figure 4**,** for contingent changes the constraints set were q12=q34 & q21=q43, as well as q13=q24 & q31=q24. The first constraint will test if change in the state of gall location is dependent on number of galls and the second constraint will test if change in the number of galls is dependent on the state of gall location. For the temporal order tests, the constraints set were q12=q13 and q42=q43 [7]. The first constraint tests that the multiple gall chambers were gained before location of the gall and the second constraint tests that multiple gall chambers were lost before the location of the gall [6]. This process is repeated for the second variation which will use all resulting trees from the highest likelihood run of the MCMC walk. The only difference is that for this variation the output of each test was logged to an output ‘.txt’ file and then inputted into a Google Sheet where the average log likelihood of all trees, and minimum and maximum log likelihood over all trees is calculated.

1. **Results**

**III.1 Phylogenetic Inference**

For the generation of the phylogenies, the results were extremely successful. After running MCMC for 100,000 generations over 2 runs, the average standard deviation of split frequencies was 0.015831. MrBayes suggests that standard deviation should be less than 0.01, if it is not then additional generations need to be run [4]. However, looking at the additional results of the MCMC run below, we can see that the 2 runs do in fact converge. In the MrBayes log output, we can see that the two runs were able to converge around an arithmetic mean of log likelihood score -3071.50. The PSRF value for TL was noted as 1.000 and PSRF for alpha was 1.001. PSRF values should be 1.000 or extremely close to it, considering it is the measure of convergence of the two runs. Moreover, the average ESS (estimated sample size) was 497.7 for TL and 306.99 for alpha. MrBayes recommends that “ESS values below 100 may indicate that the parameter is undersampled”, which is not the case in these MCMC runs [4]. Out of 1407 trees samples, 50% credible set contains 657 trees, 90% credible set contains 1257 trees, 95% credible set contains 1332 trees, and 99% credible set contains 1392 trees. The result of this analysis showed that the likelihood of run 1 was better than the likelihood of run 2 by 0.34. From run 1, we gained 1407 trees with one tree having the posterior probability of 1.0. The entire log file for the MrBayes computations can be found in the GitHub link in the appendix. The consensus tree, in figure 1 of the appendix, showed extremely strong posterior probability support for internal nodes, 95% PP support for the clade that contains all taxa between Aylax papaveris and Andricus quercusradicis. All nodes in this branch were generally about 90% PP except for 4 out of 15 nodes. MrBayes accurately generated this consensus tree phylogeny as it is almost identical to the exemplar tree by Ronquist et al. and maintains the same relationships of common ancestry.

**III.2 Comparative Methods**

In terms of the results from the comparative methods, it seems to be that the hypothesis cannot be well supported. First, if we look at the mirror tree in figure 2, we can see that there does not appear to be much overlap between the proposed hypothesis. We do not see extremely convincing evidence of correlation between single gall chambers and vegetative locations, nor multi gall chambers and reproductive locations. To convince myself of the results, further Pagel tests were conducted and all results can be visualized in the table of figure 3. Running the Omnibus test on the Bayesian consensus tree and BayesTraits compatible character matrix resulted in a log likelihood score (abbreviated as ln(L) from here on out) for the independent model being -36.527. The dependent model had an ln(L) of -34.621. This resulted in a chi-squared value of 3.812 with 4 degrees of freedom, which evaluates to a p-value of 0.432. This p-value is not less than 0.05 and therefore not significant enough for us to choose the dependent model over the independent model. However, further Pagel tests were conducted to see if constrained dependent models would be significantly preferred over independent models, implying that there is correlation. For the Contingent Changes test, the constrained model for q12=q34 & q21=q43 had a ln(L) of -34.623. This performed worse in likelihood than the dependent model with no constraints, and therefore was immediately rejected and no p-value was evaluated. The constrained model with q13=q24 & q31=q42 had an ln(L) = -34.627, which again was a worse likelihood score and immediately rejected. Thus the hypothesis has to be rejected because we do not see change in either gall chambers or location of the gall being dependent on the state of the other. However, Temporal Order tests were still conducted and the constrained model with q12=q13 had ln(L) = -34.490 and the constrained model with q42=q43 resulted in ln(L) = -34.621. While the constrained model of gain had slightly better log likelihood, by 0.2, this is not enough to prompt p-value evaluation. Thus, both Temporal Order tests were rejected as well. The same evaluation occurred when running BayesTraits with all 1407 trees from MCMC run 1 and the results can be seen in the table of figure 6. The Omnibus test resulted in an average ln(L) for the dependent model being -34.21 and the average ln(L) for the independent model was -36.466. The difference in likelihood scores is almost identical to the difference in likelihood scores for the consensus tree. Thus, further testing was not conducted and the possibility that 1407 trees could display any sort of dependence was rejected.

1. **Discussion**

As we can see from the results section, the hypothesis of this study has to be rejected. Originally, the thought process was that the galls induced by gall wasps are dependent on the nutrients that are supplied to the gall. It was thought that reproductive structures provide more carbohydrates and nutrients to the plant-host tissue lining the gall, therefore leading to the evolution of multiple chambered galls. Moreover, it was expected that galls on twigs and stems (vegetative locations) would provide less nutrients and thus maintain the state of the single gall chamber. This was not the case, instead there does not appear to be any correlation between the number of chambers in a gall and the location of the gall. Looking at the mirror tree again, we can see that for some taxa, especially the taxa in the Pha Tim Complex family show signs of correlation between the single chambered galls and vegetative locations. However, this could be attributed to the fact that the taxa in the Pha Tim Complex generally induce galls on stems and leaves, vegetative locations and can vary in number of chambers of gall [1]. It could be a pure chance that we can see the correlation for this clade. Moreover, for the rest of the phylogeny there does not seem to be correlation between the states of the two traits. Moreover, this pattern is not seen in other studies on gall wasps.

One of the key issues in this analysis could be scoring of the characters. Due to limitations of BayesTraits, polymorphic characters can be tested, however only with 3 states [6]. To perform a proper analysis of this correlation the hypothesis should be framed as such: single chambered galls will be correlated with gall locations on stems, twigs, runners, growing root tips, and leaves while the multi-chambered galls will be correlated with gall locations on inflorescence, flower heads, buds, bract, seed, or seed capsule. Moreover, some of the taxa present themselves with the capabilities of producing single and multiple chambered galls so this would need to be accounted for as well. Thus, instead of binary character mappings for both taxa we would need to incorporate 3 states for the number of gall chambers and 12 states for the location of the gall. Using this method, there is a greater probability that we would see some sort of correlation between the number of gall chambers and position of the gall. However, this cannot be done in BayesTraitsV4 and will need much more advanced software to run a 54 rate model which is also extremely computationally expensive.

While there does not seem to be correlation between the number of gall chambers and location of the gall in this study, there is still a possibility that this correlation can be explored with different methods in future work. Furthermore, there are other traits of galls that can be explored for correlation. For example, it could be investigated whether there is a correlation between the structural complexity of the gall and the location of the gall, or whether structural complexity of the gall is dependent on the number of gall chambers. There are several analyses that can be conducted in the future to learn more about galls and gall wasps, especially analyses related to the Inquilines who do not induce galls but rather embed their eggs into existing galls. There is not much research done in the field of Inquilines, which prompts investigations into what adaptive selective pressures led to the evolution of Inquilines.

**V. Literature Cited**

[1] Ronquist, Fredrik, and Johan Liljeblad. “Evolution of the Gall Wasp-Host Plant Association.” Evolution International Journal of Organic Evolution, vol. 55, no. 12, 2001, pp. 2503–2522., <https://doi.org/10.1111/j.0014-3820.2001.tb00765.x>.

[2] De Souza, Amanda Rodrigues, et al. “Longevity and Survival of Leptocybe Invasa (Hymenoptera: Eulophidae), an Invasive Gall Inducer on Eucalyptus , with Different Diets and Temperatures.” PeerJ, vol. 6, 25 July 2018, <https://doi.org/10.7717/peerj.5265>.

[3] Ronquist, Fredrik, et al. “Phylogeny, Evolution and Classification of Gall Wasps: The Plot Thickens.” PLOS ONE, vol. 10, no. 5, 2015, <https://doi.org/10.1371/journal.pone.0123301>.

[4] Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755. Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

[5] Maddison, W. P. and D.R. Maddison. 2023. Mesquite: a modular system for evolutionary analysis. Version 3.81 <http://www.mesquiteproject.org>

[6] Pagel, Mark. “Bayestraits .” Bayes Traits Manual, <http://www.evolution.reading.ac.uk/Files/BayesTraits-V1.0-Manual.pdf>.

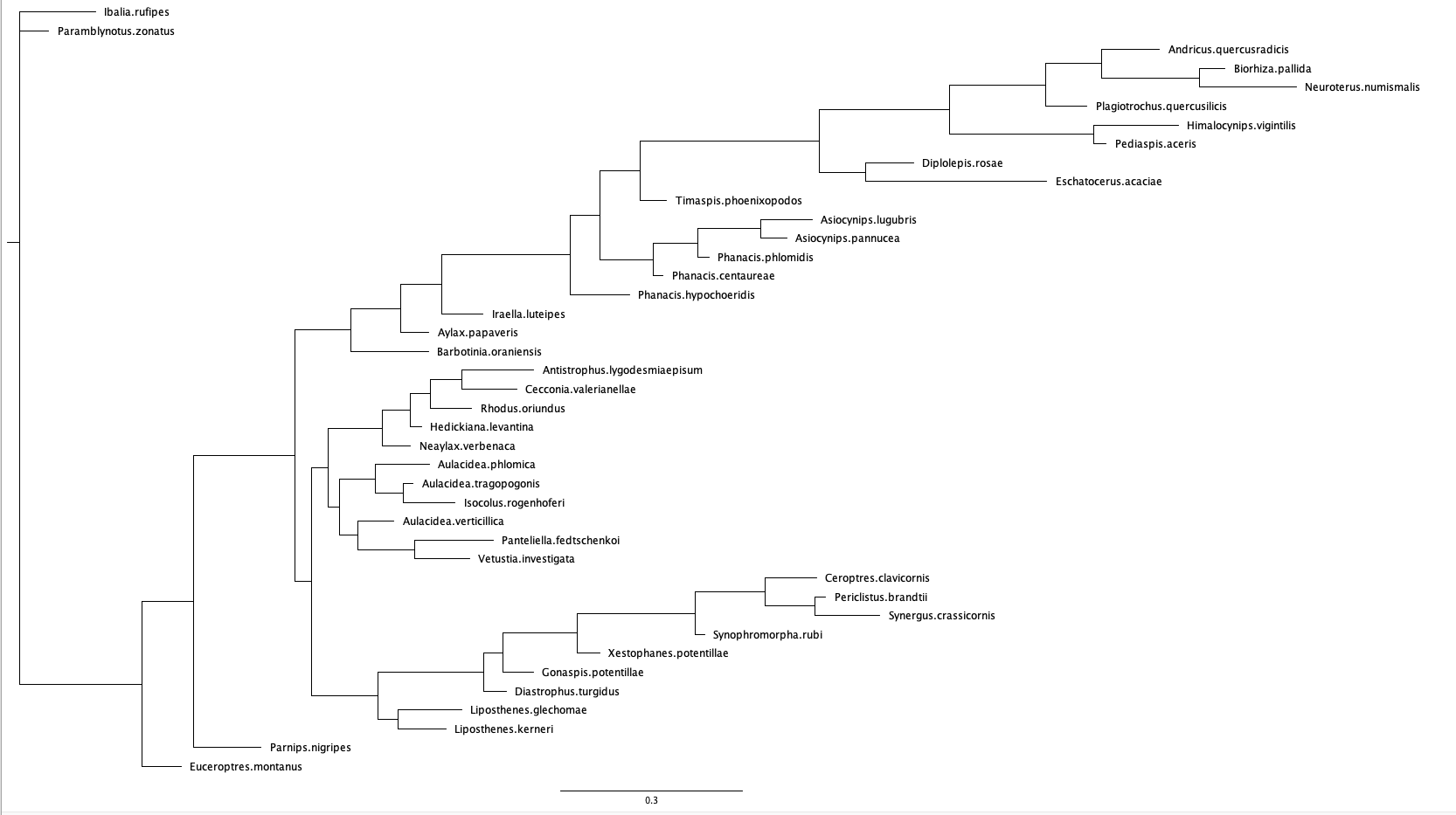
[7] Pagel, Mark. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” Proceedings of the Royal Society of London. Series B: Biological Sciences, vol. 255, no. 1342, 1994, pp. 37–45., <https://doi.org/10.1098/rspb.1994.0006>.

**VI. Appendix**

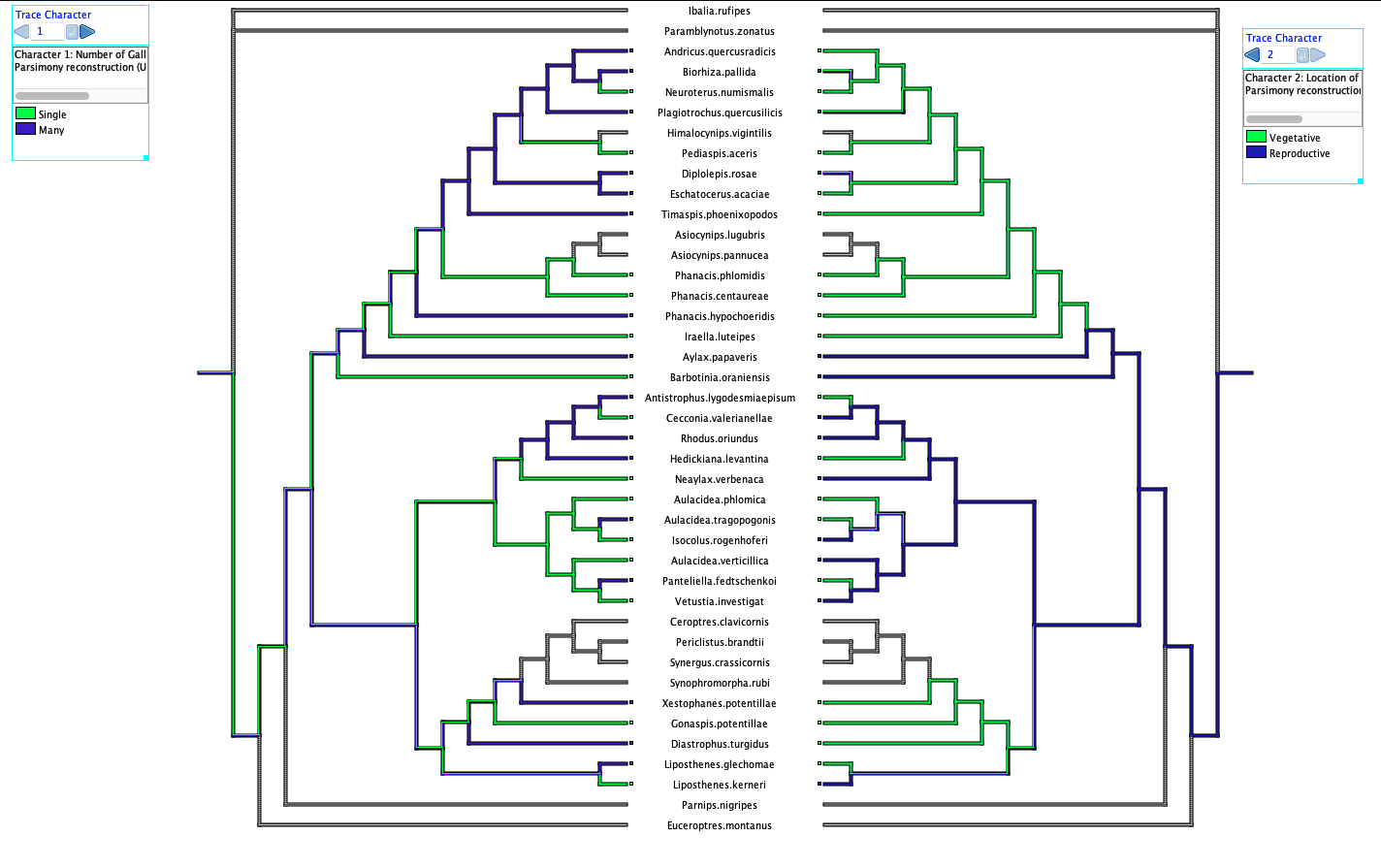
Github Repository: <https://github.com/ssarthak01/GallWaspsDiscreteTraitEvolution>

Figures listed on next page

*Figure 1- Bayesian Consensus Tree*

**

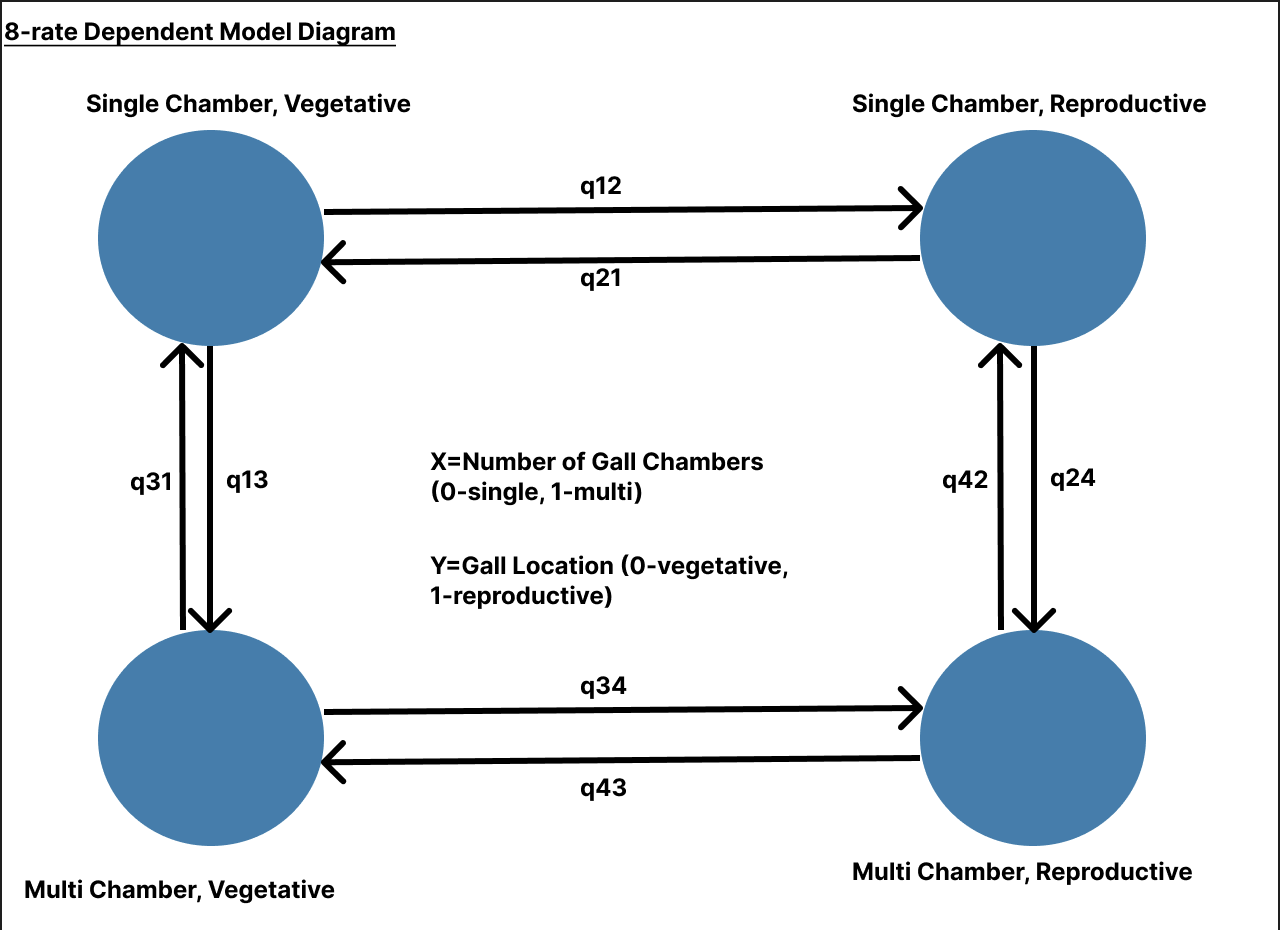
*Figure 2 - Mirror tree with mapped traits*

**

*Figure 3 - Table of Pagel Test Results for Bayesian Consensus Tree*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Pagel Tests** | **Log Likelihood** | **X^2** | **P-val** |
| **Omnibus** | **Independent Model (df=4)** | -36.527 | n/a | n/a |
| **Omnibus** | **Dependent Model (df=8)** | -34.621 | 3.812 | 0.432 |
| **Contingent Changes** | **Dependent Model w/ constraints: q12=q34 & q21=q43 (df=7)** | -34.623 | 0.004 | 0.9496 |
| **Contingent Changes** | **Dependent Model w/ constraints: q13=q24 & q31=q42 (df=7)** | -34.627 | 0.012 | 0.9128 |
| **Temporal Order** | **Dependent Model w/ constraints: q12=q13(df=7)** | -34.490 | 0.262 | 0.6087 |
| **Temporal Order** | **Dependent Model w/ constraints: q42=q43 (df=7)** | -34.621 | 0.0 | 1.0 |

*Figure 4 - Visualization of 8-rate Dependent Model*

**

*Figure 5 - Character Matrix*

**

*Figure 6 - Table of Pagel Test Results for 1407 Bayesian Trees*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Pagel tests** | **Min to Max Log Likelihood Range** | **Average Log Likelihood** | **X^2** | **P-val** |
| **Omnibus** | **Independent Model (df=4)** | [-36.53, -34.09] | -36.47 | n/a | n/a |
| **Omnibus** | **Dependent Model (df=8)** | [-34.67, -31.25] | -34.21 | 4.52 | 0.3402 |

*Figure 7 - Taxa Selection*

|  |  |
| --- | --- |
| *Taxa* | *Classification* |
| *1 Ibalia.rufipes,*  *2 Andricus.quercusradicis,*  *3 Antistrophus.lygodesmiaepisum,*  *4 Asiocynips.lugubris,*  *5 Asiocynips.pannucea,*  *6 Aulacidea.phlomica,*  *7 Aulacidea.tragopogonis,*  *8 Aulacidea.verticillica,*  *9 Aylax.papaveris,*  *10 Barbotinia.oraniensis,*  *11 Biorhiza.pallida,*  *12 Cecconia.valerianellae,*  *13 Ceroptres.clavicornis,*  *14 Diastrophus.turgidus,*  *15 Diplolepis.rosae,*  *16 Eschatocerus.acaciae,*  *17 Euceroptres.montanus,*  *18 Gonaspis.potentillae,*  *19 Hedickiana.levantina,*  *20 Himalocynips.vigintilis,*  *21 Iraella.luteipes,*  *22 Isocolus.rogenhoferi,*  *23 Liposthenes.glechomae,*  *24 Liposthenes.kerneri,*  *25 Neaylax.verbenaca,*  *26 Neuroterus.numismalis,*  *27 Panteliella.fedtschenkoi,*  *28 Paramblynotus.zonatus,*  *29 Parnips.nigripes,*  *30 Pediaspis.aceris,*  *31 Periclistus.brandtii,*  *32 Phanacis.centaureae,*  *33 Phanacis.hypochoeridis,*  *34 Phanacis.phlomidis,*  *35 Plagiotrochus.quercusilicis,*  *36 Rhodus.oriundus,*  *37 Synergus.crassicornis,*  *38 Synophromorpha.rubi,*  *39 Timaspis.phoenixopodos,*  *40 Vetustia.investigata,*  *41 Xestophanes.potentillae;* | *1. Outgroup*  *2. Ingroup*  *3. Ingroup*  *4. Ingroup*  *5. Ingroup*  *6. Ingroup*  *7. Ingroup*  *8. Ingroup*  *9. Ingroup*  *10. Ingroup*  *11. Ingroup*  *12. Ingroup*  *13. Inquiline*  *14. Ingroup*  *15. Ingroup*  *16. Ingroup*  *17. Outgroup*  *18. Ingroup*  *19. Ingroup*  *20. Ingroup*  *21. Ingroup*  *22. Ingroup*  *23. Ingroup*  *24. Ingroup*  *25. Ingroup*  *26. Ingroup*  *27. Ingroup*  *28. Outgroup*  *29. Outgroup*  *30. Ingroup*  *31. Inquiline*  *32. Ingroup*  *33. Ingroup*  *34. Ingroup*  *35. Ingroup*  *36. Ingroup*  *37. Inquiline*  *38. Inquiline*  *39. Ingroup*  *40. Ingroup* |