1.0. Introduction

1.1. *Epichloë* importance and impacts on agriculture

Epichloë is an endophytic ascomycete fungal symbiont which grows in certain cool season grass species. Species in this genus have two forms of reproduction. In sexual reproduction, a macroscopic fungal body is formed on grass inflorescences. This is termed choke disease, and can be the cause for plant death (Ruying, Bruce, & Faith, 2019). Asexual reproduction is asymptomatic as *Epichloë* inhabits the intercellular space of grass leaves. Many species are hybrids (Scott, 2001), with a variety of parental species that do not appear to have specific phylogenetic origins.

It has been demonstrated that *Epichloë* often aids pasture growth, both in the increased uptake of nutrients for the grass, as well as pest control and reduction (Lugtenberg, Caradus, & Johnson, 2016). Secondary metabolites released by the fungus may deter insect pests and certain strains produce specific chemical protection. Some strains are sold commercially for farming sectors for this reason. Conversely, toxins produced by certain *Epichloë* species may impact cattle and lamb flocks, with ergot-alkaloids causing staggers disease (Guerre, 2016). This reduces agricultural health.

In addition to alkaloid production, *Epichloë* species produce a number of other secondary metabolites. Notably, cyclic peptides, a class of proteins recently found to be involved in necessary biological function of several species of fungus including *Epichloë* (Johnson et al., 2015). GigA, one of the most highly transcribed genes *in planta Epichloë*, has a suggested bioactive role. Other cyclic peptides are often produced by proteins with specific domains (Umemura et al., 2014). These genes are evolutionarily variable, with repeat elements often neighbouring alkaloid genes. It is suggested that the genes encoding alkaloids are under diversifying selection (Schardl et al., 2013).

The development of *Epichloë* as a bioprotective product is the result of a sustained program of basic and applied research. However, this work has largely focused on a small number of well-characterised strains. Developing novel protective products or expanding the range of hosts available to *Epichloë* species will require a broader approach. Advantage must be taken of the ecological and evolutionary diversity of the genus as a whole. Determining the relationships between *Epichloë* species, and, indeed whether a single tree can represent those relationships, is a key first step taking this approach.

1.2. Epichloë phylogeny

At present, the evolutionary relationships between various *Epichloë* species are unclear. Having only recently had one full genome sequence from the genus (*Epichloë festucae*), phylogenetic relationships have not been clarified (Winter et al., 2018). Craven, Hsiau, Leuchtmann, Hollin, and Schardl (2001) utilised a multi-gene model to appropriate a species tree, with only tree genes (introns of β-tubulin, actin and translation elongation factor 1 – α) and 10 species with two outgroups (Supplementary Figures, Figure 11). This determined two distinct groups, entitled ‘main’ and the *E. typhina* species complex. It suggests gene flow between *Epichloë* lineages may lower species differentiation.

Phylogenies estimated from a few genes do not necessarily reflect the true relationship between species. Lateral gene transfer, gene duplication, gene loss, alternative coalescence or reticulate evolution (Pamilo & Nei, 1988) can all lead to gene-tree species-tree conflict. For this reason, specific species-tree estimation methods provide the most accurate way to recover among-species relationships. ASTRAL has proved to be a powerful and efficient method among these approaches. It

uses quartet topology approach frequencies and is consistent with a multispecies coalescent model (Zhang, Rabiee, Sayyari, & Mirarab, 2018).

Since the first *Epichloë* phylogeny was published, whole genome sequences have been produced for several *Epichloë* species (Schardl et al., 2013). The wealth of data provided by these sequences will allow us to take a comprehensive approach to estimating relationships between *Epichloë* species. In addition, we can also detect any gene-tree species-tree conflict in this genus. Here, I have combined data from thousands of genes shared by 24 *Epichloë* strains to estimate a species tree for 13 *Epichloë* species. Annotation of metadata was then used to explore patterns within the dataset, including geographical distribution, host taxonomy and secondary metabolite representation. Hybrid data was also explored. The tree and analysis presented here will be an important resource for further work on this genus.

2.0. Methods

2.1. Genomic data and taxonomic sampling

Whole genome sequences were obtained from previously published assemblies (Schardl et al., 2013; Winter et al., 2018) and a number of ongoing projects (http://www.endophyte.uky.edu/, UKY) in other research groups at Masset University (MU). A variety of short and long read technologies have been used to produce these assemblies and RNAseq data was available for some strains (see Table 1). In total, 24 strains of *Epichloë* representing 13 species were analysed. In addition, an assembly for the closely related species, *Claviceps purpurea,* (Craven et al., 2001), was obtained from NCBI database.

Table 1. Species and analysed strains including origin and method of gene identity.

Species	Strains	Genome Origin	RNAseq (Yes/No)	Genome Read Length (Short/Long)
E. amarillans	EmaE4668	UKY	N	S
	EmaE57	UKY	N	S
	Eam702	MU	Y	L
E. bromicola	EbroAL0426	UKY	N	S
	EbroAL0434	UKY	N	S
	EbroBfe1	MU	N	L
E.gansuensis	Ega7080	UKY	N	S
	Ega	MU	N	S
E.elymi	EelyE56	UKY	N	S
	Eel732	MU	Y	L
	FUN	MU	Y	S
E.festucae	Efe2368	UKY	Y	S
	C2857 (Fl1)	MU	Y	L
E.poae	EpoNfe76	MU	N	L
	Epo	MU	N	S
E.baconii	Ebac200745	UKY	N	S
E.brachyelytri	EbraE4804	UKY	N	S
E.aoteoroae	Eao	UKY	N	S
E.clarkii	E8Q19	UKY	N	S
E.mollis	Emo	UKY	N	S
E.sylvatica	Esy	MU	N	S
*E. Typhina**	E8Q16	MU	N	L
	Ety8	MU	Y	S
C.purpurea	CCE27021	MU	N	S

*Typhina strains were not grouped for species tree.

2.2. Gene calling and ortholog identification

Each genome was annotated using the funannotate pipeline v1.4 (Palmer, 2018), which makes use of AUGUSTUS (http://augustus.gobics.de/) to identify likely genes for all strains. The manually curated "M3" gene models derived from *E.festcuae* strain e2368 were used to train AUGUSTUS. For those genes where the RNAseq data is available, previously produced RNA alignments were used for gene discovery. Inferred protein sequences were then clustered into ortholog groups using ProteinOrtho6 with Diamond being used for homology search. Putative paralogs within species were removed via '—selfblast' arguments, while '—singles' was used to report without orthologs. Each identified orthology group was assigned an arbitrary ID for further analysis.

2.3. Estimation of gene trees

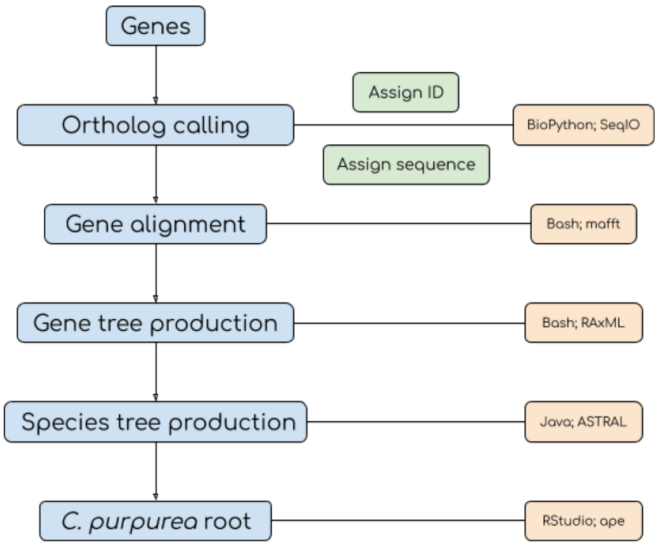

Figure 1. Method pipeline developed to analysis* Epichloë *taxa genetic information.

I developed a new pipeline to estimate gene trees for each ortholog group (Figure 1). This pipeline takes an orthology file assigning each gene in each strain to an ortholog ID as well as a series of sequence files for containing all genes from each of the strains being analysis. Making use of Python3 and the SeqIO module of Biopython (Cock et al., 2009), a series of new files are written representing unaligned sequences for each gene in an orthology group. These unaligned sequences are then aligned with mafft (v7.310-1) (Katoh, Asimenos, & Toh, 2009). Finally, a maximum likelihood gene tree is estimated for each orthology group using RAxML (v 5.6.3) (Stamatakis, 2014). Gene trees were estimated using a general time reversable (GTR) model of sequence evolution, with gamma distributed variation in evolutionary rate among sites.

I calculated summary statistics (alignment length, gap ratio, GC content and number of taxa) for each alignment using R and Biopython.

2.4. Estimation of species trees

All R analysis was carried out in RStudio using R version 3.6.1. In order to create a species tree, all aligned gene tree files were read into R (v3.6.1.) using ape (v5.1).

I then converted these gene trees into a multiphylo object and then read-out to java program ASTRAL, (v5.6.3.) (Zhang et al., 2018). Out log files were produced to ensure quality and success. It is important to note that a large number of gene trees were 'lost' here as ASTRAL requires all gene trees to have the same number of tips. Out of 6020 gene trees created, only 1852 were used to produce the strain and species tree. As ASTRAL arbitrarily roots outputs, I manually rooted the species tree at *C. purpurea*, an outgroup species. This produced a 'species tree' with all strains at tips.

I then created the species tree through the same process with the addition of a 'species map' file, with strains segregated under species headers. An example is shown in columns 1 and 2 of **Table 1**.

2.5. Visualising gene-tree conflict

I also analysed species and strain information with two functions within the Phangorn package (v2.5.5.). The topological distance between all gene trees was calculated, and the resulting distance matrix clustered using the UPGMA method. This was utilised to ensure lack of presence of major competing trees with split differences.

In addition, I created a consensusNet to explore competing splits in strain trees. The threshold was set to 0.2. Edge lengths are proportional to the frequency of the corresponding splits in the gene trees. This graphic was produced from the 'full' subset of gene trees, i.e. only those containing all 24 strains.

2.6. Metadata analysis

Once a species tree was produced, I added various metadata annotations including reproductive method, host taxonomy, geographical range, hybrid species and secondary metabolite gene presence. All metadata information was provided from localised study, the University of Kentucky website or other published papers.

In RStudio, I produced a ggtree (v1.16.6) from the species or strain phylogeny. A 'wide format' dataframe was used for all metadata categories. This contained the strain name as row names and metadata variables as column names. Either 'presence' or 'absence' was used to determine gheatmaps. I annotated the reproductive observations of *Epichloë* via geom_tiplab() function.

2.7. Monophyly discrepancies

I identified potential genes involved in transspecies polymorphisms through the use of is.monophyly() function in the R ape package (v5.3.). Four main strain clades were chosen (*E. festucae*, *E. amarillans*, *E. typhina* and *E. bromicola*) and gene trees containing all strains of a particular clade were checked for monophyly amongst the various strains. If not monophyletic, genes were isolated and combined to produce a new 'species tree'. It was hoped that other sister relationships would be recognised amongst strains of different species, and potential sites of transspecies polymorphisms would be identified. ConsensusNets were also produced to highlight inconsistencies.

2.8. Code availability

Source code developed at this project is available at https://github.com/Imogen-D/EpichloeSpeciesTreeProject.

3.0. Results

3.1. Short read genome assemblies do not contain fewer genes than long-read assemblies

I obtained a total of 188,899 thousand genes from 24 strains of *Epichloë* and the *Claviceps purpurea* outgroup. There is no significant difference in the number of genes detected in the fragmentary genomes produced from short-read data compared to the complete genomes included in this study (Figure 2). This result suggests the gene-rich portions of *Epichloë* genomes are well-represented in short-read genomes.

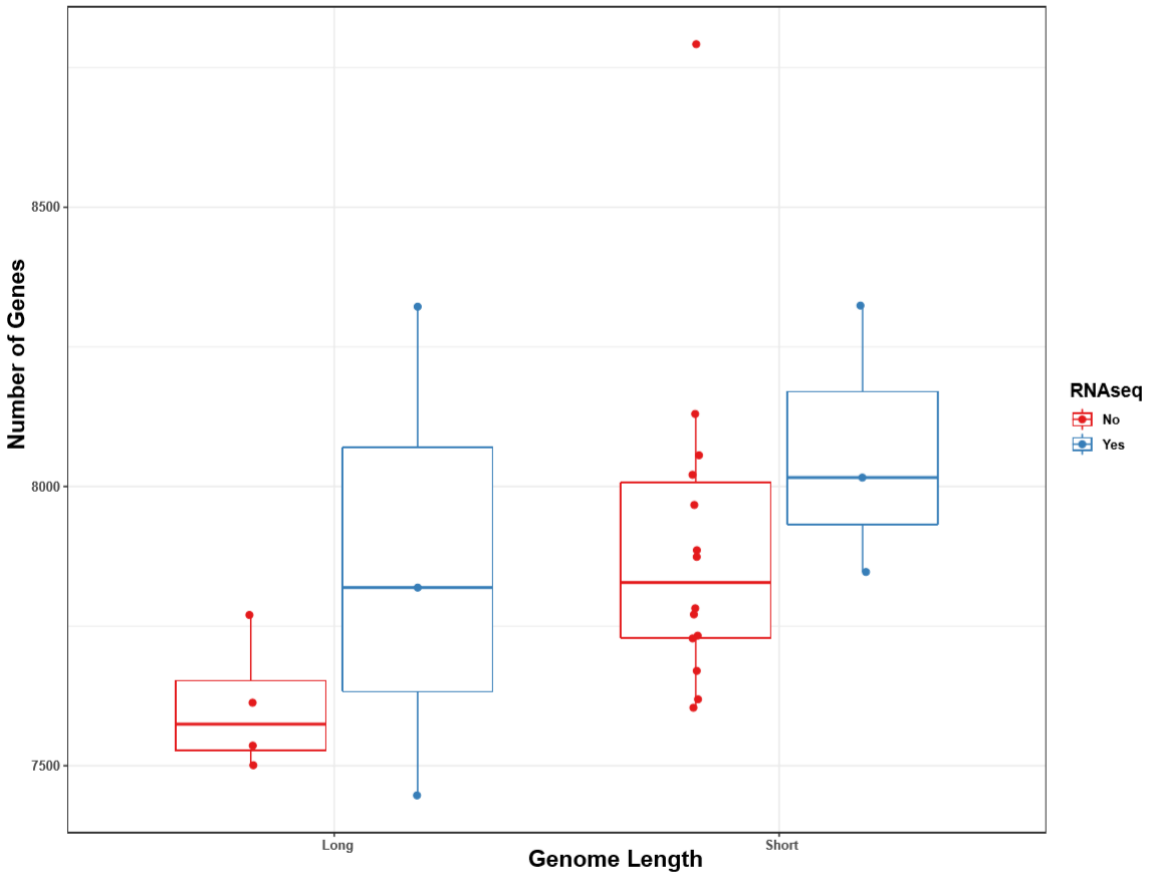

Figure 2. Red denotes no RNAseq processes while the right-hand side two charts display genomes with short reads. *It appears as though shorter reads and RNAseq may lead to more gene identification, although this is not statistically significant (P = 0.0616 and 0.1717 for length and RNAseq respectively).*

3.2. RNAseq processes do not impact quality

Figure 2. Red denotes no RNAseq processes while the right-hand side two charts display genomes with short reads. It appears as though shorter reads and RNAseq may lead to more gene identification, although this is not statistically significant (P = 0.0616 and 0.1717 for length and RNAseq respectively).Figure 2 highlights potential data quality concerns including genome read length and RNAseq identification on number of gene identified. Although the boxplots appear to have some differences, there was no statistical significance of genome reads or RNAseq use on the number of genes identified. Short genome reads have some correlation with number of genes identified (P = 0.0616).

A total of 61,042 orthologous groups were identified. Of these, 2,471 were present in all strains (including C*laviceps*) and 5,686 contained sequences of all *Epichloë* species.

6,020 alignments were produced by mafft for all orthogroups with 20 or more taxa. In this final dataset, there were an average of 22.83 taxa per alignment, and alignments had an average length of 1,910 (± 36) basepairs. The mean GC content was 55.43% (± 0.11 %) while the mean number of gaps per strain was 0.53% (± 0.014%).

3.3. A fully resolved species tree for *Epichloë*

From gene tree production with RAxML, I was able to recover a fully resolved species tree for the genus *Epichloë* through the use of ASTRAL (Figure 3).

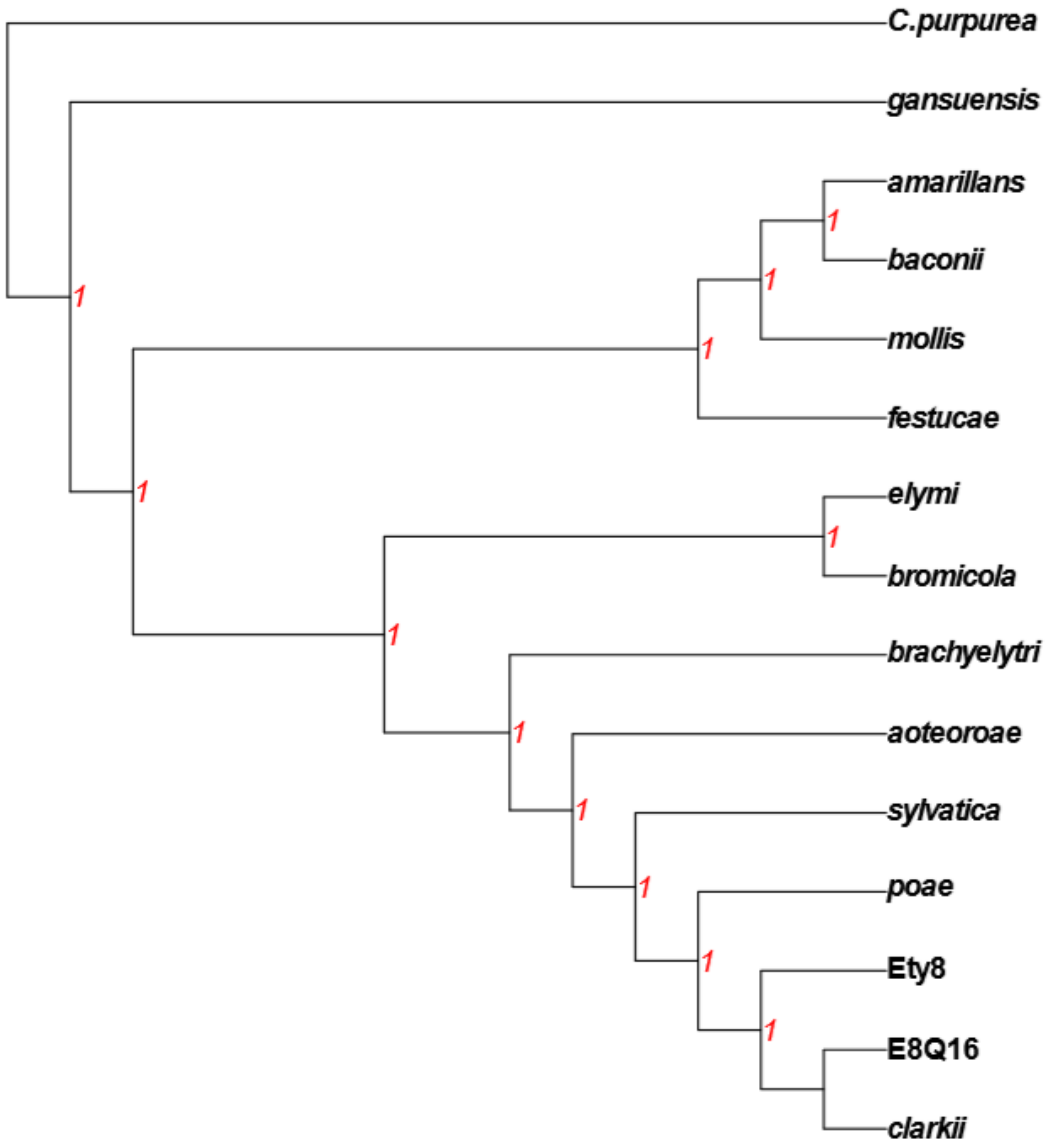

Figure 3. ASTRAL species tree of 14* Epichloë *species, rooted with* Claviceps purpurea *outgroup*. Strains of each species are combined. Confidences are denoted in red.***

Very strong node confidences were produced by ASTRAL and observable on Figure 3. This is a strongly supported species tree by 6,020 genes.

3.4. Discrepancies in data shown on consensusNets

Although the species tree in Figure 3 displays high confidences, a consensusNet was used to visualise areas of variability (Figure 4).

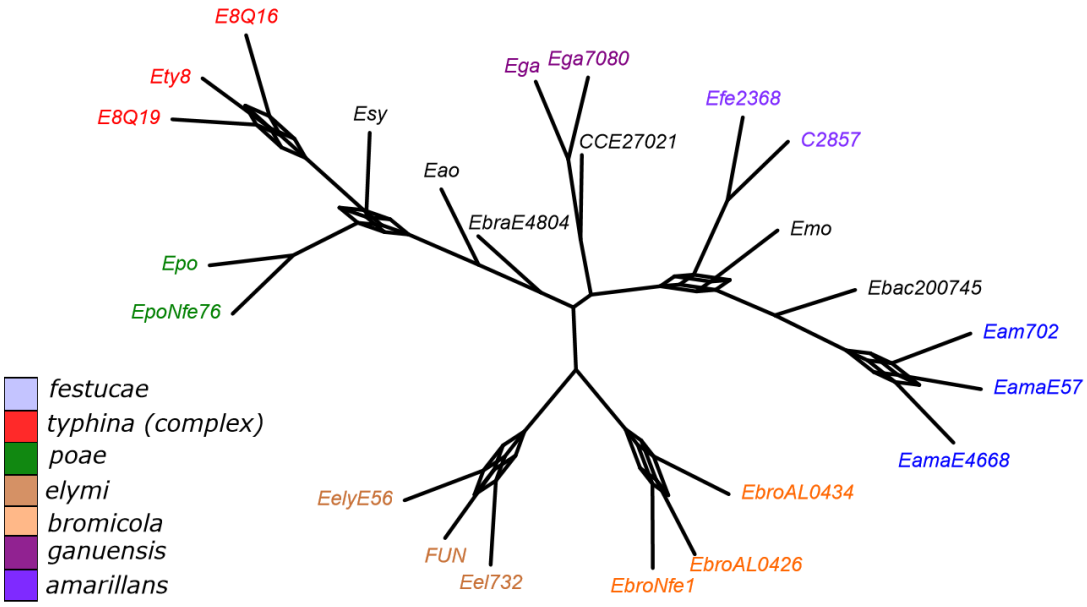

Figure 4. Phangorn consensusNet of all strains*. Various strains of major clades shown in different colours. Threshold set at 0.2 to show variability amongst clades. Note strain CCE27021 is the outgroup* C. purpurea.

Variability is shown amongst several clades including the relationships amongst *E. typhina*, *E. amarillans* and *E. elymi*. It is also interesting to note that the basal relationships amongst *E. mollis* and *E. festucae* are also more variable.

A distance topographical graphic was also produced (supplementary information Figure 13) but not suitable for publishing. The fact no deep splits exist in this cluster suggest there are no major competing topologies among our gene trees (a result which is consistent with the high degree of support for the species tree).

3.5. No single topology for trees where species are not monophyletic.

Having produced a species tree, analysis was carried out to quantify variability among gene trees. As expected, all species with multiple representatives in our dataset are monophyletic in the majority of gene trees (Table 2). Nevertheless, all species appear as para- or polyphyletic in some gene trees. We estimated new species trees and consensus networks from only these trees to determine whether they had a consistent topology which might suggest the descent from a single evolutionary event. As the data for *E. typhina* makes clear, no such consistent topology exists (Figure 5 and Figure 6). Although only non-monophyletic genes were utilised, monophyly still presents among the *E. typhina* clade (Figure 6 and Figure 6) although with less confidence than in the overarching species tree (Figure 3). Lower confidences are interestingly also observed between *E. amarillans* and *E. elymi* strains. Closer evolutionary relationships are observed between the *typhina* clade and that of *E. poae* and *E. sylvatica* (Figure 5). This result suggests those gene trees where species are not monophyletic may be the result of one-off chance events, incomplete lineage sorting or artefacts in our data.

***Table 2. Information on monophyletic genes for four main strain clades**. Non-monophyletic genes refers to the number of genes that displayed non-monophyly amongst the clade utilising the is.monophyly function. .'Full' refers to the number of gene trees containing all strains included in this study, and with non-monophyly amongst the clade in study.*

Clade/Species	Trees Containing	Monophyletic	Non-monophyletic	'Full'
E. Typhina	5584	5244	340	140
E. Festucae	5763	5571	192	53
E. Bromicola	5753	5425	328	153

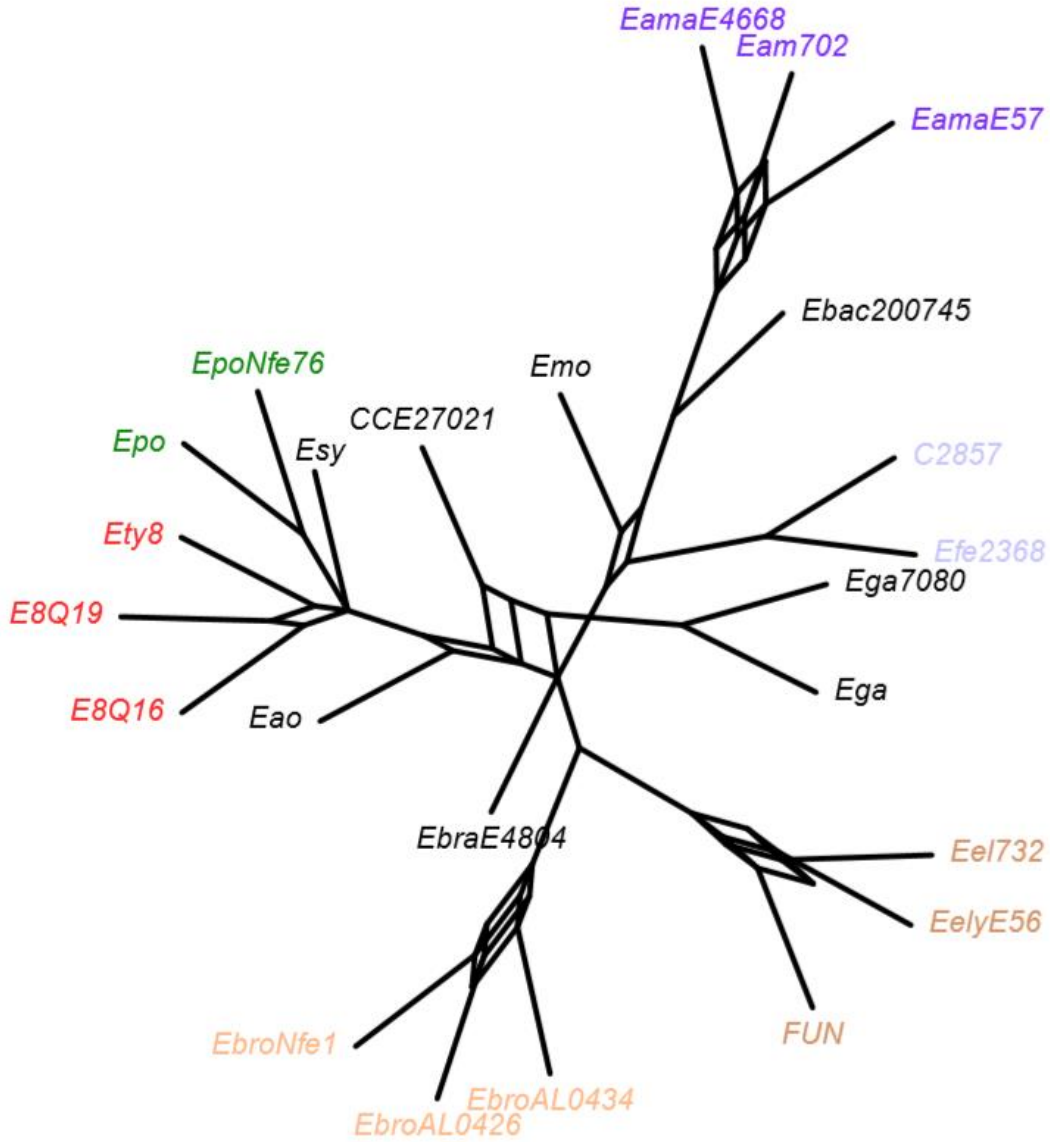

Figure 5. ConsensusNet of all genes not contributing to monophyly of the* E. typhina *clade. E. Typhina *strains are examined in red. Threshold parameters are set at 0.2.*

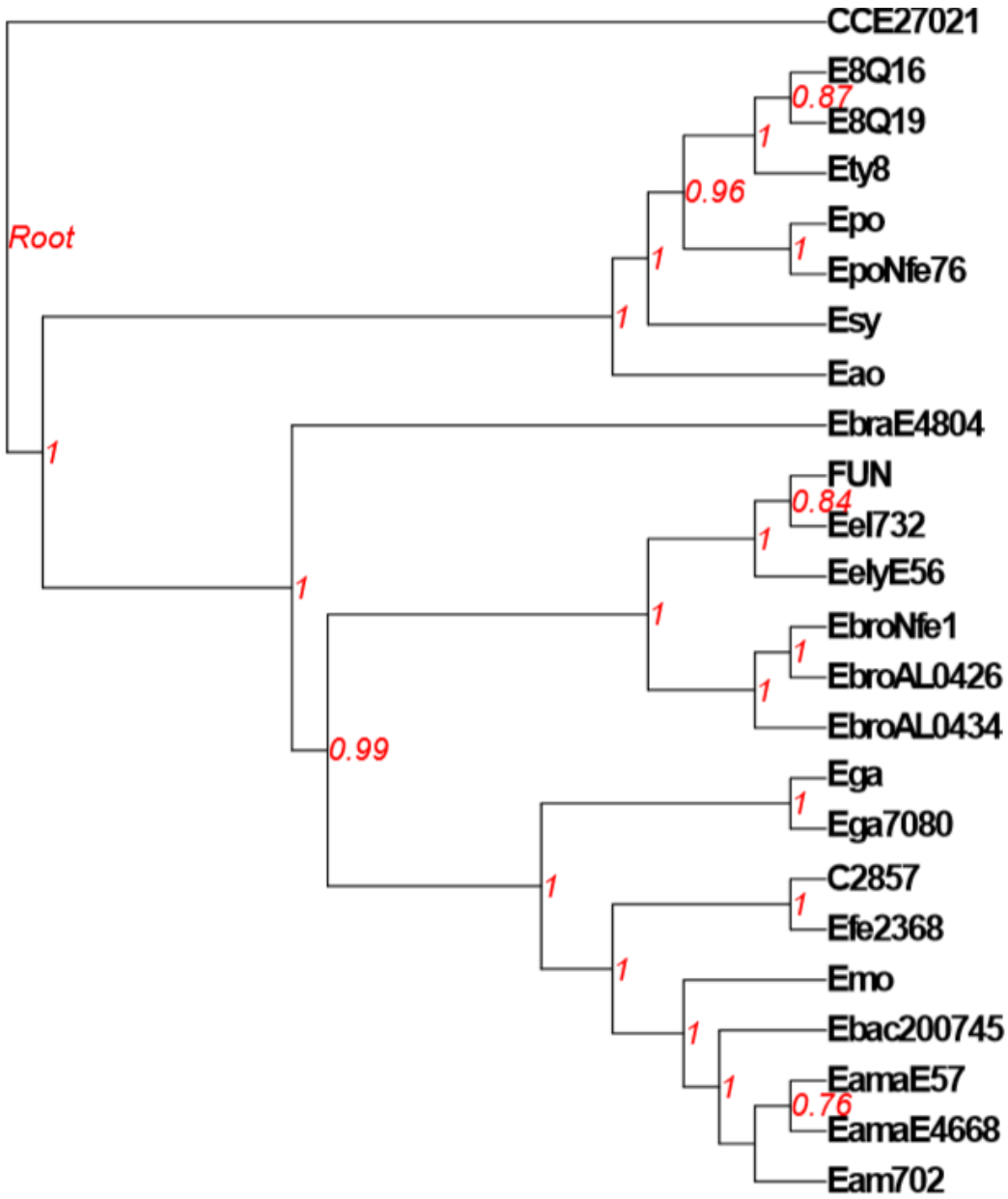

***Figure 6. Phylogenetic tree of relationships among* E. typhina *and other* Epichloë**. *This tree utilises genes that are non-monophyletic for* E. typhina.

3.6. Geography and host taxonomy do not pattern with evolutionary history

Having analysed potential qualitative drawbacks and alternative trees, I used the strongly supported species tree as a grounding for metadata annotation. This opens opportunities for further investigation. I firstly examined host information, including continental distribution.

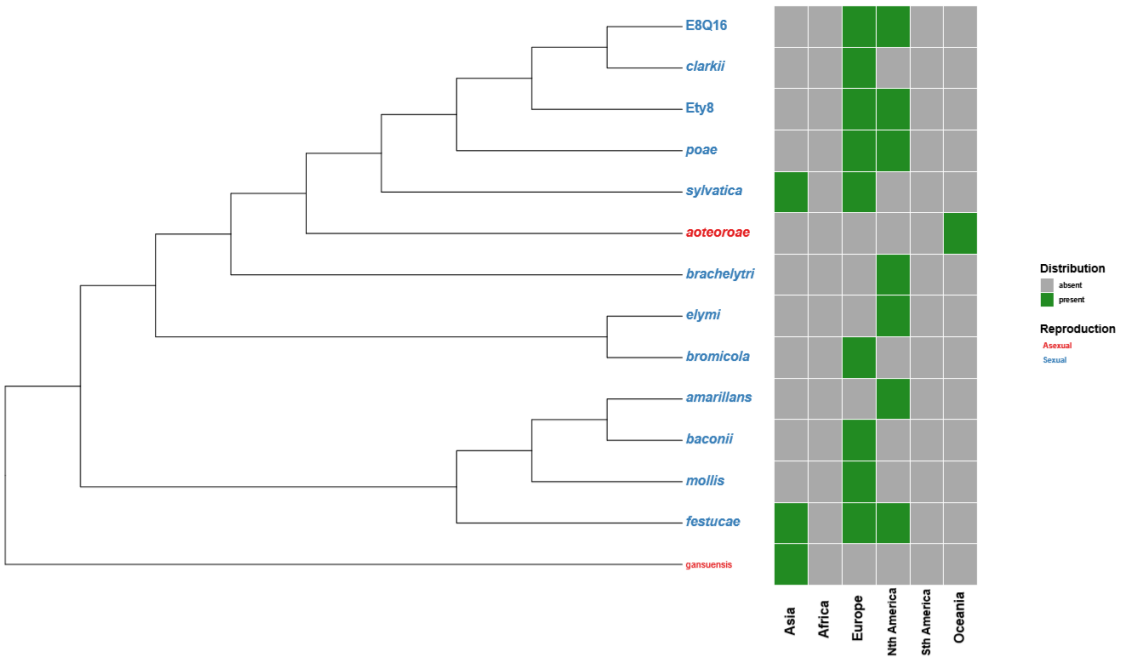

Figure 7. GGtree of all* Epichloë *species with annotation of reproduction, with attached host geography heatmap. *Tree was rooted at* Claviceps*, then root was dropped. Continental geographic distribution of hosts is displayed in green.*

Geographic distribution of hosts by continent does not appear to show patterns amongst Figure 7.

Metadata of host taxonomic information was challenging. Genera examination was too variable, while familial taxonomy was consistent among *Epichloë* hosts due to them being grasses (Poaceae). Therefore, I utilised subtribe information (**Figure 8**).

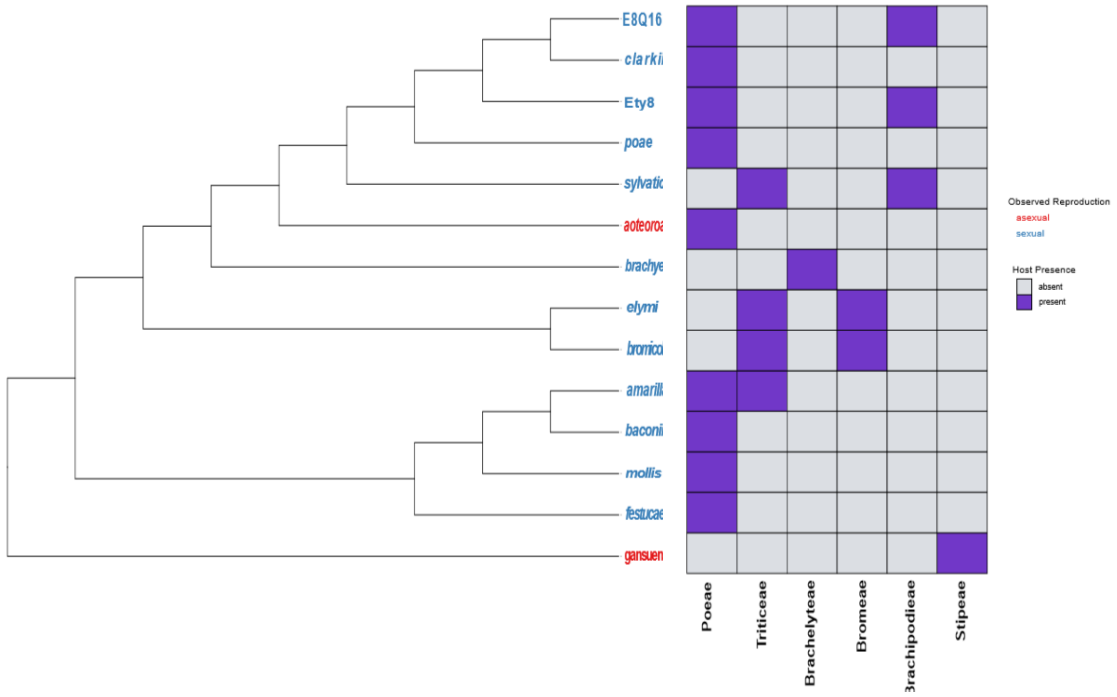

Figure 8. GGtree of all* Epichloë *species with annotation of reproduction, with attached host subtribe heatmap. *Tree was rooted at* Claviceps*, then root was dropped. Host subtribe presence is displayed in purple.*

Most *Epichloë* have hosts within the Poeae subtribe. This is the largest grass subtribe. Little patterns are observed regarding host switching or variability throughout the tree. *E.* t*yphina* strains (E8Q16 and Ety8) are both observed to have two different hosts, amongst a few other species. Other species have various strains found on different hosts. It is as yet unknown whether the same strain is found in multiple hosts.

3.7. Variation observed in origin of hybrids

While host information requires more analysis, other metadata results were more interesting. Hybrid parental species were provided and then graphed as to their parental genomes.

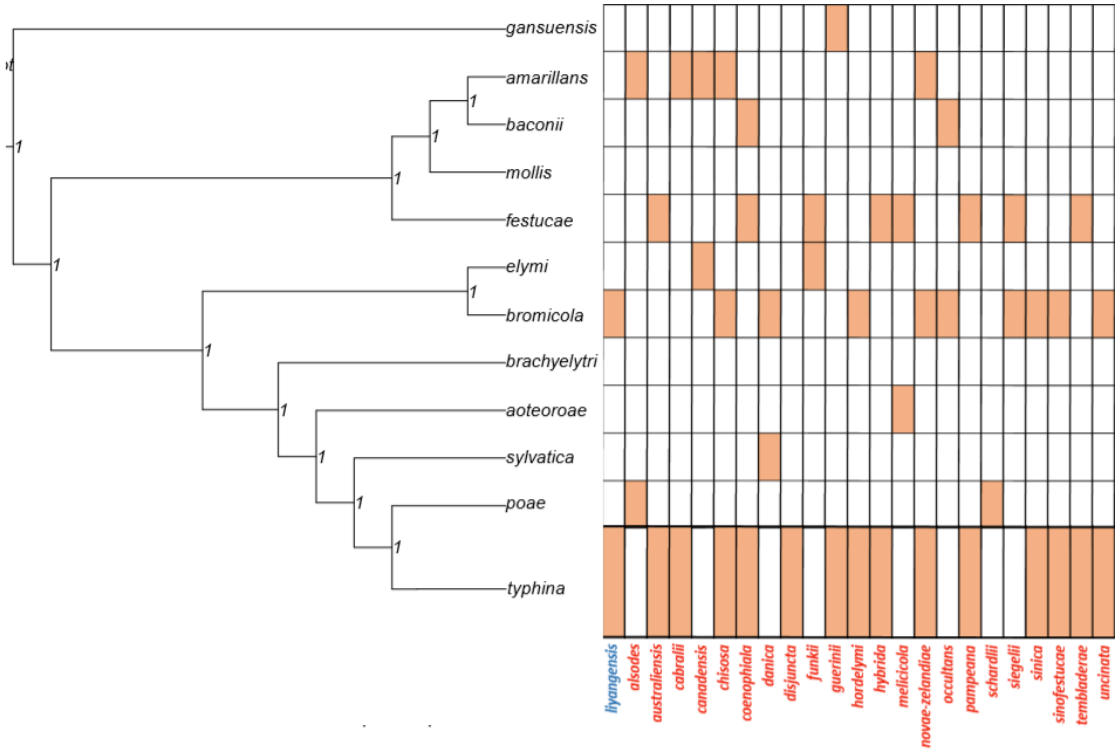

Figure 9. GGtree of all* Epichloë *species with attached hybrid heatmap. *Tree was rooted at* Claviceps*, then root was dropped. Hybrid parental presence is displayed in orange. Hybrid sexual reproduction is shown in blue.*

The *E*. *typhina* species complex is involved in 15/23 hybrids provided and shown in Figure 9. Some non-hybrid *Epichloë* species are not observed (*E. brachyleytri* and *E. mollis*) while others have only singular hybrid offspring (*E. aoteaoroae* and *E. sylvatica*). Hybrid formation does not appear to be patterned throughout the tree, although the most derived group, *E. typhina*, is the most 'involved'.

3.8. Cyclic peptides are conserved and divergent among taxa

Secondary metabolite analysis included that of cyclic peptides. I was provided ortholog identities closely associated with cyclic peptide formation and these were examined for presence within *Epichloë* strains.

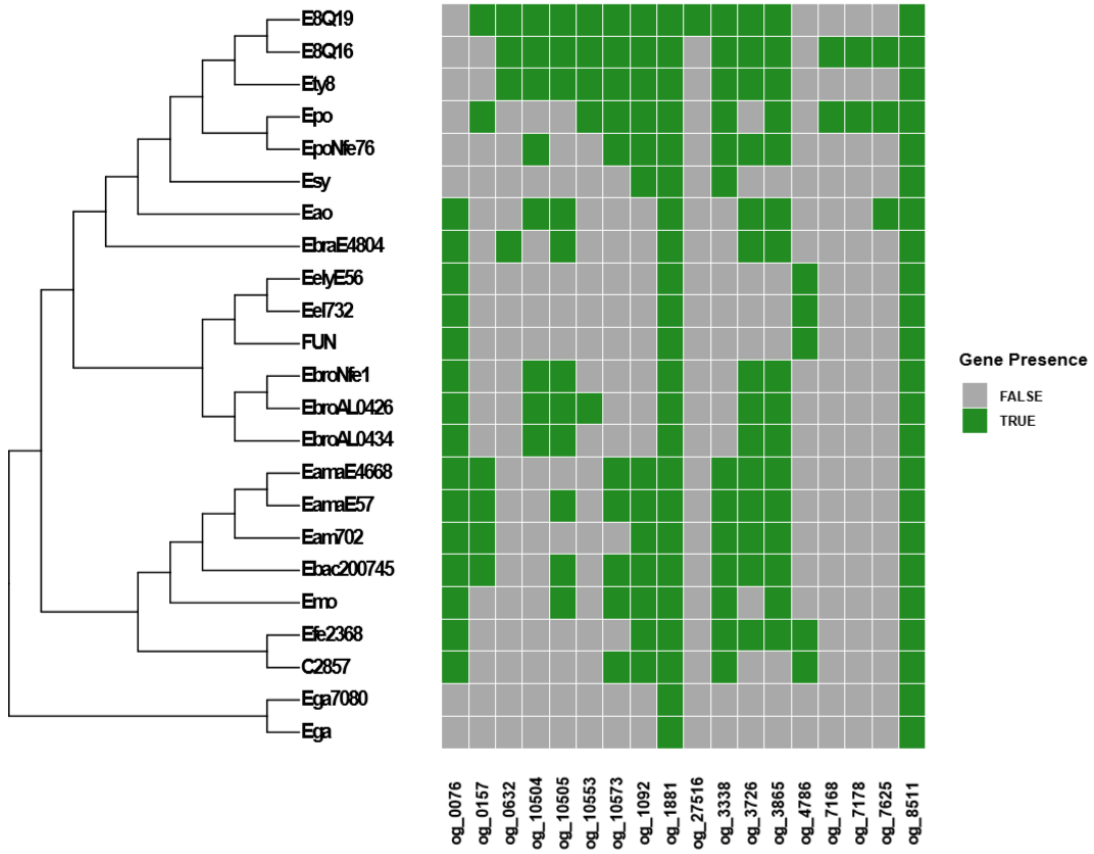

Figure 10. GGtree of all* Epichloë *strains with attached cyclic peptide gene heatmap. *Tree was rooted at* Claviceps*, then root was dropped. Ortholog presence is displayed in green.*

Og_1881 and og_8511 are highly conserved throughout Figure 10 in all strains. Other orthologs are found only in certain clades and appear to be localised to select groups. This includes several genes shared amongst *E. typhina* or *E. bromicola* respectively. Interestingly, some genes are found amongst non-monophyletic groups, such as those shared amongst E8Q16 and Epo.

4.0. Discussion

4.1. A single tree decirbes the organismal relationshpis in *Epichloë*

A highly supported species tree is a surprising and positive result from this study. Utilising ASTRAL, confidence nodes of 1 were found at all nodes on a species and strain tree. It is interesting to note that, parallel to much earlier studies on β-tubulin genes by Craven et al. (2001), two main clades are observed. These consist of *E. amarillans* and *E. festucae;* or *E. typhina* complex species, *E. elymi* and *E. bromicola* respectively. However, contrasting to earlier studies, the *E. typhina* clade is much larger with a broader range of species. This could be due to samples provided, or alternatively a better representation of genomic data. The single well-resolved tree found in this study can be used for studies focusing on host-interactions, the gain and loss of genes over evolutionary time and the biogeography of *Epichloë* lineages.

In addition to this well-resolved species tree, it is important to note that many individual gene trees do not match this topology. This is demonstrated by my consensusNet and investigation into non-monophyletic gene clades. That some gene trees differ from the species tree is an important

finding, as the inappropriate use of a single species tree for all genes in molecular evolution studies can lead to spurious results (Mendes & Hahn, 2016).

Variability amongst members of the *E. typhina* clades in the consensusNet and other graphics is not surprising. Previous research has pointed to evidence of a 'species complex' with high levels of hybridisation and transspecies polymorphisms (Leuchtmann, Bacon, Schardl, White, & Tadych, 2014). This is further evidenced by the overrepresentation of *E. typhina* in hybrid parental studies. The distinguishing of *E. clarkii, E. poae* and *E. sylvatica* are still unclear, especially with *E. clarkii* falling within other currently identified *E. typhina typhina* subspecies.

Various caveats of this species tree are present. Firstly, ASTRAL has previously discussed drawbacks including possibilities of incorrect confidences and potential to be influenced by other signals. It is however currently the best software for large scale datasets and minimises the 'anomaly zone' where the most common gene tree is otherwise not represented (Liu, Anderson, Pearl, & Edwards, 2019). In addition, the requirement of ASTRAL to rely on 'full' gene trees means that early-stage bioinformatic pipelines including parameters in ortholog identification could have had significant impact on the final species tree. In our data, many ortholog groups were lost due to software restrictions.

4.2. Metadata inclusion

Geographical data as displayed here suggests isolation in different regions is not related to evolutionary divergence. Most species displayed are found in Europe or North America. However, this potentially due to sequence availability or provided data quality. It was not possible to distinguish between introduced or natural range of the fungus from the data given. Distribution is an area of continuing research (Cagnano et al., 2019), primarily due the generally asymptomatic nature of *Epichloë* species. For the same reason, host taxonomy data may also be improved with further investigation.

Cyclic peptide distribution is a particularly exciting tangent for our data. Highly conserved ortholog presence throughout the tree, as well as areas of high variability, relates to both the high presence of genes under diversifying selection in fungi (Schardl et al., 2013) and the conservation of highly useful basal biological function genes (Ponting, 2017).

4.3. Next steps

In the realms of our study, several further steps could be focussed on. To enhance figure coherence and support for the species tree, extensions to consensusNet graphics could be made. Bootstrap support could be annotated on the net, with variation in threshold figures. This would further display areas of further interest. In addition, parameters on ortholog calling, including reducing or increasing requirements over paralog identification, would be beneficial for further study. Changing parameters will likely increase the number of genes useful for 'full' tree investigation.

The high GC content displayed by aligned sequences should be further investigated. It is noted by personal communication among other members of my lab team, also working on *Epichloë,* that genome data is AT rich. Determining the significance of aligned sequence data (being a small subset of the genome at only 6040 genes) being GC rich would be beneficial to determining gene function, conservation parameters and research involving the otherwise AT rich components of the genome.

An extension to our study should involve questions brought up by non-monophyly data. While producing trees from all non-monophyletic genes in certain clades was unsuccessful, transspecies polymorphisms should be investigated through specific gene isolation. 'Backtracking' through our methodology to extract the gene identity from the ortholog alignment would be beneficial to isolate genes of interest. Genes creating trees vastly different from the species tree should be considered as interests in functional genomics.

Further functional work should be carried out on orthologs identified with cyclic peptide protein producing related domains. Several orthologs are conserved throughout the *Epichloë* lineage as identified in figure 10. Gene identification, genome annotation, and *in vitro* lab investigation will be essential. In addition, the structure of each ortholog should be analysed for repetition and function. Investigation into hybrid origins is another future avenue from these findings. While research is extensive, incorporation of a highly supported species tree as found here is essential. Evolutionary origins and phylogenetic analysis of hybrids will be useful to further support the tree.

5.0. Bibliography

Cagnano, G., Roulund, N., Jensen, C. S., Forte, F. P., Asp, T., & Leuchtmann, A. (2019). Large scale screening of *Epichloë* endophytes infecting *Schedonorus pratensis* and other forage grasses reveals a relation between microsatellite-based haplotypes and loline alkaloid Levels. *Frontiers in Plant Science, 10*(765). doi:10.3389/fpls.2019.00765

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., . . . Wilczynski, B. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics, 25*(11), 1422-1423.

Craven, K., Hsiau, P., Leuchtmann, A., Hollin, W., & Schardl, C. (2001). Multigene phylogeny of *Epichloë* species, fungal symbionts of grasses. *Annals of the Missouri Botanical Garden*, 14-34.

Guerre, P. (2016). Lolitrem B and indole diterpene alkaloids produced by endophytic fungi of the genus *Epichloë* and their toxic effects in livestock. *Toxins*(2), 47. doi:10.3390/toxins8020047

Johnson, R. D., Lane, G. A., Koulman, A., Cao, M., Fraser, K., Fleetwood, D. J., . . . Dyer, J. M. (2015). A novel family of cyclic oligopeptides derived from ribosomal peptide synthesis of an *in planta*-induced gene, gigA, in *Epichloë* endophytes of grasses. *Fungal Genetics and Biology, 85*, 14-24. doi:10.1016/j.fgb.2015.10.005

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39-64): Springer.

Leuchtmann, A., Bacon, C. W., Schardl, C. L., White, J. F., & Tadych, M. (2014). Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia, 106*(2), 202-215. doi:10.3852/13-251

Liu, L., Anderson, C., Pearl, D., & Edwards, S. V. (2019). Modern phylogenomics: Building phylogenetic trees using the multispecies coalescent model. In M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 211-239). New York, NY: Springer New York.

Lugtenberg, B. J. J., Caradus, J. R., & Johnson, L. J. (2016). Fungal endophytes for sustainable crop production. *FEMS Microbiology Ecology, 92*(12). doi:10.1093/femsec/fiw194

Mendes, F. K., & Hahn, M. W. (2016). Gene Tree Discordance Causes Apparent Substitution Rate Variation. *Systematic biology, 65*(4), 711-721. doi:10.1093/sysbio/syw018

Palmer, J. (2018). Funannotate: Eukaryotic genome annotation pipeline. Retrieved from https://funannotate.readthedocs.io

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular biology and evolution, 5*(5), 568-583.

Ponting, C. P. (2017). Biological function in the twilight zone of sequence conservation. *BMC Biology, 15*(1), 71. doi:10.1186/s12915-017-0411-5

Ruying, W., Bruce, B. C., & Faith, C. B. (2019). Transcriptome analysis of choke stroma and asymptomatic inflorescence tissues reveals changes in gene expression in both *Epichloë festucae* and Its host plant *Festuca rubra subsp. rubra*. *Microorganisms*(11), 567. doi:10.3390/microorganisms7110567

Schardl, C. L., Young, C. A., Hesse, U., Amyotte, S. G., Andreeva, K., Calie, P. J., . . . Zeng, Z. (2013). Plant-symbiotic fungi as chemical engineers: Multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS genetics, 9*(2), e1003323. doi:10.1371/journal.pgen.1003323

Scott, B. (2001). *Epichloë* endophytes: Fungal symbionts of grasses. *Current Opinion in Microbiology, 4*(4), 393-398. doi:https://doi.org/10.1016/S1369-5274(00)00224-1

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England), 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Umemura, M., Nagano, N., Koike, H., Kawano, J., Ishii, T., Miyamura, Y., . . . Machida, M. (2014). Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in *Aspergillus flavus*. *Fungal genetics and biology: FG & B, 68*, 23-30. doi:10.1016/j.fgb.2014.04.011

Winter, D. J., Ganley, A. R. D., Young, C. A., Liachko, I., Schardl, C. L., Dupont, P.-Y., . . . Cox, M. P. (2018). Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus Epichloë festucae. *PLoS genetics, 14*(10), e1007467. doi:10.1371/journal.pgen.1007467

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics, 19*(6), 153. doi:10.1186/s12859-018-2129-y

6.0. Supplementary figures

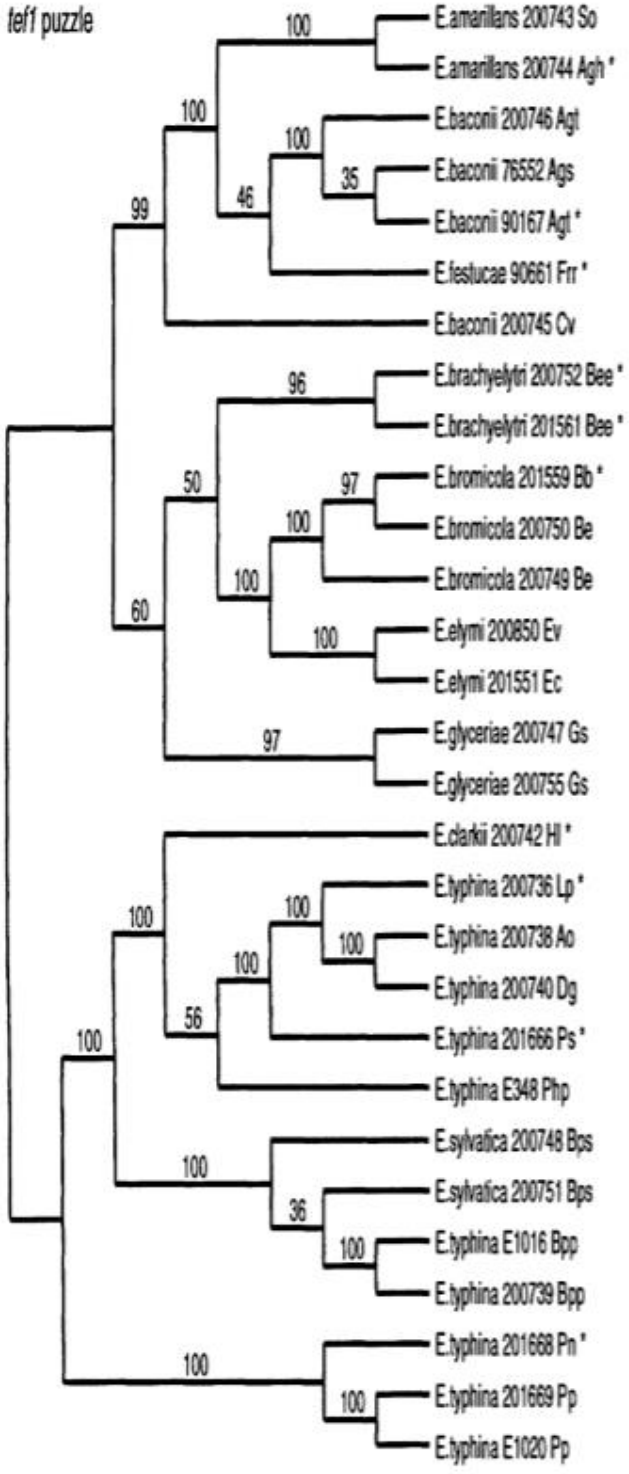

Figure 11. Epichloë spp. Tef1 gene tree based on ML puzzle analysis of introns 1-5. Tree is based on a Hasegawa-Kishino-Yano model with ts/tv = 2 (kappa = 3.97409); number of puzzling steps = 1000. Tree shown represents the majority-rule consensus, and numbers at branches are estimations of support for each branch. Note: Caption taken from source (Craven et al., 2001).

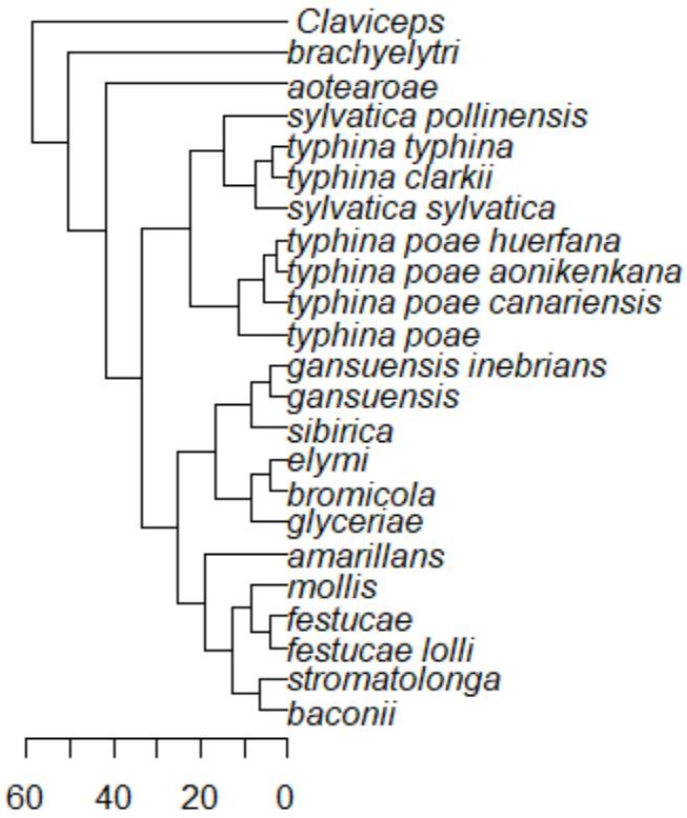

Figure 12. Draft species tree from β-tubulin marker gene information with addition of evolutionary timeline. Timeline utilises a 58.5mya estimate from Leuchtmann et al. (2019).

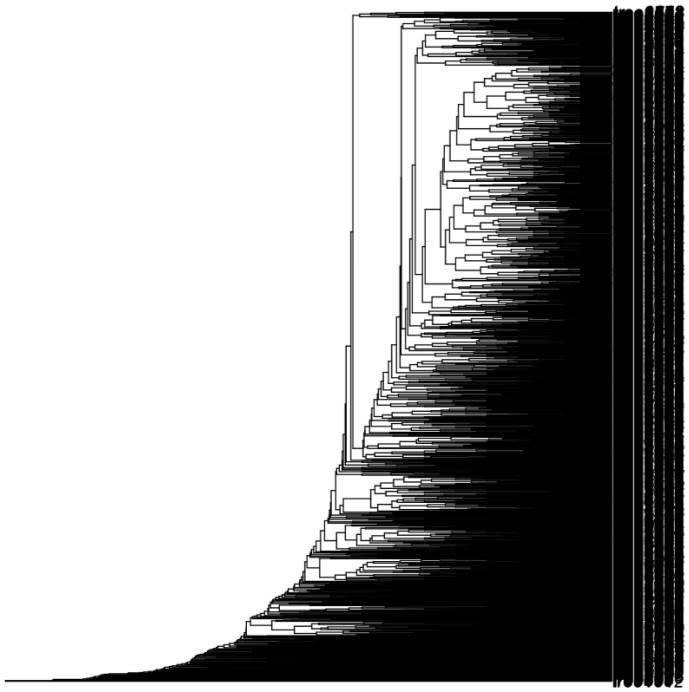

Figure 13.The topological distance between all Epichloë *gene trees was computed and graphed. The variability amongst tree branches displays that there is one main relationship amongst taxa.*