

## Research Article

# Sequence Analysis of SSR-Flanking Regions Identifies Genome Affinities between Pasture Grass Fungal Endophyte Taxa

Eline van Zijll de Jong<sup>1,2,3</sup> Kathryn M. Guthridge<sup>1,2,4</sup> German C. Spangenberg<sup>1,2,4,5</sup>  
and John W. Forster<sup>1,2,4,5</sup>

<sup>1</sup> Department of Primary Industries, Biosciences Research Division, Victorian AgriBiosciences Centre, 1 Park Drive, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia

<sup>2</sup> Molecular Plant Breeding Cooperative Research Centre, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia

<sup>3</sup> Bioprotection Research Centre, P.O. Box 84, Lincoln University, Lincoln 7647, Canterbury, New Zealand

<sup>4</sup> Dairy Futures Cooperative Research Centre, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia

<sup>5</sup> La Trobe University, Bundoora, VIC 3086, Australia

Correspondence should be addressed to John W. Forster, john.forster@dpi.vic.gov.au

Received 14 October 2010; Accepted 10 December 2010

Academic Editor: Hiromi Nishida

Copyright © 2011 Eline van Zijll de Jong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fungal species of the *Neotyphodium* and *Epichloë* genera are endophytes of pasture grasses showing complex differences of life-cycle and genetic architecture. Simple sequence repeat (SSR) markers have been developed from endophyte-derived expressed sequence tag (EST) collections. Although SSR array size polymorphisms are appropriate for phenetic analysis to distinguish between taxa, the capacity to resolve phylogenetic relationships is limited by both homoplasy and heteroploidy effects. In contrast, nonrepetitive sequence regions that flank SSRs have been effectively implemented in this study to demonstrate a common evolutionary origin of grass fungal endophytes. Consistent patterns of relationships between specific taxa were apparent across multiple target loci, confirming previous studies of genome evolution based on variation of individual genes. Evidence was obtained for the definition of endophyte taxa not only through genomic affinities but also by relative gene content. Results were compatible with the current view that some asexual *Neotyphodium* species arose following interspecific hybridisation between sexual *Epichloë* ancestors. Phylogenetic analysis of SSR-flanking regions, in combination with the results of previous studies with other EST-derived SSR markers, further permitted characterisation of *Neotyphodium* isolates that could not be assigned to known taxa on the basis of morphological characteristics.

## 1. Introduction

Fungal endophytes of the genus *Neotyphodium* and *Epichloë* are widespread in temperate grasses of the Poaceae subfamily Pooideae [1, 2]. In agronomically important pasture grasses such as perennial ryegrass (*Lolium perenne* L.) and tall fescue (*Festuca arundinacea* Schreb. [Darbysh.] syn. *L. arundinacea*), the respective symbionts (*N. lolii* [Latch, Christensen, and Samuels] Glenn, Bacon, and Hanlin and *N. coenophialum* [Morgan-Jones and Gams] Glenn, Bacon,

and Hanlin) confer both beneficial and detrimental agronomic traits [3–7]. Molecular genetic marker-based studies have contributed to knowledge of endophyte genetics, taxonomy, and phylogeny. *Neotyphodium* species were originally placed in the form (asexual) genus *Acremonium* [8], but were reclassified into the new form genus *Neotyphodium* when sequence analysis of ribosomal RNA-encoding (rDNA) genes indicated a monophyletic group with the sexual *Epichloë* species [1, 9]. Phylogenetic analysis of genes for conserved proteins such as the  $\beta$ -tubulin gene (*tub2*), translation

elongation factor 1- $\alpha$  (*tefA*), and actin (*actG*) also provided evidence for close relationships between *Neotyphodium* and *Epichloë* species [10–17].

Although taxa such as *N. lolii* are haploid in nature, other *Neotyphodium* species were shown to contain multiple gene copies and conform to heteroploid genomic constitutions [17]. The single or multiple gene copies of asexual *Neotyphodium* species appear to correspond to those of specific haploid *Epichloë* species. This observation has been interpreted to support a hybrid origin for heteroploid taxa: for instance, *N. coenophialum* has been proposed to have arisen through hybridisation and subsequent nuclear fusion events involving the extant taxa *E. typhina*, *E. baconii*, and *E. festucae*. The relative genome sizes of haploid and heteroploid endophytes (c. 30 Mb for *N. lolii*; c. 60 Mb for *N. coenophialum*) lend some support to this hypothesis [18], subject to the possibility of selective gene loss subsequent to hybridisation events. Phylogenetic relationships between endophyte taxa are hence complex and reticulated. Sequence analysis of individual gene loci may be used to infer such relationships based on affinities between shared genomes. However, performance differences between individual genes have been observed. The resolution capacity provided by rDNA and *actA* genes was low in comparison to other genes [12–15], possibly due to homoplasmy effects [13]. Heteroploid-like *Neotyphodium* species also display aneuploidy for some loci, such as the rDNA gene, limiting resolution of complete phylogenies [19]. A broader survey of gene classes is hence desirable to further clarify affinities between endophyte taxa.

Simple sequence repeats (SSRs) or microsatellites [20] have been widely used for analysis of genetic variation within and between closely related species [21]. A high rate of mutation [22] renders SSR array length polymorphism particularly useful for intraspecific genetic studies. However, sequence analysis has revealed complex mechanisms controlling allele size variation, limiting the efficiency of interspecific phylogenetic analysis. Repeat number variation is thought to arise from polymerase slippage during replication [23], but constraints on threshold size for allele expansion [24] and on allele size range [25] are evident. In addition, interruptions of the repeat structure tend to stabilise SSR loci [26]. Constraints on allele size may consequently lead to inaccurate assessment of phylogenetic divergence between taxa. Size homoplasmy of distinct alleles arising from insertions, deletions, and base substitutions in the SSR flanking regions are also common [27–30]. Changes in flanking regions appear to occur independently of changes in the SSR repeat array [28, 30]. Due to these factors, allelic variation of SSR loci, as assessed by amplicon size variation, is appropriate only for phenetic analysis and not suitable for phylogenetic reconstruction.

In contrast, several studies have performed phylogenetic interpretation through analysis of SSR-flanking sequence regions. The resolving power of evolutionary studies using individual structural genes may be constrained by limited divergence [31], and studies of a small number of gene loci may not be representative of whole-genome variation [32]. However, the abundant genomic distribution

of SSRs [33–35] permits phylogenetic assessment across the transcriptional units of multiple gene classes. SSR-flanking regions have been used for phylogenetic analysis of multiple organisms [31, 32, 36, 37], resolving relationships to otherwise inaccessible levels [31, 32, 37].

Consistent with these previous studies, gene-associated SSR loci have previously been shown to discriminate endophyte taxa based on size polymorphism [38], but did not permit phylogenetic analysis. The present study describes the comparison of sequences that flank the SSR array in 5 independently selected gene loci, across 23 distinct fungal endophyte isolates. The derived data have determined the extent of molecular variation underlying SSR size polymorphism, confirmed current models for genome affinities, inferred phylogenetic relationships and models of genome evolution (including a role for selective gene loss), and elucidated the genomic origin of several previously unclassified *Neotyphodium* taxa.

## 2. Materials and Methods

**2.1. Endophyte Isolates.** Phylogenetic analysis was performed on 20 endophyte isolates representing three *Neotyphodium* and five *Epichloë* taxa, as well as three *Neotyphodium* isolates which could not be assigned to known taxa based on their morphological characters (A. Leuchtmann, pers. comm.), and a tall fescue endophyte taxon (FaTG-2) which has yet to be allocated a Linnean name (Table 1). Endophyte isolates were cultured and DNA was extracted as described previously [38].

**2.2. DNA Sequence Analysis of EST-SSR Amplicons.** Genomic amplicons obtained with primer pairs designed to five EST-SSR loci (NCESTA1AB04, NCESTA1FH03, NCESTA1AG07, NLESTA1GF09, and NLESTA1NF04) were analysed. Amplicons were obtained as described previously [38]. Amplicons from haploid taxa were analysed by direct sequencing, with the exception of locus NLESTA1NF04 for which sequencing was performed on purified plasmids containing the cloned amplicon [38]. Sequencing reactions were performed in 10  $\mu$ l reaction volumes containing 4  $\mu$ l Sequencing Reagent Premix from the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Biosciences, Little Chalfont, UK), 0.5  $\mu$ M of forward or reverse primer for the locus of interest and 5  $\mu$ l of amplicon in a thermocycler (GeneAmp; PE Applied Biosystems, Forster City, California, USA.) programmed for 20 seconds at 92°C followed by 30 cycles of 20 s at 95°C, 15 s at 50°C, 2 minutes at 60°C, then 10 min at 60°C. Sequencing products were purified using Autoseq 96 plates (Amersham Biosciences), dried at 80°C for 30 min and resuspended in 5  $\mu$ l of sterile Milli-Q water before analysis on a ABI Prism 3700 automated sequencer (PE Applied Biosystems). For multiple products amplified in a single reaction from nonhaploid taxa, cloning was used to separate the different amplicons. Following amplification, the products were purified using a Microspin S-300 HR Column (Amersham Biosciences). The purified products were cloned into pGEM-T Easy Vector (Promega, Madison, Wisconsin,

TABLE 1: Endophyte isolates used for phylogenetic analysis.

Species or taxon	Isolate	Host species	Origin	Source
<i>N. coenophialum</i>	9309	<i>Festuca arundinacea</i>	France	ETH Zürich <sup>1</sup>
<i>N. coenophialum</i>	9920/1	<i>F. arundinacea</i>	U.S.A.	ETH Zürich
<i>N. coenophialum</i>	9920/2	<i>F. arundinacea</i>	U.S.A.	ETH Zürich
<i>N. coenophialum</i>	9920/3	<i>F. arundinacea</i>	U.S.A.	ETH Zürich
FaTG-2	8907	<i>F. arundinacea</i>	U.S.A.	ETH Zürich
<i>N. lolii</i>	9601	<i>Lolium perenne</i>	Belgium	ETH Zürich
<i>N. lolii</i>	Ellett H5837	<i>L. perenne</i>	New Zealand	DPI-Hamilton <sup>2</sup>
<i>N. lolii</i>	North African 6	<i>L. perenne</i>	Morocco	DPI-Hamilton
<i>N. lolii</i>	Victorian 2	<i>L. perenne</i>	Australia	DPI-Hamilton
<i>N. uncinatum</i>	9414	<i>F. pratensis</i>	Germany	ETH Zürich
<i>Neotyphodium</i> sp.	9303/2	<i>Elymus europaeus</i>	Switzerland	ETH Zürich
<i>Neotyphodium</i> sp.	9727	<i>F. arizonica</i>	U.S.A.	ETH Zürich
<i>Neotyphodium</i> sp.	9728	<i>L. perenne</i>	New Zealand	ETH Zürich
<i>E. baconii</i>	9707	<i>Agrostis tenuis</i>	Switzerland	ETH Zürich
<i>E. bromicola</i>	9630	<i>Bromus erectus</i>	Switzerland	ETH Zürich
<i>E. clarkii</i>	9401	<i>Holcus lanatus</i>	Switzerland	ETH Zürich
<i>E. festucae</i>	9412	<i>F. gigantea</i>	Switzerland	ETH Zürich
<i>E. festucae</i>	9436	<i>F. pratensis</i>	Switzerland	ETH Zürich
<i>E. festucae</i>	9713	<i>F. rubra</i>	Switzerland	ETH Zürich
<i>E. festucae</i>	9718	<i>F. gigantea</i>	Switzerland	ETH Zürich
<i>E. festucae</i>	9722	<i>F. rubra</i>	England	ETH Zürich
<i>E. sylvatica</i>	9301	<i>Brachypodium sylvaticum</i>	Switzerland	ETH Zürich
<i>E. typhina</i>	9635	<i>Dactylis glomerata</i>	Switzerland	ETH Zürich

<sup>1</sup>ETH Zürich: Geobotanisches Institut, ETH, Zürich, Switzerland.<sup>2</sup>DPI-Hamilton: Department of Primary Industries, Primary Industries Research Victoria, Hamilton, Victoria, Australia.

U.S.A.) and transformed into competent cells. Inserts were amplified from transformed colonies and sequenced.

Consensus sequences were derived through analysis of several independently isolated clones or direct sequencing of both strands. Sequences were compared using Sequencher (version 4.0) (Gene Codes Corporation, Ann Arbor, Michigan, U.S.A.). BLASTX (version 2.2.1 and 2.2.6) [39] was used to search for similarities between the EST sequence of SSR loci and protein sequences in the protein databases available from the National Centre for Biological Information (NR, PDB and SwissProt; <http://www.ncbi.nlm.nih.gov/BLAST/>).

**2.3. Phylogenetic Analysis of EST-SSR Amplicons.** The DNA sequences of unique amplicons were prepared for phylogenetic analysis by compilation in FastA format into a single file for sequence alignment in ClustalX (version 1.8) [40]. Manual realignment of sequences removed primer termini and polymorphic SSR arrays and converted insertion-deletion (indel) regions into single multistate characters. Sequence alignments were analysed by clustering or tree searching methods available in PHYLIP (version 3.6a3) (J. Felsenstein, University of Washington, Seattle, Washington, U.S.A., available from <http://evolution.gs.washington.edu/phylip.html>). For parsimony analysis, sequences were analysed with indels either removed, or coded as single multistate characters. Where multiple trees were resolved, the Kishino-Hasegawa-Templeton (KHT) test [41, 42] or Shimodaira-Hasegawa

(SH) test [43] was used to test for significant differences. The robustness of the trees was measured by the Bootstrap method [44] with 1000 replicates. A bootstrap value of 70% or greater was considered to be well supported. For Maximum Likelihood (ML) and distance-based analysis, sequences were analysed with indels removed. Distance matrices were obtained using the F84 model [42, 45] and clustered using the Fitch-Margoliash (FM) method [46] or Neighbor-Joining (NJ) method [47]. The transition/transversion ratio was estimated using the Tree-Puzzle program (version 5.0) [48] or by the ML method. To estimate the transition/transversion ratio by the ML method, different possible values for the transition/transversion ratio were evaluated in multiple runs to find the value with the maximum likelihood estimate. The same approach was used to estimate the among-site rate heterogeneity by the ML method. To estimate the among-site rate heterogeneity by the Minimum Evolution (ME) method, distance matrices generated for each site were analysed and the total branch length was taken as the estimated value of the rate of change for each of the sites.

### 3. Results

**3.1. Characterisation of Endophyte EST-SSR Amplicons.** Genomic amplicons from five EST-SSR loci, ranging in length from 181–385 bp, were characterised from 12

TABLE 2: Characteristics of EST-derived SSR loci used for phylogenetic analysis of endophyte isolates.

	EST-SSR locus				
	NCESTA1AB04	NCESTA1FH03	NCESTA1GA07	NLESTA1GF09	NLESTA1NF04
Product size (bp)	241–257	217–252	181–206	364–385	242–265
Number of indels	5	7	3	6	2
Number of unique SSR array variants	2	13	11	1	19
Number of unique gene variants	14	13	13	11	15
Composition of products	coding + 5'-UTR	unknown	intron + unknown	coding + intron	unknown
Length of unit for phylogenetic analysis <sup>1</sup> (bp)	195	163	116	321	166
CDS or exon	151		46	238	
UTR or intron	44		70	83	
Number of informative characters <sup>1</sup>	40	37	23	44	51
CDS or exon	30		10	30	
UTR or intron	10		13	14	
Percent informative characters <sup>1</sup>	21	21	20	14	31
CDS or exon	20		22	13	
UTR or intron	23		19	17	
Number of trees resolved	3	1	4	2	2

<sup>1</sup> Each indel was coded as a single multistate character.

*Neotyphodium* and 10 *Epichloë* isolates, as well as FaTG-2 (Table 2). The sequences of two loci (NCESTA1AB04 and NLESTA1GF09) shared amino acid sequence similarity with hypothetical or predicted proteins of *Neurospora crassa* and *Magnaporthe grisea* and were mainly composed of coding sequence (CDS) as well as 5'-untranslated region (5'-UTR) and intron sequences, respectively. The sequences of the remaining three loci did not show similarity with any proteins in public databases. Amplification products from locus NCESTA1GA07 were predominantly composed of intron sequence, based on comparison of EST and genomic DNA sequences.

Size polymorphisms between taxa for the selected loci resulted from variation at a number of indel sites (Table 2). Differences also occurred in the repeat unit number of the SSR array for three loci (NCESTA1FH03, NCESTA1GA07 and NLESTA1NF04; data not shown), accounting for the majority of the observed size polymorphisms. A number of sequence haplotypes (defined here as amplicons with multiple sequence variant content, but clearly related to a single common reference) for each locus were identified across the sample set, with identity observed between those of several *Neotyphodium* and *Epichloë* isolates. Single haplotypes for individual genes were observed for *N. lolii*, unclassified *Neotyphodium* isolate 9727, and the different *Epichloë* species, while *N. coenophialum*, FaTG-2, *N. uncinatum*, and unclassified *Neotyphodium* isolates 9303/2 and 9728 generated multiple haplotypes. The number of haplotypes present in these species varied between loci, but with a maximum of three for *N. coenophialum* and *Neotyphodium* isolates 9303/2 and 9728, and two for FaTG-2 and *N. uncinatum*. Variation was observed in cloning efficiency of different PCR products for those species possessing multiple haplotypes. Aberrant haplotypes, which were likely to be

chimeras generated by PCR-mediated recombination, were also obtained.

Pairwise comparisons identified between 23–51 informative characters for the different loci (Table 2). The proportion of informative characters ranged from 14% (locus NLESTA1GF09) to 31% (locus NLESTA1NF04), but was c. 20% for the other loci. A similar proportion of informative characters occurred in both the coding and noncoding sequences of the eligible loci.

### 3.2. Phylogenetic Analysis of Endophyte EST-SSR Amplicons.

Loci were analysed through sequence alignment (Supplementary Material: Appendices 1–5 available at doi:10.4067/2011/921312) individually, rather than as a combined dataset, due to variation of both inferred ploidy level and number of observed haplotypes between different SSR loci from heteroploid isolates. Between one and four trees were resolved for the different loci using the Parsimony method (Figures 1–5). A single tree was resolved for locus NCESTA1FH03 (Figure 2). The multiple trees obtained for loci NCESTA1AB04 (Figure 1), NLESTA1GF09 (Figure 4) and NLESTA1NF04 (Figure 5) only differed in the placement of one or two species. More variation was evident in the branching of the multiple trees identified for the locus NCESTA1GA07 (Figure 3). The trees, however, were not found to be significantly different in the KHT or SH tests. The majority of branches in the trees were supported by bootstrap analysis. Similar trees were resolved for the different loci using the ML, FM, and NJ methods (data not shown). In tests performed using the ML and ME methods, no significant differences were detected in the rate of change between coding and non-coding sequences or between exon and intron sequences for eligible loci (data not shown).

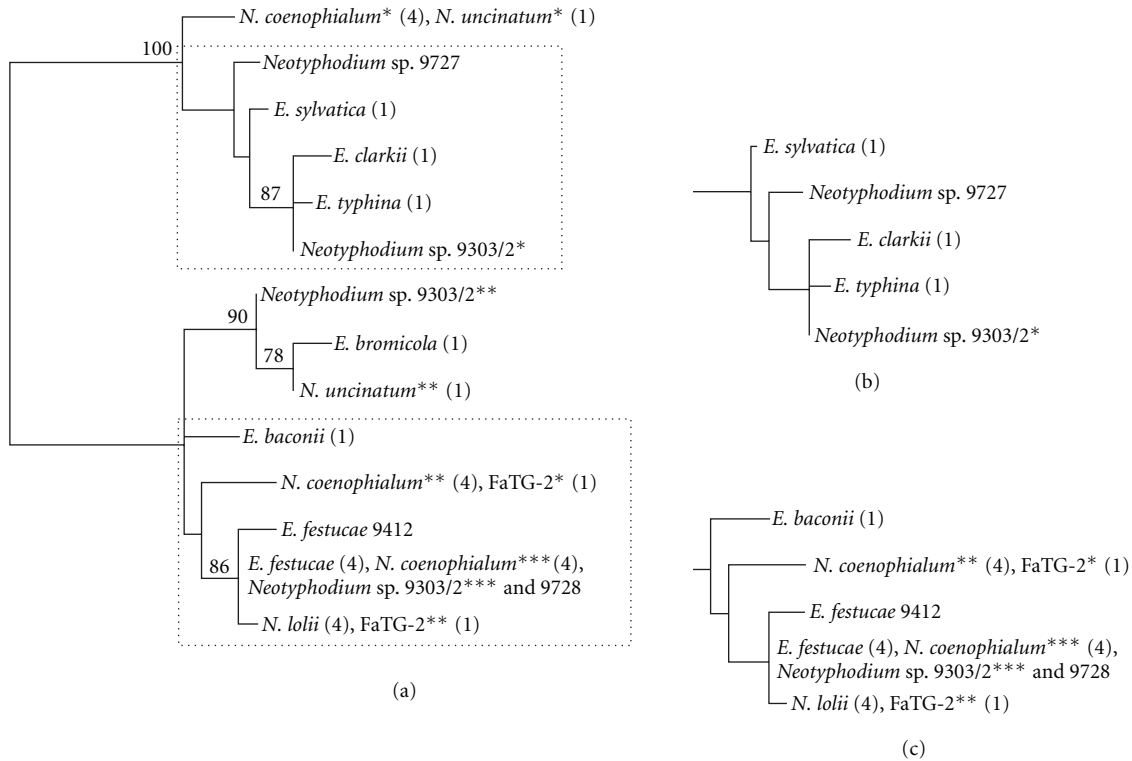


FIGURE 1: Parsimony analysis of sequence haplotypes derived from reference *Neotyphodium* and *Epichloë* isolates for the EST-SSR locus NCESTA1AB04. (a) One of three most parsimonious trees obtained. The left edge is the inferred midpoint root. Branches with bootstrap values of greater than 70% from 1000 bootstrap replications are marked. The number of isolates with an identical haplotype is indicated in the brackets following the species name. In the instances when multiple haplotypes were identified, the relevant variant number is indicated by the number of asterisks following the species name or isolate number. The boxed area indicated by the dotted line indicates the regions representing subtrees. ((b) and (c)) Subtrees showing the alternative branching of *E. sylvatica* and *E. baconii* in the other trees.

Phylogenetic analysis of the different loci revealed similar genomic relationships among endophyte species, as summarised in a network format (Figure 6). Close relationships between haplotypes from different taxa were deduced to indicate partial or complete commonality of genome content. Some locus-dependent differences in tree topology (Figures 1–5) were observed, but specific taxa consistently grouped together. In most instances, the *Epichloë* species were separated into two groups, separation being supported by bootstrap analysis. The first contained *E. festucae*, *E. baconii*, and *E. bromicola*, and the second contained *E. typhina*, *E. clarkii*, and *E. sylvatica*, *Neotyphodium* species being included within both groups. Group 1 *Epichloë* species were further divided into distinct branches according to their taxonomic classification, while Group 2 endophytes showed higher levels of genetic similarity. Close genetic relationships were evident between several *Neotyphodium* and *Epichloë* species. These relationships were observed in most of the trees and were also supported by bootstrap analysis. The single haplotypes derived from both *N. lolii* and *E. festucae* were grouped together in all trees and were identical in structure for two of the loci (NCESTA1FH03 and NLESTA1GF09). Multiple haplotypes from *N. coenophialum*, FaTG-2, and *N. uncinatum* were consistently associated with counterparts from specific *Neotyphodium* and *Epichloë* species.

*N. coenophialum* and *N. uncinatum* shared identical or very closely related haplotypes for all five loci. These variants also grouped with those from Group 2 *Epichloë* species. *N. coenophialum* also shared common haplotypes with FaTG-2 and *E. baconii*. The remaining haplotypes common to *N. coenophialum* and FaTG-2 grouped with the corresponding single haplotypes from *E. festucae* and *N. lolii*. For a subset of the target loci, *N. uncinatum*-derived sequences grouped with the corresponding haplotypes from *E. bromicola*.

Unclassified *Neotyphodium* isolates also displayed close genetic relationships with known taxa. *Neotyphodium* isolate 9727 produced single haplotypes from each locus that were either identical or very similar to the haplotypes common between *N. coenophialum* and *N. uncinatum*, and grouped most closely with those derived from *E. sylvatica*. *Neotyphodium* isolates 9303/2 and 9728 were closely related, all locus-specific haplotypes showing a high degree of sequence similarity. One of three subclasses of derived haplotypes grouped to form a distinct well-supported group with those from *E. festucae*, *N. lolii*, *N. coenophialum*, and FaTG-2. Isolates 9303/2 and 9728 exhibited a second haplotype subclass that grouped with counterparts from *E. bromicola* and *N. uncinatum*, and the same class was present in isolate over two loci. The remaining haplotype sub-class (observed in isolate 9303/2 for four loci, and in isolate 9728 for one locus)

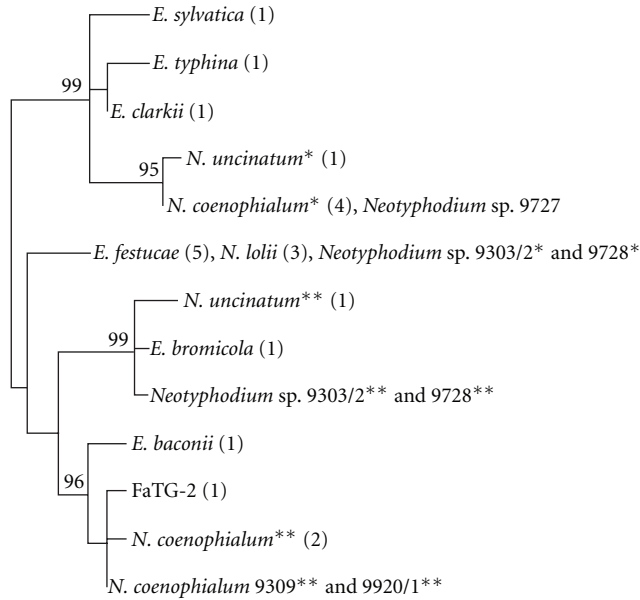


FIGURE 2: Parsimony analysis of sequence haplotypes derived from reference *Neotyphodium* and *Epichloë* isolates for the EST-SSR locus NCESTA1FH03. Diagram properties are as described in the legend to Figure 1. Note that the nucleotide sequence of the amplification product from *N. lolii* isolate North African 6 detected at this locus by autoradiography [38] was not obtained.

grouped with the equivalent haplotypes from *E. typhina* and *E. clarkii*.

## 4. Discussion

**4.1. Application of EST-SSR Loci to Endophyte Phylogenetic Analysis.** The application of SSR markers for phylogenetic analysis is limited by two main factors: complex molecular evolution of SSR loci and the occurrence of size homoplasy between distinct SSR alleles. Sequence analysis of selected SSR loci in the current study has demonstrated that these factors influence the generic inability of SSR markers to resolve phylogenetic relationships among endophyte species [38]. Changes in the SSR array repeat number appeared to be independent of flanking region changes: some of the locus NLESTA1NF04-derived haplotypes from multiple *E. festucae* isolate differed for SSR array number, but exhibited identity for flanking sequence, while others showed the converse relationship. SSR allele size homoplasy occurred between different endophyte taxa of distinct origins as a result of insertions, deletions, and base substitutions in both the SSR motif and flanking sequences, as observed for the locus NCESTA1FH03-specific *E. festucae* and *E. baconii*-related *N. coenophialum* haplotypes. Endophyte SSR locus arrays were highly variable, and differences in repeat unit number generally accounted for allele size variation between closely related endophyte species, while indel and base substitution incidence increased when comparisons were made between more distantly related taxa.

The results of this and other studies [30, 31] suggest that size variation may provide a relatively accurate measure of genetic variation between closely related species. Although homoplasy was not taken into account, SSRs have previously proven useful for genetic discrimination within and between endophyte species [38]. Presumably, the inherently variable nature of SSRs and the large number of loci analysed reduced the potential biasing effects of individual loci. The complex nature of SSR loci, however, demonstrates the critical value of sequence level analysis for phylogenetic inference.

The flanking regions of gene-associated SSRs were highly conserved within and between endophyte taxa (80%–100% sequence identity across coding and non-coding regions), supporting a common origin for these species [1, 9]. Despite this level of sequence conservation, SSR-flanking regions were informative for studying genetic relationships. The different individual loci obtained similar genetic relationships, consistent with previous studies of other genes. Differences in the power of the individual loci to resolve relationships were identified due to variation in number of informative characters and composition (exon, intron, coding, or non-coding) of amplicons. In other studies, SSR-flanking sequences from different loci have been aggregated to increase the number of informative characters and improve resolution of phylogenetic relationships [31, 32, 36, 37]. However, variation of inferred ploidy level and of number of haplotypes derived from different loci in heteroploids would potentially bias such aggregation studies. As a consequence, each locus was analysed separately in this study.

**4.2. Genome Affinities between *Neotyphodium* and *Epichloë* Species.** The close relationships between taxa were in accordance with those predicted from genes more commonly used for phylogenetic analysis such as rDNA, *tubB*, *tefA*, and *actG*. Single locus-specific haplotypes were obtained from all *Epichloë* species and from *N. lolii*, the latter being closely related to *E. festucae*. Other *Neotyphodium* species contained multiple haplotypes that were similar to those from different *Epichloë* species. Occurrence of different haplotype subclasses in *N. coenophialum*, FaTG-2, and *N. uncinatum* is consistent with the heteroploid or nonhaploid nature of these species [11, 14] and indicates the presence of multiple genomes (Figure 6).

**4.3. Relationships with Unclassified *Neotyphodium* Isolates.** *Neotyphodium* isolates that could not be assigned to known morphological classes also appear to differ from characterised taxa at the molecular level [17]. Isolate 9303/2 and isolate 9728 have been assigned to taxonomic groupings HeuTG-2 and LpTG-2, respectively, based on phylogenetic analysis of the *tefA* and *tubB* genes (A. Leuchtmann, pers. comm.; [17]). Analysis of SSR-flanking regions in this study, however, suggests that the isolates show closer affinities than formerly predicted. Moon et al. [17] reported the detection of two haplotype classes for each isolate, closely related to those from *E. bromicola* and *E. typhina* (HeuTG) and from *E. festucae* and *E. typhina* (LpTG-2). These phylogenetic affinities were also detected in the current study. However,

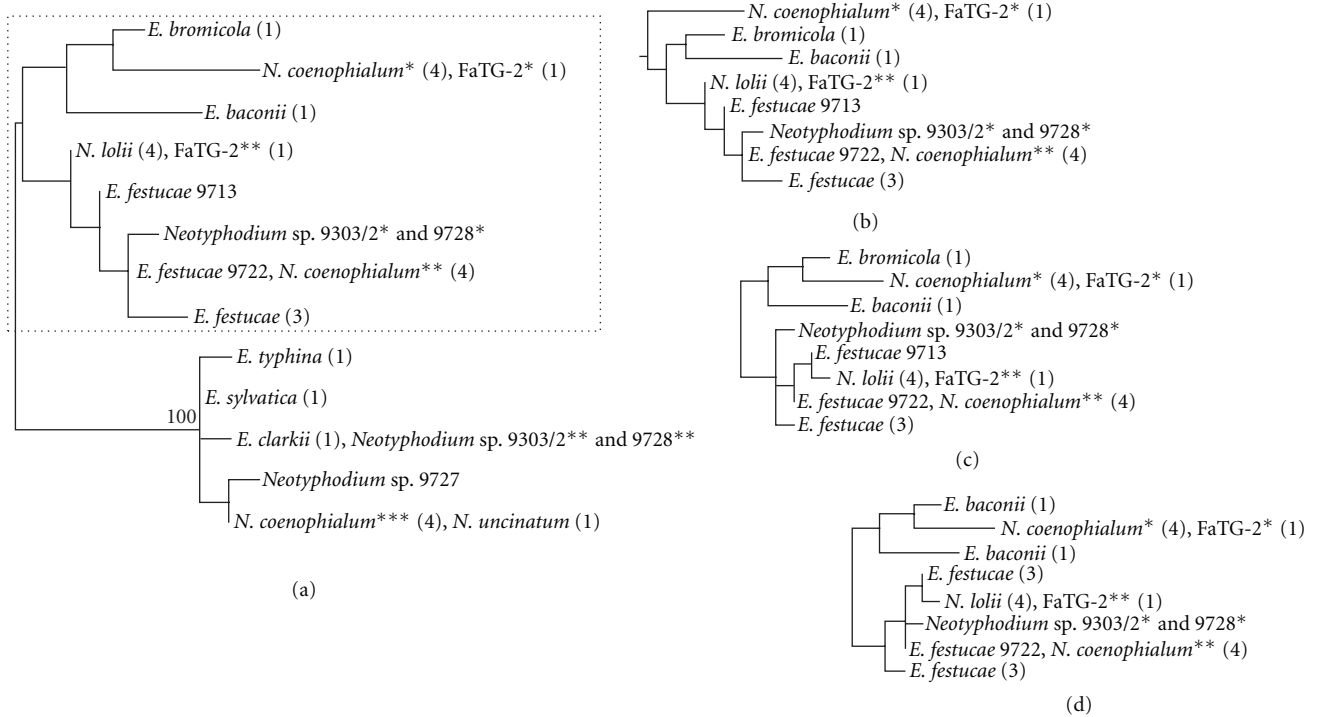


FIGURE 3: Parsimony analysis of sequence haplotypes derived from reference *Neotyphodium* and *Epichloë* isolates for the EST-SSR locus NCESTA1GA07. (a) One of four most parsimonious trees found. Diagram properties are as described in the legend to Figure 1 ((b), (c), and (d)) Subtrees showing the alternative topologies of the *E. festucae*, *E. baconii*, and *E. bromicola* clade in the other trees.

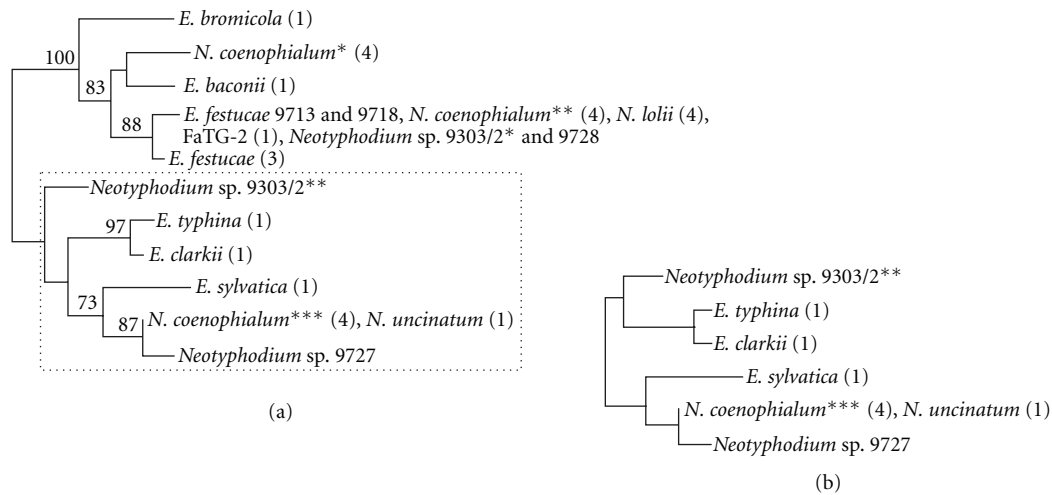


FIGURE 4: Parsimony analysis of sequence haplotypes derived from reference *Neotyphodium* and *Epichloë* isolates for the EST-SSR locus NLESTA1GF09. (a) One of two most parsimonious trees found. Diagram properties are as described in the legend to Figure 1. (b) Subtree showing the alternative topology of the *E. typhina*, *E. clarkii*, and *E. sylvatica* clade in the other tree.

both flanking sequence analysis, as well as phenetic studies based on a larger number of SSR loci [38], detected a third haplotype sub-class for both isolates and suggested common affinities with *E. festucae*, *E. bromicola*, and *E. typhina*, respectively. Accurate inference of phylogenetic relationships among *Neotyphodium* and *Epichloë* species consequently requires characterisation of a number of different genomic loci.

Although isolates 9303/2 and 9728 share common affinities, DNA-based phylogenetic and phenetic analyses suggest mutual genetic divergence and placement in taxonomic groups with different relative gene content. Although similar-sized haplotypes were detected, SSR polymorphism between these isolates was greater than that detected within *N. coenophialum*, *N. lolii*, and *E. festucae* [38]. In addition,

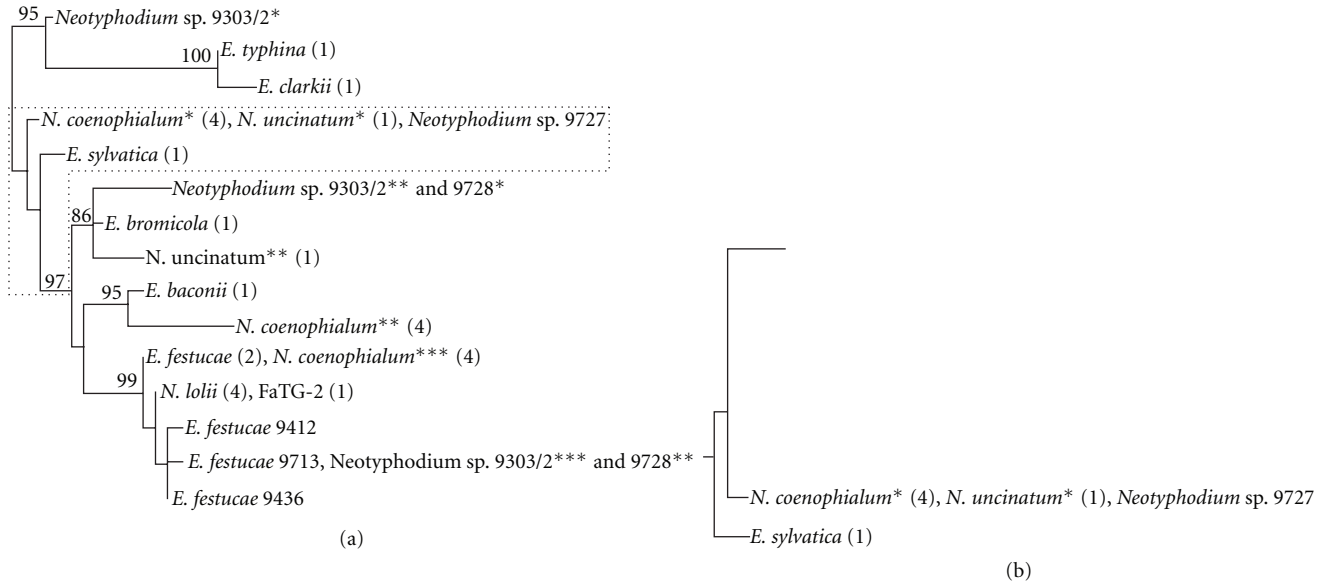


FIGURE 5: Parsimony analysis of sequence haplotypes derived from reference *Neotyphodium* and *Epichloë* isolates for the EST-SSR locus NLESTA1NF04. (a) One of two most parsimonious trees found. Diagram properties are as described in the legend to Figure 1. (b) Sub-tree showing the alternative branching of *E. sylvatica* in the other tree. Note that the nucleotide sequence of the second amplification product from FaTG-2 and the third amplification product from unidentified *Neotyphodium* isolate 9728 detected at this locus by autoradiography [38] was not obtained.

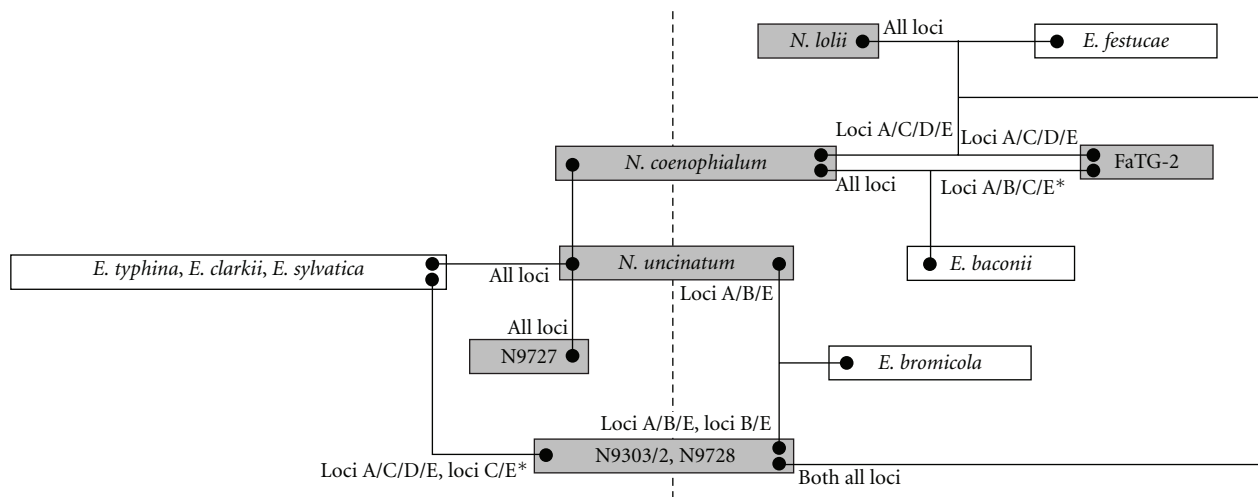


FIGURE 6: Summary of genomic affinities between *Neotyphodium* and *Epichloë* species predicted from phylogenetic analysis of the flanking regions of five SSR loci: NCESTA1AB04 (a), NCESTA1FH03 (b), NCESTA1GA07 (c), NLESTA1GF09 (d), and NLESTA1NF04 (e). Predicted partial or complete genomes, based on variant haplotype sub-classes, are indicated as black circles. Lines that connect putative common genomes, and the level of support for each inference are indicated in terms of the number of loci providing confirmatory data. This information relates to the next most adjacent taxon in the topology of the diagram. An asterisk indicates a locus-specific variant obtained by PCR, but for which the nucleotide sequence was not obtained. The dotted line defines the inferred division between the two major groups of *Epichloë* species.

9303/2 and 9728 failed to cluster together in an AFLP-derived phenogram [38], which represents a genome-wide assessment of genetic polymorphism. SSR polymorphism analysis also detected substantial differences in the number of locus-specific haplotypes: isolate 9728 produced a higher proportion of single haplotype classes. DNA sequence

analysis further demonstrated differences in both number and type of haplotype. *E. festucae*-related haplotypes were detected in both isolates for all five loci, while *E. bromicola*-like and *E. typhina*-like sequence variants were observed more frequently in isolate 9303/2 than isolate 9728 and were not represented between all loci. *Neotyphodium* species



with common phylogenetic affinities are known to occur in several different grass species. The LpTG-2, *N. tembladerae*, and *N. australiense* endophytes, which are resident in *L. perenne*, *Poa huecu*, and *Echinopogon ovatus*, respectively, all appear to be phylogenetically related to *E. festucae* and *E. typhina* [10, 16]. However, these endophytes appear to be related to different *E. typhina* strains and also differ in their genome structure [10, 13, 16]. Differences in transcript levels associated with different gene-specific sequence variants, as observed for the 60S ribosomal protein-encoding gene in this study, may also contribute to phenotypic trait variation between different heteroploid endophyte taxa.

Two asexual endophyte species, *N. huerfanum* and *N. tembladerae*, are known to occur in *Festuca arizonica* [17]. Phylogenetic analysis of the third unclassified *Neotyphodium* isolate (9727), which was also derived from *F. arizonica*, suggests that it may belong to the former taxon. Isolate 9727 produced single haplotypes and these sequences, like those of the *N. huerfanum* *tefA* and *tubB* loci [17], are closely related to the inferred *E. typhina*-related haplotypes from *N. coenophialum* and *N. uncinatum* (Section 4.2). These results were also supported by SSR polymorphism-based phenetic analysis, in which 9727 clustered with *E. typhina*, *E. clarkii*, and *E. sylvatica*, while in AFLP analysis the isolate clustered with *N. uncinatum* [38].

**4.4. Origins of *Neotyphodium* and *Epichloë* Species.** Due to their close phylogenetic relationships with specific *Epichloë* species, *Neotyphodium* species have been proposed to have originated from these sexual endophyte taxa either directly through the loss of the sexual state, or through interspecific hybridisation of distinct *Epichloë* and *Neotyphodium* species. The first process is proposed to have given rise to haploid *Neotyphodium* species such as *N. lolii*, while the heteroploid *Neotyphodium* species such as *N. coenophialum*, FaTG-2, and *N. uncinatum* may have arisen through the second evolutionary process. Because *Epichloë* species form unique mating populations [49–51] and *Neotyphodium* species are not known to sporulate *in vivo* [52], this second mode of evolution is thought to have been a parasexual process involving somatic fusion of endophyte hyphae. This hypothesis does, however, require physical collocation between endophyte taxa that generally occur in distinct host species. In addition, mechanisms of gene loss following nuclear fusion are necessary to account for the observed genomic composition of contemporary heteroploid taxa, as a range of studies [2, 10, 11, 13, 14, 16, 17] have shown that extant *Neotyphodium* species do not appear to have the full complement of genes present in phylogenetically related *Epichloë* species. Loss of genes involved in sexual reproduction and pathogenicity would be a prerequisite for such genomic rearrangement events, as well as genes vulnerable to dosage-dependent effects. It is also formally possible that sexual *Epichloë* species may have arisen from asexual *Neotyphodium* species in response to selective environmental pressures, a mechanism requiring both gene loss and gene gain, possibly through horizontal gene transfer. Mechanisms for both processes have been inferred through comparisons of different fungal taxa at

the whole genome levels [53] and may have been facilitated by structural features such as presence of conserved repetitive elements.

In conclusion, this study demonstrates the application of SSR-flanking sequences to studies of genome affinities between pasture grass fungal endophyte species for clarification of novel modes of genome evolution. The inferred affinities were consistent with those obtained from gene loci that are more commonly used in molecular phylogenetics, but provided a more extensive survey of genomic loci, that may be ultimately extended to whole genome comparisons based on second-generation sequencing technologies.

## Acknowledgments

Endophytes isolates were kindly provided by Dr. Adrian Leuchtman (ETH Zürich, Switzerland) and Nola McFarlane (DPI-Hamilton). This work was supported by the Victorian Department of Primary Industries and the Molecular Plant Breeding Cooperative Research Centre. All experiments conducted during this study comply with current Australian laws.

## References

- [1] G. A. Kuldau, J. S. Liu, J. F. White, M. R. Siegel, and C. L. Schardl, "Molecular systematics of Clavicipitaceae supporting monophyly of genus *Epichloë* and form genus *Ephelis*," *Mycologia*, vol. 89, no. 3, pp. 431–441, 1997.
- [2] K. Clay and C. Schardl, "Evolutionary origins and ecological consequences of endophyte symbiosis with grasses," *American Naturalist*, vol. 160, supplement S, pp. S99–S127, 2002.
- [3] R. T. Gallagher, E. P. White, and P. H. Mortimer, "Ryegrass staggers: isolation of potent neurotoxins lolitrem A and lolitrem B from staggers-producing pastures," *New Zealand Veterinary Journal*, vol. 29, no. 10, pp. 189–190, 1981.
- [4] S. G. Yates, R. D. Plattner, and G. B. Garner, "Detection of ergopeptine alkaloids in endophyte infected, toxic Ky-31 tall fescue by mass spectrometry/mass spectrometry," *Journal of Agricultural and Food Chemistry*, vol. 33, no. 4, pp. 719–722, 1985.
- [5] D. D. Rowan and D. L. Gaynor, "Isolation of feeding deterrents against argentine stem weevil from ryegrass infected with the endophyte *Acremonium loliae*," *Journal of Chemical Ecology*, vol. 12, no. 3, pp. 647–658, 1986.
- [6] M. Arachevaleta, C. W. Bacon, C. S. Hoveland et al., "Effect of the tall fescue endophyte on plant response to environmental stress," *Agronomy Journal*, vol. 81, no. 1, pp. 83–90, 1989.
- [7] C. Ravel, C. Courty, A. Coudret, and G. Charmet, "Beneficial effects of *Neotyphodium lolii* on the growth and the water status in perennial ryegrass cultivated under nitrogen deficiency or drought stress," *Agronomie*, vol. 17, no. 3, pp. 173–181, 1997.
- [8] G. Morgan-Jones and W. Gams, "Notes on Hyphomycetes. XLI. An endophyte of *Festuca arundinacea* and the anamorph of *Epichloë typhina*, new taxa in one of two new sections of *Acremonium*," *Mycotaxon*, vol. 15, no. 1, pp. 311–318, 1982.
- [9] A. E. Glenn, C. W. Bacon, R. Price, and R. T. Hanlin, "Molecular phylogeny of *Acremonium* and its taxonomic implications," *Mycologia*, vol. 88, no. 3, pp. 369–383, 1996.

- [10] C. L. Schardl, A. Leuchtmann, H. F. Tsai, M. A. Collett, D. M. Watt, and D. B. Scott, "Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*," *Genetics*, vol. 136, no. 4, pp. 1307–1317, 1994.
- [11] H. F. Tsai, J. S. Liu, C. Staben et al., "Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloë* species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 7, pp. 2542–2546, 1994.
- [12] C. L. Schardl, A. Leuchtmann, K. R. Chung, D. Penny, and M. R. Siegel, "Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts," *Molecular Biology and Evolution*, vol. 14, no. 2, pp. 133–143, 1997.
- [13] C. D. Moon, B. Scott, C. L. Schardl, and M. J. Christensen, "The evolutionary origins of *Epichloë* endophytes from annual ryegrasses," *Mycologia*, vol. 92, no. 1–6, pp. 1103–1118, 2000.
- [14] K. D. Craven, J. D. Blankenship, A. Leuchtmann, K. Hignight, and C. L. Schardl, "Hybrid fungal endophytes symbiotic with the grass *Lolium pratense*," *Sydowia*, vol. 53, no. 1, pp. 44–73, 2001.
- [15] K. D. Craven, P. T. W. Hsiau, A. Leuchtmann, W. Hollin, and C. L. Schardl, "Multigene phylogeny of *Epichloë* species, fungal symbionts of grasses," *Annals of the Missouri Botanical Garden*, vol. 88, no. 1, pp. 14–34, 2001.
- [16] C. D. Moon, C. O. Miles, U. Järlfors, and C. L. Schardl, "The evolutionary origins of three new *Neotyphodium* endophyte species from grasses indigenous to the Southern Hemisphere," *Mycologia*, vol. 94, no. 4, pp. 694–711, 2002.
- [17] C. D. Moon, K. D. Craven, A. Leuchtmann, S. L. Clement, and C. L. Schardl, "Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses," *Molecular Ecology*, vol. 13, no. 6, pp. 1455–1467, 2004.
- [18] G. A. Kuldau, H. F. Tsai, and C. L. Schardl, "Genome sizes of *Epichloë* species and anamorphic hybrids," *Mycologia*, vol. 91, no. 5, pp. 776–782, 1999.
- [19] C. L. Schardl, J. S. Liu, J. F. White, R. A. Finkel, Z. An, and M. R. Siegel, "Molecular phylogenetic relationships of nonpathogenic grass mycosymbionts and clavicipitaceous plant pathogens," *Plant Systematics and Evolution*, vol. 178, no. 1–2, pp. 27–41, 1991.
- [20] D. Tautz, "Hypervariability of simple sequences as a general source for polymorphic DNA markers," *Nucleic Acids Research*, vol. 17, no. 16, pp. 6463–6471, 1989.
- [21] J. L. Weber and P. E. May, "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction," *American Journal of Human Genetics*, vol. 44, no. 3, pp. 388–396, 1989.
- [22] M. D. Schug, C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. C. Mackay, and C. F. Aquadro, "The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*," *Molecular Biology and Evolution*, vol. 15, no. 12, pp. 1751–1760, 1998.
- [23] G. Levinson and G. A. Gutman, "Slipped-strand mispairing: a major mechanism for DNA sequence evolution," *Molecular Biology and Evolution*, vol. 4, no. 3, pp. 203–221, 1987.
- [24] O. Rose and D. Falush, "A threshold size for microsatellite expansion," *Molecular biology and evolution*, vol. 15, no. 5, pp. 613–615, 1998.
- [25] J. C. Garza, M. Slatkin, and N. B. Freimer, "Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size," *Molecular Biology and Evolution*, vol. 12, no. 4, pp. 594–603, 1995.
- [26] D. B. Goldstein and A. G. Clark, "Microsatellite variation in North American populations of *Drosophila melanogaster*," *Nucleic Acids Research*, vol. 23, no. 19, pp. 3882–3886, 1995.
- [27] M. C. Grimaldi and B. Crouau-Roy, "Microsatellite allelic homoplasy due to variable flanking sequences," *Journal of Molecular Evolution*, vol. 44, no. 3, pp. 336–340, 1997.
- [28] I. Colson and D. B. Goldstein, "Evidence for complex mutations at microsatellite loci in *Drosophila*," *Genetics*, vol. 152, no. 2, pp. 617–627, 1999.
- [29] I. Clisson, M. Lathuilliere, and B. Crouau-Roy, "Conservation and evolution of microsatellite loci in primate taxa," *American Journal of Primatology*, vol. 50, no. 3, pp. 205–214, 2000.
- [30] X. Chen, Y. G. Cho, and S. R. McCouch, "Sequence divergence of rice microsatellites in *Oryza* and other plant species," *Molecular Genetics and Genomics*, vol. 268, no. 3, pp. 331–343, 2002.
- [31] S. R. Santos, T. L. Shearer, A. R. Hannes, and M. A. Coffroth, "Fine-scale diversity and specificity in the most prevalent lineage of symbiotic dinoflagellates (*Symbiodinium*, Dinophyceae) of the Caribbean," *Molecular Ecology*, vol. 13, no. 2, pp. 459–469, 2004.
- [32] T. Asahida, A. K. Gray, and A. J. Gharrett, "Use of microsatellite locus flanking regions for phylogenetic analysis? A preliminary study of *Sebastes* subgenera," *Environmental Biology of Fishes*, vol. 69, no. 1–4, pp. 461–470, 2004.
- [33] D. Field and C. Wills, "Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1647–1652, 1998.
- [34] M. S. Röder, V. Korzun, K. Wendehake et al., "A microsatellite map of wheat," *Genetics*, vol. 149, no. 4, pp. 2007–2023, 1998.
- [35] M. Morgante, M. Hanafey, and W. Powell, "Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes," *Nature Genetics*, vol. 30, no. 2, pp. 194–200, 2002.
- [36] M. C. Fisher, G. Koenig, T. J. White, and J. W. Taylor, "A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis*," *Molecular Biology and Evolution*, vol. 17, no. 8, pp. 1164–1174, 2000.
- [37] M. Rossetto, J. McNally, and R. J. Henry, "Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae," *Theoretical and Applied Genetics*, vol. 104, no. 1, pp. 61–66, 2002.
- [38] E. Van Zijll De Jong, K. M. Guthridge, G. C. Spangenberg, and J. W. Forster, "Development and characterization of EST-derived simple sequence repeat (SSR) markers for pasture grass endophytes," *Genome*, vol. 46, no. 2, pp. 277–290, 2003.
- [39] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [40] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, "The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools," *Nucleic Acids Research*, vol. 25, no. 24, pp. 4876–4882, 1997.
- [41] A. R. Templeton, "Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes," *Evolution*, vol. 37, no. 2, pp. 221–244, 1983.
- [42] H. Kishino and M. Hasegawa, "Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from

- DNA sequence data, and the branching order in hominoide,” *Journal of Molecular Evolution*, vol. 29, no. 2, pp. 170–179, 1989.
- [43] H. Shimodaira and M. Hasegawa, “Multiple comparisons of log-likelihoods with applications to phylogenetic inference,” *Molecular Biology and Evolution*, vol. 16, no. 8, pp. 1114–1116, 1999.
- [44] J. Felsenstein, “Confidence limits on phylogenies: an approach using the bootstrap,” *Evolution*, vol. 39, no. 4, pp. 783–791, 1985.
- [45] J. Felsenstein and G. A. Churchill, “A Hidden Markov Model approach to variation among sites in rate of evolution,” *Molecular Biology and Evolution*, vol. 13, no. 1, pp. 93–104, 1996.
- [46] W. M. Fitch and E. Margoliash, “Construction of phylogenetic trees,” *Science*, vol. 155, no. 760, pp. 279–284, 1967.
- [47] N. Saitou and M. Nei, “The neighbor-joining method: a new method for reconstructing phylogenetic trees,” *Molecular biology and evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [48] H. A. Schmidt, K. Strimmer, M. Vingron, and A. Von Haeseler, “TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing,” *Bioinformatics*, vol. 18, no. 3, pp. 502–504, 2002.
- [49] A. Leuchtman, C. L. Schardl, and M. R. Siegel, “Sexual compatibility and taxonomy of a new species of *Epichloë* symbiotic with fine fescue grasses,” *Mycologia*, vol. 86, no. 6, pp. 802–812, 1994.
- [50] A. Leuchtman and C. L. Schardl, “Mating compatibility and phylogenetic relationships among two new species of *Epichloë* and other congeneric European species,” *Mycological Research*, vol. 102, no. 10, pp. 1169–1182, 1998.
- [51] C. L. Schardl and A. Leuchtman, “Three new species of *Epichloë* symbiotic with North American grasses,” *Mycologia*, vol. 91, no. 1, pp. 95–107, 1999.
- [52] J. F. White, “Endophyte-host associations in forage grasses. XI. A proposal concerning origin and evolution,” *Mycologia*, vol. 80, no. 4, pp. 442–446, 1988.
- [53] E. L. Braun, A. L. Halpern, M. A. Nelson, and D. O. Natvig, “Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*,” *Genome Research*, vol. 10, no. 4, pp. 416–430, 2000.