

# SCIENTIFIC REPORTS



OPEN

## Horizontal transfer of a $\beta$ -1,6-glucanase gene from an ancestral species of fungal endophyte to a cool-season grass host

Hiroshi Shinozuka<sup>1,2</sup>, Inoka K. Hettiarachchige<sup>1,2</sup>, Maiko Shinozuka<sup>1,2</sup>, Noel O. I. Cogan<sup>1,2,3</sup>, German C. Spangenberg<sup>1,2,3</sup>, Benjamin G. Cocks<sup>1,2,3</sup>, John W. Forster<sup>1,2,3</sup> & Timothy I. Sawbridge<sup>1,2,3</sup>

Molecular characterisation has convincingly demonstrated some types of horizontal gene transfer in eukaryotes, but nuclear gene transfer between distantly related eukaryotic groups appears to have been rare. For angiosperms (flowering plants), nuclear gene transfer events identified to date have been confined to genes originating from prokaryotes or other plant species. In this report, evidence for ancient horizontal transfer of a fungal nuclear gene, encoding a  $\beta$ -1,6-glucanase enzyme for fungal cell wall degradation, into an angiosperm lineage is presented for the first time. The gene was identified from *de novo* sequencing and assembly of the genome and transcriptome of perennial ryegrass, a cool-season grass species. Molecular analysis confirmed the presence of the complete gene in the genome of perennial ryegrass. No corresponding sequence was found in other plant species, apart from members of the Poeae sub-tribes Loliinae and Dactylidinae. Evidence suggests that a common ancestor of the two sub-tribes acquired the gene from a species ancestral to contemporary grass-associated fungal endophytes around 9–13 million years ago. This first report of horizontal transfer of a nuclear gene from a taxonomically distant eukaryote to modern flowering plants provides evidence for a novel adaptation mechanism in angiosperms.

The Poeae tribe of the Poaceae family is composed of a range of cool-season turf and forage grass species, including those of sub-tribes Loliinae and Dactylidinae<sup>1</sup>. Perennial ryegrass (*Lolium perenne* L.; sub-tribe Loliinae) is one of the most important pasture crop species for the dairy industry, and it has consequently been a primary target for improvement using molecular biology and genetic technologies<sup>2</sup>. Asexual fungal endophyte species of the genus *Epichloë* (syn. *Neotyphodium*) are symbionts of species belonging to the Poeae tribe genera *Lolium* and *Festuca*, and understanding of the symbiotic relationship is essential for improvement of the productivity of both forage and turf crops<sup>3</sup>. The fungal endophyte species rely on the plant host for nutrition, reproduction, and protection from abiotic and biotic stress. Benefits to the host plant include enhanced competitive abilities, tolerance to pathogens, and resistance to animal and insect herbivory<sup>4</sup>. Due to its agronomic importance, the molecular basis of the symbiosis has been the subject of detailed investigation, and deterrence of insect herbivory is largely due to the production of bioactive alkaloids, as well as *makes caterpillars floppy-like* (*mcfl*-like) gene products<sup>3,5,6</sup>.

Horizontal gene transfer (HGT) has been an important source of evolutionary novelty in both prokaryotes and eukaryotes<sup>7–9</sup>. In flowering plant species, organelle genomes have served as both donors and recipients of gene transfer events<sup>10</sup>. The group I intron in the plant mitochondrial DNA gene, *cox1*, was acquired from a fungal species through an ancient transfer event<sup>11</sup>. The intron sequence of the mitochondrial genome has been found in a wide range of angiosperms from over 200 taxonomic families, and it is therefore, likely that the sequence was spread between taxa by plant-to-plant horizontal transfer, following the primary cross-kingdom transfer event<sup>12,13</sup>. In *Amborella* (*Amborella trichopoda* Baill.), evidence for extensive ancestral horizontal transfer of mitochondrial DNA genes have been identified<sup>14</sup>. A phylogenetic approach revealed that the *Amborella* mitochondrial genome

<sup>1</sup>Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University, Bundoora, Victoria, 3083, Australia. <sup>2</sup>Dairy Futures Cooperative Research Centre, Bundoora, Australia. <sup>3</sup>School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3086, Australia. Correspondence and requests for materials should be addressed to H.S. (email: [hiroshi.shinozuka@ecodev.vic.gov.au](mailto:hiroshi.shinozuka@ecodev.vic.gov.au))



similarity at a DNA sequence identity of 82% to a *Zea mays* transposon-related gene (NCBI UI: AF434192.1) (Fig. 2a). A 11 kb-contig of *E. festucae* genome sequence (NCBI UI: EF015481), which includes the corresponding  $\beta$ -1,6-glucanase gene, was obtained from the NCBI database and was shown to contain three other genes located within a 5 kb distance from the  $\beta$ -1,6-glucanase gene. In the BAC clone-derived contigs, however, no sequences similar to these flanking gene were identified. Putative coding regions for the *E. festucae*  $\beta$ -1,6-glucanase and *LpBGNL* genes were identified (Fig. 2b). A single intron was found in both sequences, and comparison of the exonic and intronic regions identified 4 insertion-deletion (indel) variations between them. Although the position of the intron was conserved, it seemed that almost all intron sequence was replaced in perennial ryegrass, due to insertion and deletion events. Although the coding regions were relatively highly conserved, no sequence similarity was found in the flanking sequence of the coding regions. A BLAST search of 1.5-kb upstream and downstream sequences of *LpBGNL* identified partial sequence similarity to the genomes of wheat and rice, while the corresponding upstream and downstream sequences of the *E. festucae*  $\beta$ -1,6-glucanase gene included partial sequences of the flanking genes (Fig. 2b).

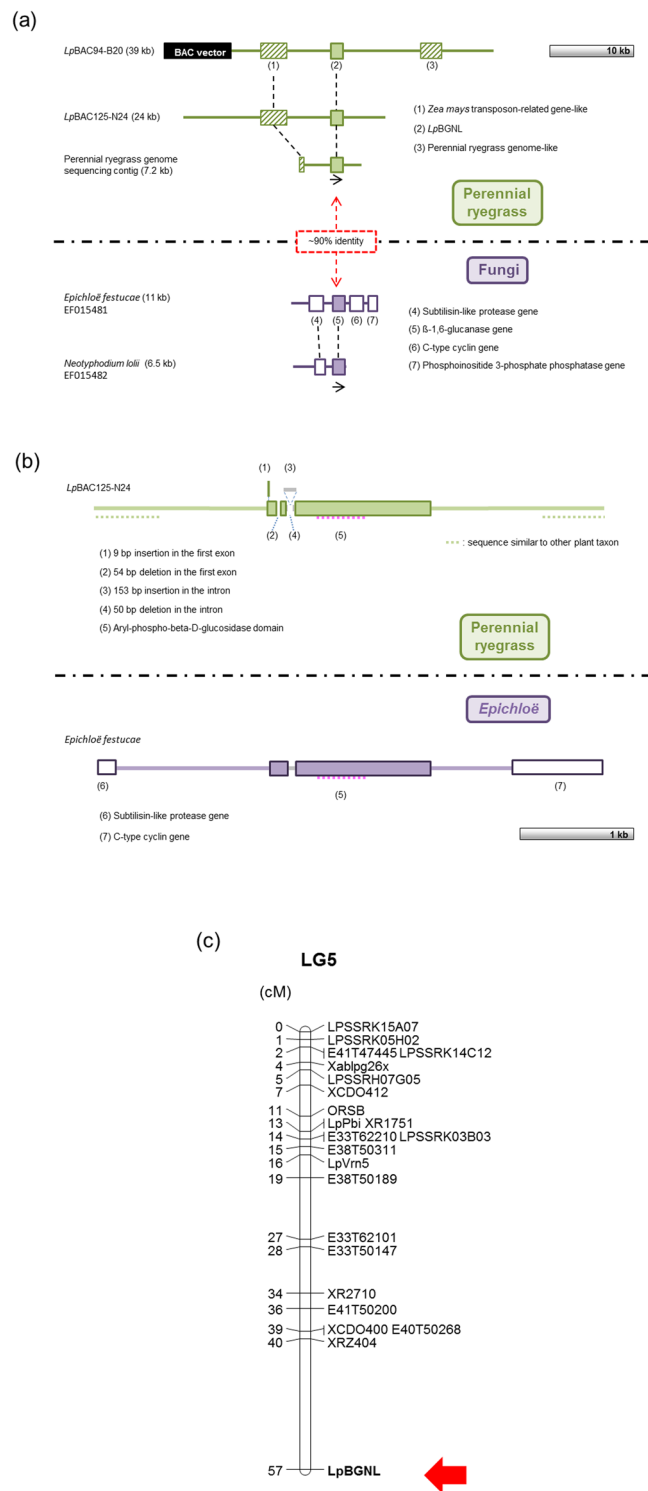
Sequencing of PCR products generated using *LpBGNL*-specific primers identified a 51-bp intron-located polymorphism between haplotypes of the heterozygous parent (C3 genotype) of the perennial ryegrass p150/112 genetic linkage mapping population (Supplementary Fig. S1), which facilitated development of an indel-based DNA marker. From the p150/112 population, 48 individuals were genotyped (Supplementary Fig. S2), and the *LpBGNL*-related marker locus was assigned to a distal region of perennial ryegrass linkage group (LG) 5 (Fig. 2c).

***LpBGNL* gene expression analysis.** Expression of *LpBGNL* was determined using data from the transcriptome sequence of the Impact<sub>04</sub> genotype (NCBI BioProject Accession: PRJNA379202). Sequencing reads corresponding to *LpBGNL* were identified from leaf, root and flower samples, and higher expression levels [based on counts per million reads (CPM)] were detected in root and flower than in leaf (Fig. 3a and Supplementary Table S2). Specificity of gene expression was examined using endophyte-infected (E<sup>+</sup>) and E<sup>-</sup> perennial ryegrass seeds and seedlings. Due to sequence divergence, sequencing reads corresponding to the plant and fungal gene could be reliably discriminated (Supplementary Table S3). Read counts were very low for both E<sup>+</sup> and E<sup>-</sup> seeds immediately after the germination treatment (Fig. 3b). Although no large-scale morphological change was observed during the following two days, the read counts substantially increased in both samples. The counts remained at relatively high levels in young seedlings at 5 and 10 days after treatment. As similar trends were observed for both E<sup>+</sup> and E<sup>-</sup> genotypes, presence of endophyte did not significantly affect *LpBGNL* expression pattern. Expression of the endogenous *E. festucae* var. *lolii*  $\beta$ -1,6-glucanase gene was observed in E<sup>+</sup> seedlings, but the read count approach revealed relatively low levels throughout the 10 days, in contrast to *LpBGNL* (Fig. 3c and Supplementary Table S3).

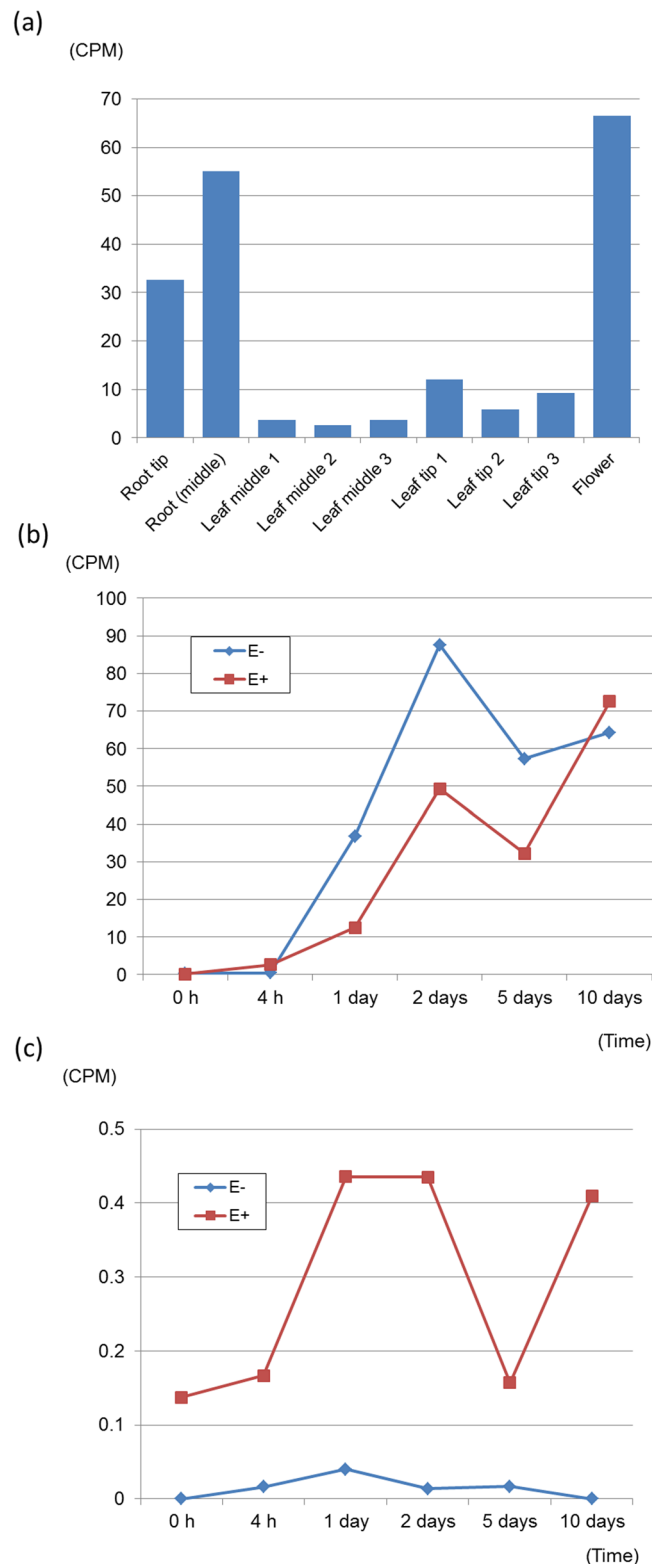
**Phylogenetic analysis of plant and fungal  $\beta$ -1,6-glucanase(-like) genes.** The presence of *LpBGNL* orthologues in other Poaceae species was determined by PCR-based screening. *LpBGNL*-specific primers were designed and short DNA fragments (178 bp in length) were amplified from genomic DNA templates of darnel (*Lolium temulentum* L.), meadow fescue (*Festuca pratensis* Huds.), tall fescue (*Festuca arundinacea* Schreb.), sheep fescue (*Festuca ovina*) (Table 1 and Supplementary Fig. S3). Products were also obtained from those of cocksfoot/orchard grass (*Dactylis glomerata* L.) and *Dactylis marina* Borrill, but not from those of coast tussock-grass [*Poa poiformis* (Labill.) Druce] or harding grass/phalaris (*Phalaris aquatica* L.). DNA fragments were not amplified from genomic DNA templates of *E. festucae* var. *lolii*, confirming the specificity of the oligonucleotide primers that were used. Alignment of *LpBGNL* and the fungal genes identified a conserved region ca. 750 bp in length, corresponding to the aryl-phospho-beta-D-glucosidase domain. For a phylogenetic analysis, DNA fragments, including the aryl-phospho-beta-D-glucosidase domain, were amplified from the selected Loliinae and Dactylidinae species. *De novo* amplicon sequence analysis and assembly obtained a single sequence contig of the 750 bp region for each *Lolium* and *Festuca* species (Supplementary Fig. S4). For each *Dactylis* species, three contigs (haplotypes) were generated. A putative premature stop codon was found in two haplotypes of *Dactylis marina*, and the haplotypes with the premature stop codon were excluded from the further analysis. Fungal  $\beta$ -1,6-glucanase gene-like sequences were obtained from the NCBI database and the Genome Project at the University of Kentucky website. Phylogenetic analysis with the maximum likelihood method was performed, and the plant species-derived sequences were found to be clustered in the phylogram with those from *Epichloë* (*Neotyphodium*) species (Fig. 4 and Supplementary Note S1). The sequences from other fungi were more distantly related to the plant sequences.

Synonymous and non-synonymous nucleotide substitution ( $K_s/K_a$ ) ratios were calculated using the 750 bp sequences. Both  $K_s$  and  $K_a$  values between the *Epichloë* and plant species were substantially lower than those between the *Epichloë* species and other fungi (*H. lixii* and *T. harzianum*) (Table 2). The  $K_s/K_a$  ratios for candidate *LpBGNL* orthologues were between 0.027–0.221, lower or equivalent to values from the fungal  $\beta$ -1,6-glucanase genes (0.057–0.275).

In order to verify presence/absence boundaries, a set of PCR primers was designed to amplify 415 bp fragments within a region highly conserved between plant and fungal  $\beta$ -1,6-glucanase(-like) genes. This assay confirmed the absence of the gene in the *Poa* and *Phalaris* species samples (Fig. 5). As a control experiment for capacity to amplify cross-species, PCR was performed with primers specific to perennial ryegrass histone H3 and candidate plant architecture genes, such as the ATP-binding cassette protein sub-family G 5 and 6 genes (*LpABCG5* and *LpABCG6*, respectively). Although all primers were designed based on perennial ryegrass sequence, PCR amplification from coast tussock-grass and harding grass/phalaris was observed, except for the combination of ABCG 5 primers and one of the *Poa* genotypes (Supplementary Fig. S5). Database searches were performed using published short-read sequencing data on the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>). Sequences significantly matching the *LpBGNL* and *E. festucae*  $\beta$ -1,6-glucanase genes were



**Figure 2.** Genome structure of the *LpBGNL* and *E. festucae*  $\beta$ -1,6-glucanase genes and genetic linkage analysis for *LpBGNL*. **(a)** Alignment of genome sequences from perennial ryegrass and *Epichloë* species. The light green-filled and striped boxes show the location of *LpBGNL* and the plant genome-related sequence, respectively. The black-filled box represents the BAC vector. The purple-filled and empty boxes show the location of the  $\beta$ -1,6-glucanase gene and flanking gene, respectively. The transcription direction of the genes is indicated with the arrow. Corresponding gene sequences are connected with black dashed lines. **(b)** Alignment of coding regions of the *LpBGNL* and *Epichloë*  $\beta$ -1,6-glucanase genes. The light green and purple lines represent non-gene coding region of perennial ryegrass and *Epichloë* species. The grey line and pink breaking line show the location of the intron and aryl-phospho-beta-D-glucosidase domain, respectively. **(c)** Genetic linkage map of perennial ryegrass LG5 with the *LpBGNL*-related locus. The *LpBGNL*-related marker locus is indicated with the red arrow. Genetic distance (cM) is shown on the right side of the genetic markers.



**Figure 3.** The expression levels of *LpBGNL* and the *E. festucae* var. *lolii* β-1,6-glucanase gene. (a) The expression level of *LpBGNL* in each tissue of perennial ryegrass. The y-axis shows normalised read count number (CPM). (b and c) The expression levels of the genes in E<sup>-</sup> and E<sup>+</sup> perennial ryegrass genotypes. The x-axis shows time since the germination treatment, and y-axis shows CPM for *LpBGNL* (b) and the *E. festucae* var. *lolii* β-1,6-glucanase gene (c).

Species		Genotype or cultivar	UI	Reference
Common name	Scientific name			
Perennial ryegrass	<i>Lolium perenne</i> L.	Impact <sub>04</sub>		Shinozuka <i>et al.</i> , 2017
Darnel	<i>Lolium temulentum</i> L.	Genotype from Aberystwyth (Great Britain)	IBERS: BA13157	Hand <i>et al.</i> <sup>37</sup>
Meadow fescue	<i>Festuca pratensis</i> Huds.	Genotype from Tadharn Moor (Great Britain)	IBERS: BF1199	Hand <i>et al.</i> <sup>37</sup>
Tall fescue	<i>Festuca arundinacea</i> Schreb.	Demeter		
Tall fescue	<i>Festuca arundinacea</i> Schreb.	Quantum		Hand <i>et al.</i> <sup>37</sup>
Sheep fescue	<i>Festuca ovina</i>	Genotype from Ponterwyd (Great Britain)	IBERS: BL2643	Hand <i>et al.</i> <sup>37</sup>
Cocksfoot (Orchard grass)	<i>Dactylis glomerata</i> L.	Currie	SARDI: 778	
( <i>Dactylis marina</i> )	<i>Dactylis marina</i> Borrill	Wild genotype from Algeria	SARDI: 38013	
Coast tussock-grass	<i>Poa poiformis</i> (Labill.) Druce	Wild genotype from Australia	SARDI: 41525	
Harding grass (Phalaris)	<i>Phalaris aquatica</i> L.	Landmaster		

**Table 1.** Plant materials used in the current study. UI denotes the unique identifier of the Genetic Resources Unit of Institute for Biological, Environmental and Rural Studies (IBERS; Aberystwyth, Wales. UK) or the South Australian Research and Development Institute (SARDI).

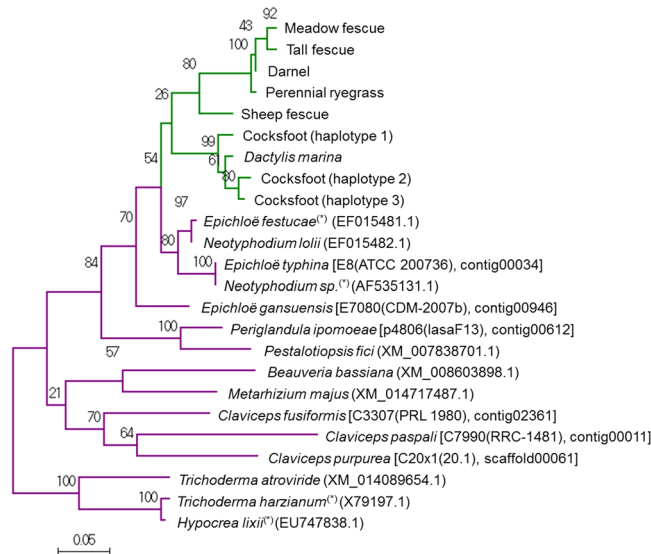
Ks, Ka (Ka/Ks)	Candidate orthologue of <i>LpBGNL</i>				$\beta$ -1,6-glucanase gene			
	Darnel ryegrass	Tall fescue	Meadow Fescue	Sheep fescue	<i>E. festucae</i>	<i>Neotyphodium sp.</i>	<i>H. lixii</i>	<i>T. harzianum</i>
Perennial ryegrass	0.0622, 0.0017 (0.0273)	0.1302, 0.0102 (0.0783)	0.0991, 0.0102 (0.1029)	0.3333, 0.0416 (0.1247)	0.1961, 0.0501 (0.2553)	0.2539, 0.0629 (0.2477)	0.6116, 0.1507 (0.2464)	0.6396, 0.1469 (0.2297)
Darnel ryegrass	—	0.1670, 0.0085 (0.0509)	0.1359, 0.0085 (0.0625)	0.3575, 0.0433 (0.1211)	0.1957, 0.0518 (0.2647)	0.2533, 0.0646 (0.2550)	0.6041, 0.1491 (0.2468)	0.6383, 0.1453 (0.2276)
Tall fescue	—	—	0.0308, 0.0068 (0.2213)	0.3313, 0.0467 (0.1411)	0.1887, 0.0535 (0.2838)	0.2369, 0.0655 (0.2766)	0.6285, 0.1470 (0.2339)	0.6563, 0.1432 (0.2182)
Meadow Fescue	—	—	—	0.3247, 0.045 (0.1387)	0.1823, 0.0519 (0.2845)	0.2305, 0.0638 (0.2770)	0.6093, 0.1487 (0.2441)	0.6494, 0.1449 (0.2231)
Sheep Fescue	—	—	—	—	0.2207, 0.0433 (0.1961)	0.2969, 0.051 (0.1717)	0.6598, 0.1442 (0.2186)	0.6805, 0.1407 (0.2068)
<i>Epichloe festucae</i>	—	—	—	—	—	0.0896, 0.0247 (0.2751)	0.6363, 0.1488 (0.2339)	0.6424, 0.1492 (0.2323)
<i>Neotyphodium sp.</i>	—	—	—	—	—	—	0.6546, 0.1533 (0.2341)	0.6670, 0.1537 (0.2304)
<i>Hypocrea lixii</i>	—	—	—	—	—	—	—	0.1186, 0.0068 (0.0572)

**Table 2.** Ks and Ka values, and Ka/Ks ratio within the Aryl-phospho-beta-D-glucosidase domain region between  $\beta$ -1,6-glucanase(-like) gene sequences.

found from *Dactylis* species, but not from *Poa* or *Phalaris* species (Table 3, Supplementary Table S4). No significantly matching sequence was obtained from Antarctic hair grass (*Deschampsia antarctica* É. Desv.), which is believed to be taxonomically closer than *Poa* species to members of the sub-tribes Loliinae and Dactylidinae<sup>1</sup> (Fig. 5b). As a control analysis, sequences similar to the *LpABCG* 5 and 6 genes were sought, leading to identification of matching sequences from all tested Poae species, including those belonging to both *Poa* and *Phalaris*.

## Discussion

The high level of DNA sequence similarity to the fungal genes, and absence of *LpBGNL*-like sequences from other representative angiosperm species suggested that *LpBGNL* was obtained from a fungal species through HGT. However, it was also possible that the *LpBGNL* sequence may have originated from a microbe associated with the plant individual, especially from a species of asexual fungal endophyte, such as *E. festucae* var. *lolii* (syn. *Neotyphodium lolii*), which are known to form symbiotic associations with perennial ryegrass<sup>3</sup>. Although an E<sup>-</sup> plant individual was selected for sequencing on the basis of negative PCR-based screening for diagnostic DNA sequences, a fungal symbiont could be present in plant material below the limit of detection. Genomic and genetic characterisation was consequently performed in order to demonstrate that *LpBGNL* is located on a perennial ryegrass chromosome. Furthermore, *LpBGNL* orthologues were found to be present in other Loliinae and Dactylidinae species. It is hence unlikely that *LpBGNL* was an assembly or annotation artefact, even though *LpBGNL* shows unusually high DNA sequence similarity (ca. 90%) to the  $\beta$ -1,6-glucanase genes



**Figure 4.** Phylogenetic tree of *LpBGNL* orthologues and fungal  $\beta$ -1,6-glucanase genes. The phylogram was generated based on amino acid sequence of the aryl-phospho-beta-D-glucosidase domain, and sequences from angiosperm and fungi species are indicated with green and purple, respectively. Asterisk (\*) denotes species from which gene products have been confirmed to have the  $\beta$ -1,6-glucanase activity. For cocksfoot, three contigs (haplotypes 1–3) were generated, and amino acid sequences from those contigs were used. For *Dactylis marina*, the contig without a putative premature stop codon was used. Strain and sequence contig (scaffold) identifiers of the Genome Project at the University of Kentucky website are shown in brackets. For the other fungi sequences, NCBI UI is shown in bracket. The clade including the sequences from *Trichoderma* and *Hypocrea* species were selected as an outer group to obtain a root of the phylogenetic tree. A figure legend generated with the MEGA7 software can be found in Supplementary Note S1.

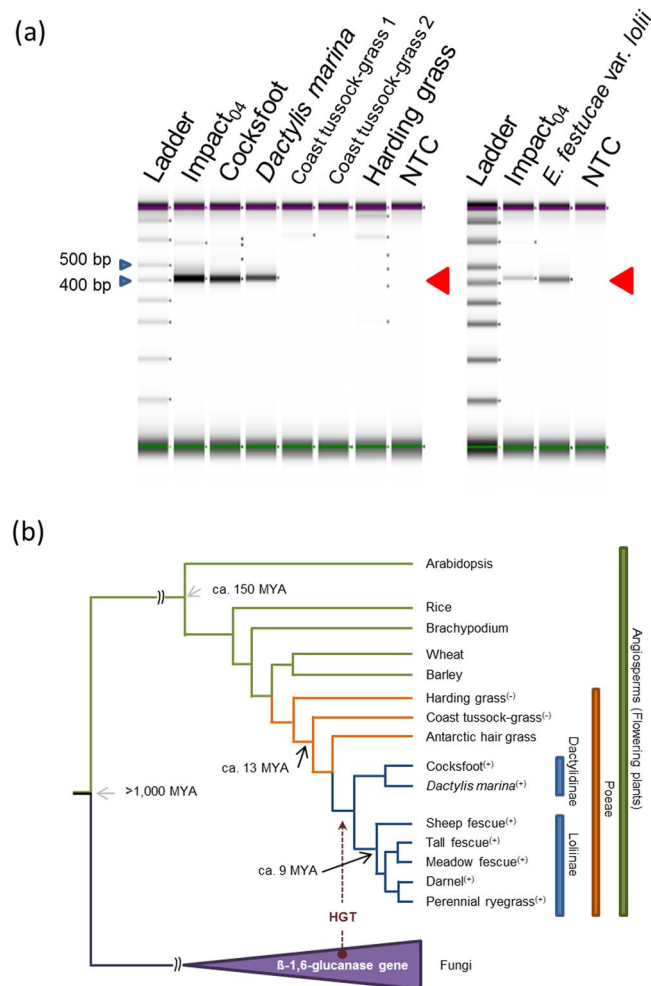
of contemporary species descended from the putative donor, when compared to other horizontally transferred genes in eukaryotes<sup>20</sup>.

The PCR-based screening and database searches suggested that the  $\beta$ -1,6-glucanase-like gene is present in only a limited number of Poae species including the genera *Lolium*, *Festuca* and *Dactylis*, which are confined to the sub-tribes Loliinae and Dactylidinae<sup>1,21</sup>. The phylogenetic analysis suggested a common origin of the *Epichloë*-derived  $\beta$ -1,6-glucanase genes and *LpBGNL* orthologues, and the close relationship between the *LpBGNL* orthologues of contemporary Loliinae and Dactylidinae grasses suggests that the gene may have been introduced into the genome of a common ancestor of the sub-tribes by a single transfer event. The HGT event may consequently have occurred between ca. 9 to 13 million years ago (MYA), based on the predicted time of divergence of the two sub-tribes from other Poae lineages (Fig. 5b)<sup>21–23</sup>.

Despite the relatively recent appearance (ca. 200–230 MYA) and diversification (ca. 100–125 MYA) of the lineage, the angiosperms currently compose the largest grouping of land plants, with nearly 300,000 species<sup>24–26</sup>. In the previous systematic study of four angiosperm genomes (*A. thaliana*, rice, sorghum and poplar) no evidence for HGT from fungal species was found<sup>16</sup>. Due to identification of two and three events for moss and lycophyte lineages, respectively, it was consequently concluded that gene transfer between fungi and angiosperms must be exceedingly infrequent. The present study, however, has demonstrated that such an event has also occurred relatively recently in evolutionary time. Considering the more distant date of divergence of the moss and lycophyte lineages (>400 MYA) compared to angiosperm species<sup>24</sup>, a longer period of evolutionary time may have provided more opportunities for fixation of transfer events in moss and lycophyte species. The specificity of the  $\beta$ -1,6-glucanase-like gene to only two sub-tribes of a single family suggests the possibility that other fungal-angiosperm HGT events could also be highly lineage-specific, and their detection would require more detailed comparisons of a larger number of fully sequenced genomes.

The Ks/Ka ratios between plant  $\beta$ -1,6-glucanase-like genes (0.027–0.221) were not substantially different from those of the fungal  $\beta$ -1,6-glucanase genes, and those of *LpABCG5* and *LpABCG6* (0.166 and 0.238)<sup>27</sup>. This suggests that *LpBGNL* may have been subjected to selection pressures. Although similar Ks/Ka ratios were obtained from the angiosperm and fungi groups, DNA mutation rates between those two groups may not be equivalent. A further characterisation is essential to confirm that both gene groups have been subjected to selection pressure at an equivalent level. From cocksfoot/orchard grass and *Dactylis marina*, three haplotypes of the aryl-phospho-beta-D-glucosidase domain were identified, and two haplotypes obtained from *Dactylis marina* contained a putative premature stop codon. As these *Dactylis* species have autopolyploid genomes<sup>28</sup>, the  $\beta$ -1,6-glucanase-like genes in these species may have been subjected to unique selection pressures due to genetic redundancy, when compared with genes in the other plant species.

As asexual symbiotic *Epichloë* species grow as hyphae between cells of vegetative aerial tissues in Poae species<sup>3,29</sup>, it is likely that the close physical proximity of both partners in the symbiosis may have facilitated an



**Figure 5.** PCR screening for the *LpBGNL* sequence, and taxonomic classification of plant species described in the current study (a) PCR amplicons from perennial ryegrass genotype *Impact*<sub>04</sub>, *Cocksfoot*, *Dactylis marina*, coast tussock-grass and harding grass, and a strain of *E. festucae* are visualised on the 2200 TapeStation instrument (Agilent Technologies, CA, USA). The target fragments are indicated with the red arrow. Two genotypes of coast tussock-grass were subjected to the screening. The purple and green lines show the position of upper and lower markers, respectively, of the D1000 Kit (Agilent). NTC denotes no-template control for the PCR assay. The blue arrows indicate the positions of 400 and 500-bp fragments of the ladder. (b) Species classified into angiosperms, the tribe Poae, and sub-tribes Loliinae and Dactylidinae are indicated by the green, orange and blue boxes, respectively, on the right side of the phylogenetic tree. The species which were positive and negative in the PCR screening step are indicated with the plus (+) and minus (–) signs, respectively. The divergent points of plant species from other species (>1,000 MYA) and the clade monocotyledon from other plant species (ca. 150 MYA) are indicated with light grey arrows. The divergent points of sub-tribes Loliinae and Dactylidinae from the remaining Poae species, especially from genus *Poa*, which is a closely related taxon (ca. 13 MYA), and fine-leaved fescue (Sheep fescue) from broad-leaved fescues (tall fescue and meadow fescue) (ca. 9 MYA) are indicated with black arrows. The purple arrow represents the putative period of HGT in the evolutionary lineage.

HGT event, similar to the physical contacts between parasitic and host plants<sup>7,9</sup>. The conservation of the intron position between the *LpBGNL* and *Epichloë*  $\beta$ -1,6-glucanase genes suggests that a part of the endophyte genome including the  $\beta$ -1,6-glucanase gene, rather than a reverse-transcription product of endophyte gene mRNA, was incorporated into the recipient genome. In prokaryotes, transformation is a prevalent mechanism of gene exchange, in which HGT occurs through uptake of exogenous double-stranded DNA by the recipient cell<sup>9</sup>. It is possible that the ancestral *LpBGNL* gene was acquired through a similar mechanism, due to endophyte cell death and transient liberation of genomic DNA in the vicinity of recipient plant cells. This scenario suggests that a transformation-like mechanism may have been active in the angiosperm lineages, as well as other HGT mechanisms such as *Agrobacterium* spp.-based conjugation and endosymbiotic gene transfer<sup>9</sup>.

In multicellular organisms, foreign genes introduced during unicellular or early developmental stages may increase the chance of successful transmission to offspring, due to subsequent potential dispersion of the genes within the individual at maturity, including to germline cells ('weak-link model')<sup>30</sup>. The life-cycle of asexual



Species	UI	Source	Data size	LpBGNL	<i>E. festucae</i> $\beta$ -1,6-glucanase gene	LpABCG5	LpABCG6
<i>Dactylis glomerata</i> L. (orchardgrass)	SRX738187	Transcriptome	32.6G bases	95% (8e-35)	99% (6e-41)	99% (3e-41)	94% (5e-32)
<i>Deschampsia antarctica</i>	SRX465632	Genome	31G bases	N.S.	N.S.	99% (7e-42)	97% (1e-38)
<i>Poa annua</i>	SRX745831	Transcriptome	25.9G bases	N.S.	N.S.	100% (5e-43)	95% (1e-33)
<i>Poa supina</i>	SRX745855	Transcriptome	8.9G bases	N.S.	N.S.	99% (8e-42)	90% (6e-27)
<i>Poa infirma</i>	SRX745858	Transcriptome	9.3G bases	N.S.	N.S.	100% (2e-43)	92% (2e-26)
<i>Phalaris aquatica</i>	SRX669405	Transcriptome	10.2G bases	N.S.	N.S.	97% (5e-39)	94% (4e-34)

**Table 3.** Database search result using NCBI BLASTN suite. The sequence identity (%) and expect (E) value of the top matching sequence are shown.

*Epichloë* endophytes involves growth within plant reproductive tillers and colonisation of inflorescences (including intimate contact with female gametophytic structures), followed by maternal transmission through seed<sup>3</sup> It is consequently possible that the  $\beta$ -1,6-glucanase-like gene was introduced into an ancestral angiosperm species from infecting fungi during an early developmental stage, resulting in successful transmission into offspring. Symbiotic and parasitic relationships are thought to have been important factors for facilitation of HGT between eukaryote genomes<sup>7,9</sup>. Considering the importance of event timing during the life-cycle of recipients, additional ancestral HGT events in angiosperm genomes may be detected if the species associates with symbiotic or parasitic partners during unicellular or early developmental stages, such as entomophilous flowers which interact with insect pollen vectors and microorganisms commensal with insects<sup>31</sup>.

Fungal  $\beta$ -1,6-glucanases have been reported to be specifically secreted into plant apoplasts during endophyte infection, and may play a role in provision of nutrition to the infecting endophyte, control of branching of the endophyte hyphae, and protection of plant tissues from infection of other fungal pathogens<sup>29</sup>. It is possible that the plant-encoded enzyme may participate in one or more of these processes, and so contribute to establishment of a stable symbiotic relationship. Although *Epichloë* endophytes do not colonise root tissues<sup>32</sup>, a relatively high level of expression of LpBGNL was observed in root tissues. It is hence possible that the plant-encoded enzyme may function to protect against infection by soil-borne fungal pathogens. Similarly, active expression in flowers may suggest the capacity to protect against fungal pathogens such as *Epichloë typhina*, which causes choke disease<sup>33</sup>. Further analysis is, however, required to test this hypothesis, as some *Festuca* and *Dactylis* species are relatively susceptible to infection by *E. typhina*, even though those species presumably also possess the  $\beta$ -1,6-glucanase-like gene<sup>4,33,34</sup>. As natural stable associations with asexual *Epichloë* endophytes are confined to the Poeae lineages that possess LpBGNL-like genes, the ancestral HGT event (which might have occurred from a sexual pathogenic *Epichloë*-like species) may have provided pre-adaptive conditions for the contemporary symbiosis. The evidence for selective pressure on the gene is suggestive, but further functional analysis is required to fully test this hypothesis.

As major symbionts, the asexual *Epichloë* species provide abiotic and biotic stress tolerance to grass species of the Poeae tribe. Tolerance to invertebrate herbivory is a well-characterised benefit to the host, partially attributable to the effects of a *makes caterpillars floppy-like (mcf-like)* gene<sup>6</sup>. The *mcf*-like gene was horizontally transferred into the endophyte genome from a bacterial species 7.2–58.8 MYA. It is hence possible that multiple horizontal transfer events, including transfer of the LpBGNL-like gene as described in the present study, have been involved in the establishment of the current stable symbiotic relationship.

The potential role of a  $\beta$ -1,6-glucanase-like gene in protection against other, pathogenic, fungal species is of particular interest. Species such as *T. harzianum* are mycoparasites of fungal phytopathogens, and this property is related to glucanase activity<sup>35</sup>. Fungal-derived genes for anti-fungal enzymes such as endochitinases and glucanases have also been used for generation of transgenic plants with enhanced pathogen resistance<sup>35,36</sup>. Further studies of the role of the LpBGNL-like gene could hence involve experimental transfer into the genomes of crop plants such as rice and wheat, and subsequent evaluation of resistance to fungal diseases.

## Methods

**Plant materials and DNA extraction.** Details of plant genotypes are summarised in Table 1<sup>19,37</sup>. Genomic DNA was extracted from young leaves of plants and fungal endophyte mycelium using the DNeasy plant mini kit (QIAGEN, Hilden, Germany).

**PCR amplification.** Locus-specific primers were designed using the Sequencher software (GENECODE, MI, USA) and the PCR primers are listed in Supplementary Table S5<sup>27,38</sup>. PCR amplification was performed with MyFi polymerase kit (BIOLINE, London, UK). PCR amplicons were visualised on the 2200 TapeStation instrument.

**Short-read sequencing of BAC clones and amplicons.** The BAC-based genomic library<sup>39</sup> was screened through use of PCR. For the phylogenomic analysis, PCR primers were designed to obtain genomic fragments from Loliinae and Dactylidinae species (Supplementary Table S5). Sequencing libraries for the MiSeq platform (Illumina, San Diego, California, USA) were prepared from the BAC clones and PCR amplicons, following the previously described MspJI-based method<sup>40</sup>. The library was characterised with the TapeStation and

Qubit instruments (Thermo Fisher Scientific). The outcome reads were assembled with the Sequencher and SOAPdenovo programs<sup>41</sup>.

**Genetic linkage analysis.** PCR primers were designed to detect the indel polymorphism within the *LpBGNL* sequence (Supplementary Table S4). Genetic linkage analysis was performed through use of the p150/112 reference genetic mapping population of perennial ryegrass using the JoinMAP 3.0 application<sup>42,43</sup>.

**Gene expression analysis.** The transcriptome sequencing reads from Impact<sub>04</sub> tissues were mapped against Impact<sub>04</sub> genome contigs (>999 bp) for filtering. The number of reads which contained *LpBGNL* sequence (no sequence mismatch for 60 bp or longer) were counted as *LpBGNL*-derived reads. For gene expression in seedlings, E<sup>+</sup> and E<sup>-</sup> seeds of perennial ryegrass cultivar Alto were subjected to germination treatment by placement on wet filter paper in the dark for 2 days followed by seedling growth under full-light conditions<sup>44</sup>. RNA was extracted with a CTAB extraction method, and sequencing libraries were prepared using the SureSelect strand-specific RNA library preparation kit (Agilent). Sequencing analysis was performed on the Illumina HiSeq 3000 platform.

**In silico analysis.** The DNA sequences of fungal  $\beta$ -1,6-glucanase genes were obtained from the NCBI (<http://www.ncbi.nlm.nih.gov/>) database and the Genome Project at the University of Kentucky website<sup>45</sup>. Putative orthologous sequences were sought in the NCBI, *Brachypodium distachyon* (<http://www.brachypodium.org/>) and Ensembl (<http://plants.ensembl.org/index.html>) databases. Non-synonymous and synonymous nucleotide substitution rates (Ka and Ks, respectively) were calculated using the Synonymous Non-synonymous Analysis Program (SNAP; <http://www.hiv.lanl.gov/>). Alignment of DNA sequences was performed with the CLUSTALW program (<http://www.genome.jp/tools/clustalw/>) with the default parameters. Phylogeny was generated with the MEGA7 program (<http://www.megasoftware.net/>).

**Data Availability.** The datasets of perennial ryegrass transcriptome sequencing are available in the NCBI SRA repository, (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA379202>). The other sequencing datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

- Soreng, R. J. *et al.* A worldwide phylogenetic classification of the Poaceae (Gramineae): Phylogenetic classification of the grasses. *Journal of Systematics and Evolution* **53**, 117–137 (2015).
- Forster, J. W. *et al.* In *Plant genotyping II: SNP technology* (ed. Henry, R. J.) 154–186 (CABI, 2008).
- Hettiarachchige, I. K. *et al.* Phylogenomics of asexual *Epichloë* fungal endophytes forming associations with perennial ryegrass. *BMC Evolutionary Biology* **15**, 72 (2015).
- Saikkonen, K., Young, C. A., Helander, M. & Schardl, C. L. Endophytic *Epichloë* species and their grass hosts: from evolution to applications. *Plant Molecular Biology* **90**, 665–675 (2016).
- Wilkinson, H. H. *et al.* Contribution of fungal loline alkaloids to protection from aphids in a grass-endophyte mutualism. *Molecular Plant-Microbe Interactions* **13**, 1027–1033 (2000).
- Ambrose, K. V., Koppenhöfer, A. M. & Belanger, F. C. Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloë* fungal symbionts of grasses. *Scientific Reports* **4**, 5562 (2014).
- Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**, 605–618 (2008).
- Schönknecht, G., Weber, A. P. M. & Lercher, M. J. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution: Insights & Perspectives. *BioEssays* **36**, 9–20 (2014).
- Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**, 472–482 (2015).
- Bock, R. The give-and-take of DNA: horizontal gene transfer in plants. *Trends in Plant Science* **15**, 11–22 (2010).
- Vaughn, J., Mason, M., Sper-Whitis, G., Kuhlman, P. & Palmer, J. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *cox1* gene of *Peperomia*. *Journal of Molecular Evolution* **41**, 563–572 (1995).
- Sanchez-Puerta, M. V., Cho, Y., Mower, J. P., Alverson, A. J. & Palmer, J. D. Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. *Molecular Biology and Evolution* **25**, 1762–1777 (2008).
- Sanchez-Puerta, M. V. *et al.* Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evolutionary Biology* **11**, 277 (2011).
- Rice, D. W. *et al.* Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**, 1468–1473 (2013).
- Davis, C. C. & Xi, Z. Horizontal gene transfer in parasitic plants. *Current Opinion in Plant Biology* **26**, 14–19 (2015).
- Richards, T. A. *et al.* Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *The Plant Cell Online* **21**, 1897–1911 (2009).
- De la Cruz, J. & Llobell, A. Purification and properties of a basic endo- $\beta$ -1,6-glucanase (BGN16.1) from the antagonistic fungus *Trichoderma harzianum*. *European Journal of Biochemistry* **265**, 145–151 (1999).
- Konno, N. & Sakamoto, Y. An endo- $\beta$ -1,6-glucanase involved in *Lentinula edodes* fruiting body autolysis. *Applied Microbiology and Biotechnology* **91**, 1365–1373 (2011).
- Shinozuka, H. *et al.* Genome and transcriptome analysis in perennial ryegrass. in *8th International Symposium on Molecular Breeding of Forage and Turf* (2014).
- Ku, C. & Martin, W. F. A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70% rule. *BMC Biology* **14**, 89 (2016).
- Charmet, G., Ravel, C. & Balfourier, F. Phylogenetic analysis in the *Festuca-Lolium* complex using molecular markers and ITS rDNA. *Theoretical and Applied Genetics* **94**, 1038–1046 (1997).
- Chang, C. C., Chen, H. L., Li, W. H. & Chaw, S. M. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution* **58**, 424–441 (2004).
- Berbee, M. L. & Taylor, J. W. Dating the molecular clock in fungi – how close are we? *Fungal Biology Reviews* **24**, 1–16 (2010).
- Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences* **107**, 5897–5902 (2010).
- Grossnickle, D. M. & Polly, P. D. Mammal disparity decreases during the Cretaceous angiosperm radiation. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20132110–20132110 (2013).

26. Christenhusz, M. J. M. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201 (2016).
27. Shinozuka, H., Cogan, N. O. I., Spangenberg, G. C. & Forster, J. W. Comparative genomics in perennial ryegrass (*Lolium perenne* L.): identification and characterisation of an orthologue for the rice plant architecture-controlling gene *OsABC5*. *International Journal of Plant Genomics* **2011**, 1–12 (2011).
28. Lumaret, R., Bowman, C. M. & Dyer, T. A. Autopolyploidy in *Dactylis glomerata* L.: further evidence from studies of chloroplast DNA variation. *Theoretical and Applied Genetics* **78**, 393–399 (1989).
29. Moy, M., Li, H. M., Sullivan, R., White, J. F. & Belanger, F. C. Endophytic fungal beta-1,6-glucanase expression in the infected host grass. *Plant Physiology* **130**, 1298–1308 (2002).
30. Huang, J. Horizontal gene transfer in eukaryotes: The weak-link model: Insights & Perspective. *BioEssays* **35**, 868–875 (2013).
31. Mohr, K. I. & Tebbe, C. C. Field study results on the probability and risk of a horizontal gene transfer from transgenic herbicide-resistant oilseed rape pollen to gut bacteria of bees. *Applied Microbiology and Biotechnology* **75**, 573–582 (2007).
32. Vázquez-de-Aldana, B. R., García-Ciudad, A., García-Criado, B., Vicente-Tavera, S. & Zabalgoeazcoa, I. Fungal endophyte (*Epichloë festucae*) alters the nutrient content of *Festuca rubra* Regardless of Water Availability. *PLoS ONE* **8**, e84539 (2013).
33. Zabalgoeazcoa, I., Ciudad, A. G., Leuchtmann, A., Vázquez de Aldana, B. R. & Criado, B. G. Effects of choke disease in the grass *Brachypodium phoenicoides*. *Plant Pathology* **57**, 467–472 (2008).
34. Pfender, W. F. & Alderman, S. C. Geographical distribution and incidence of orchardgrass choke, caused by *Epichloë typhina*, in Oregon. *Plant Disease* **83**, 754–758 (1999).
35. Emani, C. *et al.* Enhanced fungal resistance in transgenic cotton expressing an endochitinase gene from *Trichoderma virens*: Fungal-resistant transgenic cotton. *Plant Biotechnology Journal* **1**, 321–336 (2003).
36. Antony Caesar, S. & Ignacimuthu, S. Genetic engineering of crop plants for fungal resistance: role of antifungal genes. *Biotechnology Letters* **34**, 995–1002 (2012).
37. Hand, M. L., Cogan, N. O., Stewart, A. V. & Forster, J. W. Evolutionary history of tall fescue morphotypes inferred from molecular phylogenetics of the *Lolium-Festuca* species complex. *BMC Evolutionary Biology* **10**, 303 (2010).
38. Tu, Y. *et al.* Functional analyses of *caffeic acid O-methyltransferase* and *cinnamoyl-CoA-reductase* genes from perennial ryegrass (*Lolium perenne*). *The Plant Cell* **22**, 3357–3373 (2010).
39. Spangenberg, G. *et al.* Future directions in the molecular breeding of forage and turf. in *Molecular breeding for the genetic improvement of forage crops and turf* 83–97 (Wageningen Academic Publishers, 2005).
40. Shinozuka, H. *et al.* A simple method for semi-random DNA amplicon fragmentation using the methylation-dependent restriction enzyme MspI. *BMC Biotechnology* **15**, 25 (2015).
41. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
42. Van Ooijen, JoinMap 3.0: software for the calculation of genetic linkage maps (2001).
43. Jones, E., Dupal, M., Dumsday, J., Hughes, L. & Forster, J. An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics* **105**, 577–584 (2002).
44. Timothy Sawbridge. Genomic and Transcriptomic analysis of perennial ryegrass/*Epichloë* endophytes symbiote. In *International Plant & Animal Genome XXIV* (2016).
45. Schardl, C. L. *et al.* Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the *Clavicipitaceae* reveals dynamics of alkaloid loci. *PLoS Genetics* **9**, e1003323 (2013).

## Acknowledgements

The authors would like to thank Drs Melanie Hand and Jatinder Kaur for provision of DNA samples. We also acknowledge the South Australian Research and Development Institute (SARDI) for provision of orchard grass, *Dactylis marina* and coast tussock-grass seeds. This work was supported by funding from the Victorian Department of Economic Development, Jobs, Transport and Resources.

## Author Contributions

T.I.S. identified the LpBGNL sequence and contributed to the computational analysis. H.S., I.H., M.S. and N.O.I.C. contributed to the experimental work and data analysis. G.C.S., B.G.C. and J.W.F. provided overall project leadership. H.S., B.G.C. and J.W.F. prepared the primary drafts of the manuscript and contributed to finalisation of the text. N.O.I.C., G.C.S. and T.I.S. co-developed interim and final drafts of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-07886-2

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017