Horizontal gene transfer of a bacterial insect toxin gene into the Epichloë fungal symbionts of grasses

Karen V. Ambrose¹, Albrecht M. Koppenhöfer² & Faith C. Belanger¹

¹Department of Plant Biology and Pathology, Rutgers University, New Brunswick, NJ 08901, ²Department of Entomology, Rutgers University, New Brunswick, NJ 08901.

Horizontal gene transfer (HGT) is recognized as an important factor in genome evolution, particularly when the newly acquired gene confers a new capability to the recipient species. We identified a gene similar to the makes caterpillars floppy (mcf1 and mcf2) insect toxin genes in Photobacterium, bacterial symbionts of nematodes, in the genomes of the Epichloë fungi, which are intercellular symbionts of grasses. Infection by Epichloë spp. often confers insect resistance to the grass hosts, largely due to the production of fungal alkaloids. An mcf-like gene is present in all of the Epichloë genome sequences currently available but in no other fungal genomes. This suggests the Epichloë genes were derived from a single lineage-specific HGT event. Molecular dating was used to estimate the time of the HGT event at between 7.2 and 58.8 million years ago. The mcf-like coding sequence from Epichloë typhina subsp. poae was cloned and expressed in Escherichia coli. E. coli cells expressing the Mcf protein were toxic to black cutworms (Agrotis ipsilon), whereas E. coli cells containing the vector only were non-toxic. These results suggest that the Epichloë mcf-like genes may be a component, in addition to the fungal alkaloids, of the insect resistance observed in Epichloë-infected grasses.
Results

Detection of a makes caterpillars floppy-like Gene in Epichloë Genomes. Recently we reported a quantitative comparative transcriptome analysis of E. festucae-infected versus endophyte-free Festuca rubra L. subsp. rubra (strong creeping red fescue) using the high throughput sequencing approach of SOLiD-SAGE15. Strong creeping red fescue is a commercially important grass species and is often naturally infected with the fungal endophyte E. festucae16,17. One SOLiD-SAGE tag recovered was from an E. festucae gene similar to the insect toxin genes makes caterpillars floppy1 and 2 (mcf1 and mcf2) of Photorhabdus luminescens15,18 and the fitD toxin genes of some Pseudomonas spp.19. Ph. luminescens is a symbiotic bacterium of entomopathogenic nematodes. When the nematodes infest their insect host, the bacteria are regurgitated into the insect hemolymph where they release several toxins that kill the insect host20,21. The mcf1 gene was discovered in a screen of Ph. luminescens cosmid clones for insect toxicity and the name comes from the effect seen after normally non-toxic Escherichia coli expressing the gene were injected into caterpillars. mcf2, also from Ph. luminescens, and fitD, from Pseudomonas protegens Pf-5 and CHA1 isolates (previously called Pseudomonas fluorescens)22 were discovered by their sequence similarity to mcf1 and have been shown to have a similar effect on caterpillars23–25. mcf1 has a 1,623 bp N-terminal extension relative to mcf2 (Fig. 2a). All of these genes encode large proteins (2,388 to 2,996 amino acids) with a TcdA/TcdB pore-forming domain similar to that of Clostridium difficile toxin B25. Mcf1 was found to kill insects by promoting apoptosis via the mitochondrial pathway25,26. The Ps. protegens FitD toxin contributes to the oral insecticidal activity of the bacteria25. The mode of action of Mcf2 is unknown26.

The genome sequences of 13 Epichloë spp. are publicly available (15; http://www.endophyte.uky.edu/) and all contain sequences similar to the E. festucae mcf-like gene. However, similar sequences are not found in any of the hundreds of non-Epichloë fungal genome sequences currently available in NCBI. The uniqueness of the mcf-like gene to the Epichloë lineage suggests it likely arose from a HGT event from a bacterial donor. The alternative scenario in which the Epichloë mcf-like genes originated from vertical gene transmission can be rejected since this would imply that the mcf-like gene was lost in all fungal lineages except from Epichloë. Diagrams of the Epichloë mcf-like genes are shown in Fig. 2a. The annotated position of the N-terminal of the proteins was variable, but most contained genes with long open-reading frames interrupted by 2 introns. The gene structure of E. amarillans, E. bonnicii, E. brachyelytri, E. glyceriae, E. mollis, and E. aotearoeae is considered the normal structure, based on our expression analysis of the gene from E. typhina subsp. poae infecting Poa secunda subsp. juncifolia (big blue grass), designated E. typhina subsp. poae Ps125,27 (discussed below). The E. elymi, E. festucae RC (the Rose City isolate of E. festucae), E. typhina subsp. poae 5819, and E. gansuensis isolates had mutations that resulted in frame shifts and early termination codons and so would not be expected to produce an intact Mcf-like protein. The E. gansuensis var. inebrians and E. typhina E8 isolates were annotated as having N-terminal extensions relative to most of the other species. In both cases the annotated extension can be attributed to a deletion of the presumed normal coding sequence at the terminal region of the gene and were annotated as beginning at a downstream region of the gene. The remarkable amino acid similarity of the predicted Mcf-like protein from E. typhina subsp. poae Ps1 to the bacterial Mcf1 and Mcf2 proteins is illustrated in Fig. 2b. The similarity extends throughout the 1,997 amino acid long protein and is particularly striking in the TcdA/TcdB pore forming domain.

The Epichloë mcf-like coding sequences were more similar in size to those of Ph. luminescens mcf2, but the overall number of amino acid identities in the region of overlap was similar between the Epichloë sequences and mcf1, mcf2, and fitD (Supplementary Table 1). The nearly identical similarity of the Epichloë sequences to all three bacterial genes, as well as a preliminary phylogenetic analysis precluded identification of the likely donor in the HGT event (discussed more below).

As is typical of bacterial genes, there are no introns in the mcf1, mcf2, or fitD genes, but there are presumed introns in the Epichloë mcf-like genes. Since it appears the Epichloë mcf-like genes originated from a HGT event, the introns must have been acquired subsequent to the HGT. Such intron acquisition has been found in other eukaryotic genes of bacterial origin27–29. Identical intron positions in genes from different recipient species suggest a single HGT event in the oldest member of the lineage, with subsequent vertical inheritance in the more recently diverged species29. All of the Epichloë mcf-like genes had an intron in the 3' region of the gene at an identical insertion site, although there is sequence variation within the intron sequence itself. Most also had a conserved 5' intron, except the E. gansuensis var. inebrians and E. typhina E8 isolates, both of which had deletions in the 5' sequence, which resulted in annotated N-terminal extensions. The E. festucae 2368 and Fl1 isolates, which had truncated 5' regions, also lacked the 5' intron. The two identical intron positions suggest there was a single HGT insertion as well as single intron acquisition events in the original horizontally transferred mcf-like gene, which was subsequently vertically inherited throughout the speciation of the Epichloë.

Estimation of the Age of the mcf-like Gene HGT Event into Epichloë. There have been numerous phylogenetic studies of Epichloë, but none have directly addressed the order and age of species divergence. To aid in estimation of the age of the mcf-like gene HGT event into Epichloë, we undertook such an analysis. Most phylogenetic studies of Epichloë have been based on β-tubulin (tubB), elongation factor EF-1α (tefA), and γ-actin DNA sequence comparisons. In a comprehensive survey that identified optimal fungal genes for use in phylogenetic analyses, these commonly used genes were not among the best30. The two most useful genes were MCM7, a subunit of the minichromosome maintenance complex31, and TSR1, a gene required for synthesis of 40S ribosomal subunits32. MCM7 and TSR1 have been found useful in fungal phylogenies that include a wide range of taxa33–36. Other high-performing genes identified were NAD-dependent glutamate dehydrogenase and isoleucine tRNA synthase37.

Maximum parsimony phylogenetic analyses of MCM7, TSR1, NAD-dependent glutamate dehydrogenase, and isoleucine tRNA
synthase sequences from 13 Epichloë, as well as some related Clavicipitaceae, are shown in Fig. 3 and Supplementary Fig. 1. Accession numbers of the sequences are given in Supplementary Table 2. To facilitate the alignments for these analyses the intron sequences were removed and only the DNA coding sequences were used. The \textit{Metarhizium anisopliae} sequences were designated as the outgroups for rooting the trees since \textit{M. anisopliae} is in a sister clade, which diverged prior to that of \textit{Epichloë}.

In all of the trees, the two \textit{E. gansuensis} sequences were in a sister clade to the rest of the \textit{Epichloë} spp., which had \textit{E. glyceriae} at its base. Since both \textit{E. gansuensis} and \textit{E. glyceriae} have the \textit{mcf-like} gene in their genomes, the HGT event must therefore have occurred in the common ancestor of these two basal \textit{Epichloë} spp.

To estimate the age of the HGT event we used a molecular clock approach\textsuperscript{36} similar to that used to estimate the age of polyploidy in \textit{Gossypium} (cotton)\textsuperscript{37}, the divergence times of maize LTR-retrotransposons\textsuperscript{38}, and the age of a segmental duplication in maize\textsuperscript{39}. First we estimated the divergence rate for each gene by calibration to \textit{Atkinsonella hypoxylon} using the formula $T = K_s/2r$. A. hypoxylon (synonym: \textit{Balansia hypoxylon})\textsuperscript{40} is closely related to \textit{B. epichloë}\textsuperscript{35}. \textit{B. epichloë} was placed in a sister clade to the clade containing \textit{E. typhina} and the \textit{Claviceps} spp. and the divergence of the two clades was estimated to have occurred 81 Mya\textsuperscript{35}. The \textit{K_s} (substitutions at synonymous sites) and \textit{K_a} (substitutions at nonsynonymous sites) for the coding sequence of each gene relative to the \textit{A. hypoxylon} ortholog was determined (Supplementary Table 3). In this analysis glutamate dehydrogenase from \textit{E. gansuensis} var \textit{inebrians} was not included, since the gene sequence was split between two contigs, which introduced a gap into the alignment. The low \textit{K_a/K_s} ratios for these four genes indicate they are under purifying selection.

From the \textit{K_s} for each gene, the rate of divergence of synonymous sites ($r$) for each gene was calculated (Supplementary Table 3). The calculated rates varied from $5.93 \pm 10^{-9}$ to $7.72 \pm 10^{-9}$ substitutions per synonymous site per year, which are similar to reported fungal gene divergence rates\textsuperscript{42}.

The calculated rates of divergence of each gene were then used to calculate the time of divergence ($T$) between the \textit{Claviceps} clade and the \textit{Epichloë} clade as well as between \textit{E. glyceriae} and \textit{E. gansuensis} sequences (Supplementary Tables 4 and 5). The mean calculated divergence time between the \textit{Claviceps} clade and the \textit{Epichloë} clade was 58.8 Mya and the mean divergence time between \textit{E. gansuensis} and \textit{E. glyceriae} was 7.2 Mya. These estimates indicate that the single introduction of the \textit{mcf-like} gene into \textit{Epichloë} was sometime after the divergence of the \textit{Claviceps} and \textit{Epichloë} clades at 58.8 Mya and before the divergence of \textit{E. gansuensis} and \textit{E. glyceriae} at 7.2 Mya.

The \textit{Epichloë mcf-like} Gene Donor is Likely a Bacterium Related to \textit{Xenorhabdus or Photorhabdus}. Phylogenetic analysis of genes
MCM7

- 50 changes

| M. anisopliae | E. festucae 2368 |
|--------------|------------------|
| C. paspali | E. mollis |
| C. fusiformis | E. typhina subsp. poae 5819 |
| C. purpurea | E. aterose |
| A. hypoxylon | E. elymi |
| E. glyceriae | E. gansuensis var. inebrians |
| E. gansuensis | C. paspali |
| E. festucae var. inebrians | A. take |
| Ps1 (henceforth E. festucae Rc) isolate | 100 |
| Ps1 | 98 |
| Ph. luminescens subsp. laumondii | 81 |
| Ph. asymbiotica subsp. asymbiotica | 93 |
| Ph. temperata | 67 |
| Ph. luminescens | 98 |
| Ps. chlororaphis subsp. aureofaciens | 81 |
| Ps. protegens | 81 |
| E. gansuensis var. inebrians | 51 |
| X. nemataphila | 51 |
| X. szeretlariai | 98 |
| E. festucae 2368 | 98 |
| E. typhina subsp. poae Ps1 | 100 |
| V. tubiashii | 100 |

Figure 3 | Rooted 50% majority rule maximum parsimony phylogenetic tree of the MCM7 DNA coding sequences. The Metarhizium anisopliae sequences were designated as the outgroup for rooting the tree. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The tree was based upon 2,472 total characters, of which 1,549 were constant, 404 variable characters were parsimony uninformative, and 828 variable characters were parsimony informative. The unmarked nodes all had bootstrap support of 69 or higher.

Figure 4 | Rooted 50% majority rule maximum parsimony phylogenetic tree of bacterial Mcf proteins and the Mcf-like protein sequences from E. gansuensis var. inebrians and E. typhina subsp. poae. The Vibrio tubiashii sequence was designated as the outgroup for rooting the tree. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The tree was based upon 1,983 total characters, of which 391 were constant, 404 variable characters were parsimony uninformative, and 1,188 variable characters were parsimony informative.

Expression and Activity of the Epichloë Mcf-like Gene. The recovery of a SOLiD-SAGE tag from Festuca rubra infected with E. festuca\textsuperscript{27} supports the expression of the E. festuca\textsuperscript{27} mcf-like gene. To confirm expression, we amplified and sequenced the mcf-like cDNA from Festuca rubra infected with the Rose City isolate of E. festuca.

makes caterpillars floppy-like protein

derived from a HGT event can often be used to identify the donor\textsuperscript{43,44}. The best matches to the Epichloë Mcf-like amino acid sequences were Mcf1 and Mcf2 from Photorhabdus luminescens, and FitD from Pseudomonas spp. As discussed above, the overall similarity of the Epichloë Mcf-like sequences to the bacterial proteins was nearly the same (Supplementary Table 1). Similar Mcf2 proteins were also found in Xenorhabdus spp., also bacterial symbionts of entomopathogenic nematodes\textsuperscript{19}, but have not been functionally characterized. Maximum parsimony phylogenetic analysis of the bacterial Mcf proteins and Epichloë mc

Designated E. festuca\textsuperscript{RC}. The expression level was extremely low and required three rounds of PCR to obtain enough sample for sequencing. Sequence analysis of the E. festuca\textsuperscript{RC} mcf-like cDNA revealed a single base insertion resulting in a frame-shift and an early termination codon, relative to the E. festuca\textsuperscript{2368} and F1 isolates. The single base insertion was confirmed by sequencing that region of the E. festuca\textsuperscript{RC} genome. Although the mcf-like gene is expressed in the E. festuca\textsuperscript{RC} isolate, it could not produce a full-length protein.

We also analyzed expression of the mcf-like gene in E. typhina subsp. poae Ps1, the fungal endophyte infecting Poa secunda subsp. junceifolia\textsuperscript{28}. The E. typhina subsp. poae Ps1-mcf cDNA generated from infected plant leaf sheath tissue was easily amplified with one round of PCR (Fig. 5a). Sequence analysis of the amplified cDNA revealed it could produce a full-length protein. The cDNA sequence also confirmed the positions of the two annotated introns. The intron sequences were determined by sequencing of genomic DNA.

Since the mcf-like gene in E. typhina subsp. poae Ps1 (henceforth designated as E. typhina subsp. poae Ps1-mcf) was expressed and was predicted to produce a full-length protein, its activity against insects was assessed. The toxicity of the bacterial Mcf proteins was determined by injection of E. coli cells expressing the protein into insect
hemocoel8,9,22. The E. typhina subsp. poae Ps1-mcf coding sequence was expressed in E. coli by cloning into the expression vector pCold II. This is the same strategy that was used to determine the activity of the Ps. protegens FitD protein23. A band of the expected size was seen on SDS-PAGE analysis of proteins from induced E. coli cultures (Fig. 5b). The induced protein was found in the insoluble fraction as was reported for the Ps. protegens FitD protein22. The toxicity of the induced protein was assayed by injecting black cutworms (Agrotis ipsilon) with 50 million E. coli cells harboring either the pCold II vector only or cells expressing the E. typhina subsp. poae Ps1-Mcf protein. The E. coli cells containing the vector only were not toxic to the cutworms, whereas injection of the cells expressing the Mcf protein resulted in statistically significant levels of cutworm deaths when compared to the results for the vector only controls (Table 1; Fig. 6). Lower levels of E. coli cells (20 million) expressing the E. typhina subsp. poae Ps1-Mcf protein were not toxic to the cutworms.

Discussion

Here we have reported the characterization of a mcf-like gene in the Epichloë fungal endophytes of grasses that is the result of a HGT event from a bacterial donor. The gene is present in the genomes of all 13 of the Epichloë spp., for which whole genome sequence data is available, as well as 2 additional isolates used in this study, but in no other fungal species. In 11 of the 15 Epichloë sequences a long protein, containing the TcDA/Tcdb pore-forming region, is predicted and in 4 of the Epichloë sequences there was a mutation leading to an early termination codon. In 3 out of these 4 cases, multiple isolates of the same species provided examples of early termination of the mcf-like gene and also of a gene capable of producing a potentially functional protein. A similar isolate specific breakdown of some alkaloid biosynthetic genes was seen among Epichloë spp.15. Phylogenetic analysis suggests the donor bacterium may have been an as yet unknown predecessor of extant Xenorhabdus or Photorhabdus spp. The mechanisms underlying HGT events are not known, but clearly the donor and recipient species must be in close contact. It is easy to imagine that the recipient Epichloë sp. could have come in contact with the bacterial donor either in soil or in association with a plant host.

HGT events are of particular significance when the newly acquired gene confers a new capability to the recipient species. A recent example is the transfer of a bacterial mannanase gene to the insect Hypothenemus hampei, the coffee berry borer beetle44. The acquisition of the mannanase gene was considered to have led to the adaptation of the insect to a new host, the coffee berry, which is rich in mannans. Here we have demonstrated that injection of E. coli expressing the E. typhina subsp. poae Ps1-mcf gene into black cutworms (Agrotis ipsilon), a pest of various crops, results in death of the insect, similar to the insecticidal activities of the bacterial mcf1, mcf2, and fitD genes8,9,22. The bacterial Mcf proteins are considered critical components of the overall insect toxicity of the bacteria. The E. typhina subsp. poae Ps1-Mcf protein has been maintained as an active protein over millions of years since its acquisition by an ancestral Epichloë sp. This suggests it may also be a component, in addition to alkaloids that are well-established as having insecticidal activity15,45, in the insect resistance observed in plants containing this endophyte isolate46. Other Epichloë spp. have genes apparently capable of producing Mcf-like proteins that may also have activity against insects, as shown here for the E. typhina subsp. poae Ps1 isolate.

Bacterial toxin genes of the aerolysin family (unrelated to the mcf toxin genes) were found to have undergone recurrent HGT to many diverse eukaryotes, possibly due to their adaptive value47. Moran et al. 47 proposed that genes derived from HGT are more likely to be maintained if they can function alone and are of immediate benefit

Table 1 | Cumulative numbers of black cutworm larvae deaths after each treatment. Insect assays were performed by injecting either sterile water, 5 × 10⁷ total induced E. coli cells containing the pCold II vector only as a negative control or the Etp-Mcfl::pCold II plasmid in a total volume of 5 μL. The experiments also included larvae that were untreated.

| Treatment/Time | 24 h | 48 h | 72 h | 120 h |
|---------------|------|------|------|-------|
| Untreated     | 0    | 0    | 0    | 0     |
| Water         | 0    | 0    | 0    | 0     |
| pCold II vector only: E. coli | 0 | 0 | 0 | 0 |
| Etp-Mcfl::pCold II: E. coli | 5 | 6 | 1 | 1 |

Figure 5 | Expression of E. typhina subsp. poae Ps1-mcf in vivo and in E. coli. (a) PCR product of the E. typhina subsp. poae Ps1-mcf transcript using infected Poa secunda subsp. juncifolia plant cDNA generated from oligo(dT) as template. (b) SDS-PAGE analysis of insoluble E. coli proteins. Lane 1: E. coli containing the pCold II vector only subjected to induction conditions. Lane 2: E. coli containing the pCold II::E. typhina subsp. poae Ps1-Mcf with no induction of protein expression. Lane 3: E. coli containing the pCold II::E. typhina subsp. poae Ps1-Mcf subjected to induction conditions. Arrow indicates presence of the induced protein at the expected size of 223 kD.

Figure 6 | Toxicity of the E. typhina subsp. poae Ps1-Mcf protein to black cutworms. Fourth instar cutworm larvae were injected with water, 50 million E. coli cells containing the pCold II vector only, or 50 million E. coli cells expressing the E. typhina subsp. poae Ps1-Mcf protein. The live cutworms curl in response to prodding.
to the organism. These characteristics apply to the Epichloe mcf genes, also derived from HGT, since the genes may contribute to the insect toxicity of Epichloe -infected plants.

Based on phylogenetic analyses and divergence time estimates, the single HGT event into Epichloe was estimated to have occurred sometime between 7.2 and 38.8 Mya. Dating of fungal evolution is problematic due to the minimal fossil record, so as more fungal calibration points become available it will be possible to refine the divergence time estimates presented here. The rooted phylogenetic trees generated in these analyses, based on genes established as phylogenetically high-performing genes, are also revealing regarding trees generated in these analyses, based on genes established as phylogenetic divergence time estimates presented here. The rooted phylogenetic problematic due to the minimal fossil record, so as more fungal single HGT event into Epichloe and with their grass hosts. The two most basal members of this genus, E. gansuensis isolates and the other comprised of the rest of the Epichloe spp., with E. gansuensis at its base. The phylogenetic analyses presented here suggest that an as yet unknown (or possibly extinct) species was the predecessor of E. gansuensis and E. glyceriae and that the rest of the currently known Epichloe spp. were derived from E. glyceriae.

It has been proposed that Epichloe spp. evolved through coevolution with their grass hosts. The two most basal members of this genus, E. gansuensis (host grass Achnatherum inebrinum, tribe Stipeae) and E. glyceriae (host grass Glycetria striata, tribe Meliceae) are symbionts of ancient tribes of the grass subfamily Pooidae. However, E. brachyelytri, the symbiont of Brachyelytrum erectum, which is in the most ancient Pooidae tribe Brachyelytraeae, was phylogenetically closer to the other Epichloe spp. than to E. gansuensis or E. glyceriae. These results indicate that the speciation of Epichloe was not exclusively through coevolution with their hosts since, in an exclusive model of coevolution, E. brachyelytri would be expected to be the most basal Epichloe sp. The phylogenetic trees presented here suggest that E. brachyelytri evolved via cross species transfer of an Epichloe sp. infecting a more recently evolved host grass to the ancient plant genus Brachyelytrum Jackson has also concluded that the Epichloe-host phylogenies are incongruent and do not support coevidence as a characteristic of the association. Lemaire et al., proposed host specificity without co-speciation for the Burkholderia bacterial end symbionts. The Epichloe – grass symbiosis may be a similar case.

Methods

Plant and Fungal Samples. Epichloe spp. -infected plants were grown by clonal propagation and maintained in 15 cm pots in the greenhouse. Each Epichloe sp. was isolated from its host plant by plating surface-sterilized leaf sheath tissue on potato dextrose agar (Difco Laboratories, Detroit, MI).

DNA and RNA Isolation. Genomic DNA of Epichloe spp. was extracted from culture grown in potato dextrose broth for 14 days on a shaker (175 rpm) at room temperature (23–25°C). The DNA was isolated as previously described. RNA was obtained from the innermost leaf sheath tissues of Epichloe-infected plants. RNA isolation was as previously described. Nucleic acid concentration was measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Amplification and Sequencing of Epichloe mcf-like Genes and cDNAs. Initial screening to determine the presence of the mcf-like gene in E. festucae BC and E. typhina subsp. poae Ps1 fungal isolates was done using fungal genomic DNA. Upon detection of the gene, first-strand cDNA from 5 µg Epichloe-infected plant total RNA was synthesized by using SuperScript III Reverse Transcriptase (Life Technologies, Carlsbad, CA) and 500 ng of oligo(dT)12-18 primer according to the manufacturer’s instructions. PCR was performed in 50 µL with either 0.5 µg of fungal genomic DNA or 1 µL Epichloe-infected plant cDNA generated from oligo(dT)12-18, 1.25 units of Primers T7 HS DNA Polymerase, 1X PrimerSTAR Buffer with Mg2+ (Takara Bio Inc., Shiga, Japan), 200 µM each dNTP, 0.3 µM of each forward and reverse primer (Integrated DNA Technologies, Inc., Coralville, IA), 2X PCR buffer. Two-step PCR was performed by template denaturation at 98°C for 10 seconds followed by 6 minutes extension at 68°C for 3 cycles in a GeneAmp 9700 thermal cycler (Applied Biosystems, Inc., Foster City, CA). The amplification products were visualized on a 1% TBE agarose gel. The amplified genomic and cDNA PCR fragments were purified by using 5X Agencourt AMPure XP (Beckman Coulter, Inc., Brea, CA) to remove any fragments under 1 kb and then sequenced directly (Genewiz, Inc., South Plainfield, NJ). For each sequencing reaction, about 800 ng of purified PCR product in 10 µL was treated with 2 µL ExoSAP-IT (USB Corp., Cleveland, OH) to remove unincorporated primers and excess dNTPs. The ExoSAP-IT reaction was performed at 37°C for 15 min following by heating at 80°C for 15 min to inactivate the enzymes.

Expression of the E. typhina subsp. poae Ps1-mcf Gene in E. coli. To test the anti-insect activity of the E. typhina subsp. poae Ps1-Mcf protein, the cDNA was cloned into the expression vector pCold II (Takara Bio). The E. typhina subsp. poae Ps1 cDNA was amplified by PCR as described above with oligonucleotides that incorporated a Kpd site at the 5’ end and an EcoRl site at the 3’ end. The amplified Kpd site oligonucleotide and an EcoRl site oligonucleotide were simultaneously digested with Kpd and EcoRl. The digested Kpd plasmid was transformed with shrimp alkaline phosphatase (Affymetrix, Santa Clara, CA) to prevent vector religation. Overnight ligation of the digested E. typhina subsp. poae Ps1-mcf PCR fragment and pCold II plasmid was done at 3 : 1 insertvector molar ratio using T4 DNA ligase (New England Biolabs, Inc., Ipswich, MA).

Two µL of the ligation product was used to transform 20 µL of E. coli DH5α electro- tropization-competent cells. The transformed cells were incubated in SOC medium for 1 hour at 37°C followed by overnight growth of cells on LB medium supplemented with 100 µg/mL ampicillin. Transformed bacterial colonies were screened for the presence of the E. typhina subsp. poae Ps1-mcf gene insert by using PCR as described above. Selected clones were grown in 3 mL of LB + ampicillin broth overnight followed by plasmid purification using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). DNA sequencing of the plasmids were done as described above. A plasmid containing the correct sequence was transformed into E. coli BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies, Santa Clara, CA). For using a control, a transformant containing the pCold II vector only was also generated.

For expression of the E. typhina subsp. poae Ps1-Mcf protein, the BL21-CodonPlus (DE3)-RIPL transformant was induced by cold shock and isoprropyl-β-D-thiogala- toxicynonoside (IPTG). We detected loss of E. typhina subsp. poae Ps1-mcf/Epcold II recombinant plasmid in E. coli cells during the course of protein expression. This was not observed in E. coli cells containing the pCold II vector only. Previous studies have shown that carbenicillin and ampicillin, even in high concentrations, quickly lose their ability to maintain selective pressure due to saturation of the media with the antibiotic degrading enzyme β-lactamase. Extra metabolic burden placed on the bacterial cells to maintain the large recombinant plasmid combined with the steady decrease of antibiotic in culture may have contributed to the plasmid loss that we observed. In order to circumvent this problem, the culture supernatant was replaced with fresh medium at regular intervals.

To induce expression a 10 mL culture of the cells was grown for 15 at 30°C in LB and 200 µg/mL carbenicillin. A 400 µL aliquot of the culture was pelleted, resuspended in 400 µL added to 10 mL of fresh LB plus carbenicillin and incubated at 30°C to an OD of 0.4–0.5. The culture was then cold-shocked for 15 minutes on ice and subsequently pelleted in a Model J2-21 centrifuge (Beckman Coulter, Inc., Brea, CA) to set to 5°C to obtain the bacterial cells. Fresh LB broth supplemented with carbenicillin and ampicillin was used to resuscitate the bacterial pellets. Protein expression in the culture, IPTG was added to 0.1 mM and the culture was then incubated at 15°C with rotational shaking for 24–21 h. The bacterial cells were pelleted and resuspended in fresh LB and carbenicillin every 3 h for the initial 6 h and again after 12 h to prevent β-lactamase saturation in the culture supernatant. The transformant containing the pCold II vector only was treated the same way.

For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), a 1 mL each of E. coli cells expressing the E. typhina subsp. poae Ps1-Mcf protein and the vector only control cultures were pelleted and lysed using 1X FastBreak Cell Lysis Reagent (Promega, Madison, WI) following the manufacturer’s protocol. Following lysis, samples were centrifuged at 10,000 RPM for 15 minutes and the supernatants removed. The proteins remaining in the pellet were solubilized by adding 100 µL of 2X SDS sample buffer, boiled for 5 minutes, and 30 µL was then subjected to electrophoresis on a 10% gel.

Insect Assays to Determine Toxicity of E. typhina subsp. poae Ps1-Mcf Protein. BL21-CodonPlus (DE3)-RIPL cells expressing the E. typhina subsp. poae Ps1-Mcf protein and the vector only control cultures were pelleted and lysed using 1X FastBreak Cell Lysis Reagent (Promega, Madison, WI) following the manufacturer’s protocol. Following lysis, samples were centrifuged at 10,000 RPM for 15 minutes and the supernatants removed. The proteins remaining in the pellet were solubilized by adding 100 µL of 2X SDS sample buffer, boiled for 5 minutes, and 30 µL was then subjected to electrophoresis on a 10% gel.
Phylogenetic Analysis. To identify genes in addition to MCM7 and TSR1 for use in the phylogenetic analyses, the FUNBASE genome database (http://genome.jouy.inra.fr/funbase/) was searched for single copy orthology identified as phylogenetically high-performing genes by Aguiletta et al. Two genes with topological scores of 96 or above and that also had significant sequence polymorphism within Epichloë were identified, FG570 (NAD-dependent glutamate dehydrogenase) and MS444 (isoleucine tRNA synthase). DNA sequences for these genes were obtained from the Genome Project at the University of Kentucky website (http://www.endothope.uky.edu/) or from NCBI.

The Clustal-X program was used to align the DNA coding sequences. The alignments generated by Clustal-X were modified manually to minimize gaps. The phylogenetic analyses were performed with the PAUP* program, version 4.0b10 for Macintosh. The phylogenetic analyses were done by using the maximum parsimony full heuristic search option set to random sequence addition, tree-bisection-reconnection (TBR) branch swapping, and Multrees on, with 1,000 bootstrap replications. Gaps were treated as missing data.

The Ks and Ka values based on pairwise comparison between species were determined by using the MEGA5.2 program with the Nei-Gojobori, Jukes-Cantor model.

1. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9, 605–618 (2008).
2. Syvanen, M. Evolutionary implications of horizontal gene transfer. Annu. Rev. Genet. 46, 341–358 (2012).
3. Richards, T. A., Leonard, G., Soanes, M. & Talbot, N. J. Molecular evolution of urea amidolyase and urea carboxylase in fungi. BMC Evol. Biol. 11, 80 (2011).
4. Friesen, T. L. et al. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. 38, 953–956 (2006).
5. Da Lage, J.-L., Binder, M., Hua-Van, A., Janecek, S. & Multrees on, 1,000 bootstrap replications. Gaps were treated as missing data.

The Ks and Ka values based on pairwise comparison between species were determined by using the MEGA5.2 program with the Nei-Gojobori, Jukes-Cantor model.
Acknowledgments

This work was supported with funds provided by the Rutgers Center for Turfgrass Science.

Author contributions

K.V.A., A.M.K. and F.C.B. conceived and designed the experiments; K.V.A., A.M.K. and F.C.B. performed the experiments; A.M.K. conducted the insect injections; K.V.A., A.M.K. and F.C.B. analysed and interpreted results. All the authors discussed the results and wrote the manuscript.

Additional information

Data deposition: Sequence data for the E. typhina subsp. poae Ps1-mcf gene has been deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/) under accession number KJ502561.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: 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. Sci. Rep. 4, 5562; DOI:10.1038/srep05562 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/