Received: 24 May 2019 | Accepted: 11 October 2019

DOI: 10.1111/mec.15275

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Greater genetic and regulatory plasticity of retained duplicates in *Epichloë* endophytic fungi

Baojun Wu[1,2] | **Murray P. Cox**[1,2]

[1]Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

[2]Bio-Protection Research Centre, Massey University, Palmerston North, New Zealand

Correspondence
Murray P. Cox, Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North 4410, New Zealand.
Email: m.p.cox@massey.ac.nz

Funding information
Royal Society of New Zealand, Grant/Award Number: 14-MAU-007; Tertiary Education Commission via the Bio-Protection Research Centre

Abstract

Gene duplicates can act as a source of genetic material from which new functions arise. Most duplicated genes revert to single copy genes and only a small proportion are retained. However, it remains unclear why some duplicate genes persist in the genome for an extended time. We investigate this question by analysing retained gene duplicates in the fungal genus *Epichloë*, ascomycete fungi that form close endophytic symbioses with their host grasses. Retained duplicates within this genus have two independent origins, but both long pre-date the origin and diversification of the genus *Epichloë*. We find that loss of retained duplicates within the genus is frequent and often associated with speciation. Retained duplicates have faster evolutionary rates (Ka) and show relaxed selection (Ka/Ks) compared to single copy genes. Both features are time-dependent. Through comparison of conspecific strains, we find greater evolutionary rates in coding regions and sequence divergence in regulatory regions of retained duplicates than single copy genes, with this pattern more pronounced for strains adapted to different grass host species. Consistent with this sequence divergence in regulatory regions, transcriptome analyses show greater expression variation of retained duplicates than single copy genes. This suggest that *cis*-regulatory changes make important contributions to the expression patterns of retained duplicates. Coupled with supporting observations from the model yeast *Saccharomyces cerevisiae*, these data suggest that genetic robustness and regulatory plasticity are common drivers behind the retention of duplicated genes in fungi.

KEYWORDS
endophytic fungi, expression plasticity, nonsynonymous substitutions, retained duplicates, selection

1 | INTRODUCTION

Gene duplicates are considered a major driver of functional divergence (Lynch & Conery, 2000) because their genetic redundancy allows one gene copy to evolve free from the selective constraints experienced by the original copy (Ohno, 1999; Wagner, 2002). Consistent with this view, duplication events often coincide with the emergence of morphological, metabolic and physiological innovations in animals and plants (Hoegg, Brinkmann, Taylor, & Meyer, 2004; Holub, 2001; Lespinet, Wolf, Koonin, & Aravind, 2002; Maere et al., 2005; Otto & Whitton, 2000). Duplicated gene copies can not only change the sequence of their coding region, but can also alter their expression patterns. Duplicated genes in *Drosophila* have as much as five times greater expression than

single copy genes (Loehlin & Carroll, 2016), with slightly lower rates of expression observed in *Caenorhabditis elegans* (Konrad et al., 2018). However, in *Drosophila yakuba* and some mammalian species, there is little evidence for increased expression following gene duplication (Guschanski, Warnefors, & Kaessmann, 2017; Henrichsen et al., 2009; Rogers, Shao, & Thornton, 2017). This may be driven by dosage balance, where changes in the relative abundance of proteins can have deleterious effects (Antonarakis, Lyle, Dermitzakis, Reymond, & Deutsch, 2004). Expression-level regulation is reflected not only by expression abundance but also by tissue specificity (Assis & Bachtrog, 2015), suggesting that expression changes following gene duplication are part of a complex process.

In the long term, preservation of gene duplicates is rare, regardless of whether they originate through whole genome duplication or smaller-scale duplications. In jawed vertebrates, most duplicated genes have been lost over a time span of 60 million years (Myr), with more than 80% of genes returning to single copies (Inoue, Sato, Sinclair, Tsukamoto, & Nishida, 2015). *Saccharomyces cerevisiae* lost 92% of its duplicates over 100 Myr (Wolfe & Shields, 1997). The loss process of duplicates in yeast has been shown to start immediately; gene copies were lost within 25 generations (<2 days) under normal laboratory conditions when an extra copy of the *IFA38* gene was introduced artificially (Naseeb, Ames, Delneri, & Lovell, 2017). Retention and loss events of duplicated genes are thus variably dynamic, and can occur over short to long time spans.

So far, most studies of duplicated genes have focused on their functional and evolutionary divergence (Conant & Wagner, 2003; Guan, Dunham, & Troyanskaya, 2007; Kellis, Birren, & Lander, 2004; Soria, McGary, & Rokas, 2014). As such, features of selection and expression are mainly compared between paralogues. However, why some duplicated genes persist in the genome for an extended time, while others are transient, remains unclear. Hypotheses related to retention frequently invoke functional innovation—neofunctionalization and subfunctionalization (Assis & Bachtrog, 2013; Blanc & Wolfe, 2004a; Innan & Kondrashov, 2010; Ohno, 1970; Stoltzfus, 1999)—which in turn implies that retained duplicates should exhibit new sequence or regulatory properties.

However, few studies have tested this hypothesis, and almost all have been performed in model organisms, such as the yeast *S. cerevisiae* (Fares, Sabater-Munoz, & Toft, 2017; Keane, Toft, Carretero-Paulet, Jones, & Fares, 2014). Yet yeasts are not characteristic of global fungal diversity. For instance, *S. cerevisiae* S288C as a lab strain has a compact genome and has undergone vastly different selective pressures from strains in nature (Gu et al., 2005). The different genomic features and environments may bias the role of selection on duplication retention. This focus thus has limited our capacity to draw general conclusions about why some duplicates are retained, while others are lost, especially under natural conditions. Here, we address this question in the fungal genus *Epichloë*, which diverged from *S. cerevisiae* 590 million years ago (Hedges, Marin, Suleski, Paymer, & Kumar, 2015). As with yeast, we find that fast-evolving coding sequences, diversification of regulatory regions and expression plasticity are all key determinants of retained duplicate genes in this genus.

2 | MATERIALS AND METHODS

2.1 | Genome re-annotation

All genomes used in this study are available in public databases (Table S1). Contigs <1 kb were excluded from analysis. The raw transcriptome data of *Epichloë festucae* var. *lolii* AR5 (accession SRX1533738) were downloaded from the NCBI SRA database and mapped to the AR5 genome using STAR 2.6.0 (Dobin et al., 2013). The genome of AR5 was annotated using BRAKER 2.1.2 (Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2016), with both the genome sequence and mapped RNA used as input. After annotation, a training file of *Epichloë* gene models was generated and applied to the remaining genomes in the same genus using AUGUSTUS 3.3.1 (Stanke et al., 2006).

2.2 | Construction of a species tree

Single copy orthologues (SCOs) were inferred using ORTHOFINDER 2 (Emms & Kelly, 2018) with BLASTP and an expectation value of 10^{-3} (Buchfink, Xie, & Huson, 2015). The species phylogeny was constructed with FASTTREE 2 (Price, Dehal, & Arkin, 2010) using 10^3 single copy orthologues predicted by ORTHOFINDER 2.

2.3 | Identification of one-to-one retained gene duplications

Groups of orthologues shared among species were inferred with ORTHOFINDER 2 (Emms & Kelly, 2018) using the same parameters as above. A maximum-likelihood tree of each orthogroup was constructed with FASTTREE 2 (Price et al., 2010). Within each orthogroup, there are many orthology relationships, such as one-to-one or one-to-many. Reconciliation analysis using a novel duplication-loss coalescent algorithm implemented in ORTHOFINDER 2 (Emms & Kelly, 2018) was performed for each gene tree relative to the species tree, thus accounting for incomplete lineage sorting and gene tree error. After reconciliation, ORTHOFINDER 2 can predict three classes of orthology relationships among the species: one-to-one, many-to-one and many-to-many. Duplicates were counted as a retention event if a one-to-one orthology relationship was inferred among duplicates from two species (i.e., the four copies of a gene as depicted in Figure 1a). This tree-based method has two advantages compared with synteny-based methods: (a) it can readily identify one-to-one orthologues even in fragmented assemblies, where synteny is often interrupted; and (b) it can identify one-to-one orthologues between species, even where genes have been transposed over the course of evolution (Woodhouse, Pedersen, & Freeling, 2010).

However, to be rigorous, we also considered the synteny relationships of one-to-one orthologues predicted by the tree method. The protein sequences of five consecutive genes flanking each one-to-one orthologue between species were extracted, and BLASTP was

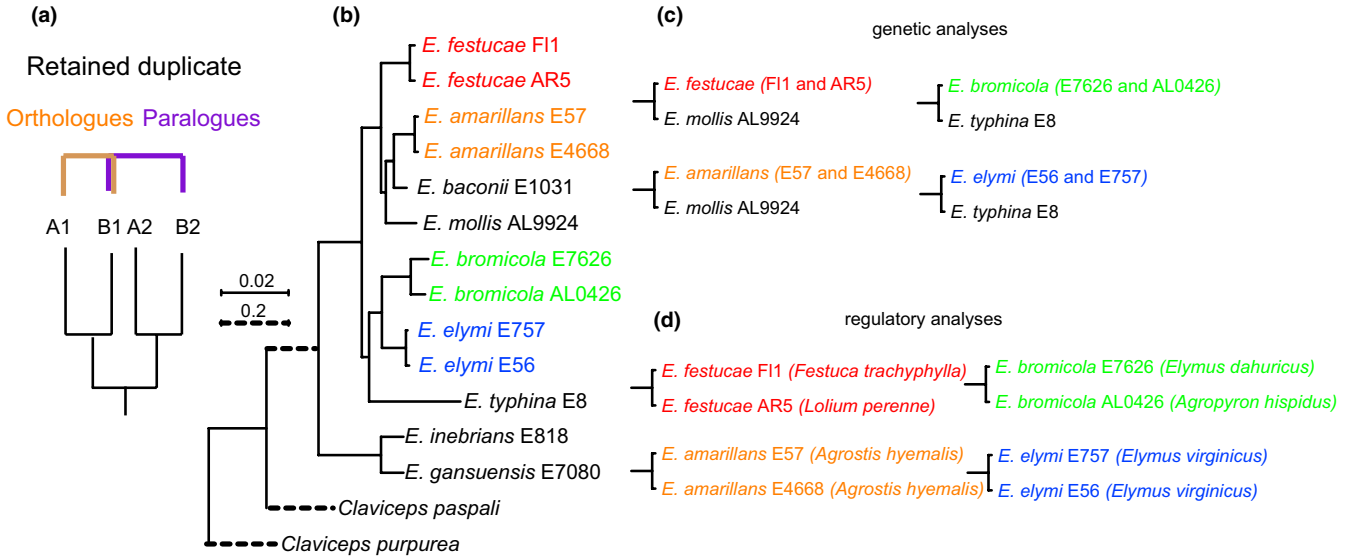

FIGURE 1 Schematic representation of the study design. (a) Explanatory tree of retained duplicate orthologues and retained duplicate paralogues, respectively. A and B are two different species; and A1/B1 and A2/B2 are one-to-one retained duplicate orthologues. The scale bars show the proportion of sequence divergence. (b) Phylogenetic relationships of *Epichloë* species and strains, with *Claviceps* outgroups. All nodes are supported by 100% bootstrap values. Groups of conspecific strains from two major clades are labelled in different sets of colours. (c) For genetic analyses (Ka and Ka/Ks between orthologues from two strains or species), conspecific strains were compared against each other or phylogenetically adjacent species. (d) For regulatory analyses (genetic divergence of upstream regions and expression differences between orthologues from conspecific strains), conspecific strains were compared against each other. Grass host species are indicated in parentheses [Colour figure can be viewed at wileyonlinelibrary.com]

performed on them with an expectation value <0.001. To accommodate fragmented genome references and associated annotation biases, the one-to-one orthologues from retained duplicates were required to have at least one flanking gene shared between species. Clustering of retained duplicates among eight strains was performed with HEATMAP3 (Zhao, Guo, Sheng, & Shyr, 2014) using Euclidean distances. All retained duplicates from the eight conspecific strains are presented in Dataset S1.

2.4 | Ka, Ks and Ka/Ks between retained duplicate orthologues

Codon alignments of each gene were generated using PAL2NAL 14 (Suyama, Torrents, & Bork, 2006). The number of nonsynonymous substitutions (Ka), the number of synonymous substitutions (Ks) and the nonsynonymous-to-synonymous substitution ratio (Ka/Ks) were calculated using the YN00 program in PAML 4.9f (Yang, 2007) with default parameters (icode = 0, weighting = 0, commonf3x4 = 0). To reduce saturation at synonymous sites, we only considered Ks < 3 when calculating Ka/Ks (Yang, 2014). Each copy of a duplicated gene is considered an independent gene with potential functional divergence between each copy (Pegueroles, Laurie, & Alba, 2013).

2.5 | Species/strain Ks

For species/strain Ks values, codon alignments of each SCO were first generated, and then all codon alignments were concatenated as a species/strain alignment. The concatenated alignment was used to calculate pairwise Ks. All other parameters are the same as described above.

2.6 | Ks of paralogues versus Ka and Ka/Ks of retained duplicate orthologues

Each duplicate pair (paralogues) has a Ks value, which reflects the relative divergence time since duplication (Kimura, 1977; Lynch & Conery, 2000) and is widely used to infer the duplication time (Blanc & Wolfe, 2004b; Qiao et al., 2019). Each copy of a duplicated pair has Ka and Ka/Ks values relative to their orthologues in phylogenetically adjacent species (Figure 1). We explored whether these Ka and Ka/Ks values are correlated with the relative age of duplication (i.e., the Ks of paralogues). Simulation studies indicate that the true age falls between the mode and upper standard deviation boundary of the estimated Ks when values are around 3 < Ks < 5 (Vanneste, Van de Peer, & Maere, 2013). Ks values exceeding 7 indicate complete saturation and do not represent true duplication events. Thus, Ks values >5 were excluded from further analysis, as in other studies (Li et al., 2016; Lynch & Conery, 2003; Qiao et al., 2019).

2.7 | Genetic distance in regulatory regions

Pairwise genetic distances of 1-kb regions upstream of the coding sequence were calculated using MEGA-CC 7 (Kumar, Stecher, Peterson, & Tamura, 2012) with the Kimura 2-parameter model. Previous

studies have indicated that transposons can insert in upstream regions (Butelli et al., 2012; Krishnan et al., 2018; Trizzino et al., 2017). To reduce the divergence bias caused by transposon insertions, we also calculated genetic distances after excluding 1-kb regions that contain transposons. We removed these sequences if: (a) two or more transposon copies were detected in a given genome; (b) the similarity between fragments was ≥80%; and (c) the expected value was <0.001.

2.8 | GO enrichment analyses

The Gene Ontology (GO) terms of gene models in each species were annotated using GO FEAT 1 (Araujo, Barh, Silva, Guimaraes, & Ramos, 2018) with an expected value of 10^{-3}. The software AGRIGO 2 (Tian et al., 2017) was used to assess the functional enrichment of retained duplicates, where retained duplicates were used as the query list and all other genes were included in the background list.

2.9 | Expression divergence

Raw transcriptome data (Campbell et al., 2017; Cox et al., 2014; Hassing et al., 2019) were downloaded from the NCBI SRA database (Table S1). The experimental conditions are described more fully in those papers, but in short, culture growth was undertaken on 2.4% potato-dextrose media for strains AR5 and Fl1, while the *in planta* case represents AR5 grown naturally in *Lolium perenne*. Reads were filtered and trimmed using BBDUK 38 (https://sourceforge.net/projects/bbmap/) such that (a) adaptors were removed, (b) reads were quality trimmed using the phred trimming method set to Q10, and (c) reads <50 bp after trimming were removed. Filtered reads from each library were aligned to the corresponding reference genome (AR5 or Fl1) using HISAT 2.2 (Kim, Langmead, & Salzberg, 2015). Based on the mapped region, coverage depth exceeds at least 25× in each sample. Gene counts were generated using FEATURECOUNTS 1.6.3 (Liao, Smyth, & Shi, 2014) using the gff3 annotations created above and mapped bam files. Only uniquely mapped reads were counted. Then the one-to-one orthology relationships were used to merge gene raw counts from AR5 and Fl1 together. Genes with differential expression between two conditions (in culture vs. *in planta*) or two strains (AR5 vs. Fl1) were determined with EDGER 3.24 (Robinson, McCarthy, & Smyth, 2010) and DESEQ2 (Love, Huber, & Anders, 2014) using a false discovery rate (FDR) < 0.05 and fold change ≥2 as cutoff values. These packages take un-normalized read counts as input, but correct for differences in the libraries during model fitting. Only differentially expressed genes (DEGs) identified by both methods were analysed further.

2.10 | Statistical analyses

The differences between Ka, Ka/Ks and genetic distance values were determined by Mann–Whitney *U* tests. The differences between proportions of genes showing differential expression were determined with Fisher's exact test. All correlations were determined using Spearman's correlation. The FDR was controlled using the Benjamini–Hochberg method. All statistics were calculated in R 3.6.1.

3 | RESULTS

We use species in the genus *Epichloë* (Figure 1b), which are ascomycete fungi that form natural and mutualistic endophytic symbioses with grasses, to interrogate the joint roles of sequence evolution and expression regulation in the retention of duplicated genes. Our study assumes that retention and loss can occur between any species, given the potentially fast loss rate of duplicates in fungi (Lynch et al., 2008; Naseeb et al., 2017). We therefore do not assume that all retained duplicates descend from the common ancestor species, although most do; for the purposes of this study, shared duplicates between any two species are treated as retained duplicate genes. In addition, our design analyses four pairs of taxa with very similar evolutionary divergences (Figure 1b), thus providing a natural control for phylogenetic distance. Because each pair is phylogenetically "separate," they largely act as independent test cases. Our aim is to compare the orthologues of retained duplicates with single copy genes (n = 4,178 across the genus) over the same timescale to answer the following two questions: Do retained duplicates have faster nonsynonymous substitution rates indicative of more relaxed selection (Figure 1c)? Do retained duplicates show more diversity of gene regulation indicative of greater transcriptional plasticity (Figure 1d)?

3.1 | Retained duplicates mostly have ancient origins with subsequent losses linked to speciation

Based on phylogenetic and synteny information, 38–76 retained duplications were identified in each genome (Figure 2a; Dataset S1). First, we estimated the synonymous substitution rate (Ks) between copies of retained duplicates (paralogues) in each genome (Figure 2a). Two peaks are apparent, one located around Ks = 2 and the other around Ks = 4, with a greater proportion of retained duplicate genes in the latter. On the basis of these two peaks, we speculate that there may have been two independent duplication events prior to the divergence of *Epichloë* species from their sister clade, *Claviceps* (genus divergence Ks = 0.987; red vertical line in Figure 2a). The key point, however, is that most retained duplicates are old and arose before the diversification of the *Epichloë* clade.

We examined the distribution of these retained duplicates among eight conspecific strains (Figure 2b), with pairs chosen specifically to control for phylogenetic distance. A sporadic pattern of retained duplicates was observed, even between conspecific strains (Figure 2b), similar to that reported for *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* (Ames et al., 2010). Cluster analyses further revealed that loss of gene duplicates is often associated with speciation, which is consistent with previous studies in yeasts (Scannell,

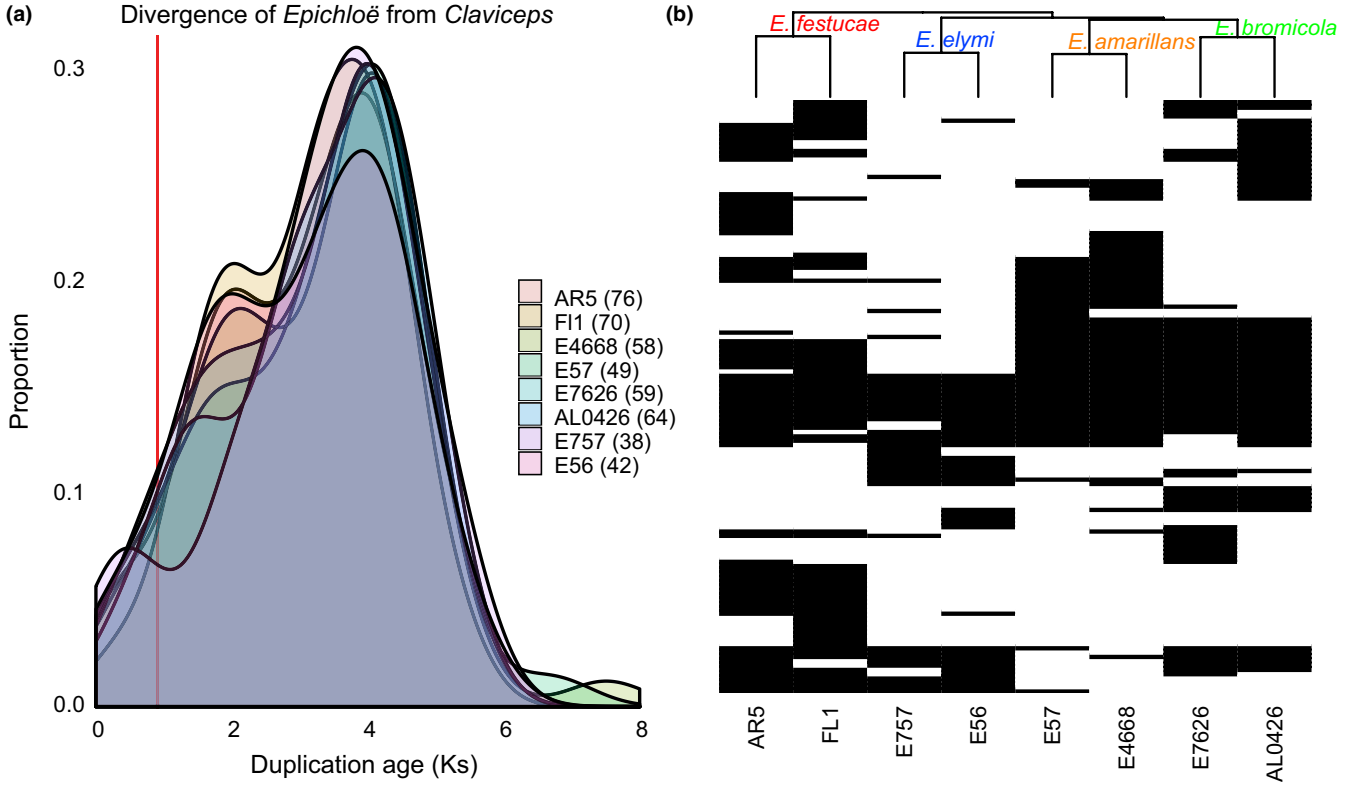

FIGURE 2 Retained duplicate genes among species. (a) Distribution of Ks values for retained duplicate genes (paralogues) in each *Epichloë* species/strain. See Figure 1 for full descriptive names. Values in parentheses indicate the number of retained duplicate genes in each genome. Divergence of the genus *Epichloë* from the genus *Claviceps* is indicated by a vertical red line. (b) The distribution of retained duplicate genes among four species pairs. Black bars represent presence, while white represents absence [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 GO enrichment analysis of retained duplicates in *Epichloë festucae* strain AR5

GO term	Definition	*p* (FDR)
GO:0004672	Protein kinase activity	1.90×10^{-8}
GO:0016772	Transferase activity	6.00×10^{-7}
GO:0016773	Phosphotransferase activity	6.00×10^{-7}
GO:0016301	Kinase activity	4.80×10^{-6}
GO:0022857	Transmembrane transporter activity	.0015
GO:0005215	Transporter activity	.0019
GO:0003824	Catalytic activity	.0027
GO:0004674	Protein serine/threonine kinase activity	.0076

Byrne, Gordon, Wong, & Wolfe, 2006). Functional enrichment finds that retained duplicates have over-represented associations with signal transduction, particularly kinase activity and transmembrane transporter activity (Table 1; Tables S2 and S3). Interestingly, these enriched functions overlap with those of retained duplicate genes in *Arabidopsis*, despite being in different kingdoms (Blanc & Wolfe, 2004a). Species-specific functions were also identified, such as lipid biosynthetic processes in *Epichloë elymi* (Table S3).

3.2 | Retained duplicates have faster evolutionary rates and relaxed selection at the protein level

Previous studies have suggested that nonsynonymous substitutions in duplicated genes are an important reservoir for maintaining genetic robustness (Hickman & Rusche, 2007; Keane et al., 2014). We tested whether duplicate genes retained in *Epichloë* have a higher rate of nonsynonymous substitution (Ka) than single copy genes. As nonsynonymous substitutions change the amino acid sequence of proteins, such substitutions are expected to be a source of functional innovation. Enrichment of nonsynonymous substitutions in retained duplicates compared to single copy genes would support this hypothesis.

In each pairwise comparison, Ka values of retained duplicate orthologues (RDOs) are significantly higher than for SCOs (Figure 3a, Mann–Whitney *U* test), suggesting a faster evolutionary rate at the protein level. Higher Ka values in retained duplicates can even be observed under relatively short-term evolution (Figure 4a), where the Ka values were estimated by comparing conspecific strains. This pattern is more pronounced for strains that are adapted to different hosts (Figure 4b). We further tested whether retained

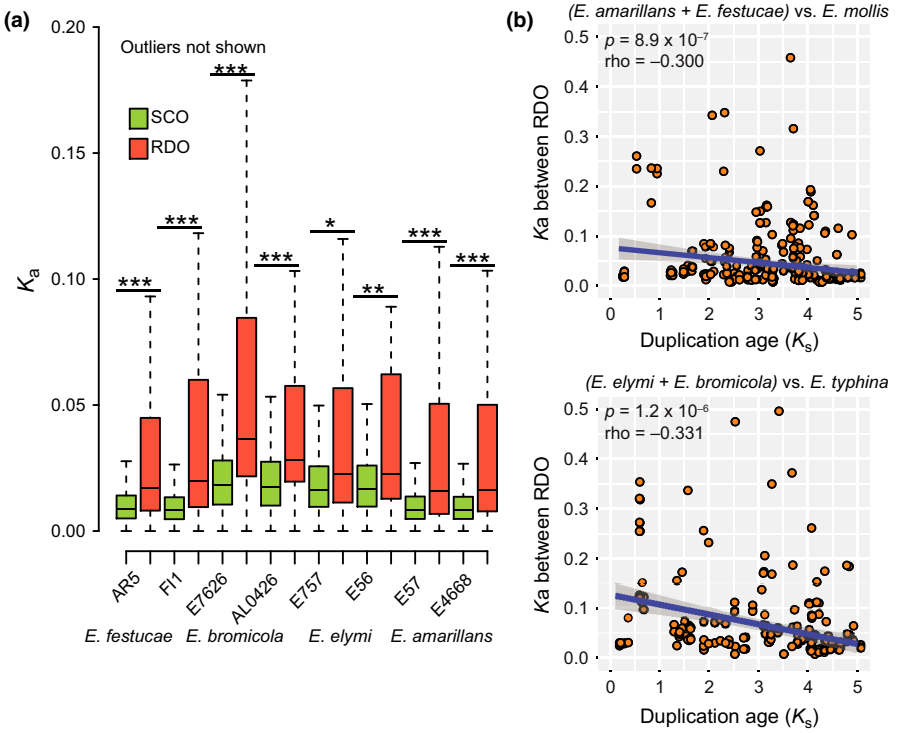

FIGURE 3 Nonsynonymous substitution rate (Ka) in retained duplicates. (a) Ka between retained duplicate orthologues (RDOs) and single copy orthologues (SCOs) in eight pairwise comparisons (detailed in Figure 1b). Outliers were removed from the image for visual clarity, but retained in calculations. Statistical significance was determined with the Mann–Whitney *U* test (*$p \le .05$, **$p \le .001$, and ***$p \le .0001$). (b) Ka against the duplication age of retained duplicate genes (Ks). The duplication age was inferred using the paralogues in one genome; Ka was calculated using the retained duplicate orthologues from species pairs (Figure 1c). Each small panel represents four conspecific strains, with one shared outgroup species. Correlations were estimated using Spearman's ρ [Colour figure can be viewed at wileyonlinelibrary.com]

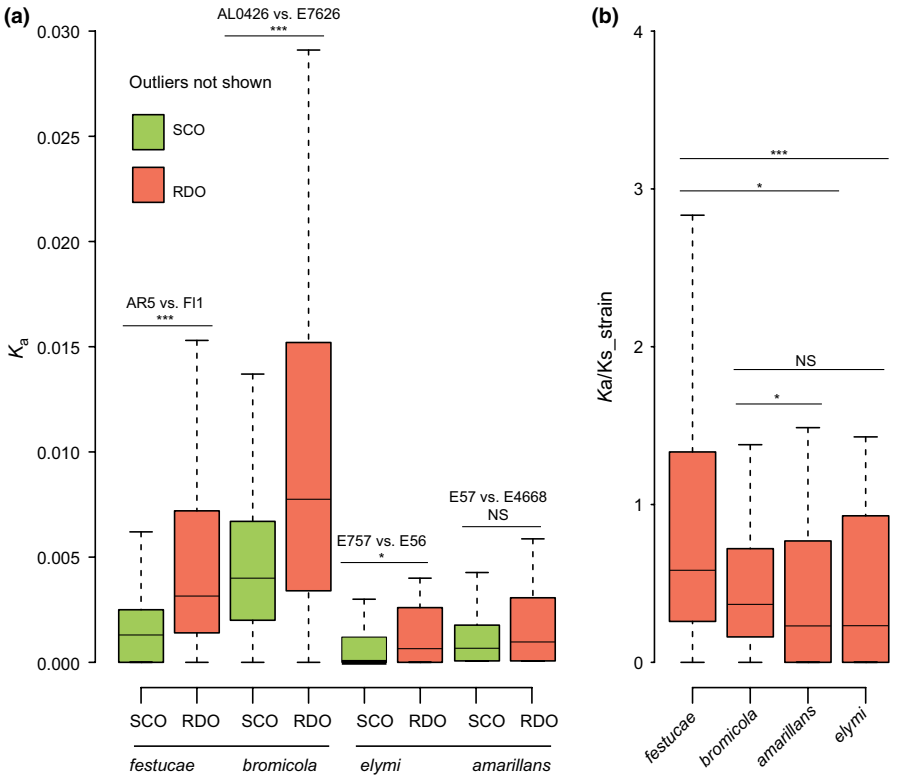

FIGURE 4 Nonsynonymous substitution rate (Ka) in retained duplicated genes between conspecific strains. (a) Ka is higher in retained duplicated genes than in single-copy genes between conspecific strains. RDO, retained duplicate orthologue; SCO, single copy orthologue. Outliers were removed from the image for visual clarity, but retained in calculations. Statistical significance was determined with the Mann–Whitney *U* test (NS indicates not significant, *$p \le .05$, **$p \le .001$ and ***$p \le .0001$). (b) Comparison of Ka between retained duplicates from conspecific-strain pairs with different (*festucae* and *bromicola*) or the same hosts (*amarillans* and *elymi*). Ka was normalized using divergence time between strains (i.e., a phylogenetic correction). Statistical significance was determined with the Mann–Whitney *U* test [Colour figure can be viewed at wileyonlinelibrary.com]

duplicates are under relaxed selective constraint compared to single copy genes by estimating the nonsynonymous-to-synonymous substitution ratio (Ka/Ks). As with Ka, Ka/Ks ratios of duplicated orthologues are greater than for single copy genes in four of eight conspecific strain pairs (Figure S1a, Mann–Whitney *U* test).

We then examined whether Ka and Ka/Ks are associated with the relative age of duplication events. Each pair of duplicates (paralogues) has a Ks value (i.e., age since duplication), which represents their relative divergence time under the assumption that synonymous substitution occurs at a relatively constant rate (Kimura, 1977). Each copy of the duplicated pair has Ka and Ka/Ks values relative to their orthologue in the phylogenetically adjacent species (Figure 1b). Plots of Ka and Ka/Ks values versus age reveal significant negative correlations (Figure 3b; Figure S1b). (Only

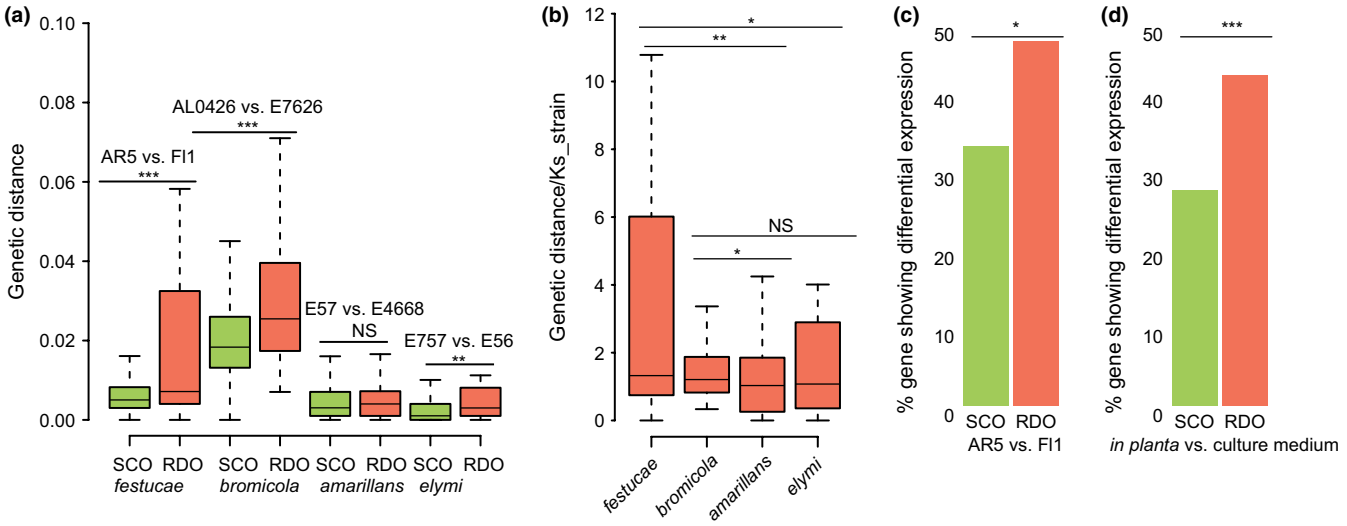

FIGURE 5 Regulatory and expression plasticity of retained duplicate genes. (a) Genetic divergence in 1-kb upstream regions in four pairs of conspecific strains (detailed in Figure 1d). Outliers were removed from the image for visual clarity, but retained in calculations. Statistical significance was determined with the Mann–Whitney *U* test (NS indicates not significant, $^{*}p \leq .05$, $^{**}p \leq .001$ and $^{***}p \leq .0001$). RDO, retained duplicate orthologue; SCO, single copy orthologue. (b) Comparison of genetic distances between retained duplicates from conspecific-strain pairs with different (*festucae* and *bromicola*) or the same hosts (*amarillans* and *elymi*). Genetic distances were normalized using the diverge time between strains (i.e., a phylogenetic correction). Statistical significance was determined with the Mann–Whitney *U* test. (c) The proportion of genes with significant expression differences between *Epichloë festucae* strains AR5 and FI1 growing under in vitro culture conditions. Statistical significance was determined with Fisher's exact test. (d) The proportion of genes with significant expression differences when strain AR5 was grown in culture medium versus *in planta*. Statistical significance was determined by Fisher's exact test [Colour figure can be viewed at wileyonlinelibrary.com]

Ks values <5 were used to minimize the effects of saturation; see Section 2 for details.) These correlations indicate that older retained duplicates have slower nonsynonymous substitution rates, probably reflecting stronger purifying selection over time.

3.3 | Retained duplicates exhibit more diversity in regulatory regions

Functional innovation of duplicated genes can also occur at the level of expression (Panchy, Lehti-Shiu, & Shiu, 2016). Under the hypothesis that retained duplicates have greater expression plasticity, duplicated genes might be expected to show higher diversity in their regulatory regions than single copy genes. The region immediately upstream of the coding sequence (1 kb) is rich in promoter elements in yeast (Dobi & Winston, 2007; Elion & Warner, 1984; Petrascheck et al., 2005; Xue et al., 2017) with similar patterns thought to hold for other fungi (Basse & Farfsing, 2006; Fischer, Durand, & Fevre, 1995). We thus examined the 1-kb upstream regions of retained duplicates and single copy genes, and compared genetic distances between conspecific strains (Figure 1d). We find greater genetic divergence in the upstream regions of retained duplicates than for single copy genes in three different comparisons (Figure 5a, Mann–Whitney *U* test). As with Ka values between conspecific strains (Figure 4b), the divergence of upstream regions is more pronounced when strains have different hosts (Figure 5b). Transposons are also increasingly recognized as key actors in regulating gene expression (Chuong, Elde, & Feschotte, 2017). In the maize genome, thousands of regulatory elements are derived from decayed transposon elements (H. Zhao et al., 2018), and repeat elements show similar, if less pronounced, regulatory effects in *Epichloë* (Winter et al., 2018). To exclude any possible effect of transposons on upstream genetic diversity, we excluded upstream regions that contain repeats and re-estimated genetic divergence; similar patterns were observed (Figure S2).

3.4 | Retained duplicates show greater expression plasticity

While genome-wide transcriptional plasticity between copies of duplicated genes has been reported (Ideker et al., 2001; Stern, Dror, Stolovicki, Brenner, & Braun, 2007), expression patterns between duplicated and single copy genes have not been well investigated. We checked the expression level (fragments per kilobase of transcript per million, FPKM) of retained duplicates in AR5 and FI1, and found that the average expression level of retained duplicates is lower than that of single copy genes (Figure S3). Based on observations of diversity in regulatory regions between conspecific strains, we expected that expression differences between retained duplicates will be greater than for single copy genes in most conspecific strain comparisons. To test this hypothesis, we compared the expression patterns of retained duplicates and single copy genes between two conspecific strains (AR5 vs. FI1) growing under in vitro culture conditions (Cox et al., 2014; Eaton et al., 2015; Hassing et al., 2019), where environmental variation is naturally controlled for because both the duplicated and the single copy genes are present in exactly the same cells. The caveat

with mapping reads to duplicated genes is that some reads may map to the wrong gene copy. To minimize this bias, only uniquely mapped reads were counted for differential expression analysis. Consistent with expectation, we find that duplicated genes are more often differentially expressed than single copy genes (Figure 5c, Fisher's exact test). We also find statistically significant positive correlations between sequence divergence in regulatory regions and gene expression divergence (Figure S4; Spearman's $\rho = 0.085$; $p = 3.23 \times 10^{-5}$), which further supports the role of genetic divergence on expression plasticity. Further, many SCOs between conspecific strains have little genetic divergence (close to zero) in both genes and promotor regions but exhibit large fold change in expression (Figure S4), suggesting that the 5′ upstream region is not the only factor determining expression. Elsewhere, *cis*-elements in the 3′ untranslated region are known to regulate transcript stability and abundance (Graber, 2003). Finally, we compared the expression patterns of retained duplicate and single copy genes from strain AR5 between culture and *in planta* (*Lolium perenne*) conditions, and found that a significantly higher proportion of retained duplicate genes show a significant difference in expression (Figure 5d, Fisher's exact test).

4 | DISCUSSION

4.1 | Nonsynonymous substitution is pervasive in retained duplicates

Gene duplication is a major force driving evolution, allowing genes, and the organisms they reside in, to acquire new functions. In fungi, retained duplicated genes are a widespread phenomenon (Cornell et al., 2007; Corrochano et al., 2016; Ma et al., 2009; Scannell et al., 2006), but few studies have addressed the reasons for their retention. A mutation accumulation experiment in *Saccharomyces cerevisiae* captured the evolution of retained duplicates over a very short timescale. In particular, they noted that nonsynonymous substitutions were more prone to occur in duplicated genes than single copy genes during a 2200-generation culture experiment (Keane et al., 2014). Our results suggest that retained duplicates are also subject to faster nonsynonymous substitution rates over much longer timescales (higher Ka in Figures 3a and 4a). In addition, the higher Ka/Ks values of retained duplicates versus single copy genes (Figure S1a) suggests that selection against nonsynonymous substitutions is relaxed, possibly enabling a longer-term presence of nonsynonymous substitutions in duplicated genes. Together, both short- and long-term evolution studies highlight the importance of nonsynonymous substitutions and higher diversity in retained duplicate genes.

These nonsynonymous substitutions emerge as random de novo mutations, but may be preferentially retained in response to environmental pressures. As most nonsynonymous substitutions are deleterious or neutral, the organism bears some burden in hosting this diversity. However, at least in *S. cerevisiae*, populations with the most substitutions adapted fastest to new environments (Li et al., 2019), suggesting that standing variation may be an important source of pre-adaptation in fungi. Alternatively, high levels of Ka and Ka/Ks in duplicated genes may not be the sole factor driving retention of duplicated genes, but may also be a predictor of duplicability (O'Toole, Hurst, & McLysaght, 2018). In other words, genes that are more robust to functional innovation may be preferentially enriched as duplicated genes, perhaps explaining the prevalence of multiple-copy gene families.

4.2 | Host selection shapes diversity in coding and regulatory regions of retained duplicates

Mutation accumulation studies and comparative genomics in *S. cerevisiae* found that duplicated genes tend to fall within mutational hotspots in the genome, both in coding and in flanking regions (Fares et al., 2017; Keane et al., 2014). Here, we observe that coding regions and *cis*-regulatory regions of retained duplicate genes are more diverse (higher Ka or genetic distance) than those of single copy genes (Figures 4a and 5a). Among the four pairs of conspecific strains in our study, two groups have different grass hosts (Figure 1d). Coding regions and upstream sequences of duplicated genes in these paired groups with different hosts have higher Ka and genetic distances, respectively, suggesting that host pressure may in part drive changes in these regions (Figures 4b and 5b). Because these changes are observed even over short time divergences between conspecific strains, they may be an important feature of ongoing host specificity and host switches.

4.3 | Different fates of ancient duplicates

Significantly negative correlations are observed between the age of duplicated genes and their evolutionary rates of change (Ka or Ka/Ks; Figure 3b; Figure S1b). Two possible explanations are proposed for these patterns. If ancient gene duplicates did not acquire new functions (neofunctionalization or subfunctionalization), these negative correlations suggest that older duplicates lose their innovation capacity at the sequence coding level. Possibly the organism has lower tolerance of changes at nonsynonymous sites, which reduce survival rates due to the fact that many nonsynonymous substitutions are slightly deleterious. We speculate that copies of these duplicated genes might preferentially be lost. In contrast, where ancient duplicates evolved new or variant functions, they may maintain their new functions through strong purifying selection, which is reflected by lower Ka and Ka/Ks. In these circumstances, the two copies of the duplicated gene will be retained for an extended time.

4.4 | Dosage balance may drive duplicated gene retention

Changing the expression levels of duplicated genes may be important for dosage balance. In mammals, increasing gene dosage has deleterious effects, and many duplicated genes thus reduce transcript levels to the one-copy baseline (Antonarakis et al., 2004). In mammals, most young duplicates appear to be down-regulated to match the expression levels of single-copy genes and to allow the survival of both copies (Lan & Pritchard, 2016). The expression level

patterns that we observe also show lower expression in retained duplicates relative to single copy genes (Figure S3). Thus, expression changes to manage dosage balance may be another important force dictating the retention of duplicated genes for an extended time in *Epichloë*.

4.5 | Expression plasticity may be a source of pre-adaptation

A key feature of ancient duplicated genes in *S. cerevisiae* is their expression plasticity (Keane et al., 2014; Mattenberger, Sabater-Munoz, Toft, & Fares, 2017; Mattenberger, Sabater-Munoz, Toft, Sablok, & Fares, 2017), where duplicated genes more often exhibit expression differences under a range of stress conditions relative to single copy genes. We compared gene expression for *Epichloë festucae* strain AR5 growing under standard in vitro culture and *in planta* conditions. As with *S. cerevisiae*, retained duplicates in *Epichloë* exhibit greater transcriptional change than single copy genes (Figure 5d). Transcriptional plasticity of retained duplicates was also observed when *Epichloë* species face new environments. Retained duplicates are more often differentially expressed than single copy genes between *E. festucae* AR5 and Fl1 when grown under unnatural in vitro culture conditions, an environment quite different to that of the grass host. These expression differences may not necessarily reflect immediate functional diversity, such as differential responses to hosts, but could potentially act as a source of pre-adaptation to new hosts.

5 | CONCLUSIONS

Here, we address a long-standing evolutionary question: why do some duplicated genes persist in the genome for an extended time? We tackle this question using the nonmodel fungal genus *Epichloë*, as a counterpoint to the extensive studies performed in yeasts. Analyses of eight strains of *Epichloë* from four species revealed 38–76 retained duplicates in each genome, most of which pre-date the origin and diversification of the genus *Epichloë*. Within the genus *Epichloë*, frequent loss of retained duplicates is closely associated with speciation. Retained duplicates exhibit faster substitution rates and relaxed selection at the protein level, and both are negatively correlated with their relative age since duplication. In upstream noncoding sequences, retained duplicates exhibit more diversity in *cis*-regulatory regions, which confers greater expression plasticity, as observed directly by transcript levels. Substitution in both coding regions and upstream noncoding sequences appears to be shaped by host pressures. Our results thus present evidence that genetic robustness and regulatory plasticity are two driving forces leading to the retention of duplicated genes in diverse fungi, *Saccharomyces* and *Epichloë*. The lessons learned here from a nonmodel fungus thus offer insight into the evolution of retained duplicates across the fungal kingdom.

ACKNOWLEDGEMENTS

We thank three anonymous reviewers for helpful comments, and David Winter, Kate Lee and Weilong Hao for expert advice. This research was supported by a Marsden grant (14-MAU-007) from the Royal Society of New Zealand and by the Tertiary Education Commission via a Bio-Protection Research Centre grant, both to M.P.C.

AUTHOR CONTRIBUTIONS

B.W. and M.P.C. designed the study and wrote the manuscript. B.W. performed the analyses.

ORCID

Murray P. Cox https://orcid.org/0000-0003-1936-0236

DATA AVAILABILITY STATEMENT

All genomic and RNA-seq data used in this manuscript are publicly available. Access details are listed in Table S1. GO terms and GO enrichment metrics of retained duplicates are listed in Tables S2 and S3 respectively. The sequences of retained duplicates from the eight conspecific strains are presented in Dataset S1.

REFERENCES

Ames, R. M., Rash, B. M., Hentges, K. E., Robertson, D. L., Delneri, D., & Lovell, S. C. (2010). Gene duplication and environmental adaptation within yeast populations. *Genome Biology and Evolution*, *2*, 591–601. https://doi.org/10.1093/gbe/evq043

Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A., & Deutsch, S. (2004). Chromosome 21 and Down syndrome: From genomics to pathophysiology. *Nature Reviews Genetics*, *5*(10), 725–738. https://doi.org/10.1038/nrg1448

Araujo, F. A., Barh, D., Silva, A., Guimaraes, L., & Ramos, R. T. J. (2018). GO FEAT: A rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific Reports*, *8*(1), 1794. https://doi.org/10.1038/s41598-018-20211-9

Assis, R., & Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(43), 17409–17414. https://doi.org/10.1073/pnas.1313759110

Assis, R., & Bachtrog, D. (2015). Rapid divergence and diversification of mammalian duplicate gene functions. *BMC Evolutionary Biology*, *15*, 138. https://doi.org/10.1186/s12862-015-0426-x

Basse, C. W., & Farfsing, J. W. (2006). Promoters and their regulation in *Ustilago maydis* and other phytopathogenic fungi. *FEMS Microbiology Letters*, *254*(2), 208–216. https://doi.org/10.1111/j.1574-6968.2005.00046.x

Blanc, G., & Wolfe, K. H. (2004a). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell*, *16*(7), 1679–1691. https://doi.org/10.1105/tpc.021410

Blanc, G., & Wolfe, K. H. (2004b). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell*, *16*(7), 1667–1678. https://doi.org/10.1105/tpc.021345

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., … Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell*, *24*(3), 1242–1255. https://doi.org/10.1105/tpc.111.095232

Campbell, M. A., Tapper, B. A., Simpson, W. R., Johnson, R. D., Mace, W., Ram, A., … Cox, M. P. (2017). *Epichloë hybrida*, sp. nov., an emerging model system for investigating fungal allopolyploidy. *Mycologia*, *109*(5), 715–729. https://doi.org/10.1080/00275514.2017.1406174

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, *18*(2), 71–86. https://doi.org/10.1038/nrg.2016.139

Conant, G. C., & Wagner, A. (2003). Asymmetric sequence divergence of duplicate genes. *Genome Research*, *13*(9), 2052–2058. https://doi.org/10.1101/gr.1252603

Cornell, M. J., Alam, I., Soanes, D. M., Wong, H. M., Hedeler, C., Paton, N. W., … Oliver, S. G. (2007). Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the fungi. *Genome Research*, *17*(12), 1809–1822. https://doi.org/10.1101/gr.6531807

Corrochano, L. M., Kuo, A., Marcet-Houben, M., Polaino, S., Salamov, A., Villalobos-Escobedo, J. M., … Grigoriev, I. V. (2016). Expansion of signal transduction pathways in fungi by extensive genome duplication. *Current Biology*, *26*(12), 1577–1584. https://doi.org/10.1016/j.cub.2016.04.038

Cox, M. P., Dong, T., Shen, G., Dalvi, Y., Scott, D. B., & Ganley, A. R. (2014). An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLoS Genetics*, *10*(3), e1004180. https://doi.org/10.1371/journal.pgen.1004180

Dobi, K. C., & Winston, F. (2007). Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, *27*(15), 5575–5586. https://doi.org/10.1128/MCB.00459-07

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Eaton, C. J., Dupont, P. Y., Solomon, P., Clayton, W., Scott, B., & Cox, M. P. (2015). A core gene set describes the molecular basis of mutualism and antagonism in *Epichloë* spp. *Molecular Plant-Microbe Interactions*, *28*(3), 218–231. https://doi.org/10.1094/MPMI-09-14-0293-FI

Elion, E. A., & Warner, J. R. (1984). The major promoter element of rRNA transcription in yeast lies 2 kb upstream. *Cell*, *39*(3 Pt 2), 663–673. https://doi.org/10.1016/0092-8674(84)90473-2

Emms, D. M., & Kelly, S. (2018). OrthoFinder2: Fast and accurate phylogenomic orthology analysis from gene sequences. *BioRxiv*, https://doi.org/10.1101/466201

Fares, M. A., Sabater-Munoz, B., & Toft, C. (2017). Genome mutational and transcriptional hotspots are traps for duplicated genes and sources of adaptations. *Genome Biology and Evolution*, *9*(5), 1229–1240. https://doi.org/10.1093/gbe/evx085

Fischer, M., Durand, R., & Fevre, M. (1995). Characterization of the "promoter region" of the enolase-encoding gene enol from the anaerobic fungus *Neocallimastix frontalis*: Sequence and promoter analysis. *Current Genetics*, *28*(1), 80–86. https://doi.org/10.1007/BF00311885

Graber, J. H. (2003). Variations in yeast 3′-processing cis-elements correlate with transcript stability. *Trends in Genetics*, *19*(9), 473–476. https://doi.org/10.1016/S0168-9525(03)00196-3

Gu, Z., David, L., Petrov, D., Jones, T., Davis, R. W., & Steinmetz, L. M. (2005). Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1092–1097. https://doi.org/10.1073/pnas.0409159102

Guan, Y., Dunham, M. J., & Troyanskaya, O. G. (2007). Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics*, *175*(2), 933–943. https://doi.org/10.1534/genetics.106.064329

Guschanski, K., Warnefors, M., & Kaessmann, H. (2017). The evolution of duplicate gene expression in mammalian organs. *Genome Research*, *27*(9), 1461–1474. https://doi.org/10.1101/gr.215566.116

Hassing, B., Winter, D., Becker, Y., Mesarich, C. H., Eaton, C. J., & Scott, B. (2019). Analysis of *Epichloë festucae* small secreted proteins in the interaction with *Lolium perenne*. *PLoS ONE*, *14*(2), e0209463. https://doi.org/10.1371/journal.pone.0209463

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, *32*(4), 835–845. https://doi.org/10.1093/molbev/msv037

Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., … Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nature Genetics*, *41*(4), 424–429. https://doi.org/10.1038/ng.345

Hickman, M. A., & Rusche, L. N. (2007). Substitution as a mechanism for genetic robustness: The duplicated deacetylases Hst1p and Sir2p in *Saccharomyces cerevisiae*. *PLoS Genetics*, *3*(8), e126. https://doi.org/10.1371/journal.pgen.0030126

Hoegg, S., Brinkmann, H., Taylor, J. S., & Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *Journal of Molecular Evolution*, *59*(2), 190–203. https://doi.org/10.1007/s00239-004-2613-z

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769. https://doi.org/10.1093/bioinformatics/btv661

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Reviews Genetics*, *2*(7), 516–527. https://doi.org/10.1038/35080508

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., … Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, *292*(5518), 929–934. https://doi.org/10.1126/science.292.5518.929

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: Classifying and distinguishing between models. *Nature Reviews Genetics*, *11*(2), 97–108. https://doi.org/10.1038/nrg2689

Inoue, J., Sato, Y., Sinclair, R., Tsukamoto, K., & Nishida, M. (2015). Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14918–14923. https://doi.org/10.1073/pnas.1507669112

Keane, O. M., Toft, C., Carretero-Paulet, L., Jones, G. W., & Fares, M. A. (2014). Preservation of genetic and regulatory robustness in ancient gene duplicates of *Saccharomyces cerevisiae*. *Genome Research*, *24*(11), 1830–1841. https://doi.org/10.1101/gr.176792.114

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, *428*(6983), 617–624. https://doi.org/10.1038/nature02424

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, *267*(5608), 275–276.

Konrad, A., Flibotte, S., Taylor, J., Waterston, R. H., Moerman, D. G., Bergthorsson, U., & Katju, V. (2018). Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. *Proceedings of the National Academy of*

Sciences of the United States of America, *115*(28), 7386–7391. https://doi.org/10.1073/pnas.1801930115

Krishnan, P., Meile, L., Plissonneau, C., Ma, X., Hartmann, F. E., Croll, D., … Sánchez-Vallet, A. (2018). Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC Biology*, *16*(1), 78. https://doi.org/10.1186/s12915-018-0543-2

Kumar, S., Stecher, G., Peterson, D., & Tamura, K. (2012). MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*, *28*(20), 2685–2686. https://doi.org/10.1093/bioinformatics/bts507

Lan, X., & Pritchard, J. K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*, *352*(6288), 1009–1013. https://doi.org/10.1126/science.aad8411

Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research*, *12*(7), 1048–1059. https://doi.org/10.1101/gr.174302

Li, J., Vázquez-García, I., Persson, K., González, A., Yue, J.-X., Barré, B., … Liti, G. (2019). Shared molecular targets confer resistance over short and long evolutionary timescales. *Molecular Biology and Evolution*, *36*(4), 691–708. https://doi.org/10.1093/molbev/msz006

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., & De Smet, R. (2016). Gene duplicability of core genes Is highly consistent across all angiosperms. *The Plant Cell*, *28*(2), 326–344. https://doi.org/10.1105/tpc.15.00877

Liao, Y., Smyth, G. K., & Shi, W. (2014). FEATURECOUNTS: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Loehlin, D. W., & Carroll, S. B. (2016). Expression of tandem gene duplicates is often greater than twofold. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5988–5992. https://doi.org/10.1073/pnas.1605886113

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *290*(5494), 1151–1155.

Lynch, M., & Conery, J. S. (2003). The evolutionary demography of duplicate genes. *Journal of Structural and Functional Genomics*, *3*(1–4), 35–44.

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., … Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9272–9277. https://doi.org/10.1073/pnas.0803466105

Ma, L.-J., Ibrahim, A. S., Skory, C., Grabherr, M. G., Burger, G., Butler, M., … Wickes, B. L. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genetics*, *5*(7), e1000549. https://doi.org/10.1371/journal.pgen.1000549

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., & Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5454–5459. https://doi.org/10.1073/pnas.0501102102

Mattenberger, F., Sabater-Munoz, B., Toft, C., & Fares, M. A. (2017). The phenotypic plasticity of duplicated genes in *Saccharomyces cerevisiae* and the origin of adaptations. *G3 (Bethesda)*, *7*(1), 63–75. https://doi.org/10.1534/g3.116.035329

Mattenberger, F., Sabater-Munoz, B., Toft, C., Sablok, G., & Fares, M. A. (2017). Expression properties exhibit correlated patterns with the fate of duplicated genes, their divergence, and transcriptional plasticity in Saccharomycotina. *DNA Research*, *24*(6), 559–570. https://doi.org/10.1093/dnares/dsx025

Naseeb, S., Ames, R. M., Delneri, D., & Lovell, S. C. (2017). Rapid functional and evolutionary changes follow gene duplication in yeast. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1861), https://doi.org/10.1098/rspb.2017.1393

Ohno, S. (1970). *Evolution by gene duplication*. London, UK: Allen & Unwin.

Ohno, S. (1999). Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Seminars in Cell & Developmental Biology*, *10*(5), 517–522. https://doi.org/10.1006/scdb.1999.0332

O'Toole, A. N., Hurst, L. D., & McLysaght, A. (2018). Faster evolving primate genes are more lkely to duplicate. *Molecular Biology and Evolution*, *35*(1), 107–118. https://doi.org/10.1093/molbev/msx270

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*, 401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, *171*(4), 2294–2316. https://doi.org/10.1104/pp.16.00523

Pegueroles, C., Laurie, S., & Alba, M. M. (2013). Accelerated evolution after gene duplication: A time-dependent process affecting just one copy. *Molecular Biology and Evolution*, *30*(8), 1830–1842. https://doi.org/10.1093/molbev/mst083

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C. P., Schaffner, W., & Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Research*, *33*(12), 3743–3750. https://doi.org/10.1093/nar/gki689

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FASTTREE 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., … Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology*, *20*, 38. https://doi.org/10.1186/s13059-019-1650-2

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). EDGER: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rogers, R. L., Shao, L., & Thornton, K. R. (2017). Tandem duplications lead to novel expression patterns through exon shuffling in *Drosophila yakuba*. *PLoS Genetics*, *13*(5), e1006795. https://doi.org/10.1371/journal.pgen.1006795

Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., & Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, *440*(7082), 341–345. https://doi.org/10.1038/nature04562

Soria, P. S., McGary, K. L., & Rokas, A. (2014). Functional divergence for every paralog. *Molecular Biology and Evolution*, *31*(4), 984–992. https://doi.org/10.1093/molbev/msu050

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server), W435–W439. https://doi.org/10.1093/nar/gkl200

Stern, S., Dror, T., Stolovicki, E., Brenner, N., & Braun, E. (2007). Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Molecular Systems Biology*, *3*, 106. https://doi.org/10.1038/msb4100147

Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *Journal of Molecular Evolution*, *49*(2), 169–181. https://doi.org/10.1007/PL00006540

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*, W609–W612. https://doi.org/10.1093/nar/gkl315

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., & Su, Z. (2017). AGRIGO v2.0: A GO analysis toolkit for the agricultural community, 2017

update. *Nucleic Acids Research*, *45*(W1), W122–W129. https://doi.org/10.1093/nar/gkx382

Trizzino, M., Park, Y. S., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., … Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Research*, *27*(10), 1623–1633. https://doi.org/10.1101/gr.218149.116

Vanneste, K., Van de Peer, Y., & Maere, S. (2013). Inference of genome duplications from age distributions revisited. *Molecular Biology and Evolution*, *30*(1), 177–190. https://doi.org/10.1093/molbev/mss214

Wagner, A. (2002). Selection and gene duplication: A view from the genome. *Genome Biology*, *3*(5), reviews1012.

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

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, *387*(6634), 708–713. https://doi.org/10.1038/42711

Woodhouse, M. R., Pedersen, B., & Freeling, M. (2010). Transposed genes in *Arabidopsis* are often associated with flanking repeats. *PLoS Genetics*, *6*(5), e1000949. https://doi.org/10.1371/journal.pgen.1000949

Xue, Y., Pradhan, S. K., Sun, F., Chronis, C., Tran, N., Su, T., … Carey, M. F. (2017). Mot 1, Ino80C, and NC2 function coordinately to regulate pervasive transcription in yeast and mammals. *Molecular Cell*, *67*(4), 594–607.e4. https://doi.org/10.1016/j.molcel.2017.06.029

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z. (2014). *Molecular evolution: A statistical approach*. Oxford, UK: Oxford University Press.

Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A. P., Wu, Y., & Jiang, J. (2018). Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiology*, *176*(4), 2789–2803. https://doi.org/10.1104/pp.17.01467

Zhao, S., Guo, Y., Sheng, Q., & Shyr, Y. (2014). Advanced heat map and clustering analysis using heatmap3. *BioMed Research International*, *2014*, 986048. https://doi.org/10.1155/2014/986048

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wu B, Cox MP. Greater genetic and regulatory plasticity of retained duplicates in *Epichloë* endophytic fungi. *Mol Ecol*. 2019;28:5103–5114. https://doi.org/10.1111/mec.15275