HAL
archives-ouvertes.fr

# Differential gene expression between fungal mating types is associated with sequence degeneration Running title: Differential gene expression correlates with sequence degeneration

Wen-Juan Ma, Fantin Carpentier, Tatiana Giraud, Michael Hood

▶ To cite this version:

Wen-Juan Ma, Fantin Carpentier, Tatiana Giraud, Michael Hood. Differential gene expression between fungal mating types is associated with sequence degeneration Running title: Differential gene expression correlates with sequence degeneration. Genome Biology and Evolution, Society for Molecular Biology and Evolution, 2020, 10.1093/gbe/evaa028/5736135 . hal-02503448

HAL Id: hal-02503448
https://hal.archives-ouvertes.fr/hal-02503448

Submitted on 10 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Differential gene expression between fungal mating types is associated with sequence degeneration

Wen-Juan Ma[1*], Fantin Carpentier[2], Tatiana Giraud[2], Michael Hood[1*]

1. Department of Biology, Amherst College, Amherst, 01002 Massachusetts, USA

2. Ecologie Systematique Evolution, Batiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400, Orsay, France

* Author for Correspondence:
Wen-Juan Ma: Department of Biology, Amherst College, Amherst, 01002 Massachusetts, USA. Tel&Fax: +1 413 542 8538, E-mail: wenjuanma84@gmail.com,

Michael Hood: Department of Biology, Amherst College, Amherst, 01002 Massachusetts, USA. Tel&Fax: +1 413 542 8538, E-mail: mhood@amherst.edu

**Running title:** Differential gene expression correlates with sequence degeneration

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

## Abstract (249 words)

Degenerative mutations in non-recombining regions, such as on sex chromosomes, may lead to differential expression between alleles if mutations occur stochastically in one or the other allele. Reduced allelic expression due to degeneration has indeed been suggested to occur in various sex-chromosome systems. However, whether an association occurs between specific signatures of degeneration and differential expression between alleles has not been extensively tested, and sexual antagonism can also cause differential expression on sex chromosomes. The anther-smut fungus *Microbotryum lychnidis-dioicae* is ideal for testing associations between specific degenerative signatures and differential expression because: 1) there are multiple evolutionary strata on the mating-type chromosomes, reflecting successive recombination suppression linked to mating-type loci, 2) separate haploid cultures of opposite mating types help identify differential expression between alleles, and 3) there is no sexual antagonism as a confounding factor accounting for differential expression. We found that differentially-expressed genes were enriched in the four oldest evolutionary strata compared to other genomic compartments, and that, within compartments, several signatures of sequence degeneration were greater for differentially-expressed than non-differentially expressed genes. Two particular degenerative signatures were significantly associated with lower expression levels within differentially-expressed allele pairs: upstream insertion of transposable elements and mutations truncating the protein length. Other degenerative mutations associated with differential expression included non-synonymous substitutions and altered intron or GC content. The association between differential expression and allele degeneration is relevant for a broad range of taxa where mating compatibility or sex is determined by genes located in large regions where recombination is suppressed.

**Key words**: mating-type chromosomes, differential gene expression, sequence degeneration, transposable elements, premature stop codon, sexual antagonism

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

## Introduction

In plants and animals with differentiated sex chromosomes, recombination is largely suppressed in the sex chromosome that is always in a heterozygous state, i.e., the Y or the W chromosome (Bachtrog et al. 2014; Charlesworth et al. 2005; Muyle et al. 2017; Wright et al. 2016). A consequence of such a lack of recombination is the diminished efficacy of selection, allowing deleterious mutations to accumulate (Bachtrog 2006, 2008; Charlesworth 1991; Rice 1996; Wright et al. 2016). The absence of recombination indeed reduces the effective population size, promotes the genetic hitchhiking of deleterious mutations with beneficial ones (Rice 1987), and prevents the purging of deleterious mutations (Bachtrog 2005; Rice 1996; Wright et al. 2016). Degenerative changes in the non-recombining regions include transposable element insertions, mutations causing alteration of protein function such as early stop codons, and even gene copy loss. Such degenerative signatures are commonly found on the non-recombining Y or W chromosomes in many plants and animals (Bachtrog 2008; Charlesworth 2002; Kaiser et al. 2017; Muyle et al. 2017).

Degenerative changes have, in particular, the potential to lead to suboptimal gene expression. Indeed, reduced allelic expression has been reported as a form of degeneration on sex chromosomes (Bachtrog et al. 2008; Bachtrog 2013; Hough et al. 2014; Konuma et al. 2013; Pucholt et al. 2017; White et al. 2015; Xu et al. 2019), although most prior studies have focused on altered expression levels in relation to dosage compensation for differences in gene copy number on sex chromosomes between sexes (Charlesworth 1996; Darolti et al. 2019; Mank 2009, 2013; Mank et al. 2011; Ohno 1966). Studies that test whether various signatures of sequence degeneration are associated with differences in expression between alleles are generally lacking. Furthermore, sexually antagonistic selection is an alternative and frequently cited cause to explain differential expression on sex chromosomes, and can render differences in gene expression between alleles challenging to interpret in the context of

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

degenerative mutations (Cannallon & Knowles 2005; Ellegren & Parsch 2007; Parsch & Ellegren 2013; Rice 1987).

Testing the association between allelic expression levels and specific degenerative signatures, particularly in systems without sexually antagonistic selection as a confounding factor, could yield insights into the evolution of gene expression in non-recombining regions. For instance, an early stop codon in an allele that truncates protein length can lead to post-transcriptional regulatory negative feedbacks upon expression (e.g. nonsense mediated decay; Montgomery et al. 2013). Transposable element insertion in upstream promoter regions, or internal to genes, has long been recognized for effects on expression (Cordaux & Batzer 2009; Feschotte 2008; Lee & Young 2013; McClintock 1942; Tirosh et al. 2009). Although less well recognized, base pair substitutions and in-frame indels (insertion or deletion mutations) can cause changes in amino-acid sequence that affect gene expression through modulation of the mRNA translation (Kimball & Jefferson 2004), or disrupt promoter regions that impact transcriptional regulation (Wray et al. 2003). Epigenetic modifications, particularly cytosine methylation, contribute both to heterochromatin formation and elevated mutation rates that reduce GC content (Bird 1980; Grummt and Pikaard 2003); thus reduced GC content could represent a signature of methylation-induced gene silencing, among multiple other factors (Galtier et al. 2001; Meunier and Duret, 2004). Shorter introns are more efficient for correcting transcription (Marais et al. 2005), such that changes in introns can influence transcription rates, nuclear export, and transcript stability (Heyn et al. 2015). These forms of degenerative changes are expected to accumulate under the reduced selection efficacy in non-recombining regions, and differential expression may occur where the mutations by chance affect one allele more than the other.

Fungal mating-type chromosomes share many features with sex chromosomes (Fraser & Heitman 2004; Hood et al. 2004), and can provide valuable insights into the relationship

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

between various degenerative signatures and differential gene expression in non-recombining regions. Particular benefits of many fungi relative to other types of organisms are that antagonistic selection is not a confounding factor, an easy access to the haploid phase where alternate mating types are expressed, and the existence of young events of recombination suppression in successive evolutionary strata in linkage to the mating-type loci (Bazzicalupo et al. 2019; Branco et al. 2017; Branco et al. 2018; Fontanillas et al. 2015; Giraud et al. 2008). The anther-smut fungi, in the genus *Microbotryum*, undergo mating in the haploid phase via isogamous yeast-like cells of opposite mating types ($a_1$ and $a_2$), which can be cultured separately to analyze expression levels of alleles (Perlin et al. 2015). The species *Microbotryum lychnidis-dioicae,* causing anther-smut disease on the plant *Silene latifolia*, carries dimorphic mating-type chromosomes that have been assembled at the chromosome-level scale (Branco et al. 2017; Hood 2002; Hood et al. 2013). These mating-type chromosomes ($a_1$, ~3.3Mb, and $a_2$, ~4.0Mb, respectively) lack recombination across 90% of their length (Hood 2002; Hood et al. 2013), and are enriched in signatures of sequence degeneration compared to autosomes (Fontanillas et al. 2015). Because mating type is determined at the haploid phase, both mating-type chromosomes are always heterozygous and non-recombining, so that both degenerate (Fontanillas et al. 2015). Importantly, evolutionary strata of different ages have been identified, i.e., regions with different levels of differentiation between mating types as a result of an expanding process of recombination suppression over the past 1.5 million years (Branco et al. 2017, 2018). The non-recombining regions of the mating-type chromosomes in *M. lychnidis-dioicae* are flanked by small recombining pseudo-autosomal regions (PARs).

In the isogamous fungus *M. lychnidis-dioicae*, there is no male or female function, so that there cannot be any sexual antagonism. Any analogous 'mating-type antagonistic selection' (*sensu* Abbate and Hood 2010) would require fitness differences associated with mating-type

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

dimorphic traits, which are necessarily expressed at the haploid stage when cells are of different mating types. However, there is only a very brief haploid stage in *M. lychnidis-dioicae,* as mating occurs readily after meiosis, within a tetrad, often before the haploid cells separate from the meiotic divisions (Day 1979; Garber & Day 1985; Giraud et al. 2008; Hood & Antonovics 2000, 2004). Moreover, a recent study on gene expression and positive selection detected no evidence for mating-type antagonistic selection (Bazzicalupo et al. 2019). This model system is therefore ideal for investigating the impact of degeneration on differential gene expression between chromosomes determining reproductive compatibility, notably without the confounding effect of sexual antagonism.

In this study, we therefore investigated whether genes that are differentially expressed between mating types were more often associated with various signatures of degeneration than non-differentially expressed genes in the genome of *M. lychnidis-dioicae*. We determined whether differential gene expression varied among genomic compartments defined as autosomes, pseudo-autosomal regions (PARs), youngest evolutionary strata on non-recombining regions of the mating-type chromosomes (i.e., the previously identified red and green evolutionary strata, Branco et al. 2017), and oldest evolutionary strata on non-recombining regions (i.e., the blue, purple, orange and black evolutionary strata, Branco et al. 2017). We studied differential gene expression between mating types only in the haploid stage because the $a_1$ and $a_2$ mating types are determined at the haploid stage: mating can only occur between haploid sporidia of opposite mating types. We determined whether differential expression was associated with greater differences in degenerative mutations between alleles, including comparisons for each degenerative trait between differentially and non-differentially expressed genes within each genomic compartment. We then assessed the possibility that the allele showing lower expression levels would have higher levels of degeneration footprints. The investigated degeneration signatures included differences

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

between alleles in the levels of non-synonymous sequence divergence, transposable element (TE) insertions, alteration of predicted protein length (via mechanisms including acquisition of indels and/or early stop codons), intron content and GC content (Bedford & Hartl 2009; Cordaux & Batzer 2009; Feschotte 2008; Hoof & Green 1996; Marais et al. 2005). Associations between mating-type specific differential expression and signatures of degeneration, while requiring further study to establish the nature of causality, would reflect an important component of gene evolution.

## Results

### Allele identification and differential gene expression between $a_1$ and $a_2$ haploid genomes

We used the two available haploid genomes of *M. lychnidis-dioicae*, sequenced from separated haploid $a_1$ and $a_2$ mating-type cells derived from a single diploid individual, assembled at the chromosome-level scale and annotated (Branco et al. 2017; Branco et al. 2018). We investigated whether differential gene expression varied among genomic compartments defined as autosomes, PARs, youngest evolutionary strata, and oldest evolutionary strata of the mating-type chromosomes, and whether the greater degeneration signatures of several types were associated with the lower expressed allele. For this goal, we used whole-genome RNA-seq data from each of two replicate cultures for separate $a_1$ and $a_2$ haploid mating-type sporidia of *M. lychnidis-dioicae,* under low nutrient conditions, that resemble the natural haploid growth environment (Perlin et al. 2015; Schäfer et al. 2010). Alleles of single-copy genes in *M. lychnidis-dioicae* were identified using the criterion of 1:1 reciprocal best BLASTp between the $a_1$ and $a_2$ haploid genomes, which was highly consistent with the results by the OrthoMCL approach, as the two methods gave nearly (99.95%) identical results. After filtering out TE-related gene sequences, we identified 371 single-copy allelic pairs in mating-type chromosomes and 9,025 in autosomes (Table S1). We retained



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

8,549 single-copy genes with significantly detectable expression (see filtering criteria and details in Materials and Methods section) for further analysis (342 on mating-type chromosomes and 8,207 on autosomes). We analyzed 8,549 genes and used the threshold for differential gene expression of $|\text{Log2}(a_1/a_2|)$ being significantly greater than zero, with false discovery rate (FDR) < 0.050. This revealed 392 genes (4.59%) that were significantly more highly expressed in the $a_1$ haploid culture, and 203 (2.37%) that were significantly more highly expressed in $a_2$ haploid culture (Fig. 1; Table S2; Fig. S1).

**Differential gene expression and multiple signatures of sequence degeneration**

Regression analysis (generalized linear model, GLM) revealed that the degree of differential expression (DE) between allele pairs of the two haploid mating types significantly increased with increasing differences between alleles (using absolute values) in the various degeneration traits examined (Table 1). We found no asymmetry in reduced allelic expression or degeneration features between $a_1$ and $a_2$ mating-type chromosomes, we combined genes with $a_1$ or $a_2$ mating-type biased expression as the set of DE genes in the following analyses. The significant main-effect predictors of differential expression included genomic compartment and differences between alleles in non-synonymous divergence (*dN*), transposable element (TE) insertion number within 20kb (up and downstream of genes), intron content (proportional to coding sequence length), and overall GC content (GC0). Differences between alleles in predicted protein length (via mechanisms including acquisition of indels and/or early stop codons) was not a significant main-effect predictor but was strongly significant as an interaction term with genomic compartment and all other traits except intron content (Table 1; Fig. 2). Differential expression indeed increased with differences between alleles in predicted protein length, but only in the oldest evolutionary strata and when associated with higher differences between alleles in *dN*, TE content, and

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

GC0 (Table 1; Fig. 2). The oldest evolutionary strata were significantly enriched in genes with differential expression between mating types compared to autosomes, while the youngest evolutionary strata or the PARs were not (Table 2). Similar patterns were observed in the comparisons using each of the $a_1$ or $a_2$ haploid genomes separately (Table S3). Further *post-hoc* assessments of degenerative traits are presented in the following sections, including whether the difference between alleles is oriented such that the more affected allele is less expressed.

**Relationship between differential expression and elevated substitution rates**

Differentially expressed genes had greater sequence divergence between alleles than non-differentially expressed (non-DE) genes within genomic compartments, specifically within the oldest evolutionary strata of the mating-type chromosomes. DE genes had significantly higher non-synonymous mutation rate (*dN*) and synonymous mutation rate (*dS*) between alleles than non-DE genes within the oldest evolutionary strata (Wilcoxon rank sum test for independent samples, *dN*: W = 1433, $P < 0.001$, *dS*: W = 1422, $P < 0.001$) (Fig. 3A, Fig. S2, Table S4). There was almost no sequence divergence (*dN* or *dS*) between alleles on either autosomes or PARs for DE or non-DE genes. The youngest evolutionary strata had only one DE gene, precluding comparison to non-DE genes within this compartment. The finding that DE genes had significantly higher *dN* and *dS* in the oldest strata held for both genes with higher expression in $a_1$ cells (*dN*: W = 1156.5, $P = 0.019$) and those with higher expression in $a_2$ cells (*dN*: W = 1968.5, $P = 0.003$; *dS*: W = 2108, $P = 1.539e\text{-}4$) when considered separately, except the *dS* in $a_1$ (W = 1028.5, $P = 0.172$; Figs. S3A and S3B). In addition, there were no significant differences in *dS* or *dN* between $a_1$-biased and $a_2$-biased gene sets (Figure S3, *dS*: W = 154, $P = 0.122$; *dN:* W = 226, $P = 0.811$). A tendency of higher *dN*/*dS* for DE genes

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

compared to non-DE genes was not significant within the oldest strata (W = 1946, $P = 0.611$) (Fig. S4).

As DE genes were found to be associated with higher *dN* than non-DE genes, we then tested within DE genes whether the allele with lower expression was associated with greater accumulation of non-synonymous (or synonymous) changes. This was assessed by computing the sequence divergence between each allele of DE genes in *M. lychnidis-dioicae* with their ortholog in *M. lagerheimii*, which has retained largely collinear and homozygous mating-type chromosomes, and was inferred to have retained an ancestral chromosomal state in the *Microbotryum* genus (Branco et al. 2017, 2018). For DE genes, the allele that had lower expression levels in *M. lychnidis-dioicae* did not show significantly greater accumulation of non-synonymous changes than the allele with higher expression. The *dN* divergence between *M. lychnidis-dioicae* alleles (either $a_1$ or $a_2$) and their ortholog in *M. lagerheimii* was not greater for alleles having lower expression levels than alleles with higher expression levels in *M. lychnidis-dioicae* (W = 1,267, smallest $P = 0.909$) (Fig. S5, Table S5).

**Relationship between differential expression and TE insertions**

Differentially expressed genes were associated with greater differences between alleles for TE insertions (within 20kb up and downstream of genes) than alleles of non-DE genes across genomic compartments. However, the difference was significant only in the autosomes (W = 313879, $P < 0.001$, Fig. 3B), not in the PARs (W = 546, $P < 0.192$) or the oldest evolutionary strata (W = 4062, $P = 0.173$); the comparison was not possible in the youngest evolutionary strata, as noted above.

The alleles with lower expression had more TE insertions than the alleles with higher expression. This was assessed by calculating the differences in TE insertion numbers (upstream and downstream of genes) between alleles, as the TE number for the allele with

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

lower expression minus the TE number for the allele with higher expression; a positive value thus represented an excess of TEs in the less expressed allele. This oriented TE number difference between alleles was tested as a predictor of the expression ratio $|Log2(a_1/a_2)|$ using a sliding window approach with a 15kb window size overlapping by 5kb. Among DE genes, oriented TE insertion difference was a significant predictor of the expression ratio only in the window covering from 10kb upstream to the gene (Fig. 4A, Fig. S6); alleles with more TE insertions had reduced expression (for this window, Wald $X^2 = 6.674$, $P = 0.010$, statistics of remaining windows in Table S6). Among non-DE genes, none of the windows was a significant predictor of variation in the expression ratio (Table S6).

**Relationship between differential expression and differences in predicted protein length**

Differential gene expression was associated with the mutational changes that affect the predicted protein length, including altered stop codon positions, indels (including those causing frameshifts). Within genomic compartments, alternate alleles of DE genes were significantly more likely to produce proteins of different lengths than alleles of non-DE genes, particularly within the oldest evolutionary strata (two-proportion Z test, z = 2.186, $P = 0.029$) and autosomes (z = 4.64, $P = 8.78e-06$, Fig. 3C, Table S7); there were too few DE genes on PARs and youngest evolutionary strata for statistical comparisons.

The various types of mutational changes that caused protein length variation between alleles differed between DE and non-DE genes, as well as among genomic compartments. Among the 258 genes with different protein sequence lengths between alleles, all had indels. However, DE genes in the oldest evolutionary strata and autosomes had significantly more indels than non-DE genes; mean indel number in oldest strata differed between alleles by 2.64 for DE genes and by 1.85 for non-DE genes (W = 2453.5, $P = 0.013$), and in autosomes

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

alleles differed by a mean of 1.19 indels for DE genes and 1.03 for non-DE genes (W = 490.5, $P = 0.025$, Fig. 5A); PARs and youngest evolutionary strata could not be analyzed.

Similarly, differences in the positions of stop codons contributed to protein length variation more for DE genes than for non-DE genes. Among genes with different protein lengths between alleles in the oldest evolutionary strata, 44.6% (N = 56) of DE genes had different stop codon positions between alleles, which was significantly higher than the 24.0% (N = 75) of non-DE genes (two-proportion z-test, $P = 0.018$, Fig. 5B). Similarly, DE genes in the autosomes were marginally significantly more likely to have different stop codon positions between alleles than non-DE genes, with 33.3% (N = 21) vs 10% (N = 40), respectively ($P = 0.057$, Fig. 5B). Only three frameshift mutations were observed among the 258 genes with different protein/coding sequence lengths, and thus frameshifts were not a distinguishing feature of DE versus non-DE genes.

The alleles with lower expression did show a pattern of truncation of protein length compared to the alleles with higher expression (i.e. by early stop codons or deletions). To assess this pattern, differences in protein lengths between alleles were calculated as the ratio of protein length for the allele with higher expression divided by the allele with lower expression; a larger ratio thus represented a shorter length for the allele with lower expression. Among DE genes, this oriented metric of protein length differences was a significant predictor of the differential expression degree as the ratio $|\text{Log2}(a_1/a_2)|$, with alleles producing shorter proteins being less expressed (Wald $X^2 = 19.326$, two-tailed $P < 0.001$, Fig. 4B). Among non-DE genes, in contrast, we found no significant relationship between the expression level ratio and the oriented protein length ratio (Wald $X^2 = 0.222$, $P = 0.638$, Fig. 4B).

**Relationship between differential expression and intron content**



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

Differential gene expression was associated with differences between alleles in intron content (the ratio of intron to coding sequence length), lower intron content being considered to be favored by selection (Marais et al. 2005). There were significantly greater intron content differences between alleles for DE genes than for non-DE genes; considering the ratio of intron to coding sequence lengths, alleles of DE genes overall differed on average by 0.008 and alleles of non-DE genes differed by 0.002 (W = 2102758, $P < 0.001$). Alleles differed in intron content more for DE than non-DE genes within the autosomes (W = 1920124, $P = 0.033$) and oldest evolutionary strata (W = 3205, $P = 0.001$) (Fig. 3D, Table S8), but not within the PARs (W = 605, $P = 0.888$); the comparison in youngest evolutionary strata was not possible.

The alleles with lower expression did not show a greater intron content than the alleles with higher expression. This was assessed by calculating differences between alleles as the value for the less expressed allele minus the value for the more expressed allele; a positive value thus represented greater intron content for the less expressed allele. This oriented metric of intron content differences between alleles was not a significant predictor of differential expression level among DE genes (Wald $X^2 = 0.350$, $P = 0.554$), or among non-DE genes (Wald $X^2 = 0.216$, $P = 0.642$).

**Relationship between differential expression and GC content**

DE genes had significantly greater overall GC0 differences between their alleles than non-DE genes within the autosomes (W = 1907831, $P < 0.001$) and oldest evolutionary strata (W = 3010, $P < 0.001$) (Fig. 3E). The comparison within the PARs was not significant (W = 578, $P = 0.318$); the comparison for youngest evolutionary strata was not possible. Analysis of third codon position GC3 provided similar patterns and levels of significance (Fig. S7, Table S9).

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

The alleles with lower expression did not show lower GC content than the alleles with higher expression. To assess this possibility, GC0 or GC3 differences between alleles were calculated as the value for the allele with higher expression minus the value for the allele with lower expression; a positive value thus represented reduced GC content for the allele with lower expression. Among DE genes, neither the oriented GC0 or GC3 differences between alleles were significant predictors of the level of differential expression (GC0: Wald $X^2$ = 1.039, $P$ = 0.308, and GC3: Wald $X^2$ = 2.226, $P$ = 0.136).

## Discussion

In the anther-smut fungus *M. lychnidis-dioicae*, the mating-type chromosomes harbor evolutionary strata of various ages and lacks sexually antagonistic selection, while sharing non-recombining and heterogametic characteristics with sex chromosomes (Branco et al. 2017; Branco et al. 2018; Giraud et al. 2008; Hood and Antonovics 2004). We found no asymmetry in reduced allelic expression or degeneration features between $a_1$ and $a_2$ mating-type chromosomes, as expected given their lack of heterozygosity asymmetry and their lack of ecological differences (Hood et al 2002). Only the oldest strata of the mating-type chromosomes were enriched in genes differentially expressed between the haploid mating types, as found previously (Bazzicalupo et al. 2019). Most importantly, our study provides evidence for associations between differential expression and several different signatures of degenerative changes in the *M. lychnidis-dioicae* mating-type chromosomes. The differentially expressed genes displayed various forms of sequence (*dN*, *dS*, or GC content) and structural (TE insertions, intron content, or protein length) heterozygosity at levels higher than non-differentially expressed genes within genomic compartments (i.e., autosomes, PARs, the youngest strata and the oldest strata). These results show that differential gene expression is strongly associated with sequence degeneration, which can result either from a direct effect



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

of the studied degenerative mutations or from relaxed selection following changes in expression levels. Our results thus support the view that differential expression should be interpreted in a context that includes degenerative mutations, which is likely a general phenomenon in non-recombining sex chromosomes in addition to antagonistic selection.

**Differential gene expression between haploid mating types**

The proportion of genes with differential expression between haploid mating types of *M. lychnidis-dioicae* was low, but slightly higher than in a previous study based on the same dataset (Fontanillas et al. 2015), likely due to an improved genome assembly and non-recombining region identification. The proportion of genes with differential expression between mating types in *M. lychnidis-dioicae* (6.96%) was similar to the proportion of genes with differential expression between sexes in plant and animal non-reproductive tissues, e.g. liver, spleen, leaves, roots (Ayroles et al. 2009; Haselman et al. 2015; Ma, Veltsos, Sermier, et al. 2018; Ma, Veltsos, Toups, et al. 2018; Meisel et al. 2017; Perry et al. 2014; Yang et al. 2006). However, this proportion was much lower than in reproductive tissues (e.g. ovaries or testes) of most animals and plants (Ellegren & Parsch 2007; Parsch & Ellegren 2013).

Differentiated sex chromosomes in anisogamous animals and plants are often enriched in differentially expressed (DE) genes (reviewed by Dean & Mank 2014; Mank et al. 2014). In fungi other than *M. lychnidis-dioicae*, such as *N. tetrasperma* and *Podospora anserina*, DE genes were also more frequently detected on non-recombining regions of the mating-type chromosomes than in autosomes, which has been interpreted as resulting from ecological differences particular to those species, i.e. differences in terms of vegetative or sexual growth between mating types (Grognet et al. 2014; Samils et al. 2013). In animals and plants, sexual antagonism occurs when trait values that increase gene transmission through the male function then decrease gene transmission through the female function, or conversely



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

(Charlesworth et al. 2014; Dean and Mank 2014; Lande 1980; Rice 1987). The linkage of sexually-antagonistic genes to the sex-determining genes in non-recombining regions is considered fundamental to the formation of evolutionary strata and the resolution of sexual conflict, for example by allowing for sex-specific or sex-biased gene expression (Charlesworth et al. 2005; Lipinska et al. 2017; Rice 1987; Otto et al. 2011). Differentially expressed genes on differentiated sex chromosomes are therefore often considered to resolve sexual antagonism or sexual conflicts between females and males, even if several studies have shown that differentially expressed genes could also be associated with degenerative mutations (Bachtrog et al. 2008; Hough et al. 2014; White et al. 2015; Wright et al. 2012; Xu et al. 2019). Our findings support the view that degenerative mutations likely contribute to levels of differential expression in sex-related chromosomes, while it is important to note that this is not exclusive of, and should be considered in addition to, the hypothesis invoking the role of sexual antagonism.

Our results thus encourage a broader view of the evolutionary forces related to DE genes and their enrichment on chromosomes determining reproductive compatibility. In *M. lychnidis-dioicae*, any differential gene expression between the alternative haploid mating types is unlikely due to 'mating-type antagonistic selection' given the lack of female and male functions and the lack of haploid phase beyond the tetrad stage. Little evidence was found for a role of antagonistic selection in driving the evolution of new evolutionary strata in mating-type chromosomes of *M. lychnidis-dioicae* (Bazzicalupo et al. 2019). Young evolutionary strata were indeed found not to be enriched in genes upregulated in the haploid mating phase compared to the dikaryotic stage, nor in genes differentially expressed between mating types, nor in genes displaying footprints of specialization (i.e. high *dN/dS*) between mating types (Bazzicalupo et al. 2019). Importantly, gene degeneration is expected to occur commonly on non-recombining sex or mating-type chromosomes due to the reduced efficacy of selection

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

caused by the absence of recombination (Charlesworth et al. 2005; Ellegren & Parsch 2007; Fontanillas et al. 2015; Lipinska et al. 2017; Rice 1987; Otto et al. 2011). The resulting mutation accumulation may generate contrasting expression levels between differently affected alleles or may accumulate where differential expression decreases purifying selection on the less expressed allele. Again, sexually antagonistic selection and degeneration are not mutually exclusive processes for explaining sex-biased gene expression. Still, the potential role of degenerative mutations remains worth highlighting in all systems. The association found between multiple types of degeneration footprints and differential expression in *M. lychnidis-dioicae*, in the absence of sexual antagonism, suggest the possibility of similar associations across diverse types of organisms. Finally, previous studies showed that there was overall very low genetic variation within *M. lychnidis-dioicae*, both on autosomes and within each of $a_1$ and $a_2$ mating-type chromosomes, due to high selfing rates and small effective population sizes (Badouin et al. 2017; Branco et al. 2017). Therefore, we expect most $a_1$ mating types to be similar in expression and most $a_2$ mating types likewise. Further studies using more individuals and species would nevertheless be interesting to see the degree to which these associations form general and consistent patterns.

**Various forms of degeneration**

One of the main insights gained in the present study was the directional associations found between differential expression and specific signatures of degeneration. The properties of non-recombining regions that reduce the efficacy of selection (reduced effective population size, hitchhiking and lack of deleterious mutation purging) can lead to the fixation of various mutations having degenerative effects, several of which were significant predictors in the overall regression model of differential expression between mating types of *M. lychnidis-dioicae*. Even within genomic compartments, DE genes had significantly higher levels of

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

degenerative mutations distinguishing the alleles than non-DE genes. Some signatures of degeneration, such as TE insertions, indels and/or premature stop codons, may be most plausibly conceived as mechanisms that reduce transcription levels. In particular, TE insertions into genes or upstream have long been recognized to alter gene expression (Britten and Davidson 1971; McClintock 1942). TEs can disrupt promoter regions or other regulatory sequences internal to genes (Cordaux & Batzer 2009; Feschotte 2008). In addition, epigenetic silencing, as a defense against TE proliferation, can tighten local chromatin structure and inhibit access of transcriptional machinery (Eichten et al. 2014; King 2015). Consistent with a direct effect upon differential expression, the relative excess of TE insertions between alleles, specifically upstream of genes, was associated with a lower expression level between alleles of DE genes. Similarly, the introduction of early stop or non-sense codons may reduce expression. Transcripts from alleles with premature stop codons are affected by nonsense mediated decay, involving the degradation of mRNA and further components of the RNAi pathway that down-regulate expression (Hoof & Green 1996). We found that an allele with a shorter predicted protein length in DE genes was indeed more likely to have lower expression. While TE insertions, indels or premature stop codons are potentially important mutations affecting differential expression between alleles, it is difficult to establish causal relationship from our current datasets, and further studies are needed to directly test the nature of causality between differential expression and specific degenerative mutations. For instance, in investigating the degeneration of the neo-Y chromosome of *Drosophila albomicans*, Zhou & Bachtrog (2012) provided evidence that TEs accumulated mostly after regulatory changes in gene expression occurred. Furthermore, the interplay between mutations and expression may be complex; for example, an early stop codon might not directly reduce expression, but it still may cause partial loss in protein function that reduces the strength of selection to maintain the expression of the allele.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

Other signatures of degeneration in *M. lychnidis-dioicae* were not directionally predictive of lower allelic expression but were nevertheless more associated with DE than non-DE genes. Most substantial among these characteristics was the degree of sequence divergence between alleles. Alleles of DE genes were distinguished by markedly more non-synonymous and synonymous base pair differences than alleles of non-DE genes. Similar results were obtained in the anisogamous, hermaphroditic ascomycete *N. tetrasperma*, showing that differential gene expression was positively correlated with sequence divergence between alleles on mating-type chromosomes (Samils et al. 2013).

**Degeneration across genomic compartments**

The different forms of genetic degeneration in *M. lychnidis-dioicae* were not equally represented across genomic compartments, perhaps reflecting the history of recombination suppression. In this system, enrichment of DE genes on the mating-type chromosomes is unlikely to be due to antagonistic selection. As a matter of fact, enrichment of DE genes was significant only in the oldest evolutionary strata and not in the younger strata, indicating it is a consequence and not a driver of recombination suppression.

Mating between different haploid sexes or mating types ensures that all diploids are heterogametic (Bull 1978), and it has long been recognized that regions linked to mating type loci can preserve heterozygosity (Mather 1942). In *M. lychnidis-dioicae*, the large non-recombining regions are in fact highly heterozygous (Branco et al. 2017). In contrast, the autosomes and PARs are largely homozygous, due to the selfing mating system of *M. lychnidis-dioicae* (Giraud et al. 2008; Hood & Antonovics 2000, 2004). Consistent with mating-type linkage preserving heterozygosity, nearly the full range of mutational changes or footprints of degeneration showed lowest levels in the autosomes and PARs and increasing from lowest levels in the youngest evolutionary strata to highest levels in the oldest

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

evolutionary strata. Importantly, however, comparisons within genomic compartments repeatedly showed that allele-distinguishing mutations occurred more in association with DE genes than non-DE genes or in the manner positively associated with levels of differential expression. This represents strong evidence for these degenerative changes being associated with changes in expression levels between alleles. Finally, the recent discovery of multiple independent mating-type linkage events across the *Microbotryum* genus (Branco et al. 2018) should allow further assessment of mutation accumulation and its consequences for gene functions.

## Conclusions

Our findings on differential gene expression, being more frequent in oldest evolutionary strata and being associated with various types of sequence degeneration, and without sexually antagonistic selection as confounding factor, shed new lights on how differentially expressed genes might evolve in non-recombining regions in general, such as sex chromosomes or mating-type chromosomes. Our study showed that the accumulation of degenerative mutations between alleles was significantly associated with the degree of differential gene expression, in a system where sexually antagonistic selection is unlikely to occur as a confounding factor. Furthermore, the genes with differential expression were highly enriched on mating-type chromosomes, as in diverse organisms where the separate sex functions have been cited as the primary cause. We further found evidence of a directional relationship between differential gene expression and some types of mutational changes, in particular TE insertions and premature stop codons, being greater in the alleles with lower expression levels, although a causal relationship remains to be demonstrated. Our results suggest an important relationship between mutation accumulation and differential expression between alleles,

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

which is relevant to a broad range of taxa where reproductive compatibility, sex or other complex traits are determined in extensive regions of recombination suppression.

## Materials and Methods

### Allele identification between $a_1$ and $a_2$ haploid genomes

In order to quantify differentially expressed genes between the two haploid genomes, we first identified the alleles between $a_1$ and $a_2$ haploid genomes for those genes. We used the $a_1$ and $a_2$ predicted coding gene sequences of the two haploid genome assemblies (accession ID: $a_1$ – PRJEB12080, ERS459551, ERZ250722 and $a_2$ - PRJEB12080, ERS1013678, ERZ250721) (Branco et al. 2017; Branco et al. 2018). To identify 1:1 single-copy homologs in each haploid genome, the Reciprocal Best BLAST(p) Hits (RBBH) python script (github.com/peterjc/galaxy_blast/tree/master/tools/blast_rbh) was applied (Camacho et al. 2009), with 50 percentage of length coverage. RBBH scripts also identified paralogs within each haploid genome. A number of protein sequence alignment identity thresholds were tested, in order to identify the best strategy for maximizing the number of allele pair identification on the non-recombining regions while avoiding spurious BLAST results with low identity percent. Increasing the percent of protein sequence identity threshold from >70% to >85% resulted in a decrease from 12.2% to 9.9% of single-copy genes on the mating-type chromosomes being identified as differentially expressed genes (detailed below), while decreasing the threshold from >70% to >30% resulted in only a marginal increase from 12.2% to 12.7%. The change in the percentages of identified alleles that were differentially expressed on autosomes was negligible, being 1.0%, 1.1% and 1.1% respectively for 80%, 70% and 30% thresholds (Fig. S8). Therefore, the threshold of >70% protein sequence identity was used. Additionally, we also have detected orthologs between $a_1$ and $a_2$ genomes

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

using OrthoMCL (Li et al. 2003). As the set of identified single-copy orthologous genes and their alleles was almost identical, differing in only five out of 9,396 genes compared to reciprocal best BLASTp analysis, we used the ortholog list generated by reciprocal best BLASTp for downstream analysis. To avoid potential bias due to paralogs for identifying differential gene expression and other downstream analysis, genes with paralogs within each haploid genome were filtered out and only single-copy allele pairs were retained for downstream analysis. Genes were assigned to genomic compartments, i.e., autosomes, pseudo-autosomal regions (PARs), youngest evolutionary strata of the mating type chromosomes (i.e., the previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (i.e., the blue, purple, orange and black strata; Branco et al. 2017). Finally, we confirmed the lack of small RNAs in our predicted coding gene sequences by checking the absence of non-coding RNAs in the *M. lychnidis-dioicae* predicted coding genes, using BLAST searches of ncRNA, tRNA and rRNA sequences from the Rfam database (https://rfam.xfam.org, Griffiths-Jones et al. 2003; Kalvari et al. 2017). No ncRNA was found in the *M. lychnidis-dioicae* predicted coding gene set; the rRNA sequences detected in the genome did not have BLAST hits in the predicted coding gene set; only one tRNA sequence returned a partial hit (17.5% alignment), which had however not been retained in our 1:1 ortholog list.

**Transposable element filtering**

Transposable element (TE) annotation of both haploid genomes of *M. lychnidis-dioicae* was published previously (Hartmann et al. 2018), and was used for analysis in this study. The coding sequence of each gene from both $a_1$ and $a_2$ haploid genomes was searched by BLAST(n) against the published annotated TE consensus sequences of the same species, and alignment >80 percent of query coverage (coding sequences) was used for identifications of

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

TEs. The BLASTn output was parsed using BASH scripts, and the coding sequences identified as TEs were removed from the gene list for all further downstream analyses.

**Identification of differentially expressed genes**

We studied differential gene expression between mating types only in the haploid stage because the $a_1$ and $a_2$ mating types are determined at the haploid stage: mating can only occur between haploid sporidia of opposite mating types. In addition, almost all genes on autosomes and PARs are homozygous due to a high selfing rate (Badouin et al. 2017; Branco et al. 2017), so that we could assign expression levels to $a_1$ or $a_2$ mating types in the diploid or dikaryotic stages only for highly differentiated alleles, i.e., on oldest evolutionary strata, which would profoundly limit and likely bias the analyses by taking into account only degenerated genes.

RNAseq samples and datasets were described previously (Perlin et al. 2015). Briefly, haploid sporidial strains of the original isolate (the same as the reference genome "Lamole strain") were generated from the meiotic products of a single tetrad. Then haploid fungal cells of either haploid $a_1$ or $a_2$ strain were grown separately on 2% water agar, each strain being grown in two replicates, with nutrient-free environment without the mating partner for two days, which essentially mimicked the natural conditions on the plant before mating and infection (Perlin et al. 2015). For RNAseq, polyA RNA was purified and a strand-specific library was constructed for each RNA sample; each library was sequenced with Illumina technology, generating on average 34.786 million 76 bp paired-end reads for $a_1$ libraries, and 35.017 million paired-end reads for $a_2$ libraries. The raw data of haploid culture growing separately in water agar conditions were downloaded from the deposited NCBI database (accession ID PRJNA246470). The RNAseq raw reads were quality assessed using FastQC v0.11.2 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and quality trimmed using Trimmomatic v0.33 with default parameters for paired-end reads (Bolger et al. 2014).

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

We filtered reads containing adaptor sequences and trimmed reads if the sliding window average Phred score over four bases was < 15 or if the leading/trailing bases had a Phred score < 3. Reads were then removed post filtering if either read pair was < 36 bases. After the filtering process, we recovered, on average per sample, 30.612 million paired-end reads for $a_1$ libraries, and 30.650 million paired-end reads for $a_2$ libraries.

Pairs of alleles from $a_1$ and $a_2$ mating types were aligned with PRANK (v170427) using the codon model (Löytynoja & Goldman 2010). To avoid possible bias for calling differential gene expression due to differences in homolog length between $a_1$ and $a_2$, gaps differing between alleles by greater than 3bp were trimmed to keep the same length, using a published custom Python script (Parker 2016). This trimming included the gaps from the ends of the alignment and inside the alignment between alleles, with inside gaps starting with the closest to the end of the alignment (greater than the minimum gap size) until there were no gaps larger than minimum gap size (Parker 2016). The trimmed allele pairs with equal length were used for read mapping and calling differential gene expression.

To quantify gene expression, we mapped the trimmed reads of haploid samples to the trimmed homolog sequences of each haploid genome respectively with Kallisto v.0.43.0 (Bray et al. 2016). Read counts of the output from Kallisto mapping (e.g. using pseudo-alignment) were imported for gene expression analysis in EdgeR v3.4 (Robinson et al. 2010; McCarthy et al. 2012). We filtered low counts and kept genes with average Log(CPM) > 0 per sample, and CPM (count per million) > 1 in half of the total samples per haploid culture. We then normalized the expression using the weighted trimmed mean of M-values (TMM) implemented in EdgeR, which is a scaling factor for library sizes that minimizes the log-fold change between samples. We explored the libraries of both haploid cultures in pairwise correlation of raw counts between replicates (Fig. S9), and two dimensions using multi-dimensional scaling (MDS) plots (Fig. S10). Normalized expression counts for each sample

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

were used to calculate differential expression between mating types using standard measures. We first identified genes with differential expression between mating types based on overall expression of the comparison group, and using Benjamini-Hochberg correction for multiple-testing with false discovery rate (FDR) of 5%. Differential expression between mating types was classified into four categories of fold changes, namely 2 (low), 2-4 (mild), 4-8 (high), and > 8 (very high), and expressed as $\log_2$ ratio of $a_1$-to-$a_2$ expression (which has negative values for genes with higher $a_2$ expression and positive values for higher $a_1$ expression). We only considered genes with fold changes > 1 (i.e. $|\log_2 FC| > 0$), as recommended (Montgomery & Mank 2016), because we worked on haploid cell cultures and there were no possible scaling nor allometry issues due to whole-body sampling. Thus, unless stated otherwise, both conditions FDR < 0.05 and $|\log_2 FC| > 0$ were met when calling mating-type bias. Finally, to investigate the expression level of differentially (with $a_1$ or $a_2$ mating-type bias) and non-differentially expressed genes, we compared normalized read counts (transcripts per million, Log2TPM, obtained from EdgeR v3.4) of significantly expressed genes at autosomal and mating-type chromosomes (filtering criteria is the same as described above) from $a_1$ and $a_2$ samples (Fig. S1).

The classification of genes as having differential expression between mating types was based on the absolute values of gene expression ratio $|\text{Log2}(a_1/a_2)|$, and was used to assess relationships to various forms of mutational changes. A generalized linear model (GLM) analysis was used to assess the predictors of the absolute values of expression ratio $|\text{Log2}(a_1/a_2)|$, with the following main effect variables: genomic compartments, the absolute value of differences between alleles for sequence divergence (*dN*), transposable element insertions number within 20kb (up and downstream), predicted protein length, intron content and GC content; we also included all two-way interactions terms in this model. The absolute value of the differences between alleles was calculated for each trait as detailed below. Model



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

family comparison was based upon minimizing Akaike's Information Criterion and over/under-dispersion using ratio of deviance/df; Tweedie, power 1.7 (approaching gamma distribution) provided the best available fit for the expression ratio response variable. A best fit model was selected using stepwise model selection, following removal of non-significant interaction terms. Other *post hoc* tests evaluating individual degeneration trait are described below. All statistical analyses were conducted in SPSS v23 (IBM Corp 2015) and R v3.4.3 (R Core Team 2017).

**Relationship between differential expression and elevated substitution rates**

Each pair of allele was aligned with PRANK (v170427), using the codon model (Löytynoja & Goldman 2010), and each alignment was then analyzed with yn00 in PAML (Yang 2007) (runmode -2) to calculate the number of nonsynonymous substitutions per nonsynonymous site (*dN*), the number of synonymous substitutions per synonymous site (*dS*), and the ratio of the two (*dN*/*dS*), the latter excluding genes with *dS* value of zero. We then compared sequence divergence between alleles using non-parametric Wilcoxon rank sum tests for DE versus non-DE genes within genomic compartments.

We also compared the allele sequences of *M. lychnidis-dioicae* to their orthologs in *M. lagerheimii*, which carries mating-type chromosomes inferred to be largely collinear and a good proxy for the ancestral state in the *Microbotryum* genus (Branco et al. 2017, 2018). Pairs of $a_1$ or $a_2$ orthologs present in *M. lychnidis-dioicae* and *M. lagerheimii* were aligned with PRANK (v170427) using the codon model, then each ortholog alignment was analyzed with codeml (runmode -2) in PAML (Yang 2007). The single-copy orthologs for $a_1$ or $a_2$ genomes between *M. lychnidis-dioicea* and *M. lagerheimii* were identified using RBBH with 70 percent protein sequence coverage identity (github.com/peterjc/galaxy_blast/tree/master/tools/blast_rbh, Camacho et al. 2009). Wilcoxon

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

rank sum test was used to assess *dN* between orthlogs in *M. lychnidis-dioicea* and *M. lagerheimii* to evaluate the hypothesis that the alleles with lower expression levels would have greater sequence divergence.

**Relationship between differential expression and TE insertions**

The TE annotation of the *M. lychnidis-dioicae* genome published previously (Hartmann et al. 2018) was used for the analysis in this study. First, the TE insertion sites were assessed for each given focal gene, upstream 0-5k, 5-10kb, 10-15kb, 15-20kb distance intervals, and downstream 0-5kb, 5-10kb, 10-15kb and 15-20kb distance intervals using Bedtools window function for each indicated distance window (https://bedtools.readthedocs.io/en/latest/content/tools/window.html). Both annotation GFF3 files of gene models and TE annotations of *M. lychnidis-dioicae* were provided as input files. The output files were parsed using Bash scripts. Wilcoxon rank sum tests were used to compare TE insertions for DE and non-DE genes within genomic compartments.

A limited GLM model was used to assess the hypothesized directional association of TE insertions and reducing allelic expression ($|Log2(a_1/a_2)|$); this model contained genomic compartment and oriented TE differences between alleles as main effects and their interaction term. Oriented TE differences between alleles were calculated as the TE number for the allele with lower expression minus the TE number for the higher expressed allele; a positive value thus represented an excess of TEs in the lower expressed allele. A sliding-window approach was used with a window size of three adjacent intervals, progressing from upstream to downstream of the genes.

**Relationship between differential expression and differences in predicted protein length**

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

We first verified whether there was no bias issue in the gene prediction model across genomic compartments, as degenerative mutation accumulation in the non-recombining region may decrease the accuracy of coding sequence prediction. The ratio of the predicted coding sequence length divided by three times the protein sequence length was consistently close to 1 and did not differ among genomic compartments (autosome, PAR, youngest strata and oldest evolutionary strata; Linear model, $R^2$ = -5.50e-05, F-statistic = 0.869, P= 0.530, Fig. S11). We therefore calculated the ratio of the predicted protein length between allele pairs, and compared the proportions of genes in DE and non-DE categories that had unequal lengths between alleles using two-proportion Z test for genes within genomic compartments. The mutational causes of unequal protein lengths was assessed by manually quantifying premature stop codons or indels using Geneious v8.1.7 (Kearse et al. 2012). A limited GLM model was used to assess the hypothesized directional association of protein truncation and reducing allelic expression ($|Log2(a_1/a_2)|$); this model contained genomic compartment and oriented predicted protein length differences between alleles as main effects and their interaction term. Oriented predicted protein length differences between alleles were calculated as the ratio for the allele with higher expression divided by the allele with lower expression; a larger ratio thus represented a shorter length for the allele with lower expression.

**Relationship between differential expression and intron content**

Using the published annotation gene models and coding sequences, we extracted the intron number and mean intron length information from the annotation gff3 file, using Perl script (https://bioops.info/2012/11/intron-size-gff3-perl/). We investigated the differences in the proportional intron content for both DE and non-DE genes within genomic compartments using Wilcoxon rank sum test. We also used a limited GLM model to test the hypothesized directional association of greater intron content and reducing allelic expression ($|Log2(a_1/a_2)|$);



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

this model contained genomic compartments and oriented intron content differences between alleles as main effects and their interaction term. Oriented intron content differences between alleles were calculated as the value for the allele with lower expression minus the value for the allele with higher expression; a positive value thus represented greater intron content for the lower expressed allele.

**Relationship between differential expression and GC content**

We calculated the total GC percentage (GC0) and the GC percentage at the third position of aminol acid (GC3) for alleles of each gene coding sequence using homemade awk scripts. We investigated the differences of GC0 and GC3 for both DE and non-DE genes within genomic compartments using Wilcoxon rank sum test. We also used a limited GLM model to test the hypothesized directional association of reduced GC content and reducing allelic expression ($|Log2(a_1/a_2)|$); this model contained genomic compartments and oriented GC content differences between alleles as main effects and their interaction term. Oriented GC content differences between alleles were calculated as the value for the allele with higher expression minus the value for the allele with lower expression; a positive value thus representing reduced GC content for the allele with lower expression.

**Ethics Statement**

N/A

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

## Acknowledgements

We thank Darren J. Parker for assistance on python scripting and valuable discussions of the results, Ricardo C. Rodríguez de la Vega for insightful comments. We thank Judith Mank and two anonymous reviewers' constructive comments to improve this manuscript. The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. This work was supported by the NIH grant R15GM119092 to M. E. H., and the Louis D. Foundation award and EvolSexChrom ERC advanced grant #832352 to T. G.

## Data deposition

We used published gene expression data to investigate the association of sequence degeneration and differential gene expression in *Microbotryum lychnidis-dioicae* (Fontanillas et al. 2015; Perlin et al. 2015, accession ID PRJNA246470). We used published genome assembly, gene predictions and assignations to genomic compartments (Branco et al. 2017, 2018). We also used published transposable elements identification in *M. lychnidis-dioicae* (Hartmann et al. 2018). All relevant scripts and data files to perform these analyses are deposited in Zenodo and Github (https://github.com/Wen-Juan/Differential_expression_associateswith_degeneration_Microbotryum_fungus), which will be released immediately upon manuscript acceptance.

## References

Abbate JL, Hood ME. 2010. Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus *Microbotryum*. J. Evol. Biol. 23:1800–1805. doi: 10.1111/j.1420-9101.2010.02036.x.

Ayroles JF et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. 41:299–307. doi: 10.1038/ng.332.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. Curr. Opin. Genet. Dev.



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

16:578–585. doi: 10.1016/j.gde.2006.10.007.
Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. Genome Res. 15:1393–1401. doi: 10.1101/gr.3543605.
Bachtrog D et al. 2014. Sex determination: why so many ways of doing it? PLoS Biol. 12:1–13. doi: 10.1371/journal.pbio.1001899.
Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. Genetics. 179:1513–1525. doi: 10.1534/genetics.107.084012.
Bachtrog D. 2013. Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14:113–124. doi: 10.1038/nrg3366.
Bachtrog D, Hom E, Wong KM, Maside X, Jong P De. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. Genome Biol. 9:R30. doi: 10.1186/gb-2008-9-2-r30.
Badouin H et al. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. Mol. Ecol. 26:2041–2062. doi: 10.1111/mec.13976.
Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. 2019. Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). Genes|Genomes|Genetics. 9:1987–1998. doi: 10.1534/g3.119.400242.
Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. Proc. Natl. Acad. Sci. 106:1133–1138. doi: 10.1073/pnas.0812009106.
Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8:1499–1504. doi: 10.1093/nar/gkl908.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi: 10.1093/bioinformatics/btu170.
Branco S et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc. Natl. Acad. Sci. U. S. A. 114:7367–7072. doi: 10.1073/pnas.1701658114.
Branco S et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. Nat. Commun. 9:2000. doi: 10.1038/s41467-018-04380-9.
Bray N, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal RNA-Seq quantification. Nat. Biotechnol. 34:525–528. doi: arXiv:1505.02710.
Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. 46:111–138. doi: 10.1016/s0140-6736(00)93717-6.
Bull JJ. 1978. Sex chromosomes in haploid dioecy: a unique contrast to Muller's theory for dipoid dioecy. Am. Nat. 112:245–250.
Camacho C et al. 2009. BLAST+: Architecture and applications. BMC Bioinformatics. 10:1–9. doi: 10.1186/1471-2105-10-421.
Charlesworth B. 1996. The evolution of chromosomal sex determination. Curr. Biol. 6:149–162. doi: 10.1002/0470868732.ch17.
Charlesworth B. 1991. The evolution of sex chromosomes. Science. 251:1030–1033.
Charlesworth B, Jordan CY, Charlesworth D. 2014. The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. Evolution. 68:1339–1350. doi: 10.1111/evo.12364.
Charlesworth D. 2002. Plant sex determination and sex choromosomes. Heredity. 88:94–101. doi: 10.1038/sj/hdy/6800016.
Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity. 95:118–128. doi: 10.1038/sj.hdy.6800697.
Connallon T, Knowles, LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends Genet. 21:495-499.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. 10:691–703. doi: 10.1038/nrg2640.
Darolti I et al. 2019. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. Proc. Natl. Acad. Sci. U. S. A. 116:19031–19036. doi: 10.1073/pnas.1905298116.
Day AW. 1979. Mating type and morphogenesis in *Ustilago violacea*. Bot. Gaz. 140:94–101.
Dean R, Mank JE. 2014. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. J. Evol. Biol. 27:1443–1453. doi: 10.1111/jeb.12345.
Eichten SR, Schmitz RJ, Springer NM. 2014. Epigenetics: beyond chromatin modifications and complex genetic regulation. Plant Physiol. 165:933–947. doi: 10.1104/pp.113.234211.
Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat. Rev. Genet. 8:689. doi: 10.1038/nrg2167.
Feschotte C. 2008. The contribution of transposable elements ot the evolution of regulatory networks. Nat. Rev. Genet. 9:397–405. doi: 10.1038/nrg2337.
Fontanillas E et al. 2015. Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. Mol. Biol. Evol. 32:928–943. doi: 10.1093/molbev/msu396.
Fraser JA, Heitman J. 2004. Evolution of fungal sex chromosomes. Mol. Microbiol. 51:299–306. doi: 10.1046/j.1365-2958.2003.03874.x.
Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics. 159:907-911.
Garber ED, Day AW. 1985. Genetic mapping of a phytopathogenic Basidiomycete, *Ustilago violacea*. Bot. Gaz. 146:449–459.
Giraud T, Yockteng R, Lo M. 2008. Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. Eukaryot. Cell. 7:765–775. doi: 10.1128/EC.00440-07.
Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nucleic Acids Res. 2003. 31:439-441.
Grognet P et al. 2014. Maintaining two mating types: Structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. Genetics. 197:421–432. doi: 10.1534/genetics.113.159988.
Grummt I, Pikaard CS. 2003. Epigenetic silencing of RNA polymerase I transcription. Nat. Rev. Mol. Cell Biol. 4:641–649. doi: 10.1038/nrm1171.
Hartmann FE, De La Vega RCR, Brandenburg JT, Carpentier F, Giraud T. 2018. Gene presence-absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure. Genome Biol. Evol. 10:1298–1314. doi: 10.1093/gbe/evy089.
Haselman JT, Olmstead AW, Degitz SJ. 2015. General and comparative endocrinology global gene expression during early differentiation of *Xenopus (Silurana) tropicalis* gonad tissues. Gen. Comp. Endocrinol. 214:103–113. doi: 10.1016/j.ygcen.2014.06.009.
Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. 2015. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. BioEssays. 37:148–154. doi: 10.1002/bies.201400138.
Hood M, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles. Heredity. 85:231–241.
Hood ME. 2002. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics. 160:457–461.
Hood ME, Antonovics J. 2004. Mating within the meiotic tetrad and the maintenance of genomic. Genetics. 166:1751–1759.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. Genetics. 168:141–146. doi: 10.1534/genetics.104.029900.
Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type chromosomes of the anther-smut fungus. Genetics. 193:309–315. doi: 10.1534/genetics.112.146266.
Hoof A van, Green PJ. 1996. Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position-dependent manner. Plant J. 10:415–424.
Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. Proc. Natl. Acad. Sci. 111:7713–7718. doi: 10.1073/pnas.1319227111.
IBM Corp. 2015. IBM SPSS Statistics for Macintosh, Version 23.0. Version 23:Armonk, NY: IBM Corp.
Kaiser VB, Zhou Q, Bachtrog D. 2017. Nonrandom gene loss from the *Drosophila miranda*. 3:1329–1337. doi: 10.1093/gbe/evr103.
Kalvari I et al. 2017. Rfam 13.0: shifting to a genomic-centric resource for non-coding RNA families. Nucleic Acids Res. 46:D335-D342.
Kearse M et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma. Appl. Note. 28:1647–1649. doi: 10.1093/bioinformatics/bts199.
Kimball SR, Jefferson LS. 2004. Amino acids as regulators of gene expression. Nutr. Metab.. 1:3. doi: 10.1186/1743-7075-1-3.
King GJ. 2015. Crop epigenetics and the molecular hardware of genotype × environment interactions. Front. Plant Sci. 6:1–19. doi: 10.3389/fpls.2015.00968.
Konuma J, Sota T, Chiba S. 2013. Quantitative genetic analysis of subspecific differences in body shape in the snail-feeding carabid beetle *Damaster blaptoides*. Heredity. 110:86–93. doi: 10.1038/hdy.2012.68.
Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution. 34:292–305.
Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. Cell. 152:1237–1251. doi: 10.1016/j.cell.2013.02.014.
Li L, Stoeckert Jr CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2178–2189. doi: 10.1101/gr.1224503.candidates.
Lipinska AP et al. 2017. Multiple gene movements into and out of haploid sex chromosomes. Genome Biol. 18:104. doi: 10.1186/s13059-017-1201-7.
Löytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics. 11:579.
Ma W-J, Veltsos P, Toups MA, et al. 2018. Tissue specificity and dynamics of sex-biased gene expression in a common frog population with differentiated, yet homomorphic, sex chromosomes. Genes. 9:294. doi: 10.3390/genes9060294.
Ma W-J, Veltsos P, Sermier R, Parker DJ, Perrin N. 2018. Evolutionary and developmental dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes. Genome Biol. 19:156. doi: 10.1186/s13059-018-1548-4.
Mank JE. 2013. Sex chromosome dosage compensation: definitely not for everyone. Trends Genet. 29:677–683. doi: 10.1016/j.tig.2013.07.005.
Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. Trends Genet. 25:226–233. doi: 10.1016/j.tig.2009.03.005.
Mank JE, Hosken DJ, Wedell N. 2011. Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. 1–12. doi: 10.1111/j.1558-5646.2011.01316.x.
Mank JE, Hosken DJ, Wedell, N. 2014. Conflict on the sex chromosomes: cause, effect, and

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

complexity. Cold Spring Harb. Perspect Biol. 6:a017715.
Marais G, Nouvellet P, Keightley PD, Charlesworth B. 2005. Intron size and exon evolution in *Drosophila*. Genetics. 170:481–485. doi: 10.1534/genetics.104.037333.
Mather K. 1942. Heterothally as an outbreeding mechanism in fungi. Nature. 149:54–56. doi: 10.1038/149054a0.
McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–4297. doi: 10.1093/nar/gks042.
McClintock B. 1942. Maize genetics. In: Carnegie institution of Washington year book.Vol. 41 pp. 181–186.
Meisel RP, Gonzales CA, Luu H. 2017. The house fly Y chromosome is young and undifferentiated from its ancient X chromosome partner. Genome Res. 27:1–10.
Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. Mol. Biol. Evol. 21:984-990.
Montgomery SB et al. 2013. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. Genome Res. 23:749–761. doi: 10.1101/gr.148718.112.
Montgomery SH, Mank JE. 2016. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. Mol. Ecol. 25:5114–5128. doi: 10.1111/mec.13824.
Muyle A, Shearn R, Marais GAB. 2017. The evolution of sex chromosomes and Dosage compensation in plants. Genome Biol. Evol. 9:627–645. doi: 10.1093/gbe/evw282.
Ohno S. 1966. *Sex Chromosomes and Sex-Linked Genes*. Springer-Verlag, Berlin Heidelberg doi: 10.1007/978-3-642-88178-7.
Otto SP et al. 2011. About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. 27:358–367. doi: 10.1016/j.tig.2011.05.001.
Parker DJ. 2016. Fastatools.pdf. Zenodo. https://zenodo.org/record/59775.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat. Rev. Genet. 14:83–87. doi: 10.1038/nrg3376.
Perlin MH et al. 2015. Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. BMC Genomics. 16:461. doi: 10.1186/s12864-015-1660-8.
Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. Mol. Biol. Evol. 31:1206–1219. doi: 10.1093/molbev/msu072.
Pucholt P, Wright AE, Conze LL, Mank JE. 2017. Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. Mol. Biol. Evol. 34:1991–2001.
R Core Team. 2017. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria doi: 10.1038/sj.hdy.6800737.
Rice WR. 1996. Evolution of the Y Sex chromosome in animals. Bioscience. 46:331–343. doi: 10.2307/1312947.
Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution. 41:911–914. doi: 10.2307/2408899.
Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Appl. Note. 26:139–140. doi: 10.1093/bioinformatics/btp616.
Samils N et al. 2013. Sex-linked transcriptional divergence in the hermaphrodite fungus *Neurospora tetrasperma*. Proc. R. Soc. B Biol. Sci. 280:20130862.
Schäfer AM, Kemler M, Bauer R, Begerow D. 2010. The illustrated life cycle of Microbotryum on the host plant *Silene latifolia*. Botany. 88:875–885. doi: 10.1139/b10-

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

061.
Tirosh I, Barkai N, Verstrepen KJ. 2009. Promoter architecture and the evolvability of gene expression. J. Biol. 8:1–6. doi: 10.1186/jbiol204.
White MA, Kitano J, Peichel CL. 2015. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. Mol. Biol. Evol. 32:1981–1995. doi: 10.1093/molbev/msv078.
Wray GA et al. 2003. The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 20:1377–1419. doi: 10.1093/molbev/msg140.
Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. Nat. Commun. 7:12087. doi: 10.1038/ncomms12087.
Wright AE, Moghadam HK, Mank JE. 2012. Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. Genetics. 192:1433–1445. doi: 10.1534/genetics.112.145102.
Xu L et al. 2019. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat. Ecol. Evol. 3:834–844. doi: 10.1101/454843.
Yang X et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. 16:995–1004. doi: 10.1101/gr.5217506.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591. doi: 10.1093/molbev/msm088.
Zhou Q, Bachtrog D. 2012. Chromosome-wide gene silencing initiates Y degeneration in *Drosophila*. Curr. Biol. 22:522–525. doi: 10.1016/j.cub.2012.01.057.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

**Table 1.** Output of a reduced best-fit generalized linear model (GLM) with differential gene expression ($|Log2(a_1/a_2)|$) as the response variable and the following predictable variables: genomic compartment and various degeneration traits, i.e., non-synonymous substitution rate (*dN*), transposable element (TE) insertions, protein length, intron content and GC content. P values <0.05 are in bold. NA: not applicable.

| Explanatory variables and interaction terms | GLM model output parameter | | | |
|---|---|---|---|---|
| | Wald Chi-Square | Degree of freedom (df) | P value | Regression coefficient |
| (Intercept) | 496.78 | 1 | **<0.001** | NA |
| Compartment | 20.151 | 3 | **<0.001** | NA |
| *dN* | 13.21 | 1 | **<0.001** | 5.081 |
| TE insertions | 8.405 | 1 | **0.004** | 0.044 |
| Protein length | 0.41 | 1 | 0.522 | 10.612 |
| Intron content | 10.209 | 1 | **0.001** | 0.768 |
| GC content | 4.233 | 1 | **0.040** | 0.499 |
| Compartment * Protein length | 24.662 | 3 | **<0.001** | NA |
| dN * Protein length | 13.36 | 1 | **<0.001** | -50.726 |
| TE insertions * Protein Length | 8.398 | 1 | **0.004** | -0.37 |
| GC content * Protein length | 10.801 | 1 | **0.001** | -3.962 |

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

**Table 2**. Numbers and percentages of genes with differential expression (DE) on the different genomic compartments on mating-type chromosomes and autosomes, and Fisher's exact test for even distribution between DE genes on autosomes and other genomic compartments, including pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). *P* values <0.05 are in bold. NA: not applicable.

| | **Autosomes** | **PAR** | **Youngest strata** | **Oldest Strata** |
|---|---|---|---|---|
| DE gene number | 507 | 12 | 1 | 74 |
| Total number | 8207 | 114 | 29 | 198 |
| Percentage | 6.18% | 10.53% | 3.45% | 37.37% |
| Fisher's exact test (*P* value) | NA | 0.085 | 1 | **2.20E-16** |

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

## Figure Legends

**Fig. 1. Heatmap and hierarchical clustering of differentially expressed genes (false discovery rate, FDR < 0.05) between haploid $a_1$ and $a_2$ cultures of *Microbotryum lychnidis-dioicae* under a low nutrient condition.**

Each column shows a replicate for each haploid cell culture. The Z-score denotes the relative gene expression level, with blue and red representing high and low expression, respectively. On each node of the clustering tree, bootstrap support values are shown based on 10,000 replicates.

**Fig. 2. Interaction plots for pairs of explanatory variables in overall GLM (generalized linear model) of differential gene expression between mating types of *Microbotryum lychnidis-dioicae*.**

Y-axes are GLM-predicted response values of differential expression ratio between alleles in $a_1$ and $a_2$ haploid genomes, and x-axes are allele differences between alleles in $a_1$ and $a_2$ haploid genomes in predicted protein length as the predictor variable, then binned into levels of interacting categorical predictor variables (i.e. panel **A**, genomic compartment) or other interacting continuous predictor variables (i.e. panels **B-D**; the lowest bin being no differences between alleles, and low and high bins being split at the median value among genes with non-zero differences between alleles). (**A**) Interaction plot between protein length differences and genomic compartment. Genomic compartments include autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata. (**B**) Interaction plots between protein length differences and differences in transposable element (TE) insertions. (**C**) Interaction plots between protein length differences and non-synonymous substitution (*dN*) rate differences. (**D**) Interaction plots between protein length differences and GC content differences.

**Fig. 3. Comparisons of differentially expressed (DE) versus non-differentially expressed (non-DE) genes between mating types of *Microbotryum lychnidis-dioicae* for various degeneration-associated traits within genomic compartments.**

(**A**) Non-synonymous sequence divergence, *dN*, between alleles of DE and non-DE genes. (**B**) Transposable element (TE) insertion number differences between alleles within 20kb (up and downstream) of DE and non-DE genes. (**C**) Proportions of differentially expressed (DE) and non-differentially expressed (non-DE) genes with different protein lengths between alleles. (**D**) Intron content proportional differences between alleles of DE and non-DE genes. (**E**) Total



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

GC content (GC0) proportional differences between alleles of DE and non-DE genes. Analyzed allele differences represent absolute value comparisons (i.e. unoriented with regard to allelic expression levels). Comparisons in panels A, C-E reflect Wilcoxon rank sum tests; panel B reflects a two-proportion z-test. Significance levels shown as, ***: $P < 0.001$, *: $P < 0.05$; non-significant test results shown in Supplementary Tables S4, S6-S9. Genomic compartments include autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata. The notation "*a*" indicates that the youngest evolutionary strata contained only one DE gene, precluding comparisons to non-DE genes within this compartment. For boxplot, the horizontal bars (from bottom to top) represent the 25% quartile, median and 75% quartile respectively.

**Fig. 4. Significant predictors of the degree of differential expression between mating types of *Microbotryum lychnidis-dioicae* testing directional effects of degeneration-associated traits.**

(**A**) Relationship between expression ratio and oriented TE insertion differences in the region from 10kb upstream to the gene, where the trait was calculated as the TE number for the allele with lower expression minus the TE number for the allele with higher expression; a positive value thus represented an excess of TEs in the allele with lower expression. (**B**) Relationship between expression ratio and oriented predicted protein length differences, where the trait was calculated as the ratio of the length for the allele with higher expression divided the length for the allele with lower expression; a larger ratio thus represented a shorter length for the allele with lower expression.

**Fig. 5. Average indel numbers and proportions of genes with different stop codon positions between alleles of differentially-expressed genes of *Microbotryum lychnidis-dioicae*.**

Among genes having alleles with different predicted protein lengths, boxplot of average indel numbers for both differentially expressed (DE, in black) and non-DE genes (in grey) across various genomic compartments (**A**), and barplots for proportions of genes with different stop codon positions for both DE and non-DE genes across genomic compartments (**B**). **: $P < 0.01$, *: $P < 0.05$, '.': $P < 0.1$, NS: not significant. Genomic compartments correspond to autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata. For

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

boxplot, the horizontal bars (from bottom to top) represent the 25% quartile, median and 75% quartile respectively.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

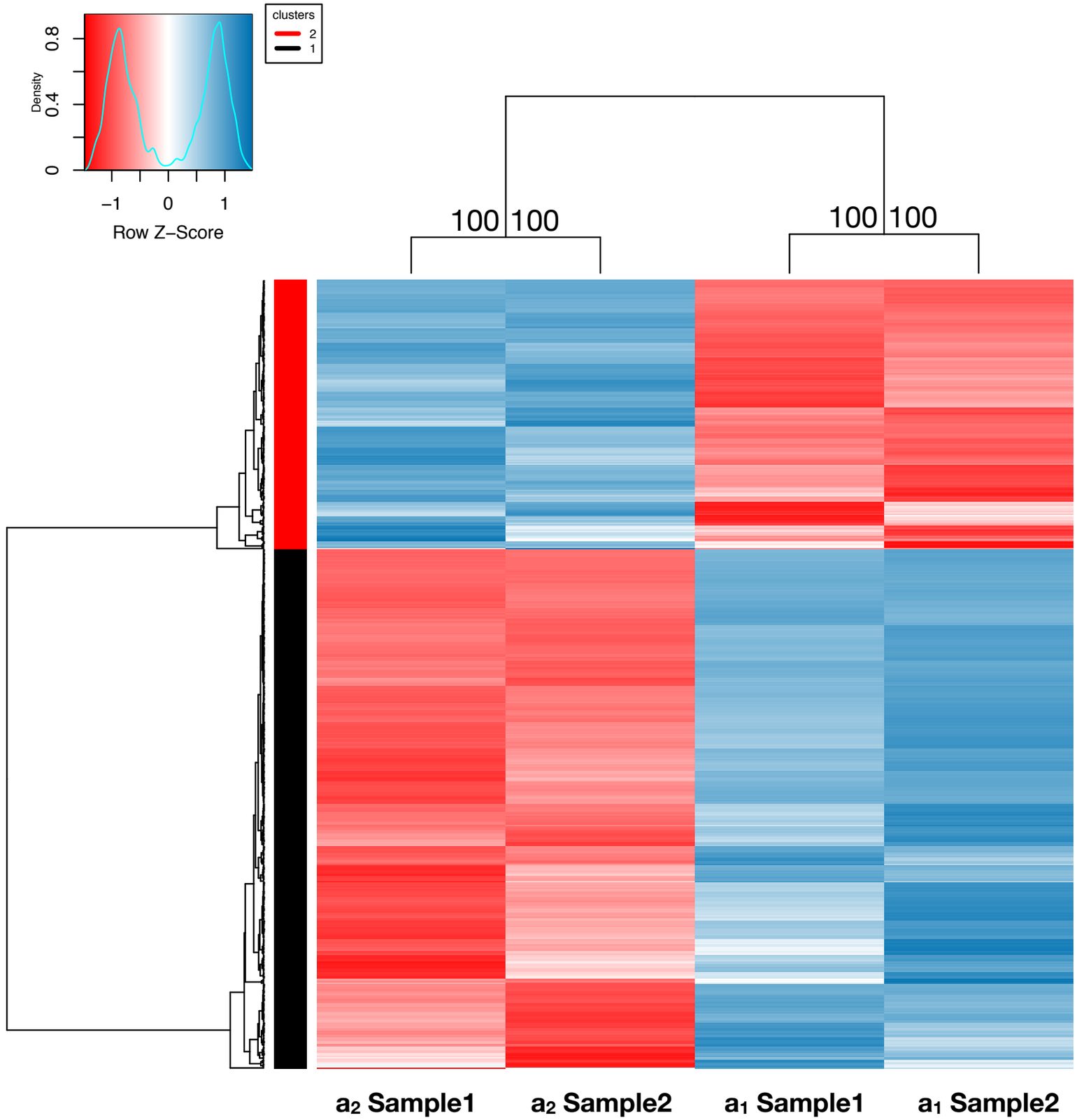
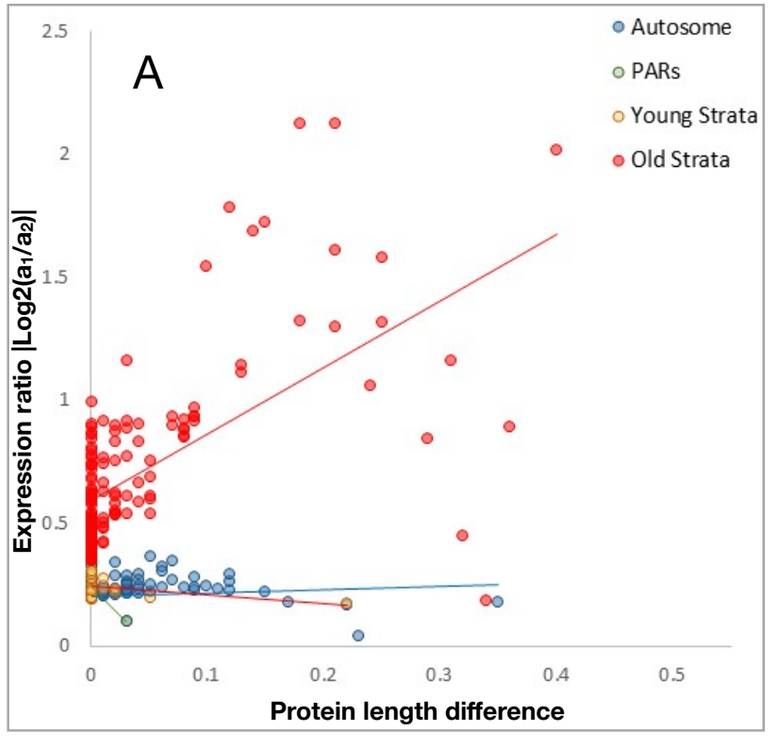
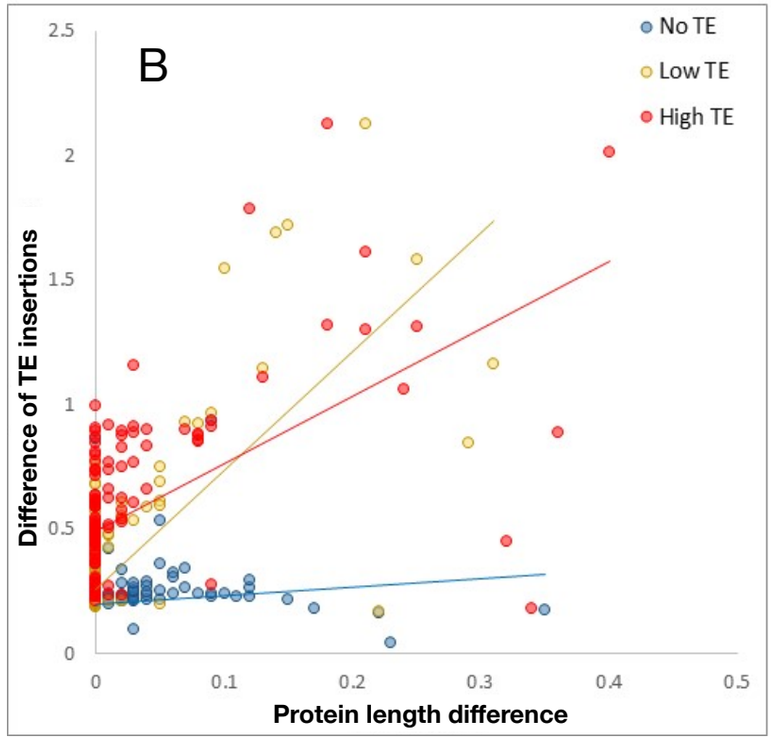


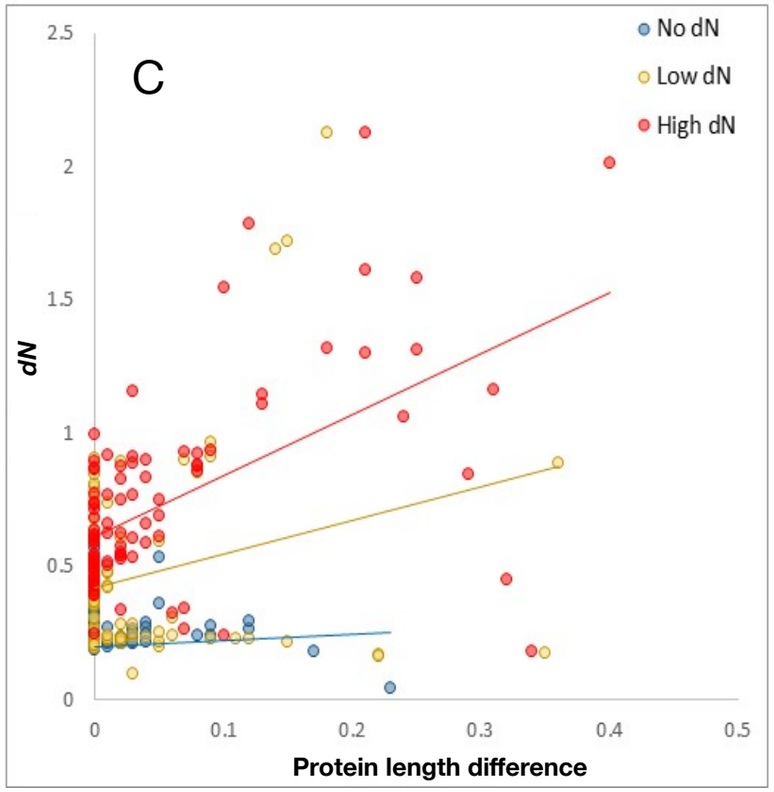

Fig 1



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020



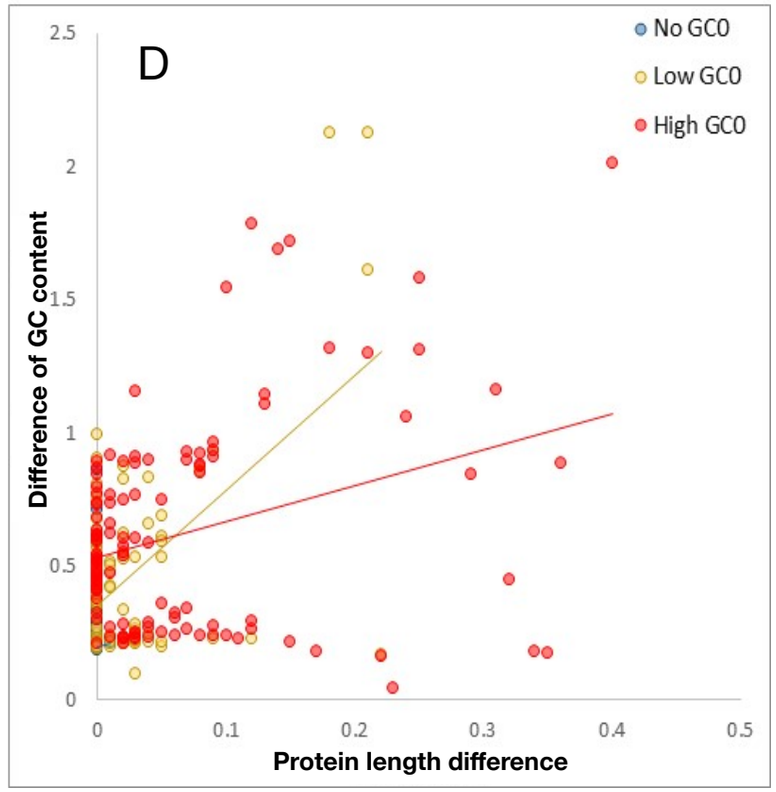




Fig 2



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

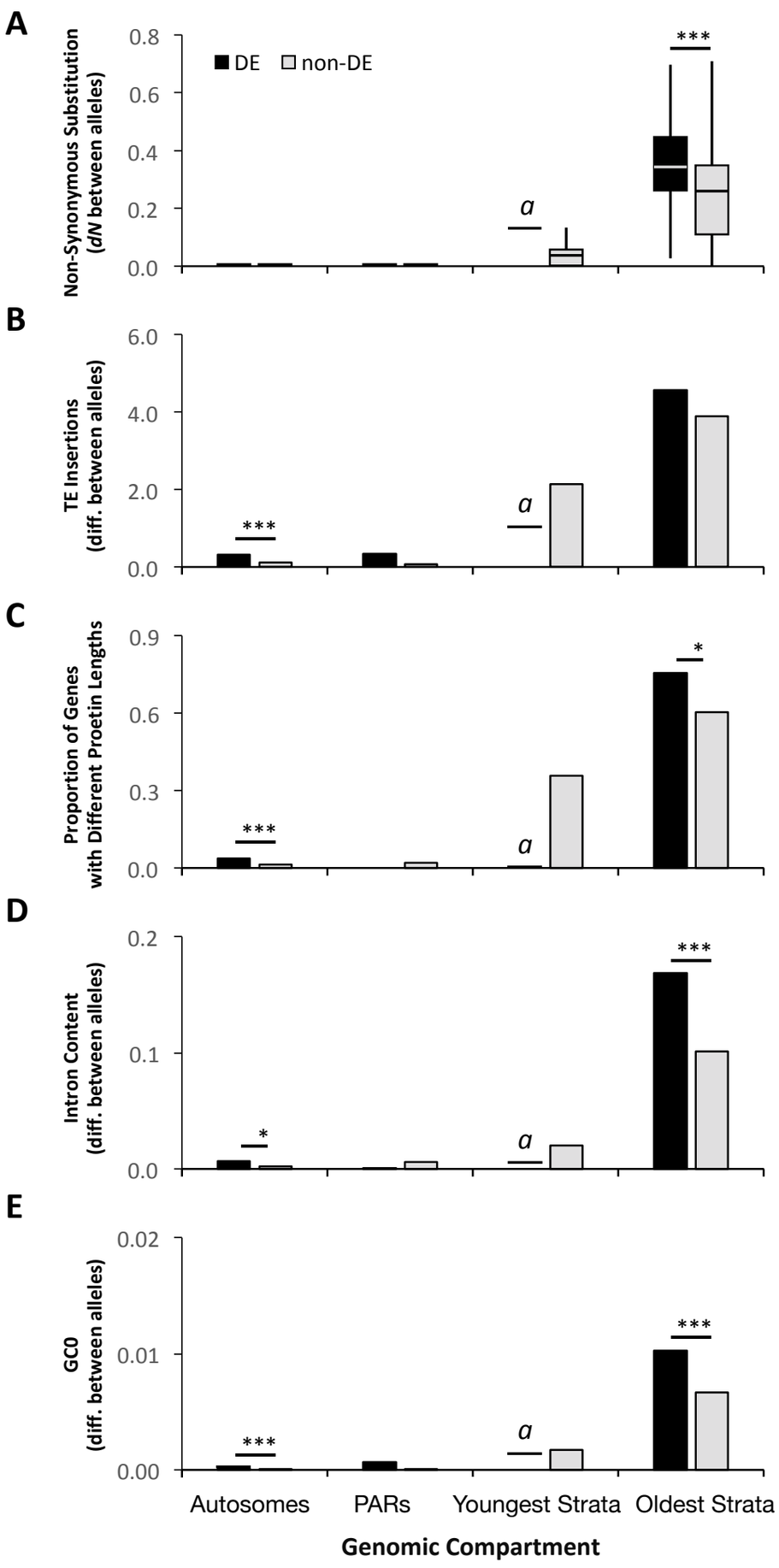



Fig 3

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

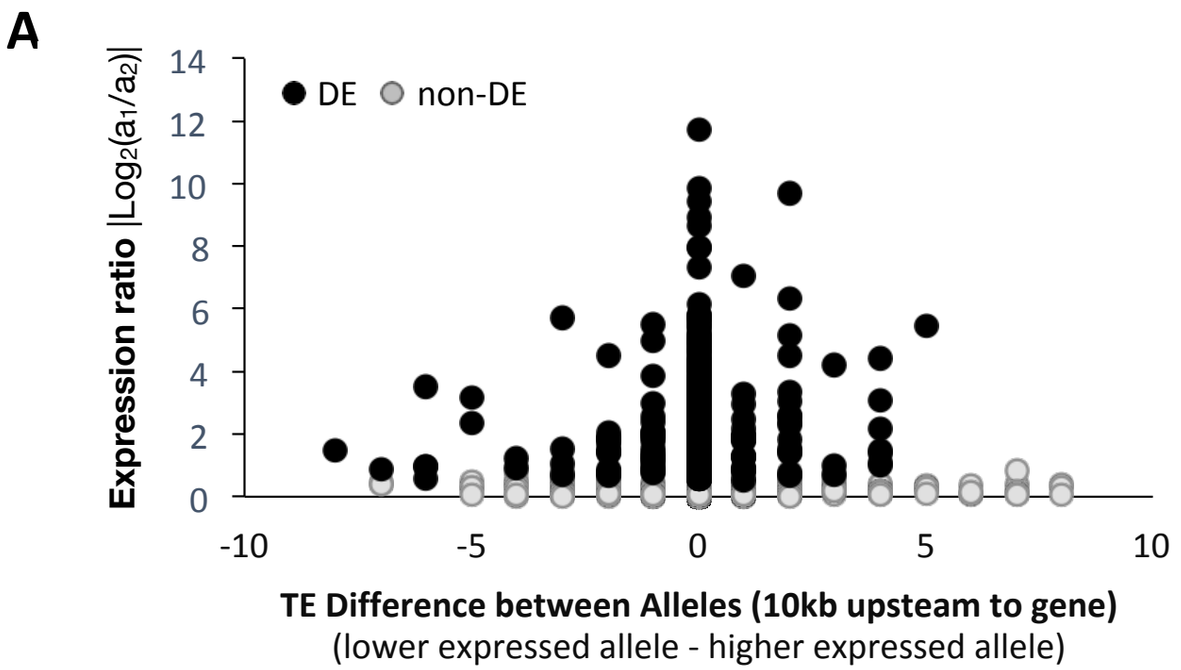



Fig 4



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020

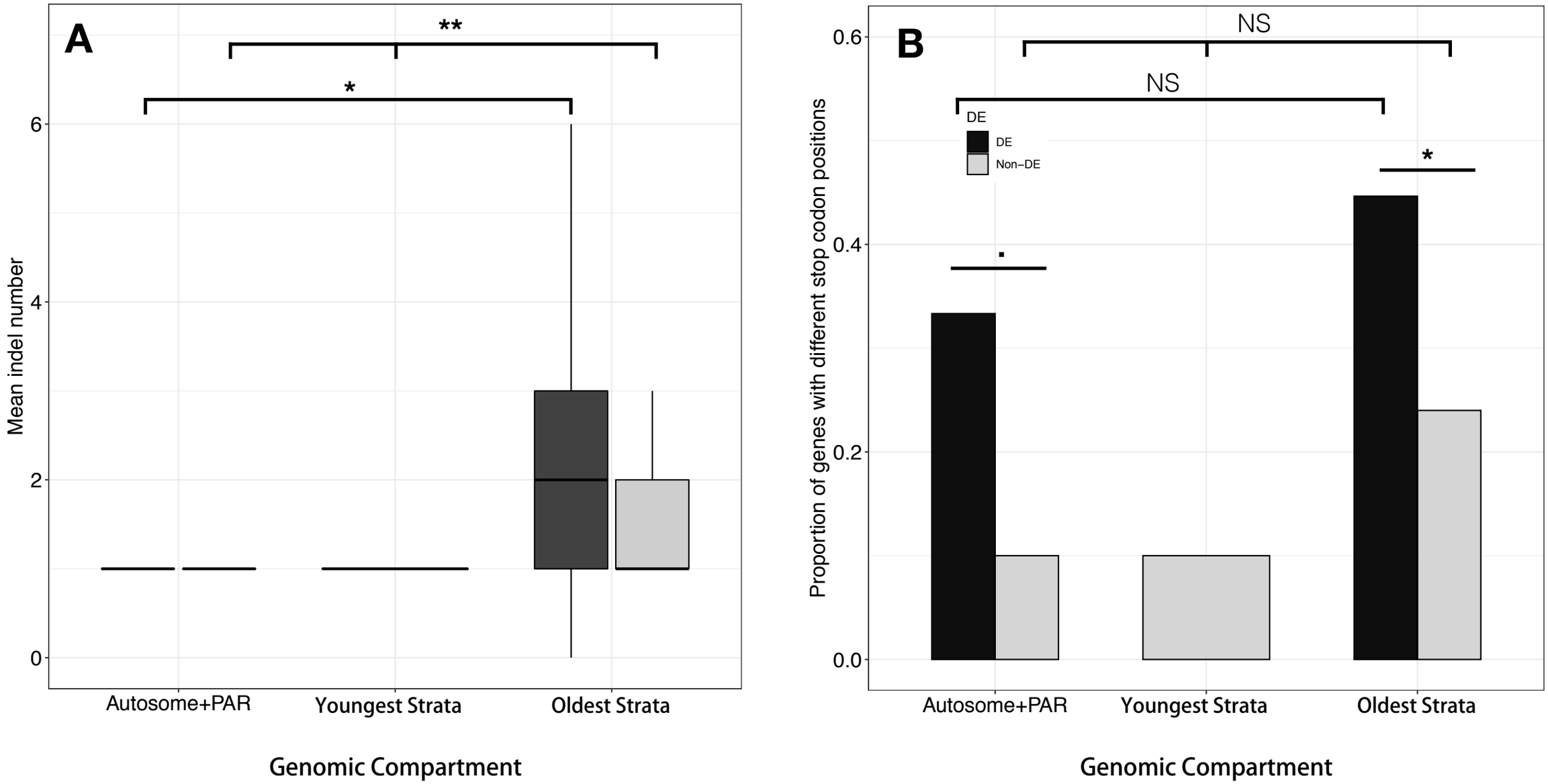



Fig 5



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa028/5736135 by Columbia University user on 10 March 2020