



Nucleus-specific expression in the multinuclear mushroom-forming fungus *Agaricus bisporus* reveals different nuclear regulatory programs

Thies Gehrman^{a,1}, Jordi F. Pelkmans^b, Robin A. Ohm^b, Aurin M. Vos^{b,2}, Anton S. M. Sonnenberg^c, Johan J. P. Baars^c, Han A. B. Wösten^b, Marcel J. T. Reinders^a, and Thomas Abeel^{a,d,1}

^aDelft Bioinformatics Lab, Delft University of Technology, 2628XE Delft, The Netherlands; ^bMicrobiology, Utrecht University, 3584CH Utrecht, The Netherlands; ^cPlant Breeding, Wageningen University and Research, 6708PB Wageningen, The Netherlands; and ^dBroad Institute of MIT and Harvard, Cambridge, MA 02142

Edited by Mitsuhiro Yanagida, Okinawa Institute of Science and Technology Graduate University, Onna-Son, Kunigami, Japan, and approved March 22, 2018 (received for review December 13, 2017)

Many fungi are polykaryotic, containing multiple nuclei per cell. In the case of heterokaryons, there are different nuclear types within a single cell. It is unknown what the different nuclear types contribute in terms of mRNA expression levels in fungal heterokaryons. Each cell of the mushroom *Agaricus bisporus* contains two to 25 nuclei of two nuclear types originating from two parental strains. Using RNA-sequencing data, we assess the differential mRNA contribution of individual nuclear types and its functional impact. We studied differential expression between genes of the two nuclear types, P1 and P2, throughout mushroom development in various tissue types. P1 and P2 produced specific mRNA profiles that changed through mushroom development. Differential regulation occurred at the gene level, rather than at the locus, chromosomal, or nuclear level. P1 dominated mRNA production throughout development, and P2 showed more differentially up-regulated genes in important functional groups. In the vegetative mycelium, P2 up-regulated almost threefold more metabolism genes and carbohydrate active enzymes (cazymes) than P1, suggesting phenotypic differences in growth. We identified widespread transcriptomic variation between the nuclear types of *A. bisporus*. Our method enables studying nucleus-specific expression, which likely influences the phenotype of a fungus in a polykaryotic stage. Our findings have a wider impact to better understand gene regulation in fungi in a heterokaryotic state. This work provides insight into the transcriptomic variation introduced by genomic nuclear separation.

nuclear-specific expression | fungi | heterokaryon | RNA-seq | quantification

Fungi are vital to many ecosystems, contributing to soil health, plant growth, and nutrient recycling (1). They are key players in the degradation of plant waste (2, 3), and they form mutually beneficial relationships with plants by sharing minerals in exchange for carbon sources (4, 5) and by inhibiting the growth of root pathogens (6, 7). They even form networks between plants, which can signal each other when attacked by parasites (8). However, many are plant pathogens (9) that threaten our crops and food security.

The genome organization of fungi is incredibly diverse and can change during their life cycle. For instance, sexual spores can be haploid with one or more nuclei or can be diploid. Sexual spores of mushroom-forming fungi are mostly haploid, and they form monokaryotic (one haploid nucleus per cell) or homokaryotic (two or more copies of genetically identical haploid nuclei) mycelia upon germination. Mating between two such mycelia results in a fertile dikaryon (one copy of the parental nuclei per cell) or heterokaryon (two or more copies of each parental nuclei) when they have different mating loci (10). In contrast to eukaryotes of other kingdoms, the nuclei do not fuse into di- or polyploid nuclei but remain side by side during the main part of

the life cycle. Only just before forming spores in mushrooms do these nuclei fuse, starting the cycle anew.

Agaricus bisporus is the most widely produced and consumed edible mushroom in the world (3). Heterokaryotic mycelia of the button mushroom *A. bisporus* var. *bisporus* (*Sylvan A15* strains) have between two and 25 nuclei per cell (11, 12) (Fig. 1A). The genomes of both ancestral homokaryons have been sequenced (1, 13), showing that DNA sequence variation is associated with different vegetative growth capabilities (1). Due to the two nuclear types, each gene exists at two alleles separated by nuclear membranes, which we call karyollele pairs. Although there have been a few studies investigating the expression of genetic variety in the transcriptome (14, 15), the differential transcriptomic activity of two (or more) nuclear types has never been systematically investigated in a heterokaryon at the genome-wide scale. Based on SNPs identified in mRNA sequencing, it has been suggested that allele-specific expression is tightly linked to the ratio of the nuclear types in a basidiomycete (16).

Allele-specific expression in mononuclear cells has been studied in fungi (17), plants (18), animals (19), and humans (20). Such

Significance

Fungi are a broad class of organisms that play crucial roles in a wide variety of natural and industrial processes. Some are also harmful, destroying crops or infecting immunocompromised patients. Many fungi, at some point during their life cycle, contain two different nuclei, each with different genetic content. We examine the regulation of genes from these nuclei in a mushroom-forming fungus. We find that these nuclei contribute differently to the regulation of the fungal cells, and may therefore have a different impact on their environment. Furthermore, these differences change throughout the development of different tissues. This work contributes to our understanding of fungal physiology by examining this process.

Author contributions: T.G., H.A.B.W., M.J.T.R., and T.A. designed research; T.G., J.F.P., R.A.O., A.M.V., A.S.M.S., J.J.P.B., H.A.B.W., M.J.T.R., and T.A. performed research; T.G. and R.A.O. analyzed data; and T.G., H.A.B.W., M.J.T.R., and T.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Methylation data have been deposited in the Sequence Read Archive (accession no. SAMN06284058).

¹To whom correspondence may be addressed. Email: t.abeel@tudelft.nl or t.gehrmann@tudelft.nl.

²Present address: Department of Biotechnology, Delft University of Technology, 2629 HZ Delft, The Netherlands.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1721381115/-DCSupplemental.

Published online April 11, 2018.

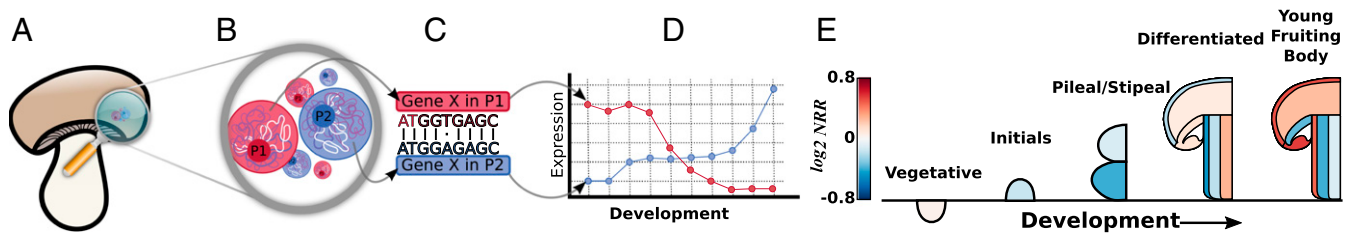


Fig. 1. Nuclear type-specific expression in *A. bisporus*. (A) *A. bisporus* mushroom is composed of different tissues that consist of hyphae composed of cellular compartments. (B) Each cellular compartment is a heterokaryon containing between two and 25 nuclei. In our strain, each nucleus is either of type P1 (red) or P2 (blue). While both nuclear types are haploid, due to the presence of multiple nuclear types, there may be multiple copies of each gene in each cell. (C) Furthermore, the gene in the two types, which we call karyolleles, may differ in their genetic sequences. (D) These differences in transcript sequence allow us to quantify expression of each karyollele in each tissue and to investigate nucleus-specific expression. (E) NRRs of *A. bisporus* throughout its development. The red color indicates higher P1 activity, and the blue color indicates higher P2 activity. The scale bar indicates the log-twofold change in activity between the P1 and P2 nuclear types.

studies have shown that allele heterogeneity is linked to differential allele expression and *cis*-regulatory effects (19–21), and even subgenome dominance (22). In many ways, *A. bisporus* is an excellent model organism with which to investigate differential karyollele expression. It only has two nuclear types in the heterokaryon in contrast to the mycorrhizae, which can have more nuclear types (23), making computational deconvolution of mRNA sequence data tractable. Additionally, the recently published genomes of the two nuclear types of *Sylvan A15* (13) exhibit an SNP density of 1 in 98 bp, allowing differentiation of transcripts in high-throughput sequencing data. Finally, bulk RNA-sequencing (RNA-seq) datasets of different stages of development and of different tissues of the fruiting bodies are available (3, 24).

Here, we show that differential karyollele expression exists in the *A. bisporus Sylvan A15* strain, which changes across tissue type and development, and affects different functional groups. Further, we show that differential karyollele expression associates with differential methylation states, suggesting that epigenetic factors may be a cause for the differential regulation of karyolleles.

Results

Detecting Karyollele-Specific Expression Through Sequence Differences.

To assign expression levels to individual karyolleles, we exploit sequence differences between karyollele pairs in the P1 and P2 homokaryon genomes of the *A. bisporus A15* strain. Briefly, the sequence differences define marker sequences for which the RNA-seq reads uniquely match to either the P1 or the P2 variant, effectively deconvolving the mRNA expression from the two nuclear types (*Materials and Methods*). In total, there are 5,090 distinguishable karyollele pairs between the P1 and P2 genomes, corresponding to ~46% of all genes. The remaining genes could not be unambiguously matched, or the karyollele pairs had too few sequence differences. Most (80%) distinguishable karyollele pairs had the same number of markers in each homokaryon. For the remaining pairs (20%), the number of markers per karyollele was different (*SI Appendix A*). Karyollele-specific expression is expressed as a read count ratio that reflects the relative abundance of mRNAs originating from the P1 or P2 nuclear type (*Materials and Methods*).

We studied *A. bisporus*' karyollele specific expression through mushroom development in two RNA-seq datasets: first, a mushroom tissue dataset comprising 17 different mycelial and mushroom tissue conditions throughout the development of the mushroom (in duplicate; *Materials and Methods* and Fig. 1E) and, second, a dataset of six stages of vegetative mycelium growth in compost throughout mushroom harvest (in duplicate; *Materials and Methods* and Fig. 2B). Measured difference in expression between nuclear types was not correlated with the number of markers ($P > 0.05$) for any of the samples, and it was not correlated with GC (guanine-cytosine) content (*SI Appendix B*). Of the 5,090 differentiable karyolleles, we observed expression of 5,029 (99%) in at least five conditions.

P1 and P2 mRNA Production Differs per Tissue and Across Development.

We assess the total mRNA production of the P1 and P2 nuclear types and their relative contributions to total sample mRNA during development. To do this, we considered the total number of reads uniquely matching to P1 with respect to P2. Fig. 1E shows that this nuclear type read count ratio (NRR; *Materials and Methods*) changes throughout development and across tissue types. For example, during the “differentiated” stage, the P2 nuclei are dominant in the skin, but in the “young fruiting body,” the P1 nuclei dominate in the skin (Fig. 1E, two rightmost panels). In contrast, the “stipe center” is dominated by P1 nuclei in the differentiated stage, while the expression of P2 nuclei dominates later.

The transcription patterns throughout mushroom development differ between the karyolleles. Based on a principal component analysis of the expression profiles of each nuclear type, we observe that the expression profiles of P1 and P2 group together in different clusters (*SI Appendix C*). This clustering is indicative of distinct regulatory programs. The first principal component appears to represent the tissue type, and the second represents the nuclear type. Interestingly, measurements of the same tissue from P1 and P2 do not have the same value for the first principal component, indicating that the difference in nuclear type does not entirely explain the variation between P1 and P2.

Within a Sample, mRNA Production of P1 and P2 Varies Between Chromosomes.

Fig. 3A shows the chromosome read count ratios (CRRs; *Materials and Methods*), demonstrating that some chromosomes are more active in P1 (e.g., chromosome 8) throughout development, while others are more active in P2 (e.g., chromosome 9). Expression of other chromosomes depends on the developmental state, changing in time (e.g., chromosome 2). The chromosome log-twofold changes lie between [−0.60, 0.79]. In the vegetative mycelium, we see less drastic differences in mRNA production throughout development than in the mushroom tissues, with expression log-twofold changes between [−0.28, 0.36] (Fig. 3A).

Gene Read Ratios Reveal a Dominant P1 Type in Mushroom Tissue, but Not in Mycelium.

To investigate whether either nuclear type is truly dominant, we correct for extremely highly expressed genes (*SI Appendices D and E*) by limiting their impact on the chromosome and tissue level ratios by using per-gene activity ratios per chromosome (CGRs; *Materials and Methods*), instead of read ratios. This revealed that in addition to P1 producing more mRNA than P2, P1 karyolleles were more frequently higher expressed than their P2 counterparts (Fig. 3B). Looking across all tissues and chromosomes, P1 is significantly dominant over P2, (i.e., the average of the log-transformed CGR is significantly larger in the P1 nuclear type than in the P2 nuclear type, following a *t* test in mushroom tissue, with $P < 0.01$; *SI Appendix F*). Using the CGR has a notable impact on chromosome 9. Although P2 produces most chromosome 9 mRNA (Fig. 3A), it is not the case that more

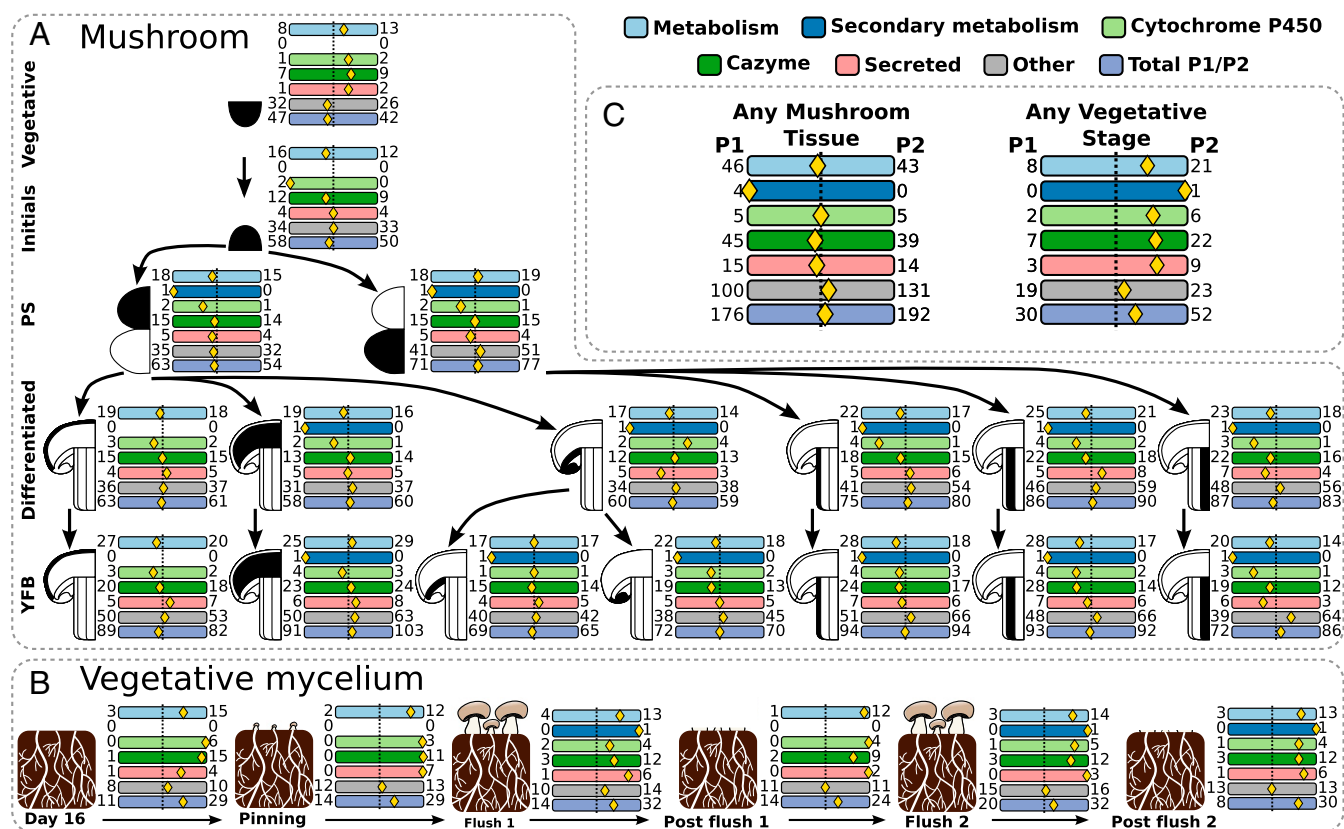


Fig. 2. Differential regulation of functional groups through mushroom development. The development of different tissues is illustrated as a tree. We investigate several functional groups. At each developmental stage, we observe how many genes of each group are differentially up-regulated in P1 (Left) and in P2 (Right). The yellow diamonds indicate the ratios of these counts. We see that the groups are more or less equally distributed between P1 and P2 (the yellow diamonds are centered), with the exception of the vegetative stage (the root node of A) and the vegetative mycelium dataset. (A) Mushroom dataset. (B) Vegetative mycelium dataset. (C) Number of differentially expressed genes across all samples in the two datasets.

P2 karyolleles are more highly expressed than P1 karyolleles. In the mycelium (Fig. 3), we do not observe such a dominance of P1 ($P > 0.05$, with a t test as in the mushroom dataset), where neither P1 nor P2 shows dominant mRNA activity.

To verify that the dominance we observe is not a result of a nuclear type imbalance, we investigated the expression of housekeeping genes across the samples. Two genes have previously been used as housekeeping genes in *A. bisporus*: 18S ribosomal RNA and *gpdII* (25). The 18S ribosomal RNA was not differentiable between the two nuclear types. The log-twofold changes of expression between the two nuclei we observed for *gpdII* did not have a mean significantly different from 0 ($P > 0.3$, by a one-sample t test), indicating that the nuclear ratio is, on average, equal across the nuclei. Further, we identified 50 genes that, within each nuclear type, had a measured expression of at least 50, and had an SD of expression within 10% of the mean across all samples. For these 50 genes, we found that the mean of the log-twofold changes between the two nuclei was not significantly different from 0 ($P > 0.6$, by a one-sample t test), supporting our observations for *gpdII*.

A Substantial Portion of Karyolleles Are Differentially Expressed. In each sample, we determined the set of karyolleles that are significantly differentially expressed between the two nuclear types. Although the dominance of the P1 nuclear type indicates a general trend of higher activity across many genes, some karyollele pairs have a much larger difference, indicating a functional role. In total, we find 411 genes that are differentially expressed (Materials and Methods) in a mushroom tissue or in vegetative mycelium throughout development (Fig. 4A); 368 genes are differentially expressed in mushroom tissues, and 82 are differentially expressed in the vegetative

mycelium. Using bisulfite sequencing of DNA isolated from vegetative mycelium, we identified methylation sites that indicate differential methylation states between the P1 and P2 nuclear types (SI Appendix G). The set of differentially expressed genes is enriched for genes with mixed methylation states (277 differentially methylated genes, and 42 of these are differentially expressed). Remarkably, if a gene is observed to be more highly expressed in P1 than in P2, it will never be observed to be more highly expressed in P2 than in P1 in other tissues, and vice versa (with only one exception, which is differentially expressed in both datasets; SI Appendix H). This indicates a distinct preference for one karyollele by the regulatory machinery across development.

Thirty-nine genes are differentially expressed between the nuclear types in both the mushroom and mycelium. In this intersection set, 10 genes had higher expression in P1 and 24 had higher expression in P2. Five were more highly expressed in P2 in the mycelium but switched to P1 in the mushroom (SI Appendix H). The small overlap is indicative of different regulatory processes during the vegetative stage and a mushroom stage. Interestingly, of the 90 named genes in *A. bisporus* (Materials and Methods), only *mnpI*, an important gene in lignin degradation, is differentially expressed and exhibits different behavior in the mushroom and the vegetative mycelium (SI Appendix I).

Although P2 up-regulates more differentially expressed genes than P1 does (Fig. 4A), more genes show a consistently higher expression in P1 than in P2 (Fig. 4B). We identify consistently higher expressed genes that show a higher expression in one nuclear type over the other across all samples (Materials and Methods). This results in a dominant mRNA contribution from P1.

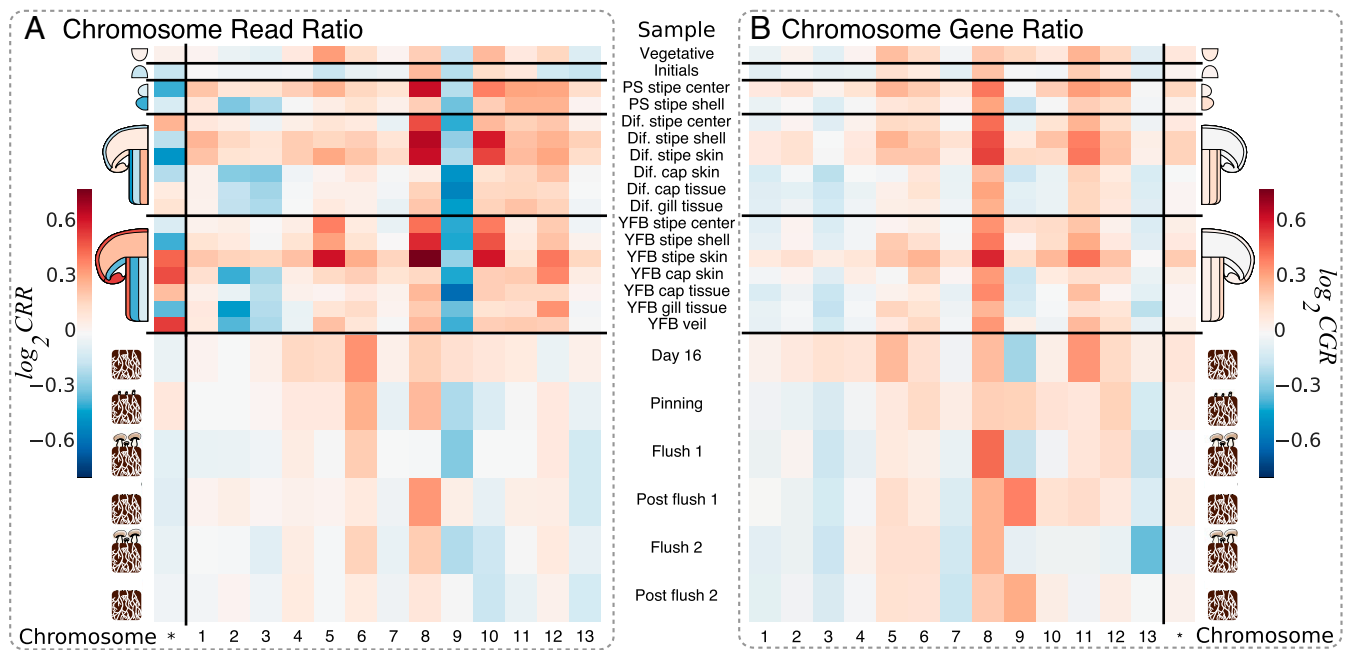


Fig. 3. P1 versus P2 expression per chromosome throughout development of the mushroom. The red color indicates higher P1 activity, and the blue color indicates higher P2 activity. Each row indicates a different developmental stage, and each column represents a different chromosome. The column specified with an asterisk shows the NRRs. Both the mushroom tissues and the mycelium datasets are shown. (A) CRRs. (B) CGRs. Dif., differentiated; PS, pileal/stipeal; YFB, young fruiting body.

Colocalized Gene Clusters Are Coregulated. To determine the level at which genes are regulated, we investigated whether there are regions where the majority of genes were consistently more highly expressed in one homokaryon than in the other. We detected many such regions, given in Fig. 4C and *SI Appendix J*, hinting toward a sub-chromosomal level of regulation. This is supported by observations in Fig. 3, where we see that chromosomes are differently regulated, even within one tissue, excluding regulation at the nuclear level. Our observation of colocalized regions being coregulated contradicts the notion of a chromosome-level regulation, where we would have observed chromosome-wide coregulation.

Coregulated regions are more frequently up-regulated for the P1 karyotype than for the P2 karyotype. This observation is in agreement with the observed P1 nuclear type dominance. We observe relatively little overlap between the mushroom and vegetative mycelium datasets (*SI Appendix J*), indicative of different regulatory programs between the vegetative mycelium and mushroom tissue cells.

Broad Range of Functionality Affected by Karyotype-Specific Expression Throughout Development. To examine the functional annotations of the differentially expressed karyotype pairs, we considered the following categories: (i) transcription factors, (ii) metabolism genes, (iii) secondary metabolism genes, (iv) cytochrome P450 genes, (v) carbohydrate active enzymes (cazymes), and (vi) secreted proteins. These categories, with the exception of secondary metabolism genes, are all enriched in the set of differentiable genes [$P < 0.05$ by a χ^2 approximation to the Fisher's exact test with false discovery rate (FDR) correction].

Fig. 2 shows the division of the 411 differentially expressed genes across the functional categories in all of the different samples. None of the differentially expressed genes were transcription factors. For the other functional categories, we saw a more or less equal amount of up-regulated karyotypes in P1 and P2 in the mushroom tissues (Fig. 2A and C, except in the vegetative stage) and a more skewed distribution of activity in the mycelium dataset (Fig. 2B and C), as well as in the vegetative stage of the mushroom dataset. In these cases, P2 had more differentially expressed genes in these functional categories (Fig. 2C).

The P2 type had higher expression of significantly more karyotypes than P1 in the mycelium (*SI Appendix K*). In the mycelium, P2 had an enriched expression of cytochrome P450 genes, secondary metabolism genes, and cazymes ($P < 0.05$, with an FDR-corrected χ^2 approximation to the Fisher's exact test). Furthermore, cazymes and metabolism genes in the mycelium were more likely to be more highly expressed in P2 ($P < 0.05$, with an FDR-corrected binomial test).

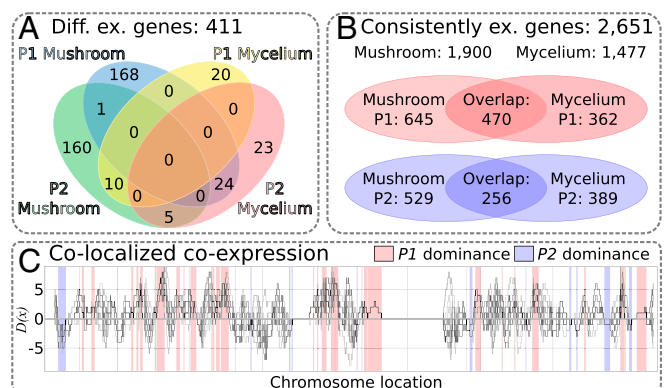


Fig. 4. (A) Sets of differentially (Diff.) expressed genes in P1 and P2 across the mushroom and vegetative mycelium datasets. Each set represents the number of genes up-regulated in one nuclear type per dataset. (B) Relationship between sets of consistently expressed (ex.) genes between the mushroom and vegetative mycelium datasets. (C) Colocalized genes are often coregulated. Pictured here are the colocalized and coregulated gene clusters along chromosome 10 in the mushroom tissue dataset. Along the x axis is the genomic coordinate. For each sample (gray lines), we plot the difference between the number of genes more highly expressed by P1 and by P2 (a value of 0 indicates an equal distribution). We also highlight the regions that are consistently up-regulated in P1 (red regions) and the number of genes that are consistently up-regulated in P2 (blue regions). $D(x)$, the number of consistently expressed genes in P1/P2 (*SI Appendix J*).

Nineteen of the 39 previously identified differentially expressed genes shared between the mycelium and mushroom datasets had the following functional annotations: 14 were annotated as metabolism genes, 14 as cazyms, 5 as secreted proteins, and 2 as cytochrome P450s (some genes have multiple annotations). Additionally, five of these 39 genes have different domain annotations, indicating different functional properties between the P1 and P2 karyolleles.

An analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations of the differentially expressed genes revealed several genes involved in the aminoacyl-tRNA biosynthesis (M00359) pathway, suggesting a differential ability to produce tRNAs (*SI Appendix L*).

Next, we studied whether differential expression of a karyollele also resulted in the production of a functionally different protein due to sequence differences between the karyolleles. In total, 216 of the 5,090 distinguishable karyolleles had sequence differences that led to an alternative protein domain annotation, and 36 of these are differentially expressed between P1 and P2 (*SI Appendix M*).

Discussion

Differently from most eukaryotes, nuclei remain side by side during most of the life cycle of basidiomycete fungi. Whether each nucleus is contributing equally to the phenotype and, if not, how this is regulated are largely unknown. To understand this, we studied the expression of alleles in both constituent nuclei of the button mushroom. From the observed average gene expression, we conclude that the expression of nuclear type P1 of the *A. bisporus* *Sylvan A15* strain is dominant. Remarkably, this dominance is present across all developmental stages in the heterokaryon. We can link this phenomenon to human fibroblasts (26), where it has been shown that individual cells preferentially express one allele over the other, which is not evident over a collection of many cells. Whereas in a diploid genome, the cell must rely on heterochromatin DNA packing or RNAi regulatory pathways (27), heterokaryotic cells could instead control the energy usage of a specific nuclear type.

In the mushroom tissue dataset, the number of up-regulated karyolleles in P1 is approximately equal to that in P2, but in the vegetative mycelium dataset, P2 has more up-regulated karyolleles. The contrast between a dominant P1, yet more differentially higher expressed genes in P2 in mushroom tissue is paradoxical. However, many genes show a consistently higher expression in either P1 or P2, with more genes showing a consistently higher expression in P1. Is it possible that the P1 homokaryon is responsible for the basal mRNA production, while P2 plays a more reactive regulatory role? Mechanisms for this kind of regulation are not known. In plants, subgenome dominance may be linked to methylation of transposable elements (22). Does something similar happen in *A. bisporus*?

Although an imbalance in the number of nuclei could explain the dominance of P1, housekeeping genes suggest an equal proportion of nuclear types, on average. Further, we showed that genes that are consistently more highly expressed in one of the karyolleles do colocalize in subchromosomal regions. If there were more P1 nuclei than P2 nuclei, we would have instead expected generally higher expression of genes of one nuclear type across all chromosomes.

For many differentially expressed genes, protein sequence differences between the two karyolleles in the two nuclear types encode for different protein domains. This suggests a functional impact of karyollele-specific expression. We also observe a broad range of differentially expressed functionality between the P1 and the P2 nuclear types. For example, the P2 up-regulation of cazyms and metabolism genes in P2 in compost highlights the importance of the P2 homokaryon in development. H97, one of the homokaryons in the cultivar *Horst U1*, from which *Sylvan A15* is derived, displays stronger vegetative growth characteristics than its counterpart H39 (1). This metabolic strength may be passed down from the H97 homokaryon to the *Sylvan A15* P2 homokaryon, and the differentially expressed karyolleles may in part be responsible for this. The gene *mnp1*, for example, is an

important gene for growth on compost, and P2 has indeed inherited the relevant chromosome 2 from H97 (13). Such characteristics are relevant for breeding strategies.

A significant proportion of differentially methylated karyolleles were also differentially expressed. However, methylation can only explain at most 10% of the observed differential expression. This may be due to a limitation of our methylation dataset (only vegetative growth), but it may also hint toward other regulatory mechanisms.

In addition to methylation, we observe colocalization of coexpressed genes. This may be indicative of a difference in genome organization, whereby the DNA is less accessible in certain regions in P1 than in P2 through different levels of chromatin compaction. It has been shown that gene expression is strongly linked to DNA availability and, further, that such chromatin organization is heritable (28).

The sequences of a pair of karyolleles need to be sufficiently different for our algorithm to be able to uniquely assign reads to each karyollele. These sequence differences between nuclear types may have an effect on various regulatory mechanisms of transcription, such as transcription factor binding efficiencies, transcription efficiency, differences in mRNA stability, or differences in epigenetic factors. Future research might shed light on whether these differences are related to observed differential karyollele expression.

Conclusion

We showed that karyolleles, the different copies of a gene separated by nuclear membranes in a heterokaryon, are differentially expressed between the two different nuclear types in the *A. bisporus* *Sylvan A15* strain. Each nuclear type contributes varying amounts of mRNA to the cell, and thereby the functional contributions also change. The mechanisms behind this process need to be investigated, but we show that it could be partially explained by epigenetic mechanisms.

Heterokaryotic fungi have major impact in clinical and biotechnological applications, and impact our economy and society as animal pathogens such as *Cryptococcus neoformans* (29); plant pathogens such as *Ustilago maydis* (30); plant and soil symbionts such as mycorrhizal fungi (23); bioreactors such as *Schizophyllum commune* (31); and, of course, the subject of this study, the cultivated, edible mushroom *A. bisporus* (13). It is known that different homokaryons in these species will produce different phenotypes (3), which no doubt need to be treated, nourished, or utilized differently. With this work, we hope to draw attention to the impact of sequence and regulatory variation in different nuclei on the function and behavior of the cell to further our understanding of the role of fungi in our environment.

Materials and Methods

RNA-Seq Data. We used two RNA-seq datasets from the *A. bisporus* (*A15*) strain: (i) tissue samples through mushroom development (BioProject accession no. PRJNA309475) from Pelkmans et al. (24) and (ii) vegetative mycelium samples taken from compost through mushroom development (BioProject accession no. PRJNA275107) from Patyshakuliyeva et al. (3). Mushrooms were grown under the same environmental conditions in both studies. Throughout this study, when we refer to the mushroom tissue dataset, we refer to all samples in dataset *i*, including the first sample, which is technically a vegetative mycelium sample. The compost dataset exhibited high amounts of PCR duplicates (*SI Appendix N*), which we removed using FastUniq (32).

Homokaryon Genome and Annotations. The P1 and P2 genomes (13) were annotated with BRAKER1 (33) and AUGUSTUS 3.0.2 (34) using the pooled RNA-seq data from the mushroom tissue dataset (*SI Appendix O*). Named genes are provided in *SI Appendix P*. Functional annotations are provided in *SI Appendix Q*.

Karyollele Pair Discovery. The genome annotations were used to produce predicted mRNA sequences for each gene. The genes in the two parental genomes were matched using a reciprocal best BLAST (35) hit. Hits that had E-values greater than 10^{-100} were removed. This resulted in a conservative

orthology prediction between the two homokaryons that are our set of karyolleles. Pairs with 100% sequence identity were removed, as it would be impossible to identify distinguishing markers for these identical pairs.

Marker Discovery. For each discovered karyollele pair, we identify markers that uniquely identify each element of the pair. This is done by constructing all possible kmers for each sequence, resulting in two sets per pair. The kmers overlapping in these sets are removed, resulting in distinguishing pairs of markers. Once distinguishing markers have been discovered for all pairs, we remove all nonunique markers. Finally, the set of markers is made nonredundant by scanning the position-sorted list of markers from left to right and removing any marker that overlaps with the previous marker. Finally, we ensure that the markers are unique throughout the whole genome by removing markers that are present anywhere else in either genome. To guarantee sufficient evidence across the whole gene, we remove karyollele pairs that do not have at least five markers each.

Marker Quantification. We scan all RNA-seq reads for the detected markers using the Aho–Corasick algorithm (36). We insert all markers and their reverse complements into an Aho–Corasick tree and count each marker only once for each fragment (a marker may be present twice if the read mates overlap). We calculate a gene expression score as the average of each marker count for a gene. This results in an expression score E_h for each gene g in each sample t for each replicate r , per homokaryon h :

$$E_h(r, s, g) = \frac{1}{|M_h(g)|} \sum_{m \in M_h(g)} C_h(r, s, m), \quad [1]$$

where $M_h(g)$ is the set of markers in a gene g in homokaryon h and $C_h(r, s, m)$ is the count for marker m in replicate r , sample s . Differential expression was tested with DESeq (37) (SI Appendix R).

Read Ratio Calculation. Using the normalized read counts from DESeq, we calculate the ratio of the number of reads originating from the two homokaryons at the gene [gene read ratio (GRR)], chromosome (CRR), and nuclear type (NRR) levels. Formulas are provided in SI Appendix S.

Gene Ratio Calculation. We calculate the geometric mean of the GRR of all genes in a chromosome to describe the expression activities at the chromosome level (CGR) and, additionally, across all genes for the nuclear type level (nuclear gene ratio). The geometric mean is more suitable than the arithmetic mean for averaging ratios. Formulas are provided in SI Appendix S.

Software and Code Availability. Marker discovery and abundance calculations were done in Scala, and downstream analysis was performed in Python using Ibdas (38). All source code, together with a small artificial example dataset, is available at <https://github.com/thiesgehrmann/Homokaryon-Expression>.

Data Availability. The RNA-seq data were previously generated and can be found at Bioproject accession nos. PRJNA309475 and PRJNA275107. The bisulfite sequencing data can be accessed in the Sequence Read Archive (SAMN06284058).

ACKNOWLEDGMENTS. We thank Brian Lavrijssen for providing the bisulfite sequencing data to determine the differential methylation states. The sequence and annotation data of *A. bisporus* H97 version 2 were produced by the US Department of Energy Joint Genome Institute (www.jgi.doe.gov) in collaboration with the user community. This research is supported by the Dutch Technology Foundation (STW), which is part of the Netherlands Organization for Scientific Research (NWO) and is partly funded by the Ministry of Economic Affairs.

- Morin E, et al. (2012) Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc Natl Acad Sci USA* 109:17501–17506.
- Ohm RA, et al. (2010) An efficient gene deletion procedure for the mushroom-forming basidiomycete *Schizophyllum commune*. *World J Microbiol Biotechnol* 26:1919–1923.
- Patyshakuliyeva A, et al. (2015) Uncovering the abilities of *Agaricus bisporus* to degrade plant biomass throughout its life cycle. *Environ Microbiol* 17:3098–3109.
- Pawlowska TE (2005) Genetic processes in arbuscular mycorrhizal fungi. *FEMS Microbiol Lett* 251:185–192.
- ud din Khanday M, et al. (2016) Arbuscular mycorrhizal fungi boon for plant nutrition and soil health. *Soil Science: Agricultural and Environmental Perspectives* (Springer International, Cham, Switzerland), pp 317–332.
- Sun C, et al. (2014) The beneficial fungus *Piriformospora indica* protects *Arabidopsis* from *Verticillium dahliae* infection by downregulation plant defense responses. *BMC Plant Biol* 14:268.
- Harrach BD, Baltruschat H, Barna B, Fodor J, Kogel K-H (2013) The mutualistic fungus *Piriformospora indica* protects barley roots from a loss of antioxidant capacity caused by the necrotrophic pathogen *Fusarium culmorum*. *Mol Plant Microbe Interact* 26:599–605.
- Babikova Z, et al. (2013) Underground signals carried through common mycelial networks warn neighbouring plants of aphid attack. *Ecol Lett*, 16, pp 835–843.
- Dean R, et al. (2012) The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 13:414–430.
- Specht CA (1995) Isolation of the B alpha and B beta mating-type loci of *Schizophyllum commune*. *Curr Genet* 28:374–379.
- Saksena KN, Marino R, Haller MN, Lemke PA (1976) Study on development of *Agaricus bisporus* by fluorescent microscopy and scanning electron microscopy. *J Bacteriol* 126:417–428.
- Craig GD, Newsam RJ, Gull K, Wood DA (1979) An ultrastructural and autoradiographic study of stipe elongation in *Agaricus bisporus*. *Protoplasma* 98:15–29.
- Sonnenberg ASM, et al. (2016) A detailed analysis of the recombination landscape of the button mushroom *Agaricus bisporus* var. *bisporus*. *Fungal Genet Biol* 93:35–45.
- Todd RB, Davis MA, Hynes MJ (2007) Genetic manipulation of *Aspergillus nidulans*: Heterokaryons and diploids for dominance, complementation and haploidization analyses. *Nat Protoc* 2:822–830.
- Boon E, Zimmerman E, Lang BF, Hijri M (2010) Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome. *J Evol Biol* 23:1519–1527.
- James TY, Stenlid J, Olson A, Johannesson H (2008) Evolutionary significance of imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus *Heterobasidion parviporum*. *Evolution* 62:2279–2296.
- Muzzey D, Sherlock G, Weissman JS (2014) Extensive and coordinated control of allele-specific expression by both transcription and translation in *Candida albicans*. *Genome Res* 24:963–973.
- Wei X, Wang X (2013) A computational workflow to identify allele-specific expression and epigenetic modification in maize. *Genomics Proteomics Bioinformatics* 11:247–252.
- Crowley JJ, et al. (2015) Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat Genet* 47:353–360.
- Buckland PR (2004) Allele-specific gene expression differences in humans. *Hum Mol Genet* 13:R255–R260.
- Pant PVK, et al. (2006) Analysis of allelic differential expression in human white blood cells. *Genome Res* 16:331–339.
- Edger PP, et al. (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkey-flower. *Plant Cell* 29:2150–2167.
- Lin K, et al. (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS Genet* 10:e1004078.
- Pelkmans JF, et al. (2016) The transcriptional regulator c2h2 accelerates mushroom formation in *Agaricus bisporus*. *Appl Microbiol Biotechnol* 100:7151–7159.
- Vos AM (2017) Compost degradation and growth of *Agaricus bisporus*. Dissertation (Utrecht University, Utrecht, The Netherlands). Available at <https://dspace.library.uu.nl/handle/1874/351375>. Accessed January 27, 2018.
- Borel C, et al. (2015) Biased allelic expression in human primary fibroblast single cells. *Am J Hum Genet* 96:70–80.
- Volpe TA, et al. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833–1837.
- McDaniell R, et al. (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* 328:235–239.
- Loftus BJ, et al. (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307:1321–1324.
- Kämper J, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101.
- Shu CH, Hsu HJ (2011) Production of schizophyllan glucan by *Schizophyllum commune* ATCC 38548 from detoxified hydrolysate of rice hull. *J Taiwan Inst Chem Eng* 42:387–393.
- Xu H, et al. (2012) FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLoS One* 7:e52249.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-et and AUGUSTUS. *Bioinformatics* 32:767–769.
- Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24:637–644.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Aho AV, Corasick MJ (1975) Efficient string matching: An aid to bibliographic search. *Commun ACM* 18:333–340.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.
- Hulsman M, Bot JJ, Vries AP de, Reinders MJT (2013) Ibdas: Querying flexible data structures to explore heterogeneous bioinformatics data. *Data Integration in the Life Sciences, DILS 2013*. Lecture Notes in Computer Science (Springer, Berlin, Heidelberg), Vol 7970, pp 23–37.