1 Nucleus specific expression in the multinucleated mushroom-

2 forming fungus *Agaricus bisporus* reveals different nuclear

3 regulatory programs

4 Thies Gehrmann[1,*], Jordi F. Pelkmans[2], Robin A. Ohm[2], Aurin M. Vos[2,#], Anton S.M.

5 Sonnenberg[3], Johan J.P. Baars[3], Han A. B. Wösten[2], Marcel J. T. Reinders[1], Thomas Abeel[1,4,$]

6 [1]Delft Bioinformatics Lab, Delft University of Technology, The Netherlands

7 [2]Microbiology, Utrecht University, The Netherlands

8 [3]Plant Breeding, Wageningen University and Research, The Netherlands

9 [4]Broad Institute of MIT and Harvard, USA

10 [*]Current position: CBS/KNAW Westerdijk Fungal Biodiversity Institute, The Netherlands

11 [#]Current position: Department of Biotechnology, Delft University of Technology, The

12 Netherlands

13 [$]Corresponding author: T.Abeel@tudelft.nl

14

Abstract

Motivation: Fungi are essential in nutrient recycling in nature. They also form symbiotic, commensal, parasitic and pathogenic interactions with other organisms including plants, animals and humans. Many fungi are polykaryotic, containing multiple nuclei per cell. In the case of heterokaryons, there are even different nuclear types within a cell. It is unknown what the different nuclear types contribute in terms of mRNA expression levels in fungal heterokaryons. Each cell of the cultivated, mushroom forming basidiomycete *Agaricus bisporus* contains 2 to 25 nuclei of two nuclear types, *P1* or *P2,* that originate from two parental strains. Using RNA-Seq data, we wish to assess the differential mRNA contribution of individual nuclear types in heterokaryotic cells and its functional impact.

Results: We studied differential expression between genes of the two nuclear types throughout mushroom development of *A. bisporus* in various tissue types. The two nuclear types, produced specific mRNA profiles which changed through development of the mushroom. The differential regulation occurred at a gene and multi-gene locus level, rather than the chromosomal or nuclear level. Although the P1 nuclear type dominates the mRNA production throughout development, the P2 type showed more differentially upregulated genes in important functional groups including genes involved in metabolism and genes encoding secreted proteins. Out of 5,090 karyolelle pairs, i.e. genes with different alleles in the two nuclear types, 411 were differentially expressed, of which 246 were up-regulated by the P2 type. In the vegetative mycelium, the P2 nucleus up-regulated almost three-fold more metabolic genes and cazymes than P1, suggesting phenotypic differences in growth. A total of 10% of the differential karyollele expression is

36   associated with differential methylation states, indicating that epigenetic mechanisms may be

37   partly responsible for nuclear specific expression.

38   **Conclusion:** We have identified widespread transcriptomic variation between the two nuclear

39   types of *A. bisporus*. Our novel method enables studying karyollelle specific expression which

40   likely influences the phenotype of a fungus in a polykaryotic stage. This is thus relevant for the

41   performance of these fungi as a crop and for improving this species for breeding. Our findings

42   could have a wider impact to better understand fungi as pathogens. This work provides the first

43   insight into the transcriptomic variation introduced by genomic nuclear separation.

Introduction

45   Fungi are vital to many ecosystems, contributing to soil health, plant growth, and nutrient

46   recycling[1]. They are key players in the degradation of plant waste[2,3], form mutually beneficial

47   relationships with plants by sharing minerals in exchange for carbon sources[4,5] and by inhibiting

48   the growth of root pathogens[6,7]. They even form networks between plants, which can signal each

49   other when attacked by parasites[8]. Yet, some are plant pathogens responsible for huge economic

50   losses in crops[9–11].

51   The genome organization of fungi is incredibly diverse and can change during the life cycle. For

52   instance, sexual spores can be haploid with one or more nuclei or can be diploid. Sexual spores

53   of mushroom forming fungi are mostly haploid and they form monokaryotic (one haploid

54   nucleus per cell) or homokaryotic (two or more copies of genetically identical haploid nuclei)

55   mycelia upon germination. Mating between two such mycelia results in a fertile dikaryon (one

56   copy of the parental nuclei per cell) or heterokaryon (two or more copies of each parental nuclei)

57   when they have different mating loci[12]. In contrast to eukaryotes of other kingdoms, the nuclei

58   do not fuse into di- or polyploid nuclei but remain side by side during the main part of the life

59   cycle. Only just before spores are formed in mushrooms, do these nuclei fuse, starting the cycle

60   anew.

61   **Figure 1**
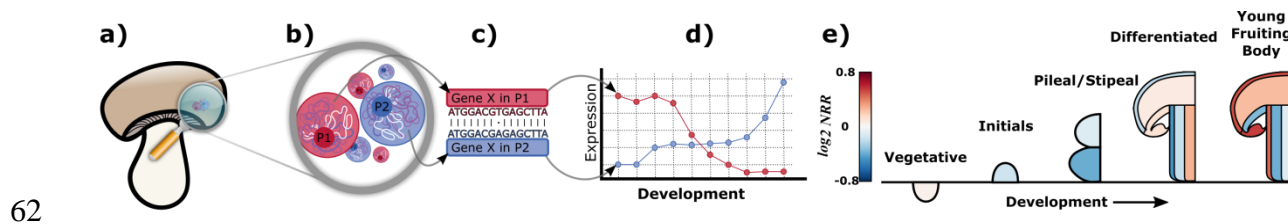
62

63   *Figure 1: Nuclear type specific expression in A. bisporus. a) The A. bisporus mushroom is composed of different tissues that*
64   *consist of hyphae comprised of cellular compartments. b) Each cellular compartment is a heterokaryon containing between 2 and*
65   *25 nuclei. In our strain, each nucleus is either of type P1 (red) or P2 (blue). Both nuclear types are haploid, and contain exactly*
66   *one copy of each gene. However, because there are multiple nuclei, there may be multiple copies of each gene in the cell. c)*
67   *Furthermore, the gene in the two types, which we call karyolleles, may differ in their genetic sequences. d) These differences in*
68   *transcript sequence allow us to quantify expression of each karyollele in each tissue and to investigate nucleus specific*
69   *expression. e) Read count ratios at the nuclear type level (Equation 5) of Agaricus bisporus throughout its development. Red*
70   *colour indicates higher P1 activity, blue colour higher P2 activity. The scale bar indicates the log2 fold change in activity*
71   *between the P1 and P2 nuclear types. We observe a differential mRNA activity in different mushroom tissues.*

72

73   *Agaricus bisporus* is the most widely produced and consumed edible mushroom in the world[2].

74   Heterokaryotic mycelia of the button mushroom *Agaricus bisporus* var. *bisporus* (Sylvan A15

75   strains) have between 2 and 25 nuclei per cell[13,14] (Figure 1a-d). The genomes of both ancestral

76   homokaryons have been sequenced[1,15] showing that DNA sequence variation is associated with

77   different vegetative growth capabilities[1]. Due to the two nuclear types, each gene exists at two

78   alleles separated by nuclear membranes, which we call karyolleles. Although there have been a

79   few studies investigating the expression of genetic variety in the transcriptome[16,17], the

80   differential transcriptomic activity of two (or more) nuclear types has never been systematically

81   investigated in a heterokaryon at the genome wide scale. Based on SNPs identified in mRNA

82   sequencing, it has been suggested that allele specific expression is tightly linked to the ratio of

83   the nuclear types in a basidiomycete[18].

84   Allele specific expression in mononuclear cells has been studied in fungi[19], plants[20], animals[21],

85   and humans[22]. Such studies have shown that allele heterogeneity is linked to differential allele

86   expression and cis-regulatory effects[21–23], and even sub-genome dominance[24]. *A. bisporus* is in

87   many ways an excellent model organism to investigate differential karyollele expression. It only

88   has two nuclear types in the heterokaryon contrasting to the mycorrhizae that can have more

89   nuclear types[25,26], making computational deconvolution of mRNA sequence data intractable with

90   currently available tools. Additionally, the recently published genomes of the two nuclear types

91   of *Sylvan A15*[15] exhibit a SNP density of 1 in 98 bp allowing differentiation of transcripts in high

92   throughput sequencing data. Finally, bulk RNA-Seq datasets of different stages of development

93   and of different tissues of the fruiting bodies are available[2,27].

94   Here, we show that differential karyollele expression exists in *Agaricus bisporus Sylvan A15*

95   strain, which changes across tissue type and development and affects different functional groups.

96   Further, we show that differential karyollele expression associates with differential methylation

97   states, suggesting that epigenetic factors may be a cause for the differential regulation of

98   karyolleles.

99   **Results**

100   **Karyollele specific expression through sequence differences**

101   To assign expression levels to individual karyolleles, we exploit sequence differences between

102   karyollele pairs in the P1 and P2 homokaryon genomes of *A. bisporus A15* strain (Materials).

103   Briefly, the sequence differences define marker sequences for which the RNA-Seq reads

104   uniquely match to either the P1 or the P2 variant, effectively deconvolving the mRNA

105   expression from the two nuclear types (see Methods). There are a total of 5,090 distinguishable

106   karyollele pairs between the *P1* and *P2* genomes, corresponding to ~46% of all genes. The

107   remaining genes could not be unambiguously matched, or the karyollele pairs had too few

108   sequences differences. Most (80%) distinguishable karyollele pairs had the same number of

109   markers in each homokaryon. For the remaining pairs (20%), the number of markers per

110   karyollele was different (see Supplementary Material Note A). This variation can be explained

111   by the non-symmetric number of markers produced by the different kinds of variation. While a

112   SNP will result in one marker in each karyollele, an indel (if longer than 21bp) will result in one

113   marker in one karyollele, and at least two in the other. Karyollele specific expression is

114   expressed as a read count ratio that reflects the relative abundance of mRNAs originating from

115   the P1 or P2 nuclear types (Equation 3, Methods).

116   We studied *A. bisporus*' karyollele specific expression for different tissues and development in

117   two RNA-Seq datasets, one studying the mycelium in compost throughout mushroom harvest,

118   and one studying different mushroom tissues throughout mushroom formation (Figure 1e,

119   Supplementary Material Notes B, and Materials). Measured difference in expression between

120   nuclear types is not correlated with the number of markers ($p > 0.05$) for any of the samples, nor

121   is it correlated with CG content (see Supplementary Material Notes C).

122   **P1 and P2 mRNA production differs per tissue and across development**

6

123   First, we assess the total mRNA production of the P1 and P2 nuclear types and their relative

124   contributions during development. To do this, we considered the total number of reads uniquely

125   matching to P1 with respect to P2. Figure 1e shows that this nuclear type read count ratio (NRR,

126   see Equation 5, Methods) changes throughout development and across tissue types. For example,

127   during the '*Differentiated*' stage, the P2 nuclei are dominant in the skin, but in the '*Young*

128   *Fruiting Body*', the P1 nuclei dominate the skin (two right most panels in Figure 1e). In contrast,

129   the '*Stipe Center*' is dominated by P1 nuclei in the differentiated stage, while later the expression

130   of P2 nuclei dominates.

131   The transcription patterns throughout the mushroom development differ between the karyolleles.

132   Based on a principal component analysis of the expression profiles of each nuclear type, we

133   observe that the expression profiles of P1 and P2 group together in different clusters, based on

134   the first and second principal components (Supplementary Material Note D). This clustering is

135   indicative of distinct regulatory programs. It appears as though the first principal component

136   represents the tissue type, and the second represents the nuclear type. Interestingly,

137   measurements of the same tissue from P1 and P2 do not have exactly the same value for the first

138   principal component, indicating that the difference in nuclear type does not entirely explain the

139   variation between P1 and P2.

140   **Within a sample, mRNA production of P1 and P2 vary between chromosomes**

141
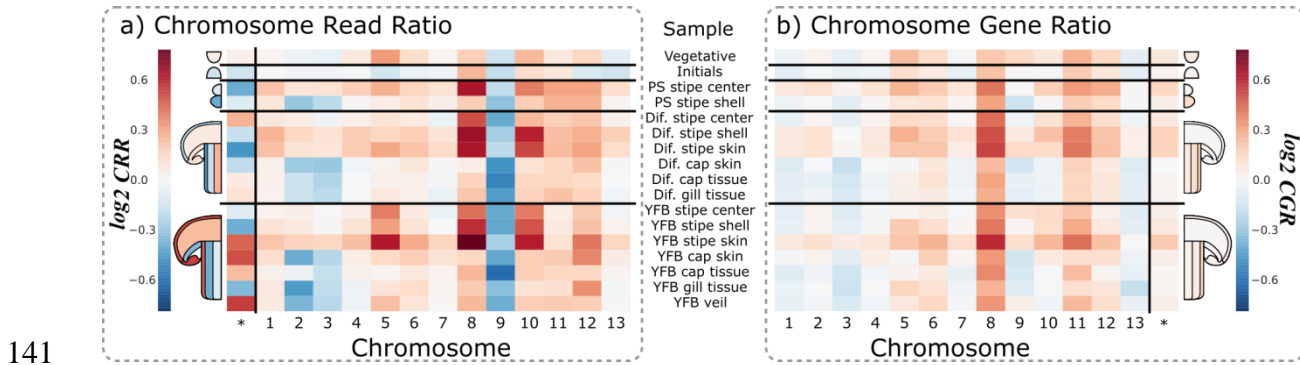
*Figure 2: P1 versus P2 expression per chromosome throughout development of the mushroom. A red color indicates a higher P1 activity and a blue color indicates a higher P2 activity. Each row indicates a different developmental stage, and each column represents a different chromosome. The column noted with an asterisk is the ratio at the nuclear type level. **a)** The read count ratios at the chromosome level (CRR, Equation 4). Supplementary Material Note G provides the read count ratios at the chromosome level in the vegetative mycelium dataset. **b)** The Chromosome Gene Ratios (CGR, Equation 6). See Supplementary Material Note H for the gene ratio measures in the vegetative mycelium dataset.*

149   Figure 2a shows the Chromosome Read count Ratios (CRR, Equation 4), demonstrating that

150   some chromosomes are more active in P1 (e.g. chromosome 8) throughout development, while

151   others are more active in P2 (e.g. chromosome 9). Expression of other chromosomes depend on

152   the developmental state, changing in time (e.g. chromosome 2). The chromosome log2 fold

153   changes lie between [-0.60, 0.79]. In the vegetative mycelium we see less drastic differences in

154   mRNA production throughout development than in the mushroom tissues, with expression log2

155   fold changes between [-0.28, 0.36] (see Supplementary Material Notes B).

Gene read ratios reveal a dominant P1 type in mushroom tissue, but not in mycelium

157   To investigate whether either nuclear type is truly dominant we correct for extremely highly

158   expressed genes (Supplementary Material Note E-F) by limiting their impact on the chromosome

159   and tissue level ratios by using per-gene activity ratios per chromosome (CGR, Equation 6),

160   instead of read ratios. This revealed that, in addition to P1 producing more mRNA than P2, P1

161   karyolleles were also more frequently higher expressed than their P2 counterpart (Figure 2b).

8

162   Looking across all tissues and chromosomes, P1 is significantly dominant over P2, i.e. the

163   average of the log-transformed CGR is significantly larger in the P1 nuclear type than the P2

164   nuclear type, following a t-test in mushroom tissue, with $p < 0.01$, (see Supplementary Material

165   Note G). Using the Chromosome Gene Ratio has a notable impact on chromosome 9. Although

166   P2 produces most chromosome 9 mRNA (Figure 2a), it is not the case that more P2 karyolleles

167   are more highly expressed than P1 karyolleles.

168   We do not observe such a dominance of P1 in the mycelium ($p > 0.05$, with t-test as in

169   mushroom dataset), where neither P1 nor P2 show a dominant mRNA activity (see

170   Supplementary Material Note H).

171   **A substantial portion of karyolleles are differentially expressed**

172   In each tissue, we determined the set of karyolleles which are statistically significantly

173   differentially expressed between the two nuclear types. Although the dominance of the P1

174   nuclear type indicates a general trend of higher activity across many genes, some karyollele pairs

175   have a much larger difference pointing towards a functional role. In total, we find 411 genes that

176   are differentially expressed (see Methods) in a mushroom tissue or in vegetative mycelium

177   throughout development (Table 1); 368 genes are differentially expressed in mushroom tissues,

178   and 82 in the vegetative mycelium. The set of differentially expressed genes is enriched in the set

179   of genes with mixed methylation states (Methods, Supplementary Material Notes I).

180   Interestingly, when a karyollele pair is differentially expressed, with only a few exceptions (see

181   Supplementary Material Notes J), it will always be observed to be more highly expressed in the

182   same nuclear type, i.e. if a gene is observed to be more highly expressed in P1 than in P2, than it

183   will never be observed to be more highly expressed in P2 than in P1 in other tissues, and vice

9

184   versa. The only exceptions to this rule lies in the set of genes that are differentially expressed in

185   both the mushroom dataset and the mycelium dataset.

186   The set of differentially higher expressed genes between the nuclear types in mushroom and

187   mycelium sets overlap with only 39 genes. In this intersection set, more genes are higher

188   expressed in P2 than in P1. Ten genes had a higher expression in P1, and 24 had a higher

189   expression in P2. Five were more highly expressed in P2 in the mycelium, but switched their

190   origin of primary expression to P1 in the mushroom (see Supplementary Material Notes J). The

191   lack of a substantial overlap of differentially expressed genes between the two nuclear types is

192   indicative of different regulatory processes during the vegetative stage and a mushroom stage.

193   Although P2 upregulates more differentially expressed genes than P1 does, more genes show a

194   consistently higher expression in P1 than in P2. We identify consistently higher expressed genes

195   that show a higher expression in one nuclear type over the other across all samples (Methods). In

196   the mushroom tissue dataset, we find 1,115 genes that are consistently higher expressed in P1,

197   and 785 genes that are consistently higher expressed in P2. Similarly, in the vegetative

198   mycelium, we find 832 genes that are consistently higher expressed in P1 and 645 that are

199   consistently higher expressed in P2. The two datasets overlap with 470 and 256 genes for P1 and

200   P2, respectively. Interestingly, Of the 90 named genes in *S. commune* (Methods), only *mnp1* is

201   differentially expressed and exhibits different behavior in the mushroom and the vegetative

202   mycelium (see Supplementary Material Note K).

10

203   *Table 1: Karyolleles differentially expressed between P1 and P2 in mushroom tissue and vegetative mycelium across*
204   *development. In the first row we indicate the number of differentially expressed genes that are higher expressed in the different*
205   *nuclear types for the two datasets (columns). The second row gives the total number of differentially expressed genes in the two*
206   *different datasets. Row three shows the number of differentially expressed genes in a dataset that are not differentially expressed*
207   *in the other dataset. In the last row, we show the number of differentially expressed genes that overlap between the two datasets.*

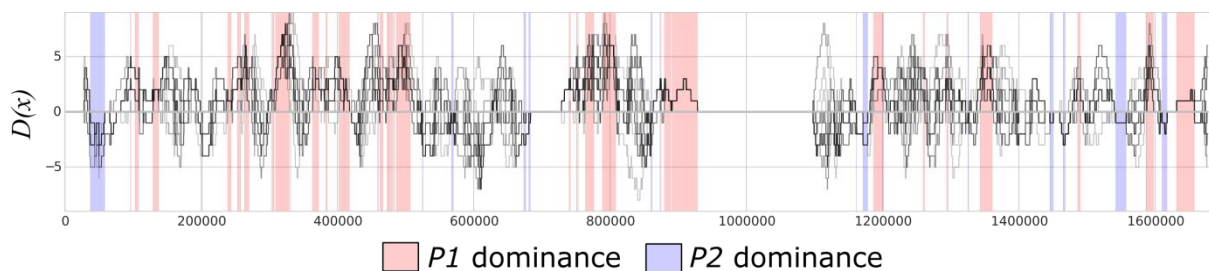|	**Mushroom tissue dataset**		**Mycelium tissue dataset**	
	P1 up	**P2 up**	**P1 up**	**P2 up**
Diff. ex.	176	193	30	52
Total/dataset	368		82	
Unique/dataset	329		43	
Overlap	39 (411 total)			

208

209

210

211

212 **Co-localized gene clusters are co-regulated**

213

214   *Figure 3: Co-localized genes are often co-regulated. Pictured here are the co-localized and co-regulated gene clusters along*
215   *chromosome 10 in the mushroom tissue dataset. Along the x-axis is the genomic co-ordinate. For each sample (gray lines), we*
216   *plot the difference between the number of genes more highly expressed by P1 and the number of genes more highly expressed by*
217   *P2 (Equation 8, a value of 0 indicates an equal distribution). We also highlight the regions that are consistently upregulated in*
218   *P1 (red regions) and the number of genes that are consistently upregulated in P2 (blue regions). See Supplementary Material*
219   *Note J for other chromosomes.*

220

11

221 To investigate the level at which genes are regulated, we investigated whether there are regions

222 where the majority of genes were consistently higher expressed in one homokaryon than in the

223 other. We detected many of such regions, given in Table 2 and Figure 3 (Methods,

224 Supplementary Material Note L), hinting towards a sub-chromosomal level of regulation. This is

225 supported by observations in Figure 2, where we see that within one tissue chromosomes are

226 differently regulated, excluding a regulation at the nuclear level. Because we observe that co-

227 regulated gene are co-localized in regions, regulation can also not occur at the chromosome

228 level, because then we would have expected regions of co-regulation of the size of whole

229 chromosomes.

230 Co-regulated regions are more frequently upregulated for the P1 karyollele than for the P2

231 karyolleles. This observation is in agreement with the observed P1 nuclear type dominance. We

232 observe relatively little overlap between the Mushroom and Vegetative Mycelium datasets

233 (Table 2), indicative of different regulatory programs between the vegetative mycelium and

234 mushroom tissue cells.

235 *Table 2: The number of regions in which the majority of the genes are coregulated (Methods), across the mushroom and*
236 *mycelium datasets and with the number of genes in these regions. P1 and P2 columns indicate whether the region is consistently*
237 *higher in for the P1 kayollele or the P2 karyollele, respectiverly. Row Both indicates overlapping regions between the mushroom*
238 *and vegetative mycelium datasets. Supplementary Material Note L offers detailed expression profiles of these regions.*

	P1		P2	
Dataset	**#Regions**	**#Genes**	**#Regions**	**#Genes**
Mushroom	207	741	73	233
Vegetative Mycelium	414	1955	43	140
Both	151	484	7	17

239

12

240 **Broad range of functionality affected by karyollele specific expression throughout**

241 **development**

242 Next, we set out to examine the functional annotations of the differentially expressed karyollele

243 pairs, considering the following categories: (i) transcription factors, (ii) metabolic genes, (iii)

244 secondary metabolism genes, (iv) cytochrome P450 genes, (v) carbohydrate active enzymes

245 (cazymes) and (vi) secreted proteins. These categories, with the exception of secondary

246 metabolite genes, are all enriched in the set of differentiable genes ($p < 0.05$ by a chi-squared

247 approximation to the fisher's exact test with FDR correction).

248 Figure 4 show the division of the 411 differentially expressed genes across the functional

249 categories in all the different samples. None of the differentially expressed genes were

250 transcription factors. For the other functional categories, we saw a more or less equal amount of

251 up-regulated karyolleles in P1 and P2 (Figure 4a) in the mushroom tissues (except the vegetative

252 stage), and a more skewed distribution of activity in the mycelium dataset (and the vegetative

253 stage of the mushroom dataset). In these cases, P2 had more differentially expressed genes in

254 these functional categories (Figure 4b).

13
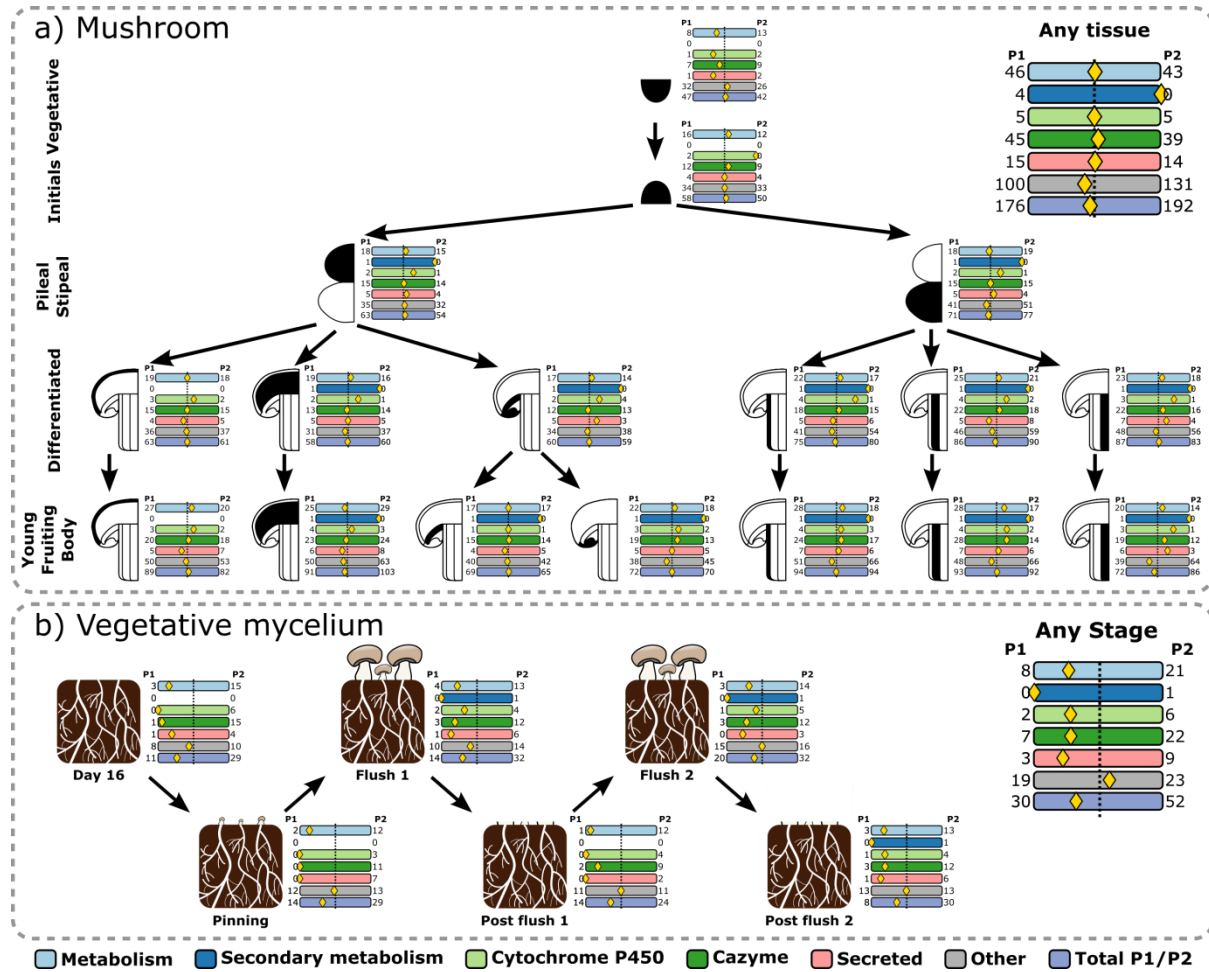
Figure 4: Differential regulation of functional groups through mushroom development. The development of different tissues is illustrated as a tree. We investigate metabolic genes (light blue), secondary metabolic genes (dark blue), cytochrome P450 genes (light green), carbohydrate active enzymes (dark green), secreted protein genes (light red), and all others not fitting into any of the previous groups (grey). At each developmental stage, we observe how many genes of each group are differentially upregulated in P1 (left) and in P2 (right). The yellow diamond indicates the ratio of these counts. **a)** For the mushroom dataset. **b)** For the vegetative mycelium dataset. We see that the groups are more or less equally distributed between P1 and P2 (the yellow diamond is centered), with the exception of the vegetative stage (the root node of Figure 4A), and the vegetative mycelium dataset.

The P2 type had a higher expression of significantly more karyolleles than P1 in mycelium (see

Supplementary Material Notes M). In the mycelium, P2 had an enriched expression of

cytochrome P450 genes, secondary metabolite genes, and cazymes ($p < 0.05$, with an FDR

corrected chi-squared approximation to the fisher's exact test). Furthermore, cazymes and

14

269   metabolic genes in mycelium were more likely to be more highly expressed in P2 ($p < 0.05$, with

270   an FDR corrected binomial test).

271   Nineteen of the 39 previously identified differentially expressed genes that are shared between

272   the mycelium and mushroom datasets had the following functional annotations: 14 were

273   annotated as metabolic genes, 14 as cazymes, five as secreted proteins, and two as cytochrome

274   P450s (some genes have multiple annotations). Additionally, five of these 39 overlapping genes

275   have different domain annotations, indicating different functional properties between the P1 and

276   P2 karyolleles.

277   To further elucidate the functional impact of the 411 differentially expressed genes, we mapped

278   them onto the KEGG pathway database. Sixteen of the genes that are differentially expressed in

279   mushroom tissue or vegetative mycelium samples are found in 20 pathways. Interestingly, three

280   differentially expressed genes are found in the Aminoacyl-tRNA biosynthesis (M00359)

281   pathway (Supplementary Material Notes N). Two genes belong to valine and methionine tRNAs

282   pathways and were upregulated in P1. One gene in the pathway producing aspartamine tRNAs

283   pathway was upregulated in P2. Together, this suggests that P1 is able to produce more valine

284   and methionine tRNAs than P2.

285   Next we studied whether differential expression of a karyollele also resulted in the production of

286   a functionally different protein due to sequence differences between the karyolleles. 216 of the

287   5,090 distinguishable karyolleles had sequence differences that led to an alternative protein

288   domain annotation, and 36 of these 216 have alternative domain annotations. 36 of these 216

289   karyollele pairs are differentially expressed between P1 and P2 (see Supplementary Material

290   Notes O).

15

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 is largely unknown. In an attempt to understand this, we studied the expression of alleles in both constituent nuclei (P1 and P2) of the button mushroom cultivar Sylvan 15. From the observed average gene expression, we conclude that the expression of nuclear type P1 of the *Agaricus bisporus Sylvan A15 strain* is dominant over nuclear type P2. Remarkably, this dominance is present across all developmental stages in the heterokaryon. We can link this phenomenon to the human case, where in fibroblasts[29], 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 and RNAi regulatory pathways[30], 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 those in P2, but in the vegetative mycelium dataset, P2 has more up-regulated karyolleles relative to P1. The contrast between a dominant P1, yet more differentially over-expressed genes in P2 in mushroom tissue is paradoxical. However, there are many genes that show a consistently higher expression in either P1 or P2, with more genes showing a consistently higher expression for 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, sub-genome dominance may be linked to methylation of transposable elements[24]. Might it be possible that something similar happens in *A. bisporus*?

16

313   Although an imbalance in the number of nuclei could very well explain the dominance of P1, we

314   have shown that genes that are consistently higher expressed in one of the karyolleles do co-

315   localize in sub-chromosomal regions. If there were more P1 nuclei than P2 nuclei, we would

316   have expected a general higher expression of genes of one nuclear type across all chromosomes,

317   which we do not observe.

318   For many differentially expressed genes, the protein sequence differences between the two

319   karyolleles in the two nuclear types encode for different protein domains. This suggests a

320   functional impact of karyollele specific expression. We also observe a broad range of

321   functionality being differentially expressed between the P1 and the P2 nuclear types. For

322   example, the P2 upregulation of cazymes and metabolic genes in P2 in compost highlight the

323   importance of the P2 homokaryon in development. H97, one of the homokaryons in the cultivar

324   Horst U1, from which Sylvan A15 is derived, displays stronger vegetative growth characteristics

325   than its counterpart H39[1]. This metabolic strength may be passed down from the H97

326   homokaryon to the Sylvan A15 P2 homokaryon, and the differentially expressed karyolleles may

327   in part be responsible for this. *mnp1*, for example, is an important gene for growth on compost

328   and P2 has indeed inherited the relevant chromosome 2 from H97 (Sonnenberg et al., 2016).

329   Such characteristics are relevant for breeding strategies.

330   Surprisingly, *mnp1* is expressed and even up-regulated in the mushroom tissues. *mnp1* is known

331   to be involved in lignin degradation, which occurs in the vegetative mycelium[2,28]. In compost,

332   the abundance decreases dramatically throughout development (Supplementary Material Note

333   K). Therefore, the abundance of *mnp1* in the stipe of the fruiting body is unexpected, although it

334   has been shown that proteins produced in the mycelium can find their way into the mushroom[31].

17

335   However, it does not explain the fact that the P1 karyollele exists in higher abundance in the

336   mushroom tissues, while the P2 karyollele is higher expressed in the vegetative mycelium.

337   Transport of the P2 karyollele from the vegetative mycelium into the mushroom conflicts with

338   the abundances of the P1 karyollele observed in the mushroom tissues.

339   A significant proportion of differentially methylated karyolleles were also differentially

340   expressed, most differentially expressed genes are not observed to be methylated. The overlap

341   we observe between methylated genes and differentially expressed genes in different

342   developmental stages explain an effect in the mushroom tissue. However, we cannot link the

343   methylation to a preference of nuclear type. For example, the five differentially expressed genes

344   between compost and mushroom that change their nuclear dominance are not methylated.

345   Although, methylation seems to play a role in the differential use of nuclear type for mRNA

346   production, it only explains 10% of the observed differential expression. This may be due to a

347   limitation of our methylation dataset, (which only comprises vegetative growth), but it may also

348   hint towards other regulatory mechanisms.

349   In addition to methylation, we also observe co-localization of co-expressed genes. This may be

350   indicative of a difference in genome organization, whereby the DNA is less accessible in certain

351   regions in P1 than in P2 through different levels of chromatin compaction. It has been shown that

352   gene expression is strongly linked to DNA availability, and further, that such chromatin

353   organization is heritable[32].

354   The sequences of a pair of karyolleles need to be sufficiently different for our algorithm to be

355   able to uniquely assign reads to each karyollele. These sequence differences between nuclear

356   types may have an effect on various regulatory mechanisms of transcription, such as

18

357   transcription factor binding efficiencies, transcription efficiency, differences in mRNA stability,

358   or differences in epigenetic factors. Future research might shed light on whether these

359   differences are related to observed differential karyollele expression.

360   Causative mechanisms of karyollele specific expression can further be elucidated by population

361   studies across multiple spore isolates. Sylvan A15 is derived as a heterokaryotic single spore

362   isolate from Horst U1. In such heterokaryons, non-sister nuclei are paired in one spore.

363   Combined with the restriction of recombination to chromosome ends, such heterokaryons are

364   genetically very similar to the parent and differ only in the distribution of parental type

365   chromosomes over both nuclei. Karyollele expression could thus be studied in different

366   heterokaryotic single spore isolates having different distributions of otherwise very similar

367   chromosomes over both nuclei. If the expression patterns are consistent with nuclear

368   chromosome organization across different single spore isolates, it will suggest that expression of

369   specific karyolleles can be controlled by selecting isolates where karyolleles lie in the desired

370   nuclei.

Conclusion

372   We show that karyolleles, the different copies of a gene separated by nuclear membranes in a

373   heterokaryon, are differentially expressed between the two different nuclear types in the

374   *Agaricus bisporus Sylvan A15 strain*. Each nuclear type contributes varying amounts of mRNA

375   to the cell, and differential expression occurs at the gene level. Despite a dominant P1 type, we

376   see no evidence that would suggest an imbalance in the number of copies P1 and P2 nuclei in

377   any cell type, though it may vary from cell to cell.

378   Genes with various vital functions are differentially expressed. The P2 homokaryon significantly

379   up-regulates cazymes and metabolic genes, which may indicate a difference in vegetative growth

380   strengths. This corroborates what was observed in the constituent homokaryons of the Horst U1

381   cultivar from which P1 and P2 are essentially derived.. Manganese peroxidase is one of the

382   differentially expressed genes, and exhibits interesting, previously unknown behavior. The cause

383   of these differential regulations is still not known, but it is possible that epigenetic mechanisms,

384   like methylation, play a role.

385   The biological gene regulation mechanisms between heterokaryons need to be investigated.

386   Unfortunately, such research is hindered by current mRNA isolation procedures. As mRNA

387   transcripts are secreted from the nuclei and mixed in the cytoplasm of the cell, traditional

388   sequencing methods will be unable to generate a full resolution of both homokaryon expression

389   from full cell isolates. Single nucleus sequencing[33,34] would circumnavigate this issue by

390   isolating mRNA from individual nuclei. As we have shown that the two nuclear types exhibit

391   distinguishable regulatory programs, it will be possible to distinguish them based on their

392   expression profiles.

393   The impact of differential expression between nuclei of heterokaryotic organisms is

394   underappreciated. Heterokaryotic fungi have major impact in clinical and biotechnological

395   applications, and impact our economy and society as animal pathogens such as *Cryptococcus*

396   *neoformans*[35], plant pathogens such as *Ustilago maydis*[36], plant and soil symbionts such as

397   mycorrizal fungi[26], bioreactors such as *Schizophyllum commune*[37], and of course the subject of

398   this study, the cultivated, edible mushroom *Agaricus bisporus*[15]. It is known that different

20

399   homokaryons in these species will produce different phenotypes[2] which no doubt need to be

400   treated, nourished or utilized differently.

401   We have demonstrated differential nuclear regulation of a fungal organism and we showed that

402   variation between homokaryons results in functional differences that were previously unknown.

403   With this work, we hope to draw attention to the impact of sequence and regulatory variation in

404   different nuclei on the function and behavior of the cell in order to further our understanding of

405   the role of fungi in our environment.

Materials and Methods

407   **RNA-Seq data:** We used two RNA-seq datasets from the *Agaricus bisporus (A15)* strain: (1)

408   tissue samples through mushroom development (BioProject: PRJNA309475)[27], and (2)

409   vegetative mycelium samples taken from compost through mushroom development (BioProject

410   PRJNA275107)[2]. Throughout the text, when we refer to the mushroom tissue, we also refer to all

411   samples in dataset (1), including the first sample, which technically is a sample of the vegetative

412   mycelium. The compost dataset exhibited high amounts of PCR duplicates (Supplementary

413   Material Note P). This can be attributed to the difficulty in isolating RNA from soil. To remedy

414   the biases involved with this, we removed all PCR duplicates using FastUniq[38].

415   **Methylation data:** A sample of vegetative stage mycelium of A15 was treated with the EpiTect

416   Bisulphite conversion and cleanup kit and sequenced with the Illumina HiSeq 2000. Raw reads

417   were trimmed using TRIMMOMATIC[39] and aligned to the A15 P1 genome using Bismark[40] and

418   bowtie2[41]. Methylated bases were analyzed with Methylkit[42]. Only bases which had a minimum

419   coverage of 10 were retained. For samples with mixed methylation states, we will observe what

420 appear to be incomplete conversions of unmethylated cytosines but in reality represents the

421 mixed methylation states of those bases. Therefore, to include only differentially methylated

422 bases between the two nuclei (i.e. methylated in one homokaryon, but not in the other), we

423 considered only those bases which were measured to be methylated between 40 and 60% of all

424 reads (Supplementary Material Notes I). While 164,290 bases had an indication of methylation

425 signal, 10,325 bases had methylation signals of about 50%, suggestive of differential methylation

426 states. Methylated bases were mapped to genes when between the start and stop codons, or

427 1000bp up/downstream (Supplementary Material Note Q).

428 **Homokaryon genome and annotations:** The P1 and P2 genomes[15] were annotated with

429 BRAKER1[43] using the pooled RNA-seq data described above. In order to prevent chimeric genes

430 (neighboring genes that are erroneously fused into one predicted gene) the following procedure

431 was used. After the first round of gene prediction, predicted introns were identified that were at

432 least 150 bp in size and not supported by RNA-seq reads. The midpoint of these introns were

433 labeled as intergenic regions in the next round of gene prediction using AUGUSTUS 3.0.2[44] and

434 the parameter set produced in the first round of gene prediction. The SNP density between the

435 genomes was estimated using MUMMER's[45] show-snps tool.

436 **Karyollele pair discovery:** The genome annotations were used to produce predicted mRNA

437 sequences for each gene. The genes in the two parental genomes were matched using a reciprocal

438 best BLAST [46] hit. Hits which had E-values greater than 10^{-100} were removed. This resulted in a

439 conservative orthology prediction between the two homokaryons that are our set of karyolleles.

440 Karyollele pairs which have a 100% sequence identity were removed, as it would be impossible

441 to identify distinguishing markers for these identical pairs.

442   **Marker Discovery:** For each discovered karyollele pair, we identify markers that uniquely

443   identify each element of the pair. This is done by constructing all possible kmers for each

444   sequence, resulting in two sets per pair. The kmers overlapping in these sets are removed,

445   resulting in distinguishing pairs of markers. Once distinguishing markers have been discovered

446   for all pairs, we remove all non-unique markers. Finally, the set of markers is made non

447   redundant by scanning the position-sorted list of markers from left to right and removing any

448   marker that overlaps with the previous marker. Finally, we ensure that the markers are unique

449   throughout the whole genome by removing markers that are present anywhere else in either

450   genome. In order to guarantee sufficient evidence across the whole gene, we remove karyollele

451   pairs which do not have at least five markers each.

452   **Marker quantification:** We scan all RNA-Seq reads for the detected markers using the Aho-

453   Corasick algorithm[47]. We insert all markers and their reverse complements into an Aho-Corasick

454   tree and count each marker only once for each fragment (a marker may be present twice, if the

455   read mates overlap). We calculate a gene expression score as the average of each marker count

456   for a gene. This results in an expression score E_h for each gene g in each sample t for each

457   replicate r, per homokaryon h:

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

458

459   where $M_h(g)$ *is* the set of markers in a gene g, and $C_h(r,s,m)$ is the count for marker m in replicate

460   r, sample s.

461   **Differential expression:** Using DE-Seq[48], we perform a differential expression test for each

462   karyollele pair in a tissue, i.e. we test if a gene has a differential expression in P1 or P2. DESeq

23

463   requires a size factor to be calculated, which normalizes for the library sizes of each sample.

464   Since however, the counts from P1 and from P2 originate from the same sample, these must have

465   the same size factor. Size factors are therefore calculated manually, by counting the total number

466   of reads for each sample, and dividing it by the largest value for any sample (Equation 2).

$$sf(s,r) = \frac{\displaystyle\sum_{h}\sum_{m \in M_h(g)} C_h(r,s,m)}{\displaystyle\max_{(s',r')}\left(\sum_{h}\sum_{m \in M_h(g)} C_h(r',s',m)\right)} \quad (2)$$

467

468   The P1 and P2 counts originating from the same sample will then be assigned the same size

469   factor. The expression counts for each gene in each replicate in each tissue (equation 1) are

470   provided to DE-Seq with the provided size factor (Equation 2). The normalized read counts per

471   gene $D_h(s,g)$ are returned by DE-Seq, together with significance values for each test. We select

472   only differentially expressed genes that have a q-value < 0.05, and a fold change of at least three.

473   **Read ratio calculation:** Using the normalized read counts from DE-Seq [48], we calculate the

474   ratio of the number of reads originating from the two homokaryons at the gene (GRR),

475   chromosome (CRR) and nuclear type level (NRR).

476

24

$$GRR(s,g) = \frac{D_{P1}(s,g)}{D_{P2}(s,g)} \qquad (3)$$

$$CRR(s,c) = \frac{\sum\limits_{g \in c} D_{P1}(s,g)}{\sum\limits_{g \in c} D_{P2}(s,g)} \qquad (4)$$

$$NRR(s) = \frac{\sum\limits_{c \in C} \sum\limits_{g \in c} D_{P1}(s,g)}{\sum\limits_{c \in C} \sum\limits_{g \in c} D_{P2}(s,g)} \qquad (5)$$

477

478   **Gene ratio calculation:** Using the normalized read counts from DESeq [48], we calculate the ratio

479   of the number of reads originating from the two homokarons at the gene level, and use those

480   ratios to calculate the geometric mean of the relative expression activities at the chromosome

481   (CGR, Equation 6) and nuclear type level (NGR, Equation 7). The geometric mean is more

482   suitable than the arithmetic mean for averaging ratios.

$$CGR(s,c) = \sqrt[|c|]{\prod_{g \in c} GRR(s,g)} \quad (6)$$

$$NGR(s) = \sqrt[|C|]{\prod_{c \in C} CGR(s,c)} \quad (7)$$

483

484   **Identifying consistent genes:** For each gene, we observe the relative expression in each sample

485   (Equation 3). We refer to a gene as being consistently expressed if it is more highly expressed in

486   the same nuclear type in each sample. I.e. the GRR is always greater than one, or always less

487   than 1.

488   **Identifying co-regulated clusters:** We slide a window of size 20,001bp (10,000- up and down-

489   stream) across each chromosome. In this window, we count the number of genes that are more

490   highly expressed by P1 and by P2, and calculate the difference per sample. I.e.

$$D(x,s) = \sum_{g \in W(x-10000,\,x+10000)} \begin{cases} 1 & \text{if } GRR(g,s) > 1 \\ -1 & \text{if } GRR(g,s) < 1 \end{cases} \quad (8)$$

491

492   where $W(x,y)$ is the set of genes between genomic location x and y, and s is a sample. This

493   difference is shown in Figure 3. Next, we identify regions where each sample in the dataset

494   shows consistent regulation. That is to say, in these regions, $D(x,s) > 0 \ \forall \ s \in S$, or $D(x,s) < 0 \ \forall$ s

495    $\in S$, where S is the set of all samples. These regions contain co-localized genes that are co-

496   regulated across all samples.

497   **Functional predictions:**

498   *PFAM:* Conserved protein domains were predicted using PFAM version 27[49,50].

499   *Transcription factor definitions*: Predicted proteins with a known transcription factor-related

500   (DNA-binding) domain (based on the PFAM annotations) were considered to be transcription

501   factors.

502   *Carbohydrate-active enzymes prediction:* Using the Cazymes Analysis Toolkit (CAT) [51], we

503   predicted carbohydrate-active enzymes based on the original gene definitions. If a gene's protein

504   sequence was predicted to be a cazyme by either the sequence-based annotation method or the

505   PFAM-based annotation method then we considered it a cazyme.

506   *Secreted Proteins prediction:* We used the same procedure as [52] to predict secreted proteins.

507   Briefly, genes with SignalP [53] signal peptides, or a TargetP [54] Loc=S were kept. The remaining

508   genes were further filtered with TMHMM [55], keeping only genes with zero or one

26

509   transmembrane domains. Finally, genes were filtered using Wolf PSort [56] to select genes with a

510   Wolf PSort extracellular score greater than 17.

511   *Metabolic and Cytochrome P450 gene groups*: Genes with the GO annotation "metabolic

512   process" (annotation ID: GO:0008152) were called as metabolism genes. Genes with the PFAM

513   annotation PF00067 were used as Cytochrome P450 genes.

514   *KEGG:* KEGG annotations were made with the KAAS KEGG [57] annotation pipeline, using

515   genes from all available fungi, with the exception of leotiomycetes, Dothideomycetes, and

516   Microsporidians, due to the limitation of the number of species (Selected organisms by ID: cne,

517   cgi, ppl, mpr, scm, uma, mgl, sce, ago, kla, vpo, zro, cgr, ncs, tpf, ppa, dha, pic, pgu, lel, cal, yli,

518   clu, ncr, mgr, fgr, nhe, maw, ani, afm, aor, ang, nfi, pcs, cim, cpw, pbl, ure, spo, tml). The

519   GHOSTX and BBH options were selected. Predictions were made individually for both the P1

520   and P2 genomes, using the translated protein sequences.

521   *Named genes*: Named genes for *Agaricus bisporus* version 2 were downloaded from the JGI

522   DOE Genome Portal (http://genome.jgi.doe.gov/pages/search-for-

523   genes.jsf?organism=Agabi_varbisH97_2) by searching for genes with 'Name' in the 'user

524   annotations' attribute. Gene names were transferred from *A. bisporus* v. 2 using reciprocal best

525   blast hit to P1 and P2, and then selecting the best match (in the single case of an ambiguity). See

526   Supplementary Material Note R.

527   **Software and code availability:** Marker discovery and abundance calculations was done in

528   Scala, while downstream analysis was performed in python using the ibidas data query and

529   manipulation suite [58]. All source code, together with a small artificial example dataset is

530   available at: https://github.com/thiesgehrmann/Homokaryon-Expression

531   **Data Availability:** The RNA-Seq data was previously generated and can be found at bioprojects

532   PRJNA309475 and PRJNA275107. The bisulphite sequencing data can be accessed at

533   SAMN06284058.

534   **Supplementary information:** Together with this manuscript, we provide a file of

535   Supplementary Notes, and Supplementary Tables 1-4 to support our findings.

Acknowledgements

537   The authors would like to thank Brian Lavrijssen for providing the bisulphite sequencing data to

538   determine the differential methylation states. The sequence and annotation data of *A. bisporus*

539   H97 version 2 were produced by the US Department of Energy Joint Genome Institute

540   http://www.jgi.doe.gov/ in collaboration with the user community. This research is supported by

541   the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for

542   Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

Author contributions

544   TG, HABW, MJTR and TA wrote the manuscript. JFP performed the experiments. TG, HABW,

545   MJTR and TA designed the analyses. RAO created the gene and functional annotations. TG

546   performed the analyses. All authors aided in biological interpretation of the results. All authors

547   reviewed the manuscript.

Conflict of interest

549   The authors declare no conflicts of interest

550   **References**

551   1.   Morin, E. *et al.* Genome sequence of the button mushroom Agaricus bisporus reveals

552   mechanisms governing adaptation to a humic-rich ecological niche. *Proc. Natl. Acad. Sci.*

553   **109,** 17501–17506 (2012).

554   2.   Patyshakuliyeva, A. *et al.* Uncovering the abilities of Agaricus bisporus to degrade plant

555   biomass throughout its life cycle. *Environ. Microbiol.* **17,** 3098–3109 (2015).

556   3.   Ohm, R. a *et al.* Genome sequence of the model mushroom Schizophyllum commune.

557   *Nat. Biotechnol.* **28,** 957–63 (2010).

558   4.   Pawlowska, T. E. Genetic processes in arbuscular mycorrhizal fungi. *FEMS Microbiol.*

559   *Lett.* **251,** 185–192 (2005).

560   5.   ud din Khanday, M. *et al.* in *Soil Science: Agricultural and Environmental Prospectives*

561   317–332 (Springer International Publishing, 2016). doi:10.1007/978-3-319-34451-5_14

562   6.   Sun, C. *et al.* The beneficial fungus Piriformospora indica protects Arabidopsis from

563   Verticillium dahliae infection by downregulation plant defense responses. *BMC Plant*

564   *Biol.* **14,** 268 (2014).

565   7.   Harrach, B. D., Baltruschat, H., Barna, B., Fodor, J. & Kogel, K.-H. The mutualistic

566   fungus Piriformospora indica protects barley roots from a loss of antioxidant capacity

567   caused by the necrotrophic pathogen Fusarium culmorum. *Mol. Plant. Microbe. Interact.*

568   **26,** 599–605 (2013).

569   8.   Babikova, Z. *et al.* Underground signals carried through common mycelial networks warn

570   neighbouring plants of aphid attack. *Ecol. Lett.* n/a-n/a (2013). doi:10.1111/ele.12115

571   9.   Collins, C. *et al.* Genomic and proteomic dissection of the ubiquitous plant pathogen,

572   armillaria mellea: Toward a new infection model system. *J. Proteome Res.* **12,** 2552–2570

573   (2013).

574   10.   Khoshraftar, S. *et al.* Sequencing and annotation of the Ophiostoma ulmi genome. *BMC*

575   *Genomics* **14,** 162 (2013).

576   11.   Guo, L. *et al.* Genome and transcriptome analysis of the fungal pathogen fusarium

577   oxysporum f. Sp. Cubense causing banana vascular wilt disease. *PLoS One* **9,** (2014).

578   12.   Specht, C. A. Isolation of the Ba and Bb mating-type loci of Schizophyllum commune.

579   *Curr. Genet.* **28,** 374–379 (1995).

580   13.   Saksena, K. N., Marino, R., Haller, M. N. & Lemke, P. a. Study on development of

581   Agaricus bisporus by fluorescent microscopy and scanning electron microscopy. *J.*

582   *Bacteriol.* **126,** 417–428 (1976).

583   14.   Craig, G. D., Newsam, R. J., Gull, K. & Wood, D. A. An ultrastructural and

584   autoradiographic study of stipe elongation inAgaricus bisporus. *Protoplasma* **98,** 15–29

585   (1979).

586   15.   Sonnenberg, A. S. M. *et al.* A detailed analysis of the recombination landscape of the

587   button mushroom Agaricus bisporus var. bisporus. *Fungal Genet. Biol.* **93,** 35–45 (2016).

588   16.   Todd, R. B., Davis, M. a & Hynes, M. J. Genetic manipulation of Aspergillus nidulans:

589   heterokaryons and diploids for dominance, complementation and haploidization analyses.

590   *Nat. Protoc.* **2,** 822–830 (2007).

591   17.   Boon, E., Zimmerman, E., Lang, B. F. & Hijri, M. Intra-isolate genome variation in

592   arbuscular mycorrhizal fungi persists in the transcriptome. *J. Evol. Biol.* **23,** 1519–1527

593   (2010).

594   18.   James, T. Y., Stenlid, J., Olson, ??Ke & Johannesson, H. Evolutionary significance of

595   imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus

596   Heterobasidion parviporum. *Evolution (N. Y).* **62,** 2279–2296 (2008).

597   19.   Muzzey, D., Sherlock, G. & Weissman, J. S. Extensive and coordinated control of allele-

598   specific expression by both transcription and translation in Candida albicans. *Genome Res.*

599   **24,** 963–973 (2014).

600   20.   Wei, X. & Wang, X. A computational workflow to identify allele-specific expression and

601   epigenetic modification in maize. *Genomics. Proteomics Bioinformatics* **11,** 247–52

602   (2013).

603   21.   Crowley, J. J. *et al.* Analyses of allele-specific gene expression in highly divergent mouse

604   crosses identifies pervasive allelic imbalance. *Nat. Genet.* **47,** 353–360 (2015).

605   22.   Buckland, P. R. Allele-specific gene expression differences in humans. *Hum. Mol. Genet.*

606   **13,** 255–260 (2004).

607   23.   Pant, P. V. K. *et al.* Analysis of allelic differential expression in human white blood cells.

608   *Genome Res.* **16,** 331–339 (2006).

609   24.   Edger, P. P., Smith, R., Mckain, M. R., Cooley, A. M. & Vallejo-marin, M. Subgenome

610   dominance in an interspecific hybrid , synthetic allopolyploid , and a 140 year old

611   naturally established neo-allopolyploid monkeyflower. *bioRxiv* 1–27 (2016).

612   doi:10.1101/094797

613   25.   Horton, T. R. The number of nuclei in basidiospores of 63 species of ectomycorrhizal

614   Homobasidiomycetes. *Mycologia* **98,** 233–238 (2006).

615   26.   Lin, K. *et al.* Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei

616   of an Endomycorrhizal Fungus. *PLoS Genet.* **10,** (2014).

617   27.   Pelkmans, J. F. *et al.* The transcriptional regulator c2h2 accelerates mushroom formation

618   in Agaricus bisporus. *Appl. Microbiol. Biotechnol.* **2,** (2016).

619   28.   Bonnen, A. M., Anton, L. L. H., Orth, A. B., Anton, L. L. H. & Ortht, A. N. N. B. Lignin-

620   degrading enzymes of the commercial button mushroom, Agaricus bisporus. *Appl.*

621   *Environ. Microbiol.* **60,** 960–965 (1994).

622   29.   Borel, C. *et al.* Biased allelic expression in human primary fibroblast single cells. *Am. J.*

623   *Hum. Genet.* **96,** 70–80 (2015).

624   30.   Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9

625   methylation by RNAi. *Science* **297,** 1833–7 (2002).

626   31.   Woolston, B. M. *et al.* Long-distance translocation of protein during morphogenesis of the

627   fruiting body in the filamentous fungus, agaricus bisporus. *PLoS One* **6,** (2011).

628   32.   McDaniell, R. *et al.* Heritable Individual-Specific and Allele-Specific Chromatin

bioRxiv preprint doi: https://doi.org/10.1101/141689; this version posted July 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

629   Signatures in Humans. *Science (80-.).* **328,** 235–239 (2010).

630 33. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA

631   sequencing of the human brain. *Science (80-.).* **352,** 1586–1590 (2016).

632 34. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome

633   of postmortem neurons. *Nat. Protoc.* **11,** 499–524 (2016).

634 35. Loftus, B. J. *et al.* The genome of the basidiomycetous yeast and human pathogen

635   Cryptococcus neoformans TL - 307. *Science (80-.).* **307 VN-,** 1321–1324 (2005).

636 36. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen

637   Ustilago maydis. *Nature* **444,** 97–101 (2006).

638 37. Shu, C. H. & Hsu, H. J. Production of schizophyllan glucan by Schizophyllum commune

639   ATCC 38548 from detoxificated hydrolysate of rice hull. *J. Taiwan Inst. Chem. Eng.* **42,**

640   387–393 (2011).

641 38. Xu, H. *et al.* FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads.

642   *PLoS One* **7,** 1–6 (2012).

643 39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina

644   sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

645 40. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for

646   Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).

647 41. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*

648   **9,** 357–359 (2012).

33

649   42.   Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide

650   DNA methylation profiles. *Genome Biol.* **13,** R87 (2012).

651   43.   Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1:

652   Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and

653   AUGUSTUS: Table 1. *Bioinformatics* **32,** 767–769 (2016).

654   44.   Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically

655   mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24,** 637–644

656   (2008).

657   45.   Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5,**

658   R12 (2004).

659   46.   Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment

660   search tool. *J. Mol. Biol.* **215,** 403–10 (1990).

661   47.   Aho, A. V. & Corasick, M. J. Efficient string matching: an aid to bibliographic search.

662   *Commun. ACM* **18,** 333–340 (1975).

663   48.   Anders, S. *et al.* Differential expression analysis for sequence count data. *Genome Biol.*

664   **11,** R106 (2010).

665   49.   Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36,** D281–D288

666   (2008).

667   50.   Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42,** 222–230

668   (2014).

669   51.   Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The

670   carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490–

671   D495 (2014).

672   52.   Morais do Amaral, A., Antoniw, J., Rudd, J. J. & Hammond-Kosack, K. E. Defining the

673   Predicted Protein Secretome of the Fungal Wheat Leaf Pathogen Mycosphaerella

674   graminicola. *PLoS One* **7,** 1–19 (2012).

675   53.   Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating

676   signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786 (2011).

677   54.   Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting Subcellular

678   Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J. Mol. Biol.*

679   **300,** 1005–1016 (2000).

680   55.   Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane

681   protein topology with a hidden markov model: application to complete genomes. *J. Mol.*

682   *Biol.* **305,** 567–580 (2001).

683   56.   Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35,**

684   W585–W587 (2007).

685   57.   Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: An automatic

686   genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35,** 182–185

687   (2007).

688   58.   Hulsman, M., Bot, J. J., Vries, A. P. de & Reinders, M. J. T. Ibidas: Querying Flexible

689   Data Structures to Explore Heterogeneous Bioinformatics Data. *Data Integr. Life Sci.* 23–

690   37 (2013). doi:10.1007/978-3-642-39437-9_2

691

692