SI Appendix for *"*Nucleus specific expression in the multinucleated mushroom *Agaricus bisporus* reveals different nuclear regulatory programs*"*

Thies Gehrmann, Jordi F. Pelkmans, Robin A. Ohm, Aurin M. Vos, Anton S.M. Sonnenberg, Johan J. P. Baars, Han A. B. Wösten, Marcel J. T. Reinders, Thomas Abeel

SI Appendix A: The number of markers per karyollele pair can be asymmetric

Most karyollele pairs have the same number of markers per karyollele. This is illustrated by large points along the diagonal of figure SF.A.1. However, some markers have different numbers of markers. Figure SF.A.2 illustrates an example of how this asymmetry occurs due to differences in the variant boundaries. Furthermore, asymmetry can also be introduced by duplication events which make markers non-unique across the rest of the genome, or even domain sequences with a mutation in one homokaryon, but not in the other.

This variation can be explained by the non-symmetric number of markers produced by the different kinds of variation. While a SNP will result in one marker in each karyollele, an indel (if longer than 21bp) will result in one marker in one karyollele, and at least two in the other.

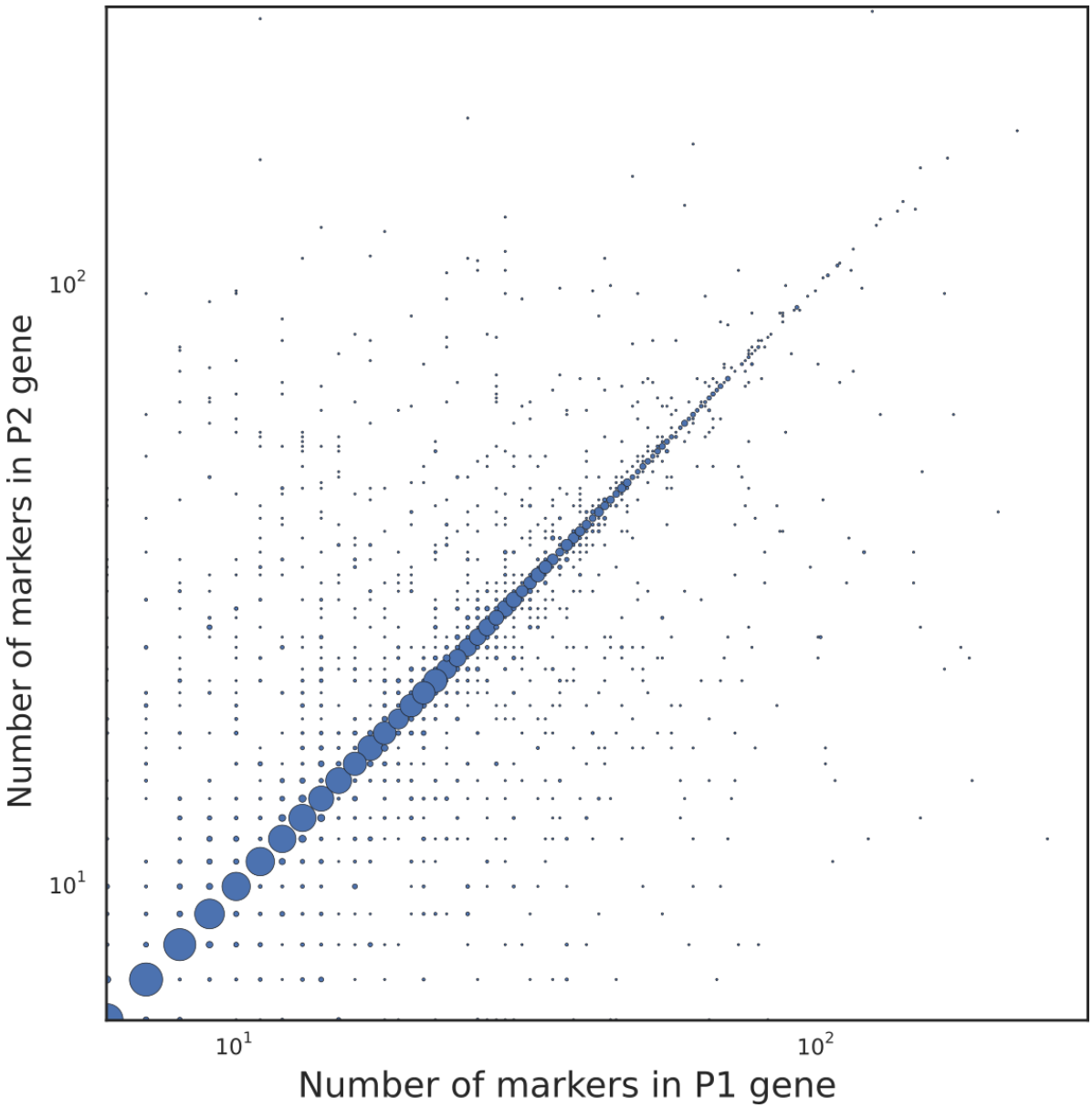

Figure SF.A.1: The number of karyollele pairs with a given number of markers. On the x-axis and y-axis is given the number of markers discovered in P1, and P2 respectively. The size of a point represents the number of karyollele pairs with that number of markers per homokaryon.

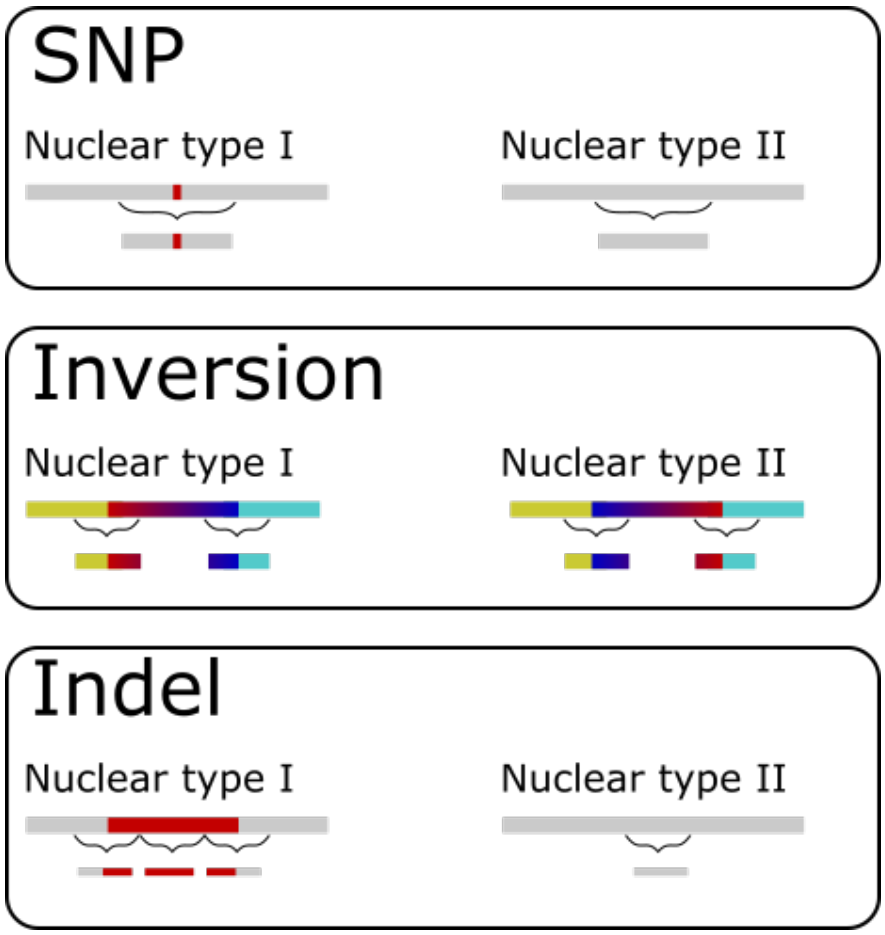

Figure SF.A.2: Different variants cause different marker counts per nuclear type. SNPs and inversions will result in an equal number of markers per nuclear type, whereas an indel can result in many more, due to the increased number of variant boundaries, and the novel sequence introduced.

SI Appendix B: Differences in CG content is not correlated to differences in expression

Figure SF.B.1 shows the relationship between CG content and karyollele expression. For each karyollele pair, we calculate the average CG content and the average expression across all markers for each karyollele. We calculate the difference between CG contents. Additionally, the difference in read depth and normalize it to lie within [1,-1] (1 represents an entirely P1 expressed karyollele, and -1 an entirely P2 expressed karyollele, 0 represents equal expression). The difference in karyollele expression and CG content of its markers is not correlated (p=0.05)

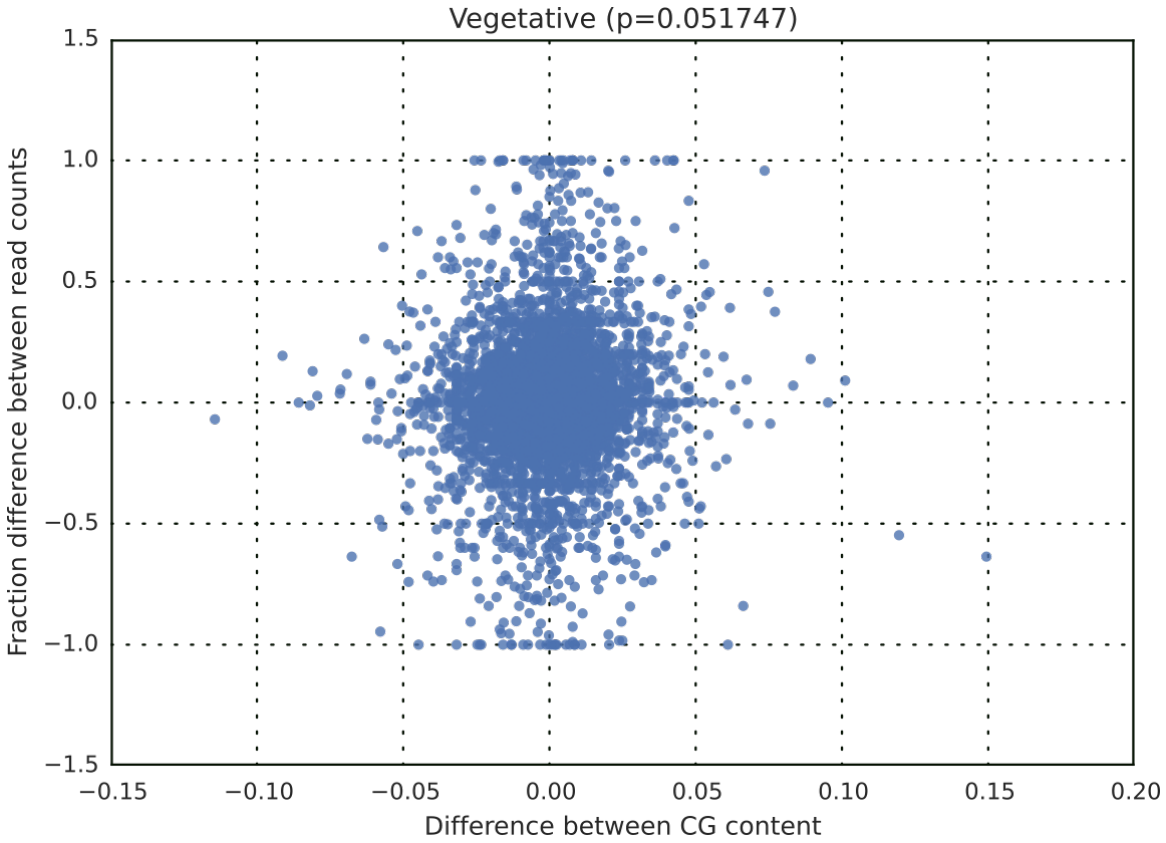

Figure SF.B.1: Relationship between the difference in gene expression between karyollele pairs (y-axis) and average marker CG content per karyollele (x-axis). The average expression and average CG content of markers do not seem to be related. Here shown only for the vegetative stage in the mushroom dataset; Other samples show the same behaviour.

SI Appendix C: PCA plots of tissue and compost samples

For each sample, we construct two vectors of size 5,060, describing the expression of each karyollele in P1 and P2, respectively. With this, we perform a PCA to observe the relationships between the different samples and nuclear types. Figure SF.C.1 shows a divergence of P1 (shown in red) and P2 (shown in blue) expression measurements in mushroom samples, indicating that P1 homokaryons are more similar to P1 homokaryons than to any P2 homokaryon, and vice versa. Figure SF.C.2 shows the same for compost data.

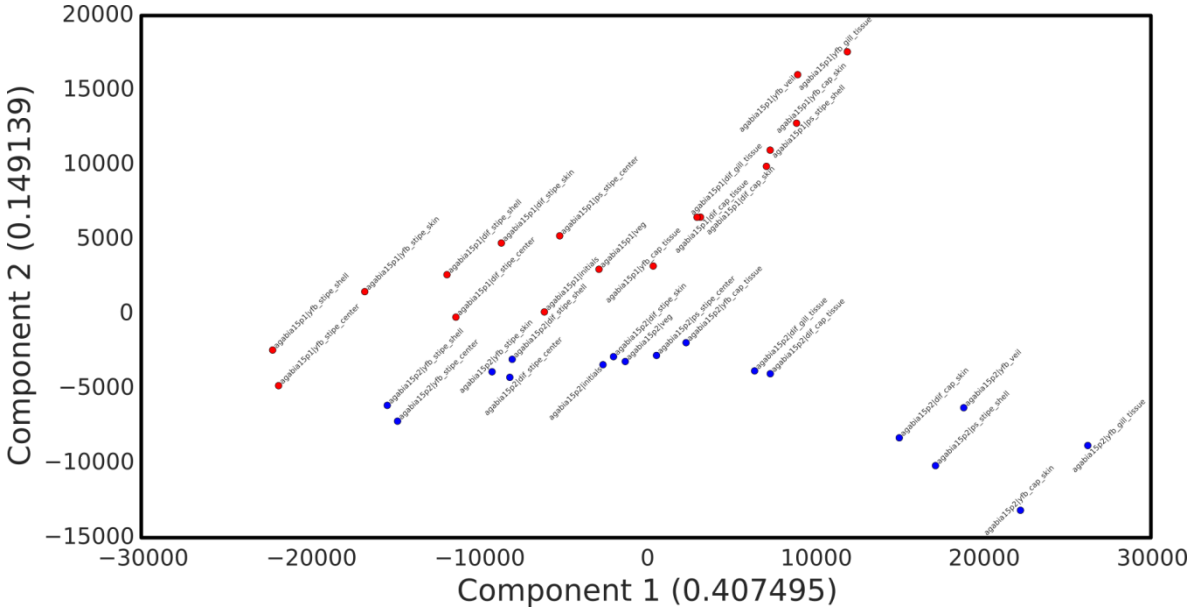

Figure SF.C.1: PCA dimensionality reduction of karyollele expression data. The first and second components are plotted on the X and Y axes, respectively. P1 homokaryon samples are shown in red, and P2 homokaryons are shown in blue.

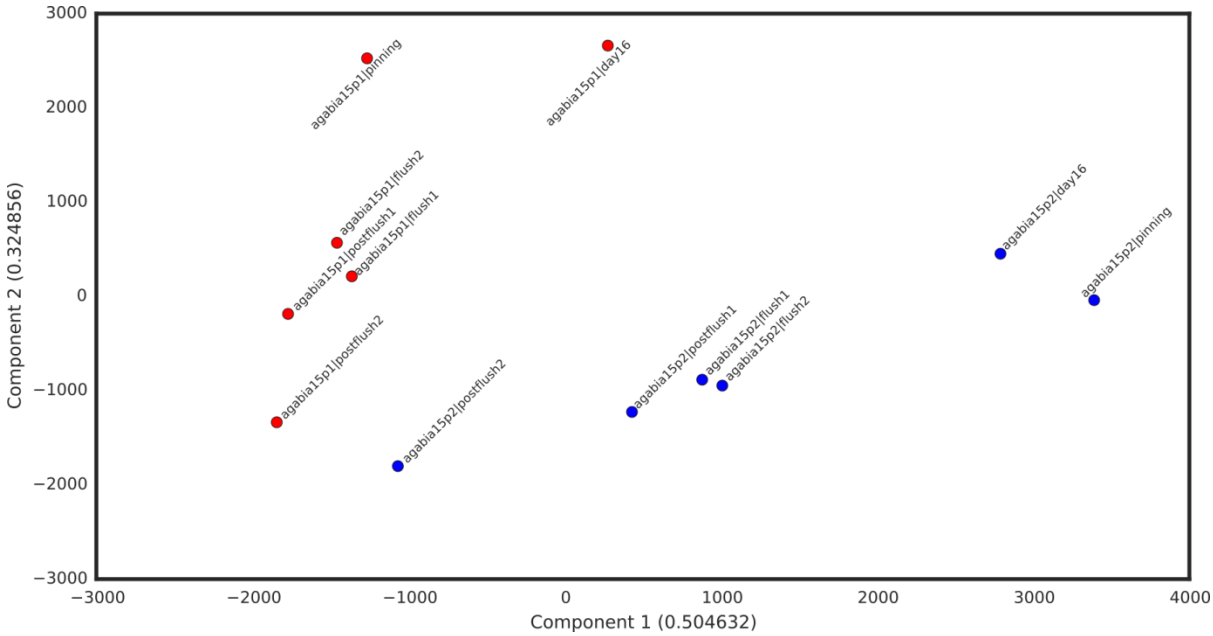

Figure SF.C.2: The first and second components of a PCA are shown on the X and Y axes, respectively for compost data.

SI Appendix D: Bias of extreme genes

Figure SF.D.1 shows the influence of highly expressed genes on the Chromosome Read Ratio (CRR measure). We sort the genes on each chromosome based on the sum of their expression in the two nuclear types. Sorting the genes by their expression (lowest to highest), and starting with the lowest expressed gene, we calculate the CRR and CGR (Equations 4 and 6, respectively) ratios per chromosome, for increasingly larger sets of genes. We see that the CRR (red line) considerably changes once we consider highly expressed genes. The CGR (black line) is more stable towards highly expressed genes.

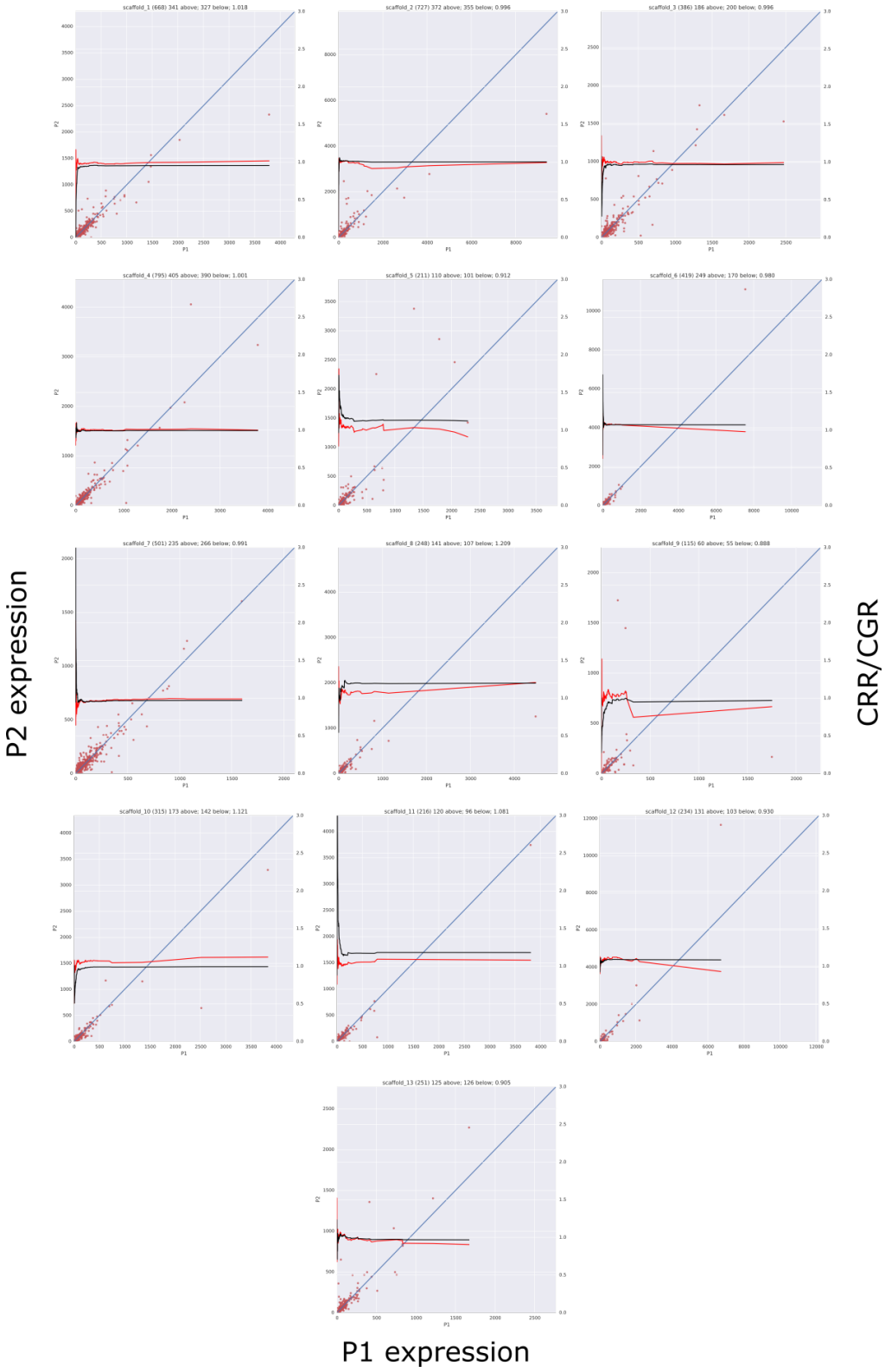

Figure SF.D.1: The bias of highly expressed genes. X-axis represents P1 expression, and Y-axis represents P2 expression. The red points represent the genes on each chromosome and their expression values in the P1 and P2 nuclear types. The blue line is the identity line; points on or near this line have near-identical expression. The red and black lines are the CRR and the CGR, with a separate y-axis on the right hand side, calculated by continuously considering the next most highly expressed gene. Read ratio is very affected by the highest expressed genes. Each chromosome is plotted individually.

SI Appendix E: Extreme genes

In Figure 2 of the main text, we showed that more mRNA originates from P2 in the case of, for example, chromosome 9 than from P1. This was in part due to a few genes which were very highly expressed. These highly expressed genes skew the read count ratios (see Supplementary Material Note E). The differences can be quite extreme; In one case, a P1 karyollele accounted for <1% of all reads originating from chromosome 9, while its P2 karyollele accounted for 21% of all the chromosome 9 reads. Hence, most of the observed differences for chromosome 9 (Figure 2a) is explained by such highly expressed genes.

In total, we identified 22 genes whose contribution exceeds 10% of the total expression of the chromosome it is located on. Most of these genes are differentially expressed between the two nuclear types, with 16 showing fold changes larger than 2. These genes are primarily metabolic.

These genes, and their contribution to the CRR scores are provided in data sheet SI Dataset S1.

SI Appendix F: Nuclear type Gene Ratio (NGR) measures in the mushroom dataset

Figure SF.F.1 shows the Nuclear type Gene Ratio measures for the mushroom dataset. It becomes clear that on average, the P1 nuclear type is dominant over the P2 nuclear type. This is statistically significant (the log-transformed chromosome gene ratios are significantly > 0, following a t-test in mushroom tissue, with $p < 0.01$).

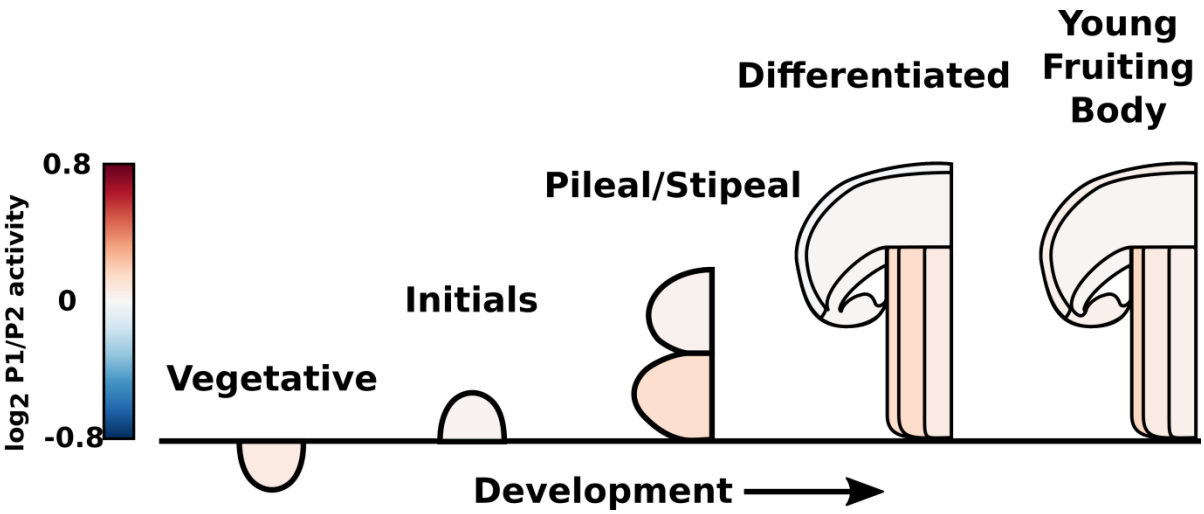

Figure SF.F.1: The Nuclear type Gene Ratio measures in the mushroom dataset.

SI Appendix G: Methylation

A sample of vegetative stage mycelium of A15 was treated with the EpiTect Bisulphite conversion and cleanup kit and sequenced with the Illumina HiSeq 2000. Raw reads were trimmed using TRIMMOMATIC(1) and aligned to the A15 P1 genome using Bismark(2) and bowtie2(3). Methylated bases were analyzed with Methylkit(4). Only bases which had a minimum coverage of 10 were retained. For samples with mixed methylation states, we will observe what appear to be incomplete conversions of unmethylated cytosines but in reality represents the mixed methylation states of those bases. Therefore, to include only differentially methylated bases between the two nuclei (i.e. methylated in one homokaryon, but not in the other), we considered only those bases which were measured to be methylated between 40 and 60% of all reads (Supplementary Material Notes I). While 164,290 bases had an indication of methylation signal, 10,325 bases had methylation signals of about 50%, suggestive of differential methylation states. Methylated bases were mapped to genes when between the start and stop codons, or 1000bp up/downstream (Figure SF.G.1)

To investigate the biological mechanism causing differential expression, we measured methylation on the A15 genome. Assuming that the relative Cystosine/Thymine coverage at each base relates to a differential methylation state between the two nuclear types, we conclude that 277 genes are differentially methylated (Methods). 42 of these 277 genes were also found to be differentially expressed between the two nuclear types at some point in development. Although this is a significant proportion ($p < 0.05$, χ^2 test), methylation only explains at most 10% of the differential expression we observe. Noteworthy is that 40 of the 42 differentially expressed and differentially methylated genes are differentially expressed in mushroom tissues), whereas only 9 are differentially expressed in the vegetative mycelium. This indicates that the largest impact of differential methylation is much later in mushroom development, suggesting that methylation has a delayed effect on expression.

Table ST.G.1 shows these overlaps for the different sets of differentially expressed genes. We find that the methylated genes overlap mostly with the genes which are differentially expressed in the mushroom data.

Table ST.G.1: Overlap of differentially expressed genes and methylated genes. The p-values of a chi-squared approximation of the fisher's exact test have been bonferroni corrected (q-values). Significant corrected values have been highlighted in bold.

	Differentially expressed		Not Differentially Expressed			
Set	Methylated	Not Methylated	Methylated	Not Methylated	p-value	q-value
Vegetative mycelium + Mushroom	42	369	235	4440	1.465E-05	8.791E-05
Vegetative mycelium	9	73	268	4736	4.780E-02	2.868E-01
Mushroom	40	328	237	4481	3.473E-06	2.084E-05
Overlap	7	32	270	4777	1.937E-03	1.162E-02
Unique Vegetative mycelium	2	41	275	4768	9.150E-01	1.000E+00
Unique Mushroom	33	296	244	4513	2.493E-04	1.496E-03

Figure SF.G.1 shows the regions on the genome which are differentially methylated in green. These mostly overlap with regions that are represented by repetitive elements (Sonnenberg et al. 2016), where we cannot distinguish genes based in sequence. In Figure SF.G.2, we show the fraction of methylated and unmethylated cytosines for a given base. This figure is indicative of an organism with mixed methylation statuses, which is to be expected in our case.

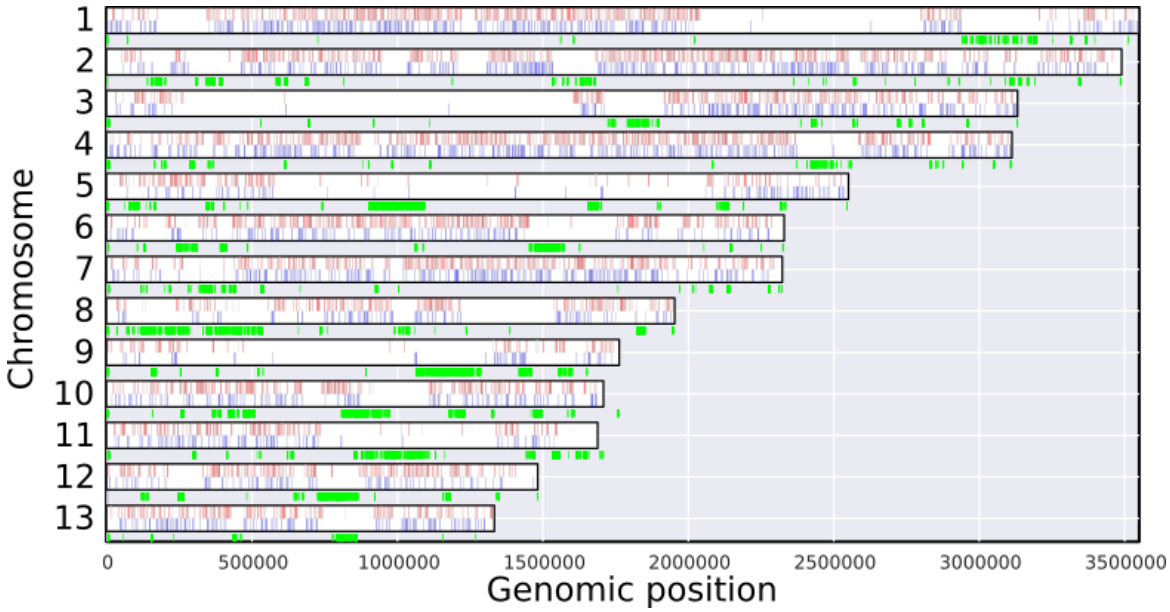

Figure SF.G.1: Differentially methylated bases in the A15 genome. As in figure 2 of the main text, the red and blue marks indicate genes with upregulation in the P1 and P2 homokaryons, respectively. The green marks indicate bases which are methylated at that point.

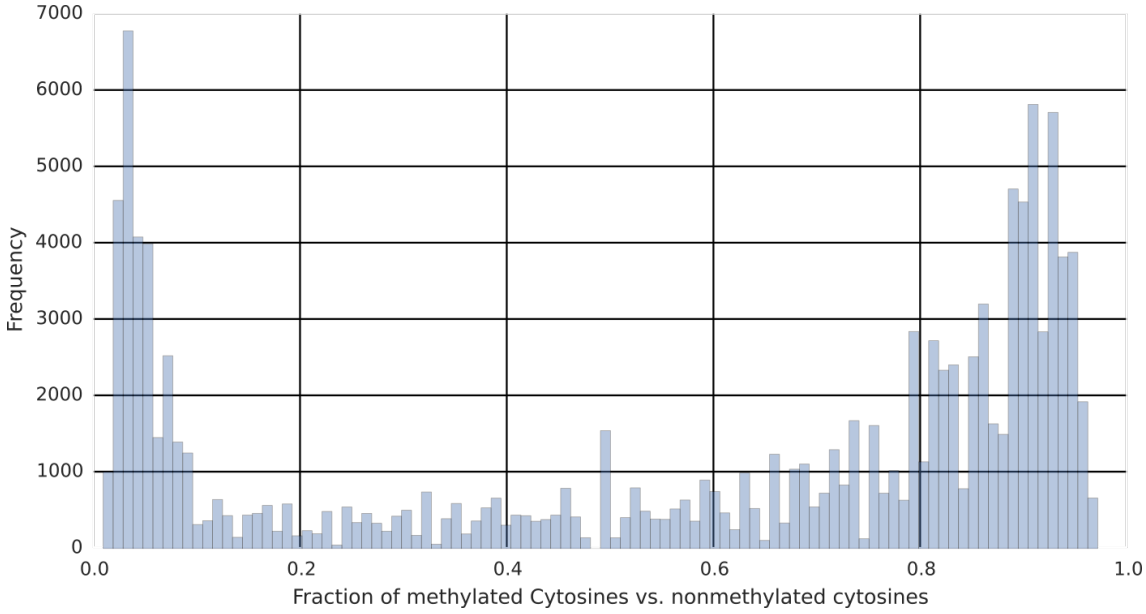

Figure SF.G.2: Frequency distribution of methylated vs. non methylated cytosines.

SI Appendix H: Overlapping differentially expressed genes that changed origin

5 genes changed their differential expression between P1 and P2 between the vegetative mycelium and the mushroom datasets. These are provided in SI Dataset S2. For each karyollele pair, we provide the samples for which the differential expressions were significant. The blue rows indicate the samples in which they were more highly expressed by the P2 nuclear type in the vegetative mycelium, and the red rows indicate the samples in which they were more highly expressed by the P1 nuclear type in the mushroom tissue.

SI Appendix I: Manganese Peroxidase

Of 90 genes with named annotations in *A. bisporus* (see SI Appendix R), 42 were identified as differentiable karyollele pairs, and one, manganese peroxidase (*mnp1)* was differentially expressed between the two nuclear types in any stage of development. *mnp1* is known to be highly expressed in early stages of development, and drops to much lower levels (log fold change of -2.8) after mushroom formation(5, 6). In our datasets, the individual contributions of P1 and P2 to *mnp1* expression are largely different. In the vegetative mycelium, we find that P2 produces four-fold more *mnp1* immediately before mushroom formation than P1. In the mushroom tissue, however, *mnp1* is expressed on average 4.2-fold higher by P1 in the stem of the fruiting body throughout development. Whether this switching behavior is functionally relevant remains unclear, as two karyolleles of *mnp1* have the same protein domain annotations in the P1 and P2 homokaryon genomes. The Gene Read Ratio (GRR) in mushroom tissues is provided in figure SF.I.1, and the total read counts in compost are provided in figure SF.I.2. Notice that the scale of figure SF.I.1 is different from figure 2 in the text.

mnp1 is known to be involved in lignin degradation, which occurs in the vegetative mycelium(5, 6). In compost, the abundance decreases dramatically throughout development. Therefore, the abundance of *mnp1* in the stipe of the fruiting body is unexpected, although it has been shown that proteins produced in the mycelium can find their way into the mushroom(7). However, it does not explain the fact that the P1 karyollele exists in higher abundance in the mushroom tissues, while the P2 karyollele is higher expressed in the vegetative mycelium. Transport of the P2 karyollele from the vegetative mycelium into the mushroom conflicts with the abundances of the P1 karyollele observed in the mushroom tissues.

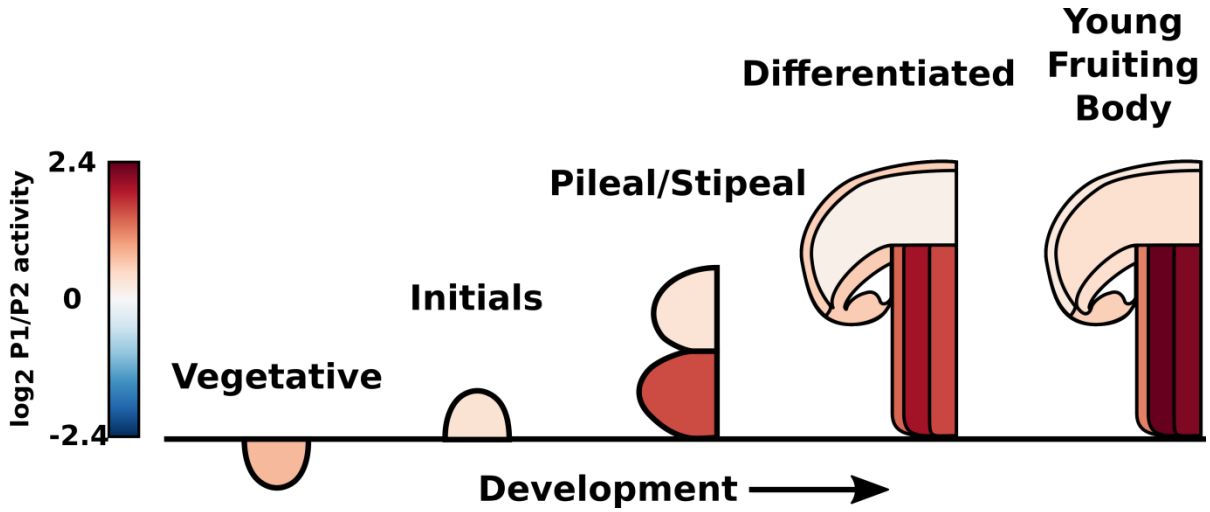

Figure SF.I.1: Gene Read Ratios for Manganese Peroxidase in mushroom tissues. Red indicates a higher contribution of P1, and blue indicates a higher contribution of P2. Expression values range between [26.2,51.1] in *P1,* and between [9.3,38.2] in *P2*.

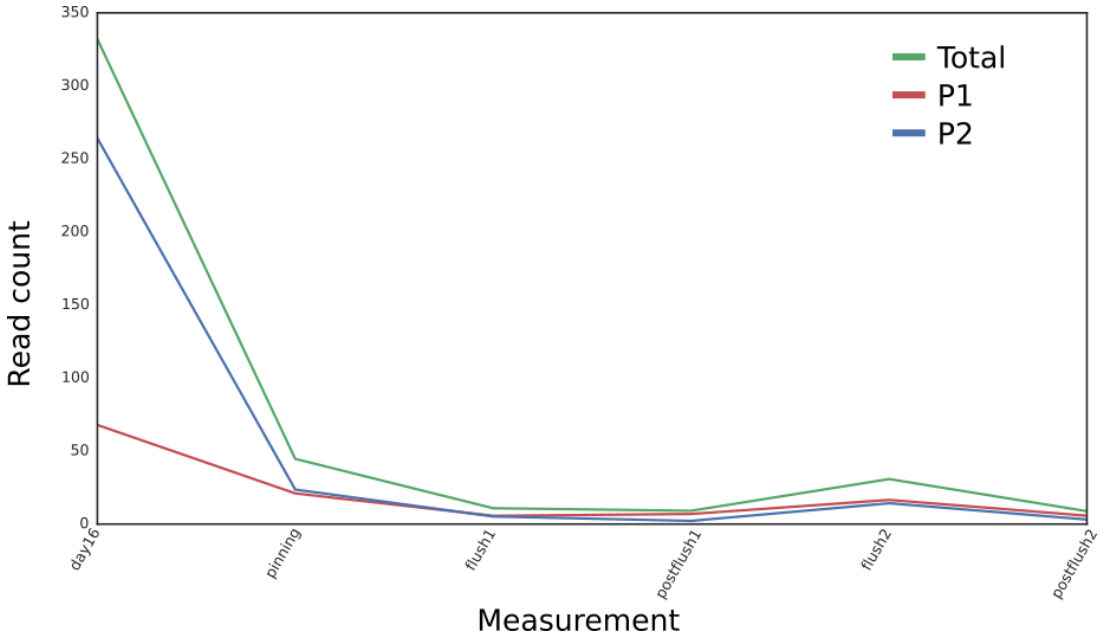

Figure SF.I.2: Manganese peroxidase expression in the vegetative mycelium dataset. P2 expression is dominant in the vegetative growth of mycelium, but drops shortly after that, in concordance with previous literature(5)**.**

SI Appendix J: Co-localized, co-regulated clusters

To detect co-regulated clusters, we slide a window of size 20,001bp (10,000- up and down-stream) across each chromosome. In this window, we count the number of genes that are more highly expressed by P1 and by P2, and calculate the difference per sample. I.e.

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

where *W(x,y)* is the set of genes between genomic location *x* and *y*, and *s* is a sample. This difference is shown in Figure 3c. Next, we identify regions where each sample in the dataset shows consistent regulation. That is to say, in these regions, $D(x,s) > 0 \; \forall$ s ∈ S, or $D(x,s) < 0 \; \forall$ s ∈ S, where S is the set of all samples. These regions contain co-localized genes that are co-regulated across all samples.

Table ST.J.1 indicates the number of regions that are consistently upregulated in P1 and P2. We find that more genes are consistently upregulated in P1, in a larger number of regions than in P2. This observation consolidates the conflicting observations of a dominant P1 nuclear type in terms of mRNA production, and more upregulated, differentially expressed genes in P2.

Figure 3c of the main manuscript showed the co-localized and co-regulated regions on chromosome 10. In Supplementary Figures SF.J.1-13, we provide the same figures for the remaining chromosomes. Additionally, in SI Dataset S3 we provide the exact regions for each dataset.

Table ST.J.1: The number of regions in which the majority of the genes are coregulated (SI Appendix L) across the mushroom and mycelium datasets and with the number of genes in these regions. P1 and P2 columns indicate whether the region is consistently higher in for the P1 or P2 karyollele, respectiverly. Row Both indicates overlapping regions between the mushroom and vegetative mycelium datasets.

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

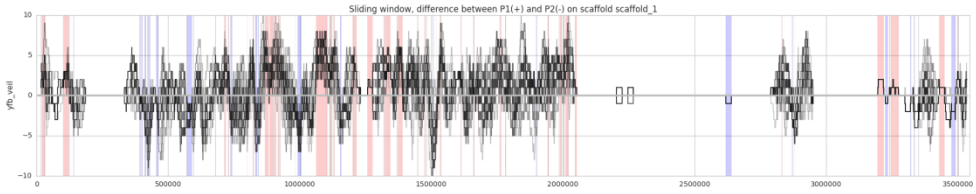

Figure SF.J.1: Equation 8 plotted for all datasets on scaffold 1. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

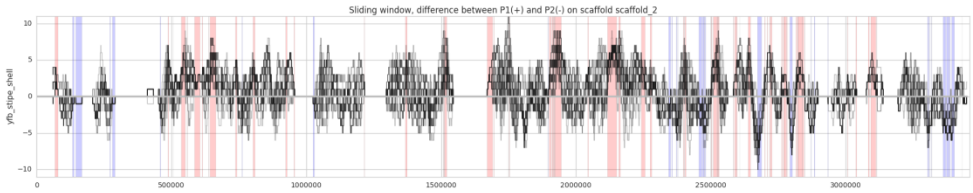

Figure SF.J.2: Equation 8 plotted for all datasets on scaffold 2. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

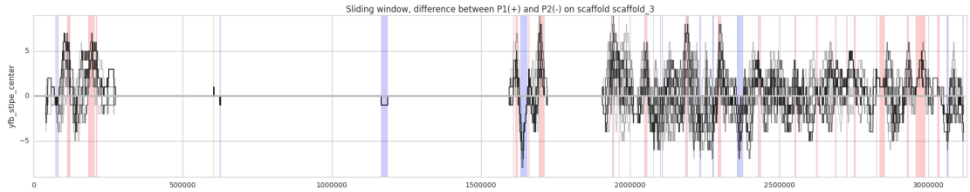

Figure SF.J.3: Equation 8 plotted for all datasets on scaffold 3. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

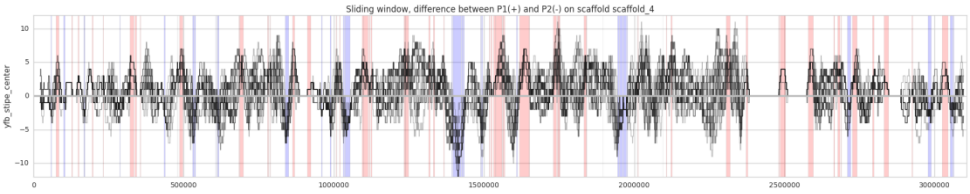

Figure SF.J.4: Equation 8 plotted for all datasets on scaffold 4. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

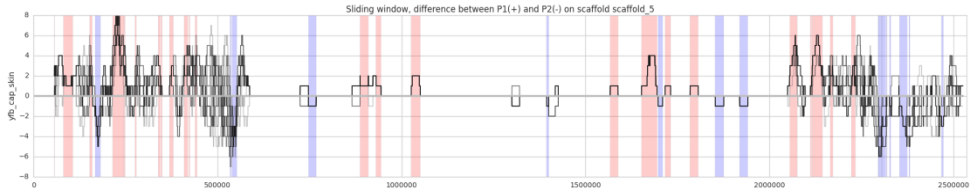

Figure SF.J.5: Equation 8 plotted for all datasets on scaffold 5. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

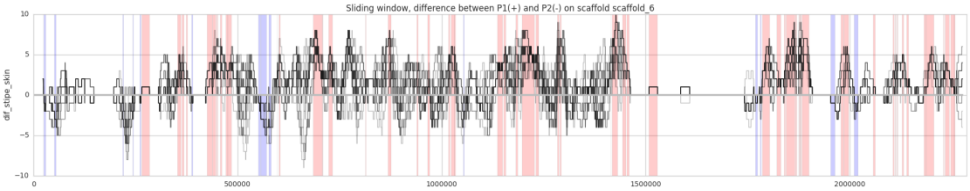

Figure SF.J.6: Equation 8 plotted for all datasets on scaffold 6. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

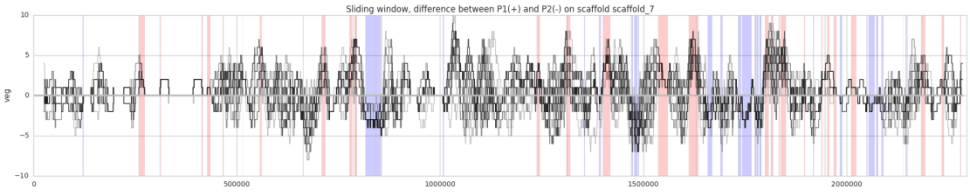

Figure SF.J.7: Equation 8 plotted for all datasets on scaffold 7. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

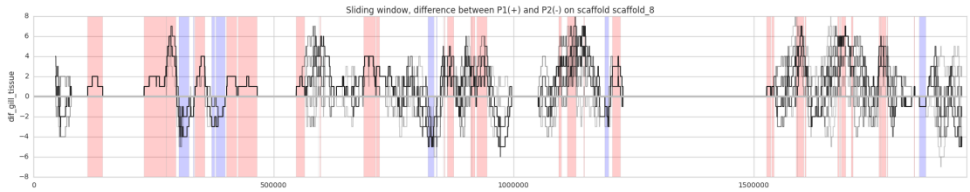

Figure SF.J.8: Equation 8 plotted for all datasets on scaffold 8. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

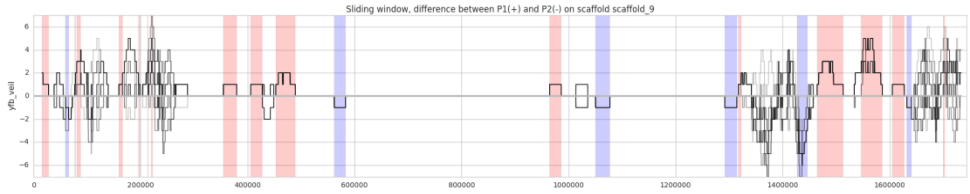

Figure SF.J.9: Equation 8 plotted for all datasets on scaffold 9. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

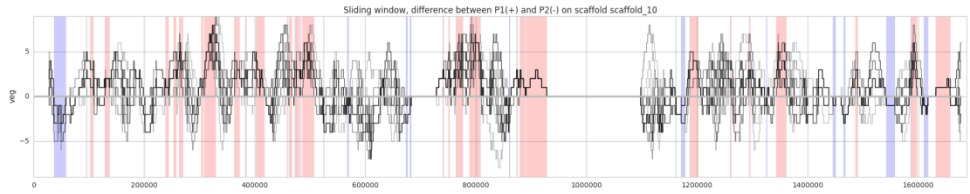

Figure SF.J.10: Equation 8 plotted for all datasets on scaffold 10. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

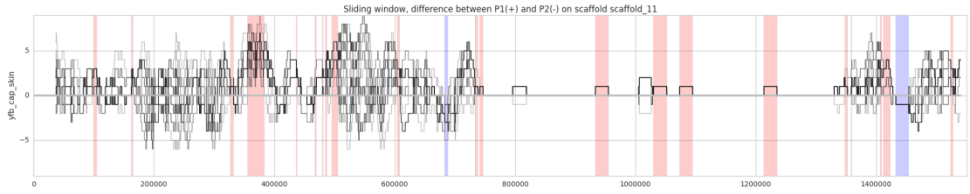

Figure SF.J.11: Equation 8 plotted for all datasets on scaffold 11. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

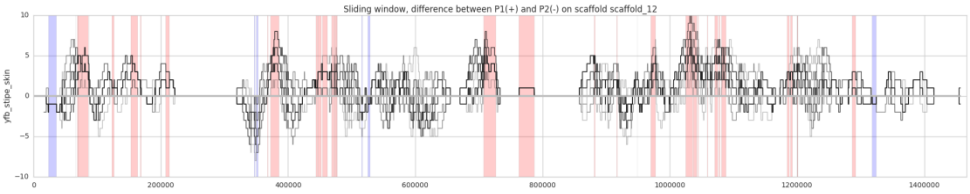

Figure SF.J.12: Equation 8 plotted for all datasets on scaffold 12. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

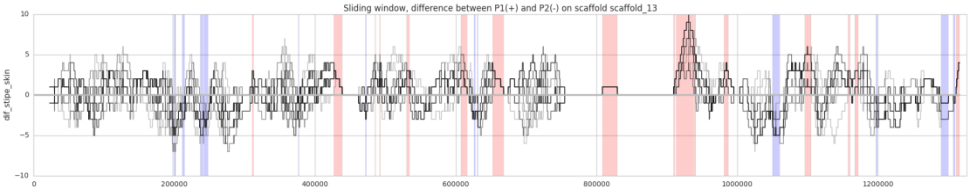

Figure SF.J.13: Equation 8 plotted for all datasets on scaffold 13. Red regions indicate regions of P1 predominance, and blue regions indicate regions of P2 predominance.

SI Appendix K: Probability of observing differential expression imbalance

We can judge the likelihood of observing an imbalance in the number of differentially expressed genes on the two homokaryons. For a given time point with x upregulated genes in P1, and y upregulated genes in P2, we can determine the probability of observing *max(x, y)* positive trials within *x+y* trials, under the null hypothesis of there being no difference in chance of a gene being upregulated in either homokaryon (i.e. $p = 0.5$). The probability of observing a value *max(x,y)* or greater can be calculated as $1\text{-}P(x < max(x,y))$. If this probability is sufficiently small, we may reject the underlying assumption that the probability of being upregulated is the same in both homokaryons (i.e. $p \neq 0.5$). P-values are corrected by controlling for a *0.05* FDR. The difference was only significant in some vegetative growth time points, shown underlined in ST.K.1.

Table ST.K.1: The number of differentially expressed genes that are more highly expressed in P1 or P2 in each different sample, together with the significance of this difference

Condition	Up in P1	Up in P2	pvalue	qvalue
Day 16	11	29	0.001111	0.027768
Flush 1	14	32	0.002267	0.028337
Total vegetative dataset	30	52	0.005319	0.044323
Pinning	14	29	0.006859	0.044323
Flush 2	20	32	0.035197	0.148658
Post Flush 2	18	30	0.029732	0.148658
Post Flush 1	14	24	0.036476	0.148658
YFB Stipe Center	72	86	0.116317	0.363492
Initials	58	50	0.193286	0.487041
Total Mushroom dataset	176	192	0.18777	0.487041
YFB Cap Tissue	91	103	0.175335	0.487041
PS Stipe Shell	63	54	0.17764	0.487041
Dif. Stipe skin	75	80	0.314999	0.504822
Dif. Cap Skin	58	60	0.391278	0.504822
YFB Veil	72	70	0.40067	0.504822
YFB Cap Skin	89	82	0.270406	0.504822
Dif. Cap Skin	63	61	0.393873	0.504822
Dif. Gill Tissue	60	59	0.427315	0.504822
Dif. Stipe Center	87	83	0.350741	0.504822
Dif. Stipe Shell	86	90	0.353185	0.504822
YFB Gill Tissue	69	65	0.332975	0.504822
YFB Stipe Skin	94	94	0.470943	0.504822
PS Stipe Center	71	77	0.282586	0.504822
Vegetative	47	42	0.262507	0.504822
YFB Stipe Shell	93	92	0.441575	0.504822

SI Appendix L: KEGG pathways with differentially expressed genes

We examined the KEGG annotations of the 411 differentially expressed genes to further elucidate their functional impact. Sixteen of these genes were found in 20 pathways. Interestingly, three differentially expressed genes were found in the Aminoacyl-tRNA biosynthesis (M00359) pathway (Supplementary Material Note L). Two genes belonged to valine and methionine tRNAs pathways and were upregulated in P1. One gene in the pathway producing aspartamine tRNAs pathway was upregulated in P2. Together, this suggests that P1 is able to produce more valine and methionine tRNAs than P2.

We overlay the differentially expressed genes on KEGG pathways using KAAS pipeline(8). Table ST.N.1 shows the pathways which contained differentially expressed genes, together with the genes we identify.

Table ST.L.1: KEGG pathways with differentially expressed genes.

Pathway ID	Pathway	P1 Upregulated	P2 Upregulated
M00360	Aminoacyl-tRNA biosynthesis, prokaryotes	AgabiA15p1|1761,AgabiA15p2|1770 AgabiA15p1|2808,AgabiA15p2|2763	AgabiA15p1|660,AgabiA15p2|686
M00359	Aminoacyl-tRNA biosynthesis, eukaryotes	AgabiA15p1|1761,AgabiA15p2|1770 AgabiA15p1|2808,AgabiA15p2|2763	AgabiA15p1|660,AgabiA15p2|686
M00074	N-glycan biosynthesis, high-mannose type	AgabiA15p1|10048,AgabiA15p2|10102	AgabiA15p1|10047,AgabiA15p2|10101
M00073	N-glycan precursor trimming	AgabiA15p1|10048,AgabiA15p2|10102	AgabiA15p1|10047,AgabiA15p2|10101
M00009	Citrate cycle (TCA cycle, Krebs cycle)	AgabiA15p1|2884,AgabiA15p2|2813	
M00121	Heme biosynthesis, glutamate => protoheme/siroheme	AgabiA15p1|4282,AgabiA15p2|4299	
M00173	Reductive citrate cycle (Arnon-Buchanan cycle)	AgabiA15p1|2884,AgabiA15p2|2813	
M00172	C4-dicarboxylic acid cycle, NADP - malic enzyme type		AgabiA15p1|9348,AgabiA15p2|9524
M00395	Decapping complex		AgabiA15p1|10082,AgabiA15p2|10134
M00179	Ribosome, archaea		AgabiA15p1|9326,AgabiA15p2|9502
M00035	Methionine degradation	AgabiA15p1|8424,AgabiA15p2|8562	
M00010	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	AgabiA15p1|2884,AgabiA15p2|2813	
M00293	DNA polymerase zeta complex	AgabiA15p1|615,AgabiA15p2|7069	
M00079	Keratan sulfate degradation		AgabiA15p1|9068,AgabiA15p2|9247
M00740	Methylaspartate cycle	AgabiA15p1|2884,AgabiA15p2|2813	
M00042	Catecholamine biosynthesis, tyrosine => dopamine => noradrenaline => adrenaline	AgabiA15p1|7887,AgabiA15p2|8061	
M00128	Ubiquinone biosynthesis, eukaryotes, 4-hydroxybenzoate => ubiquinone		AgabiA15p1|2051,AgabiA15p2|2031
M00338	Cysteine biosynthesis, homocysteine + serine => cysteine	AgabiA15p1|8424,AgabiA15p2|8562	

M00169	CAM (Crassulacean acid metabolism), light		AgabiA15p1|9348,AgabiA15p2|9524
M00178	Ribosome, bacteria		AgabiA15p1|9326,AgabiA15p2|9502
M00295	BRCA1-associated genome surveillance complex (BASC)	AgabiA15p1|1458,AgabiA15p2|1537	
M00012	Glyoxylate cycle	AgabiA15p1|2884,AgabiA15p2|2813	

SI Appendix M: Differential protein domain annotations

Figure SF.M.1 shows the number of karyollele pairs in which karyollele pairs exhibit different protein domain annotations. While most (4287) karyollele pairs have the same number of annotations (see the diagonal), a few (215) have different protein domain annotations. This is a result of sequence differences.

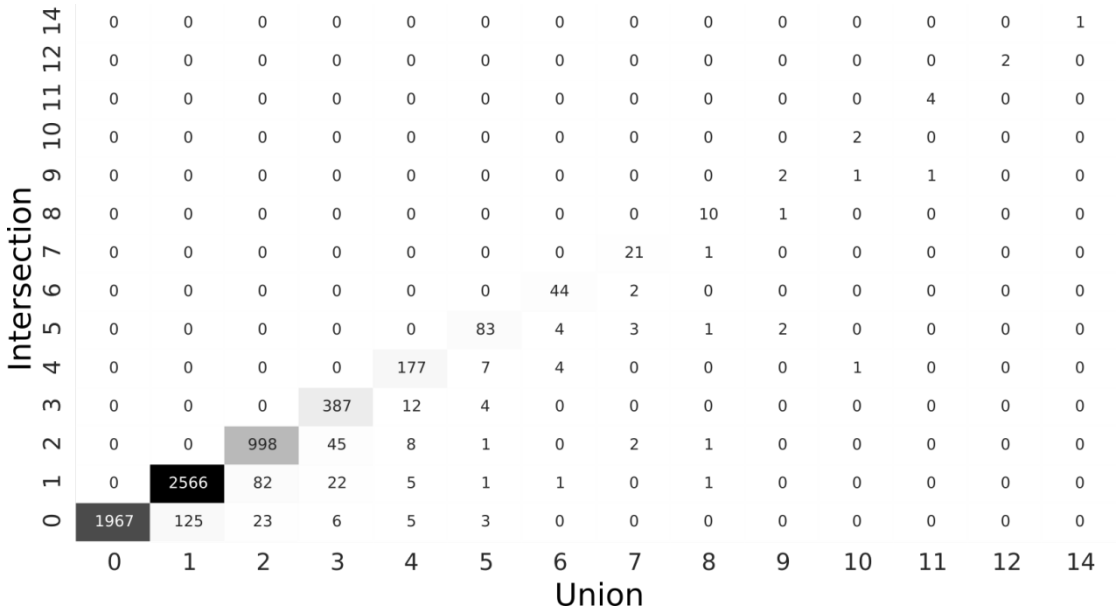

Figure SF.M.1: Implications of sequence deviations between karyollele pairs on domain predictions. The number of unique annotations are given on the x-axis, and the number of annotations in common between the two karyolleles is given on the y-axis. If the numbers are not equal, that means that both karyolleles are not annotated with the same domains. In the corresponding cells are given the number of genes with this combination of annotations. Most genes do not exhibit alternative functionality (see diagonal), but quite a number do (see below the diagonal).

SI Appendix N: PCR duplicates

PCR duplicates in mushroom tissues dataset and the vegetative mycelium dataset. We find a large amount of PCR duplication in the compost data. This can be attributed to the difficulty in isolating RNA from soil. See SI Dataset S4, where we provide the original read counts for each sample, the reads which remain after PCR duplicate removal, and the percent removed and remaining. We find that the vegetative mycelium dataset contains upwards of 50% PCR duplicates.

SI Appendix O: Homokaryon Genome annotations

The P1 and P2 genomes(9) were annotated with BRAKER1(10) and using the pooled RNA-seq data from the mushroom tissue dataset. In order to prevent chimeric genes (neighboring genes that are erroneously fused into one predicted gene) the following procedure was used. After the first round of gene prediction using BRAKER1, predicted introns were identified that were at least 150 bp in size and not supported by RNA-seq reads. The midpoint of these introns were labelled as intergenic regions in the next round of gene prediction using AUGUSTUS 3.0.2(11) and the parameter set produced in the first round of gene prediction. The SNP density between the genomes was estimated using MUMMER's(12) show-snps tool.

SI Appendix P: Named genes in *Agaricus bisporus* v. 2

Named genes for *Agaricus bisporus* version 2 were downloaded from the JGI DOE Genome Portal (http://genome.jgi.doe.gov/pages/search-for-genes.jsf?organism=Agabi_varbisH97_2) by searching for genes with 'Name' in the 'user annotations' attribute. Gene names were transferred from *A. bisporus* v. 2 using reciprocal best blast hit to P1 and P2, and then selecting the best match (in the single case of an ambiguity).

We mapped all transcripts from agabi2 to transcripts P1 with a bidirectional best nucleotide blast hit. In one case there was an ambiguous mapping, and we selected the mapping with the highest percent sequence identity (the e-values were identical). Table ST.O.1 provides these named genes, together with their mapping values.

Table ST.O.1: Named genes in version 2, and their corresponding karyollele pairs in A15.

P1	P2	agabi2	name	evalue	pident
AgabiA15p1|9363	AgabiA15p2|9538	152135	AOX	0	98.69
AgabiA15p1|8297	AgabiA15p2|8431	193168	ATP1	0	100
AgabiA15p1|3836	AgabiA15p2|3811	192355	ATP16	0	99.39
AgabiA15p1|8927	AgabiA15p2|9066	139908	ATP17	3.00E-174	98.83
AgabiA15p1|9641	AgabiA15p2|9823	194020	ATP20	0	100
AgabiA15p1|2115	AgabiA15p2|2093	194800	ATP3	0	100
AgabiA15p1|3238	AgabiA15p2|3147	138704	ATP4	0	99.29
AgabiA15p1|6514	AgabiA15p2|6684	135403	ATP7	3.00E-166	98.48
AgabiA15p1|5614	AgabiA15p2|5757	115586	CAT1	0	100
AgabiA15p1|5615	AgabiA15p2|5758	200291	CAT3	0	100
AgabiA15p1|5438	AgabiA15p2|5587	190684	CDC5	0	98.71
AgabiA15p1|3434	AgabiA15p2|3379	121800	CIPB	0	97.33
AgabiA15p1|1432	AgabiA15p2|1515	195535	COX4	0	100
AgabiA15p1|6236	AgabiA15p2|6400	177982	CYP63	0	98.82
AgabiA15p1|7984	AgabiA15p2|8153	135048	CytC2	3.00E-173	100
AgabiA15p1|6030	AgabiA15p2|6190	188638	HMG1	0	99.21
AgabiA15p1|9296	AgabiA15p2|9473	224131	Hpt	0	100
AgabiA15p1|3435	AgabiA15p2|3380	121801	INH1	3.00E-128	98.45
AgabiA15p1|4008	AgabiA15p2|3983	221245	MnP	0	99.06
AgabiA15p1|2172	AgabiA15p2|2148	226574	NDE1	0	100
AgabiA15p1|2932	AgabiA15p2|2853	227697	NDE2	0	100
AgabiA15p1|8485	AgabiA15p2|8625	136834	NUXM	0	100
AgabiA15p1|4154	AgabiA15p2|4172	192611	NUZM	0	98.67
AgabiA15p1|6172	AgabiA15p2|6334	195692	NdufA1	3.00E-123	97.29
AgabiA15p1|6031	AgabiA15p2|6191	139455	NdufA13	0	100
AgabiA15p1|2905	AgabiA15p2|2828	138930	NdufA2	4.00E-133	100
AgabiA15p1|6526	AgabiA15p2|6695	202899	NdufA4	1.00E-122	99.58
AgabiA15p1|646	AgabiA15p2|672	189651	NdufA5	0	98.97
AgabiA15p1|3259	AgabiA15p2|3172	195108	NdufA6	0	99
AgabiA15p1|9358	AgabiA15p2|9534	193806	NdufA9	0	99.45
AgabiA15p1|674	AgabiA15p2|700	214389	NdufB11	7.00E-170	99.39

AgabiA15p1|10070	AgabiA15p2|10123	208065	NdufB7	0	99.72
AgabiA15p1|6069	AgabiA15p2|6226	139429	NdufB9	2.00E-176	99.7
AgabiA15p1|2042	AgabiA15p2|2023	194758	NdufS1	0	100
AgabiA15p1|1088	AgabiA15p2|1111	190005	NdufS3	0	100
AgabiA15p1|9451	AgabiA15p2|9621	193877	NdufS4	0	98.39
AgabiA15p1|601	AgabiA15p2|629	133027	NdufS6	0	99.03
AgabiA15p1|4161	AgabiA15p2|4179	192620	NdufS7	0	100
AgabiA15p1|6033	AgabiA15p2|6193	188636	NdufS8	0	99.16
AgabiA15p1|690	AgabiA15p2|716	133109	Ndufab1	0	98.6
AgabiA15p1|3782	AgabiA15p2|3756	135814	OSCP/ATP5	0	99.08
AgabiA15p1|4327	AgabiA15p2|4346	192776	PAL1	0	99.4
AgabiA15p1|4243	AgabiA15p2|4260	192690	PAL2	0	99.63
AgabiA15p1|3454	AgabiA15p2|3399	210545	QCR2	0	98.58
AgabiA15p1|2199	AgabiA15p2|2174	138465	QCR8	1.00E-161	100
AgabiA15p1|6694	AgabiA15p2|6862	116951	RIM15	0	99.31
AgabiA15p1|3383	AgabiA15p2|3330	195170	SDH4	0	98.02
AgabiA15p1|6899	AgabiA15p2|7121	149788	SSK1	0	100
AgabiA15p1|869	AgabiA15p2|896	114317	STK/HK	0	100
AgabiA15p1|1449	AgabiA15p2|1531	228355	Tco1	0	100
AgabiA15p1|8256	AgabiA15p2|8388	143539	Tco5	0	100
AgabiA15p1|7003	AgabiA15p2|7226	230069	c2h2	0	99.86
AgabiA15p1|4651	AgabiA15p2|4717	203612	frt2	0	97.08
AgabiA15p1|9024	AgabiA15p2|9205	223670	fst4	0	98.9
AgabiA15p1|4515	AgabiA15p2|4547	192934	geranylgeranyl diphosphate synthase	0	97.73
AgabiA15p1|4280	AgabiA15p2|4297	192725	hom2	0	99.28
AgabiA15p1|5529	AgabiA15p2|5676	190759	hspA	0	99.62
AgabiA15p1|4376	AgabiA15p2|4397	192819	hspC	0	99.3
AgabiA15p1|3389	AgabiA15p2|3335	195173	hspD	0	98.77
AgabiA15p1|10959	AgabiA15p2|10986	120944	lcc10	0	91.17
AgabiA15p1|4698	AgabiA15p2|4766	184993	lcc12	0	98.61
AgabiA15p1|1413	AgabiA15p2|1497	139148	lcc2	0	100
AgabiA15p1|4686	AgabiA15p2|4752	184981	lcc3	0	99.89
AgabiA15p1|10961	AgabiA15p2|10987	194714	lcc9	0	85.85

SI Appendix Q: Functional predictions

When performing functional enrichment tests, we used as a background the overlap of the differentiable karyollele pairs and the annotated genes. This prevents an enrichment bias from the karyollele pairs when examining functional enrichment of the differentially expressed genes.

PFAM: Conserved protein domains were predicted using PFAM version 27(13, 14).

Transcription factor definitions: Predicted proteins with a known transcription factor-related (DNA-binding) domain (based on the PFAM annotations) were considered to be transcription factors.

Carbohydrate-active enzymes prediction: Using the Cazymes Analysis Toolkit (CAT) (15), we predicted carbohydrate-active enzymes based on the original gene definitions. If a gene's protein sequence was predicted to be a cazyme by either the sequence-based annotation method or the PFAM-based annotation method then we considered it a cazyme.

Secreted Proteins prediction: We used the same procedure as (16) to predict secreted proteins. Briefly, genes with SignalP (17) signal peptides, or a TargetP (18) Loc=S were kept. The remaining genes were further filtered with TMHMM (19), keeping only genes with zero or one transmembrane domains. Finally, genes were filtered using Wolf PSort (20) to select genes with a Wolf PSort extracellular score greater than 17.

Metabolic and Cytochrome P450 gene groups: Genes with the GO annotation "metabolic process" (annotation ID: GO:0008152) were called as metabolism genes. Genes with the PFAM annotation PF00067 were used as Cytochrome P450 genes.

KEGG: KEGG annotations were made with the KAAS KEGG (8) annotation pipeline, using genes from all available fungi, with the exception of leotiomycetes, Dothideomycetes, and Microsporidians, due to the limitation of the number of species (Selected organisms by ID: cne, cgi, ppl, mpr, scm, uma, mgl, sce, ago, kla, vpo, zro, cgr, ncs, tpf, ppa, dha, pic, pgu, lel, cal, yli, clu, ncr, mgr, fgr, nhe, maw, ani, afm, aor, ang, nfi, pcs, cim, cpw, pbl, ure, spo, tml). The GHOSTX and BBH options were selected. Predictions were made individually for both the P1 and P2 genomes, using the translated protein sequences.

SI Appendix R: DE-Seq differential expression size factor

Using DE-Seq(21), we perform a differential expression test for each karyollele pair in a tissue, i.e. we test if a gene has a differential expression between P1 and P2. DE-Seq requires a size factor to be calculated, which normalizes for the library sizes of each sample. Since in our case the counts from P1 and from P2 originate from the same sample, these must have the same size factor. Size factors are therefore calculated manually, by counting the total number of reads for each sample, and dividing it by the largest value for any sample.

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

The P1 and P2 counts originating from the same sample will then be assigned the same size factor. The expression counts for each gene in each replicate in each tissue (Equation 1 of Main Text) are provided to DE-Seq with the provided size factor.

The normalized read counts per gene $D_h(s,g)$ are returned by DE-Seq, together with significance values for each test. We select only differentially expressed genes that have a q-value < 0.05, and a fold change of at least three.

SI Appendix S: Formulas for GRR, CRR, NRR, CGR and NGR

In the main text we introduced measures that describe the relative expression between the P1 and P2 nuclei at the Gene, Chromosome and Gene levels. Here, we provide the formulas and notation for these measures.

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

The Gene Read Ratio (GRR) for gene *g* in sample s is defined as the ratio between the normalized expression (from DE-Seq) measured in P1, vs P2. $D_h(s,g)$ is the normalized expression returned by DE-Seq.

$$CRR(s,c) = \frac{\sum_{g \in c} D_{P1}(s,g)}{\sum_{g \in c} D_{P2}(s,g)}$$

The Chromosome Read Ratio (CRR) for a chromosome *c* in sample s is defined as the ratio of the sums of the normalized expression of each gene on chromosome *c*, in P1 and P2.

$$NRR(s) = \frac{\sum_{c \in C} \sum_{g \in c} D_{P1}(s,g)}{\sum_{c \in C} \sum_{g \in c} D_{P2}(s,g)}$$

Similarly, the Nuclear Read Ratio (NRR) for a sample *s* is defined as the ratio between the sum of all the normalized expression counts across all chromosomes *c* and all genes *g* on chromosome *c*, between the two nuclear types P1 and P2. *C* is the set of all chromosomes.

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

The Chromosome Gene Ratio (CGR) for a chromosome *c* in sample *s* is the geometric mean of the *GRR* for all genes *g* on chromosome *c*.

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

The Nuclear Gene Ratio (NGR) for a sample *s* is defined as the geometric mean of the *CGR* across all chromosomes. *C* is the set of all chromosomes.

References

1. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

2. Krueger F, Andrews SR (2011) Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.

3. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

4. Akalin A, et al. (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13(10):R87.

5. Patyshakuliyeva A, et al. (2015) Uncovering the abilities of Agaricus bisporus to degrade plant biomass throughout its life cycle. *Environ Microbiol* 17(8):3098–3109.

6. Bonnen AM, Anton LH, Orth AB (1994) Lignin-Degrading Enzymes of the Commercial Button Mushroom, Agaricus bisporus. *Appl Environ Microbiol* 60(3):960–5.

7. Woolston BM, et al. (2011) Long-distance translocation of protein during morphogenesis of the fruiting body in the filamentous fungus, agaricus bisporus. *PLoS One* 6(12). doi:10.1371/journal.pone.0028412.

8. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(SUPPL.2):182–185.

9. 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.

10. 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(5):767–769.

11. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5):637–644.

12. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.

13. Finn RD, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36(Database issue):D281–D288.

14. Finn RD, et al. (2014) Pfam: The protein families database. *Nucleic Acids Res* 42(D1):222–230.

15. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(D1):D490–D495.

16. Morais do Amaral A, Antoniw J, Rudd JJ, Hammond-Kosack KE (2012) Defining the Predicted Protein Secretome of the Fungal Wheat Leaf Pathogen Mycosphaerella graminicola. *PLoS One* 7(12):1–19.

17. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.

18. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol* 300(4):1005–1016.

19. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 305(3):567–580.

20. Horton P, et al. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35(Web Server):W585–W587.

21. Anders S, et al. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.