Supporting information

Supplementary methods

File S1

I. Calculation of SNPs and construction of consensus sequence

a) Quality control and trimming of Illumina raw data

Correction and trimming of the raw Illumina HiSeq reads with a size of 100 nt obtained from the ~300 bp PE library sequencing of Pc3 were performed with Trimmomatic v0.30 (Bolger *et al.*, 2014) using the following parameters.

```
java –jar trimmomatic-0.30.jar PE -phred33
NG-6796_Pc3_lib26819_1663_8_1.fastq NG-6796_Pc3_lib26819_1663_8_2.fastq
NG-6796_Pc3_lib26819_1663_8_1_trimmed_paired.fastq
NG-6796_Pc3_lib26819_1663_8_1_trimmed_unpaired.fastq
NG-6796_Pc3_lib26819_1663_8_2_trimmed_paired.fastq
NG-6796_Pc3_lib26819_1663_8_2_trimmed_unpaired.fastq
ILLUMINACLIP:Illumina-PE-gDNA.fa:2:30:10
LEADING:10 TRAILING:10 SLIDINGWINDOW:4:18 MINLEN:50
```

The file Illumina-PE_gDNA.fa contains all used Illumina adapters and primers (PCR- and seq-Primer). The quality of the processed reads was checked with FastQC.

```
./fastqc ./input/NG-6796_Pc3_lib26819_1663_8_1.fastq
```

For the correction and trimming of the raw Illumina HiSeq reads with a size of 50 nt obtained from single-end library sequencing of ascospore isolates AS25 and AS25-3 the parameters were adapted to library type and read length. An example is given for the reads obtained from AS25-3.

```
java -jar trimmomatic-0.30.jar SE -phred33 ./input/NG-
6879_P2_x_25_AS_3_lib28637_1722_7_1.fastq NG-6879_P2_x_25_AS_3_trimmed.fastq
ILLUMINACLIP:TruSeq-DNA-SE.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:18
MINLEN:20
```

b) Mapping of reads on a reference genome sequence

The processed reads obtained from the next-generation sequencing of Pc3, AS25, and AS25-3 were mapped to the P2niaD18 genome sequence (Chaudhuri *et al.*, 2011). Furthermore, processed reads of ascospore isolate AS25 were mapped on the *de novo* assembled contigs of Pc3 (see section II for assembly details) and reads obtained from ascospore isolate AS25-3 were mapped to an afore predicted consensus sequence of AS25. Mapping of the 100 nt long reads obtained from Pc3 genome sequencing was performed with bowtie2 v2.1.0 (Langmead & Salzberg, 2012) with the following parameters. Sorted bam files and index files were created with samtools (Li *et al.*, 2009). The sorted bam file was used for SNP calling (see section I d)).

```
bowtie2-build ./input/P2_niaD18.fa P2_niaD18.build

bowtie2 -q -p 4 --phred33 -I 50 -X 1000 --fr P2_niaD18.build -1 ./input/NG-
6796_Pc3_lib26819_1663_8_1_trimmed_paired.fastq -2 ./input/NG-
6796_Pc3_lib26819_1663_8_2_trimmed_paired.fastq -S Pc3_500PE_vs_P2_niaD18.sam
```

```
samtools view -bS Pc3_500PE_vs_P2_niaD18.sam > Pc3_500PE_vs_P2_niaD18.bam

samtools sort Pc3_500PE_vs_P2_niaD18.bam Pc3_500PE_vs_P2_niaD18_sorted

samtools index Pc3_500PE_vs_P2_niaD18_sorted.bam
```

For the mapping of the 50 nt long reads obtained from the single-end libraries of AS25 and AS25-3 were mapped with parameters adapted on single-end libraries. An example is given below. The sam file was sorted and indexed with samtools like mentioned before.

```
bowtie-build2 ./input/P2_niaD18.fa P2_niaD18.build

bowtie2 -q -p 4 --phred33 P2_niaD18.build ./input/ NG-6879_P2_x_25_AS_3_trimmed.fastq -S
AS3_vs_ P2_niaD18.sam
```

d) SNP-calling

SNP calling was performed with mpileup, which is a part of samtools, for all combinations listed in Table 1. The resulting vcf files were used as input for custom-made perl scripts. The varFilter option was used to filter ambiguous loci and low coverage regions. The value for varFilter was set twice as high as the average coverage, calculated from the mapping. On overview of the settings of varFilter used for the different analyses is given in the table below.

```
samtools mpileup -uD -f P2_niaD18.fa Pc3_500PE_vs_P2_niaD18_sorted.bam | bcftools view -bvgc -
> Pc3_500PE_vs_P2_niaD18_samtools.raw.bcf

bcftools view Pc3_500PE_vs_P2_niaD18_samtools.raw.bcf | vcfutils.pl varFilter -D70 >
Pc3_500PE_vs_P2_niaD18_samtools.vcf
```

reads of	mapped on	coverage	varFilter	total SNPs
Pc3	P2niaD18	31.4	-D70	33,956
AS25	P2niaD18	22.2	-D50	31,166
AS25	Pc3_assembly	22.8	-D50	4,448
AS25-3	P2niaD18	27.8	-D60	32,426
AS25-3	AS25_consensus	28.0	-D60	17,694

e) Construction of consensus sequences

The consensus sequences were constructed with mpileup and bcftools, both parts of samtools, on the basis of afore produced bam files. An example is given below, including the used settings for mpileup and bcftools. The consensus sequences of Pc3, AS25, and AS25-3 were calculated based on the P2niaD18 reference sequences. The consensus sequences were used for comparative SNP analyses as described in section II.

```
samtools mpileup -uf P2_niaD18.fa Pc3_500PE_vs_P2_niaD18_sorted.bam | bcftools view -cg - |
vcfutils.pl vcf2fq > Pc3_500PE_vs_P2_niaD18_consensus.fq
```

II. *De novo* assembly of the Pc3 genome sequence

The corrected reads of Pc3 (see section I) were used for an assembly with Velvet 1.2.10 (Zerbino & Birney, 2008). The files containing the paired ends were shuffled and printed into a single file.

```
shuffleSequences_fastq.pl NG-6796_Pc3_lib26819_1663_8_1_trimmed_paired.fastq
NG-6796_Pc3_lib26819_1663_8_2_trimmed_paired.fastq
NG-6796_Pc3_lib26819_1663_8_shuffled_trimmed_paired.fastq
```

Unpaired reads were combined to a single file using the cat command inside the Unix prompt.

```
cat NG-6796_Pc3_lib26819_1663_8_1_trimmed_unpaired.fastq
NG-6796_Pc3_lib26819_1663_8_2_trimmed_unpaired.fastq > Pc3_500PE_trimmed_unpaired.fastq
```

To define the best kmer size for Velvet, we tested values between 31 and 75.

```
./velveth output 33,81,4 -short -fastq ./input/Pc3_500PE_trimmed_unpaired.fastq -shortPaired -fastq
./input/NG-6796_Pc3_lib26819_1663_8_shuffled_trimmed_paired.fastq
```

```
./velvetg output_39 -exp_cov auto -cov_cutoff auto -min_contig_lgth 1000 -ins_length 300
```

Afterwards, we used the longest *de novo* assembled contig for mapping of the Pc3 reads, in order to predict the average insert length of the library. The insert length was calculated by analyzing the bam-file with FastQC v.0.10.1 (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc)*. Median insert length was calculated with 194 bp and mean insert with 182 bp. The median insert length was used to optimize parameters for Velvet.

```
./velvetg output_41 -exp_cov auto -cov_cutoff 3 -min_contig_lgth 1000 -ins_length 192
```

Usage of k-mer sizes greater than 45 shows a decreased value for N50. A k-mer size of 41 works best for the 100 nt paired end reads obtained from Pc3.

k-mer size	N50 in bp	Nmax in bp	# of contigs	total length in bp	number of gaps	length of all gaps in bp	GC-content
33	137,880	444,071	1,158	31,851,520	1,289	31,064	49.0%
35	137,684	439,015	1,175	31,973,965	1,250	31,747	49.0%
37	138,626	550,706	1,155	32,068,751	1,241	33,726	49.0%
39	142,105	550,645	1,142	32,160,114	1,233	35,093	49.0%
41	142,118	663,571	1,151	32,239,728	1,202	36,360	49.0%
43	142,132	492,109	1,162	32,333,719	1,177	36,779	49.0%
45	145,008	634,809	1,173	32,409,287	1,145	37,923	48.9%
47	127,258	634,930	1,178	32,477,933	1,155	40,837	48.9%
49	117,202	634,987	1,222	32,553,588	1,161	41,677	48.9%
53	124,135	663,665	1,253	32,718,008	1,255	47,281	48.9%
57	102,205	435,216	1,307	32,832,883	1,348	54,153	48.9%
61	82,036	410,780	1,432	32,957,557	1,564	65,839	48.9%
65	66,267	342,584	1,555	33,045,434	2,131	90,480	48.8%

III. Comparison of SNPs using custom-made Perl scripts

In order to identify differently inherited genomic regions inside the genome of ascospore isolate AS25-3 we first calculated the SNPs of the parental strains by mapping AS25 on P2niaD18. After this step, we predicted an AS25-3 consensus sequence based on P2niaD18 and compared the SNPs in the parental strains with the consensus sequence of the progeny by two custom-made Perl scripts, both are available upon request.

The first Perl script (SNP_comparison_vcf2vcfo.pl) reads in the vcf output file form mpileup, containing all SNPs in the parental strains AS25 and P2niaD18, and the sequence file containing the consensus sequence of the progeny AS25-3, which was constructed on the genome sequence of P2iaD18. In order to analyze the four consensus chromosomes of AS25-3 separately, we wrote each of the four consensus sequences into four online formatted fasta files. The perl script compares each SNP of the parental strains with the equivalent position of the consensus sequence of the progeny. Regions

inside the consensus sequence that show a low coverage (<5×) or an unclear base calling due to an inconsistent mapping were marked by mpileup with small letters or “n”, instead of capitalized letters. To avoid a comparison with those regions, the text-based comparison differentiated between capitalize and non-capitalized letters. In addition, only parental SNPs with a quality score greater than 50 and an allele frequency of 1.0, hereafter referred to as high quality SNPs, were used for further analysis. The perl script was designed to compare all parental SNPs to their corresponding position inside the AS25-3 consensus sequence, using the reference and alternative nucleotide in column four and five of the vcf file. After comparison, the script adds two columns at the end of the vcf file, containing the name (P2niaD18 or AS25) of the parental strain that matched to the nucleotide of the consensus sequence and the compared nucleotide that was extracted from the AS25-3 consensus sequence. SNPs that did not match to one or the other parental strain were marked with a question mark “?”, representing an ambiguous result. In addition, a question mark tags all SNPs that were located into regions with a low sequencing coverage of AS25-3. The extended vcf file, containing the two additional columns with the information of the origin of the SNPs inside the parental strains and the nucleotide determined from the consensus sequence of the progeny was renamed into vcfo.

The second Perl script (sliding-window-vcfo_stats_origin.pl) uses the vcfo file as input in order to perform a sliding window analysis by counting all SNPs inside a predetermined window. The script prints out the number of SNPs assigned to each parental strain and all inconsistent results inside each window and performs a statistical survey. In order to determine the best sliding window sizes for our analyses, we performed analyses with 500 bp, 1 kb, 5 kb, 10 kb, 20 kb and 50 kb windows. For 10 kb windows, best resolution was obtained. 500 bp windows were used to analyze putative recombination sites in detail. Based on the results of the 10 kb sliding window analysis, a score for each windows was determined in order to assign the corresponding sequence to one or the other parental strain. The score for each window was defined as the difference of SNPs assigned to each parental strain divided by the amount of total SNPs inside the window. Therefore, the value of the score ranges between -1 and 1. The resulting score for each window was used as an indication of the origin of the genomic region. All sliding windows that show a score of -1.0 or 1.0 were clearly and unambiguously assigned to one or the other parental strain. Sliding windows with scores between -0.5 and 0.5 were considered as not reliable to predict the origin of the sequence. Only windows with scores ranging between -1.0 and -0.5 or 0.5 and 1.0, respectively, were shown in the figures.

Supplementary Tables

Table S1. List of plasmids used in this study

Plasmid	Characteristics	Reference or source
pKOMAT	5’ flanking region of *MAT1-2-1*, *trpC* promoter of *A. nidulans*, *ble* resistance gene of *S. hindustanus*, 3’ flanking region of *MAT1-2-1*	Böhm *et al.*, 2013
pT3T7-gpd	*gpd* promoter of *A. nidulans*	Böhm *et al.*, 2013
pDrive-ptrA	*ptrA* resistance gene of *A. oryzae*	Böhm *et al.*, 2013
pPgpd-MAT-2-ptrA	*gpd* promoter of *A. nidulans*, *MAT1-2-1* gene of *P. chrysogenum*, *ptrA* resistance gene of *A. oryzae*	This study
pKOku70	5’ flanking region of *Pcku70*, *trpC* promoter of *A. nidulans*, *nat1* resistance gene of, 3’ flanking region of *Pcku70*	Hoff *et al.*, 2010b
pKOku70MAT2	5’ flanking region of *MAT1-2-1* and *MAT1-2-1* gene, *trpC* promoter of *A. nidulans*, *nat1* resistance gene of, 3’ flanking region of *Pcku70*	This study
pKompMAT2	5’ flanking region of *MAT1-2-1* and *MAT1-2-1* gene, *trpC* promoter of *A. nidulans*, *nat1* resistance gene of, 3’ flanking region of *MAT1-2-1*	This study

Table S2. List of bacterial and fungal strains used in this study

Strain	Characteristics and genotype	Mating type	Reference or source
Pc3 (syn. IB 08/921)	Wild type	*MAT1-2-1*	Böhm *et al.*, (2013)
NRRL 1951	Wild type	*MAT1-1-1*	Gailey *et al.*, (1946)
Q176	Derivative of NRRL 1951	*MAT1-1-1*	Gailey *et al.*, (1946)
Wisconsin 54-1255	Derivative of Q176	*MAT1-1-1*	Gailey *et al.*, (1946)
P2niaD18	*niaD*$^-$	*MAT1-1-1*	Hoff *et al.*, (2008)
AS 25	progeny from Q176 x Pc3	*MAT1-2-1*	Böhm *et al.*, (2013)
AS25-3	progeny from AS25 x P2niaD18	*MAT1-1-1*	This study
ΔMAT1-2-1 T10	*MAT1-2-1Δ::ble*		This study
ΔMAT1-2-1 T1	*MAT1-2-1Δ::ble*; *Pcku70Δ::nat1*		This study
ΔMAT1-2-1 T4	*MAT1-2-1Δ::ble*; *Pcku70Δ::nat1*		This study
ΔMAT1-2-1::MAT1-2-1 T24	*MAT1-2-1Δ::ble*; *MAT1-2-1*; *nat1*	*MAT1-2-1*	This study
ΔMAT1-2-1::MAT1-2-1 T26	*MAT1-2-1Δ::ble*; *MAT1-2-1*; *nat1*	*MAT1-2-1*	This study
Pc3::MAT1-2-1 T20	P*gpd*::*MAT1-2-1*; *ptrA*	*MAT1-2-1*	This study
Pc3::MAT1-2-1 T23	P*gpd*::*MAT1-2-1*; *ptrA*	*MAT1-2-1*	This study
Pc3::MAT1-2-1 T26	P*gpd*::*MAT1-2-1*; *ptrA*	*MAT1-2-1*	This study
P2::MAT1-2-1 T1	P*gpd*::*MAT1-2-1*; *ptrA*; *niaD*$^-$	*MAT1-1-1* *MAT1-2-1*	This study
P2::MAT1-2-1 T2	P*gpd*::*MAT1-2-1*; *ptrA*; *niaD*$^-$	*MAT1-1-1* *MAT1-2-1*	This study
P2::MAT1-2-1 T5	P*gpd*::*MAT1-2-1*; *ptrA*; *niaD*$^-$	*MAT1-1-1* *MAT1-2-1*	This study
XL1-Blue K12	recA1, endA1, gyrA96, thi1, hsdR17, supE44, relA1, lac-, [F' proAB, laclqZΔM15, Tn10(tetr)]		Bullock *et al.*, (1987)

Table S3. List of oligonucleotides used in this study

#	Oligonucleotide	Sequence (5'-3')	Specificity
1	PtrpC_anti1	GGCATTCATTGTTGACCTCCACTAG	*trpC* promoter
2	Tn5-phleo	GCGCCTGATACAGAACGAATTGC	*Ble* resistance gene
3	ApaI-MAT2	GGGCCCATGATGGCGAAAACCCTC	*MAT1-2-1* gene
4	MluI-MAT2	ACGCGTTTAGAACACGCTGTTCATAG	*MAT1-2-1* gene
5	APN1	ACTTTCATCTGGGCCAGCGAGTGG	5' flank *MAT1-2* locus
6	SLA2	GCCCGCCAGCGTCTGGGCGAAATG	3' flank *MAT1-2* locus
7	MAT1-2-1_f	GCCTTGCCTCAATGGAGTTC	*MAT1-2-1* gene
8	PcMAT2_f	CCTTGGTGCACGCTGGAACAAT	*MAT1-2-1* gene
9	PcMAT2_r	CCTGATCGTAGAGAATCATCCACTT	*MAT1-2-1* gene
10	Pc848_2170_r	CTTGAGTTCATCAGCCAGGTGGGTGAACTG	*MAT1-2-1* gene
11	Pc_HMG_f	CGATGGCGTTCTTGACCTGG	*MAT1-2-1* gene
12	Pc_HMG_r	GCCTGAACAAAAGGCAGG	*MAT1-2-1* gene
13	PclA_f	TGTGGTATTACCGGGAAGTC	*pclA* (Pc22g14900) gene
14	PclA_r	ACAATTCGTGCCTCGACTCC	*pclA* (Pc22g14900) gene
15	Chry1	GAGTTTGACTCGGGTCTTCG	Pc21g16000 gene
16	Chry4	AGCCAATTCCATCTGCTCTG	Pc21g16000 gene
17	PcpenDE_f	CTGCCACCAAAGAGATGATCC	*PcpenDE* gene
18	PcpenDE_r	CCTGGCGTTGAGCGCAGACCT	*PcpenDE* gene
19	pcbC_f	CACCCATGGCTTCCACCCCCAAGGCCAATG	*PcpcbC* gene
20	pcbC_r	GTGCCATGGCTGTCTGGCCGTTCTTGTTGATTAGAC	*PcpcbC* gene
21	PcFluG_f	CCACCATGCCCATAACCTATTGAA	*PcfluG* gene
22	PcFluG_r	GCAAATTTCCGATACAAAACCAAC	*PcfluG* gene
23	flbB_f	GCCAATGGCATGGACCACTC	*PcflbB* gene
24	flbB_r	TGACCAAGTGCTGTCAAGAG	*PcflbB* gene
25	nsdD_f	TAGCGTGGCTTCGCCTAATG	*PcnsdD* gene
26	nsdD_r	TAGAGCACCGAGTAGGGAAG	*PcnsdD* gene
27	Pc23g00420_f	TAGAGACCACGGTGCCGAAC	Pc2300420 gene
28	Pc23g00420_r	CATTCAAGCGTGCTAGATCC	Pc2300420 gene
29	Pc24g01940_f	GTCAGCACGCCTATAGACAC	Pc24g01940 gene

30	Pc24g01940_r	CTCAGCAACCGGGATATTTC	Pc24g01940 gene
31	Pc12g12190_f	GCCTTCCAGCTATGCCTACG	*PcflbC* gene
32	Pc12g12190_r	ACGCCCAGGTCTAGCGAAAG	*PcflbC* gene
33	Pc13g04920_f	TGAGCAACCAGCGCTCAATG	*PcstuA* gene
34	Pc13g04920_r	CTTGCGCCTAGTTCTCCTCC	*PcstuA* gene
35	RT-MAT1-2-1_f	TGCCTTTTGTTCAGGCTGATT	*MAT1-2-1* gene
36	RT-MAT1-2-1_r	AGTGAAAGGGGGAGAGAGTGG	*MAT1-2-1* gene
37	RT_fphA_f	CGCTCCCCAGTATTCGGGTT	*PcfphA* gene
38	RT_fphA_r	TTCAGGATCGGGGTGTTCGG	*PcfphA* gene
39	RT_lreB_f	GGAACTACGGGGTGCAAGGT	*PclreB* gene
40	RT_lreB_r	CGACGGGTTGGCATAGGTGA	*PclreB* gene
41	RT_cryA_f	CGCAGCCCTATTTCCGCATC	*PccryA* gene
42	RT_cryA_r	GGCCTGGGGTAGCCATTCTT	*PccryA* gene
43	RT_PcbrlA_f	GTGACCCCTCCTTCTTCGTC	*PcbrlA* gene
44	RT_PcbrlA_r	GTCTGACCCTGAGGGAGTAC	*PcbrlA* gene
45	NcSSU1	ATCCAAGGAAGGCAGCAGGC	*small subunit ribosomal* RNA
46	NcSSU2	TGGAGCTGGAATTACCGCG	*small subunit ribosomal* RNA
47	Pc3-contig_24_f	ACTTGGTTGTGCCTGTGTTG	Chromosomal translocation
48	Pc3-contig_24_r	GGCTGTGGCGTAACTGTATG	Chromosomal translocation
49	Pc3-contig_35_f	AGTGGAAGTTCGGCCTGAAC	Chromosomal translocation
50	Pc3-contig_35_r	TCTACAGGACACGTGAGCAC	Chromosomal translocation
51	AS25_chrI_1_f	TTCCTTGTCGGGCTTGTACC	Verification of recombination
52	AS25_chrI_1_r	AGAGGATCGTGATGACCGAG	Verification of recombination
53	AS25_chrII_1_f	CCTACCAACCTTGTTCAGAC	Verification of recombination
54	AS25_chrII_1_r	AGTTACTCCACCGAGTCTGC	Verification of

			recombination
55	AS25_chrII_2_f	ATAGTCATCGGCAACCAGTC	Verification of recombination
56	AS25_chrII_2_r	ATGGGTCTCATGATTCCACG	Verification of recombination
57	SNP_check1_f	GACCAACACATCGCCAAACG	Verification of SNPs
58	SNP_check1_r	GGATGACAAGGTACCTGTGC	Verification of SNPs
59	SNP_check2_f	CTTCTGCATCCTGCGATACC	Verification of SNPs
60	SNP_check2_r	AGATCTGGCCATAGGAGTGC	Verification of SNPs
61	SNP_check3_f	AGTCAACGTCACCTCGACAG	Verification of SNPs
62	SNP_check3_r	CTACACCACGAACGAGAGTC	Verification of SNPs
63	SNP_check4_f	ACATTCCGGAACGAGGAAGC	Verification of SNPs
64	SNP_check4_r	CTCGGCATGTCAAAATCTGG	Verification of SNPs
65	SNP_check5_f	CATACCGCACCAAACACCTG	Verification of SNPs
66	SNP_check5_r	ACCCTCCTCCTCTTCAACTG	Verification of SNPs
67	SNP_check6_f	CTCTGCCATTCCTCTTGGTC	Verification of SNPs
68	SNP_check6_r	GCTGTGGATCAAAGGCAAGC	Verification of SNPs
69	SNP_check7_f	TATCTGGCATGGGCTCTTCG	Verification of SNPs
70	SNP_check7_r	TGCAAGTGCCATTCTGGGTC	Verification of SNPs
71	SNP_check8_f	ACGTTCCGATGACTCGCATC	Verification of SNPs
72	SNP_check8_r	ATTGCCTGCGGTCACAGATG	Verification of SNPs
73	SNP_check9_f	TCAAGGAGAAGGTCGACGAC	Verification of SNPs
74	SNP_check9_r	GAGGCAGCAACTGTTCCAAG	Verification of SNPs
75	SNP_check10_f	GTCGATACACCGTCAACGTG	Verification of SNPs
76	SNP_check10_r	GATACGGCCTTACCTTGTCC	Verification of SNPs
77	SNP_check11_f	GCGTATTGAATCCGCTTAGC	Verification of SNPs
78	SNP_check11_r	GATAGCTATAGCGGGACAGC	Verification of SNPs
79	SNP_check12_f	GAAGCTAGGATATGGGAGAG	Verification of SNPs
80	SNP_check12_r	CACCTCGCTTGTGAAATACG	Verification of SNPs
81	SNP_check13_f	GCAACATCTGGATAGGCACG	Verification of SNPs
82	SNP_check13_r	GCAATGTGGCGTGTATGACC	Verification of SNPs
83	5FMAT2inf_f	CGGTACCACGCATGCTGCAGGTTAATTGAGCACGGAGAG	5’ flank and *MAT1-2* locus
84	5FMAT2inf_r	CACGAATTCTGGATCGCTGAGCCGAG	5’ flank and *MAT1-2*

		TCGAAATTG	locus
85	3FMAT2inf_f	GCCCATCTGAATTCGTCGAATGTCCC ATTCACAAT	3’ flank *MAT1-2* locus
86	3FMAT2inf_r	CGAGAAGCTTGTCGATGTATCTTTCC CTCTTGGCT	3’ flank *MAT1-2* locus

142

143

Supplementary Figures

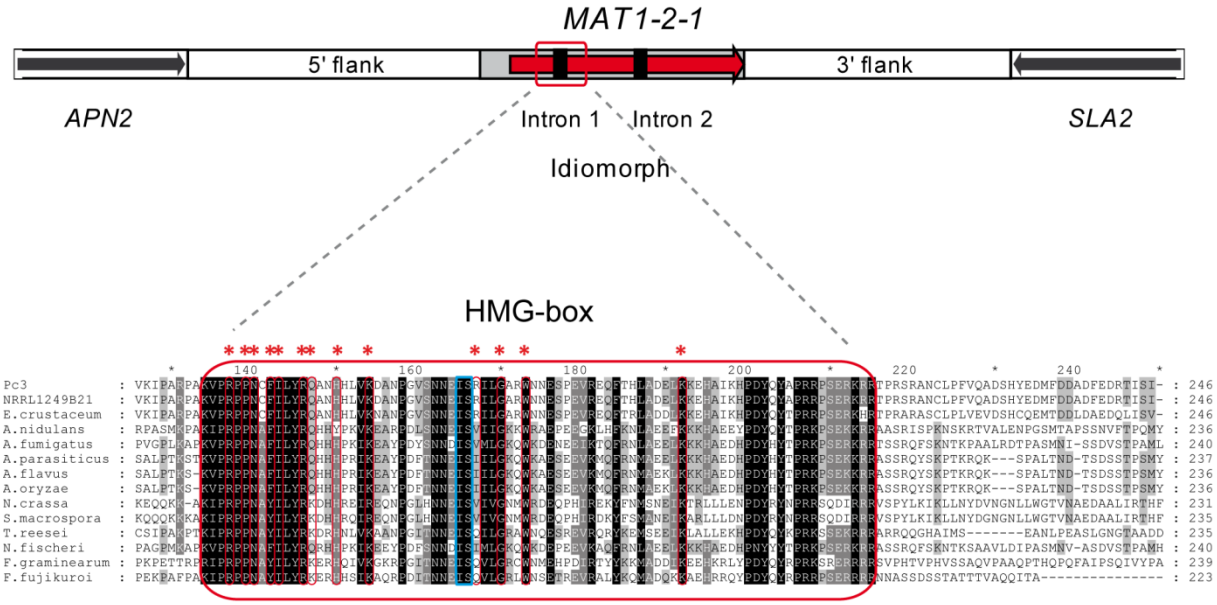


Fig. S1. Genomic organization of the *MAT1-2* locus and a multiple alignment of the HMG domain proteins. The MAT1-2 protein sequence of Pc3 is compared to HMG mating-type proteins from the Fleming isolate NRRL1249B21 and other ascomycetes. Red: the HMG-box domain. Asterisks: DNA binding sites. Blue: the conserved isoleucine-serine motif, upstream of the conserved intron.

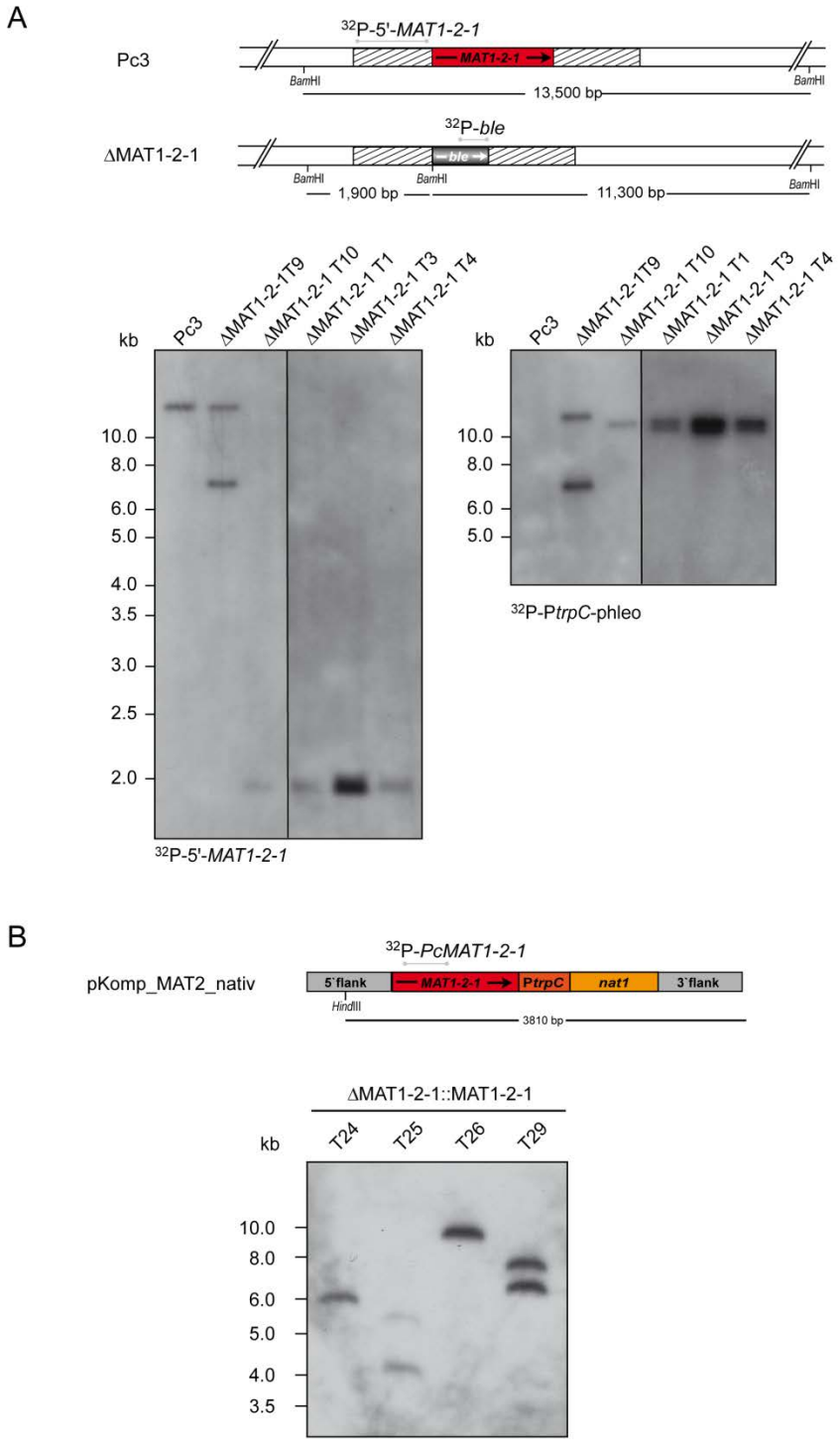


Fig. S2. Construction of ΔMAT1-2-1 and complementation strains. A. Organization of the *MAT1-2* locus in the reference strain, Pc3, and the derived knockout strains. The corresponding Southern hybridization confirmed the knockout and the phleomycin resistance cassette. B. Southern hybridization with *MAT1-2-1*-specific probes to verify successful rescue of the deletion strain.

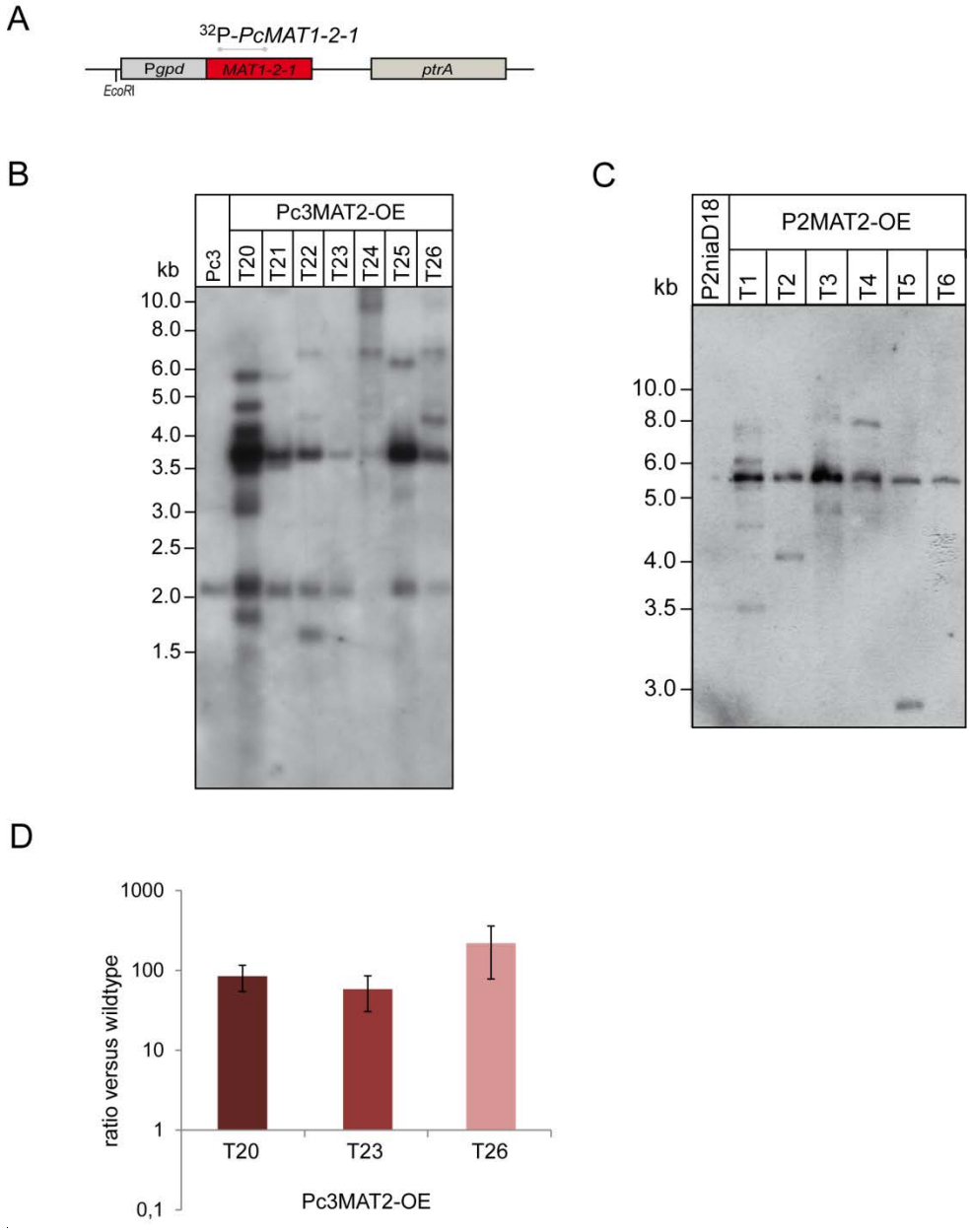
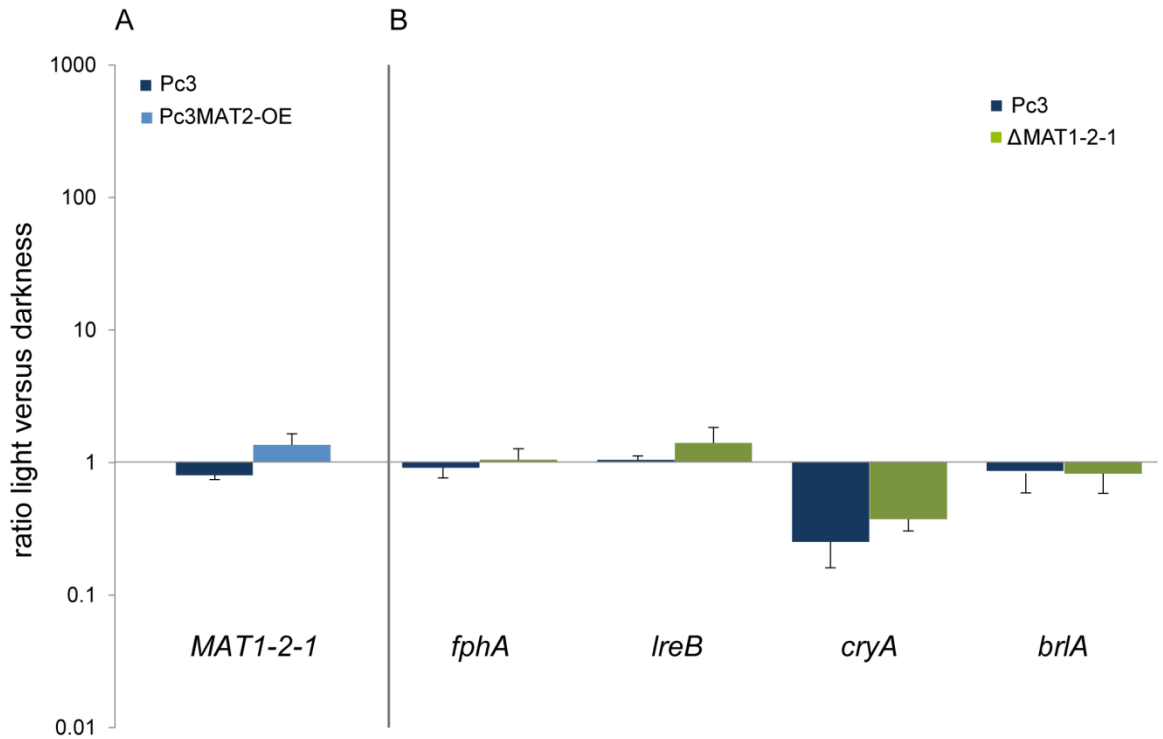


Fig. S3. Construction of *MAT1-2-1* overexpression strains. A. Overexpression construct generated to complement *MAT1-2-1* deletion strains. The *MAT1-2-1* gene is under the control of the *gpd* promoter of *A. nidulans*. B. Southern hybridization verifies ectopic integration of the *MAT1-2-1* gene into strain Pc3 with ^{32}P-labeled probes specific for *MAT1-2-1*. C. Southern hybridization with ^{32}P-labeled probes specific for *MAT1-2-1* verifies ectopic integration of *MAT1-2-1* in P2niaD18. D. qRT-PCR analysis of *MAT1-2-1* in overexpression strains. Transformants were grown in liquid shaking cultures for 72 h. Values represent average expression ratios of at least two biological replicates from three *MAT1-2-1* overexpression (reddish) strains (n = 2), relative to reference Pc3.



169

170 **Fig. S4.** Analysis of light-dependent transcriptional expression. qRT-PCR was conducted to
171 measure *MAT1-2-1* (A) or *fphA*, *lreB*, *cryA* or *brlA* (B) expression in strains as indicated.
172 Values represent average expression ratios of at least two biological replicates from three
173 independently derived strains (n = 3).

174

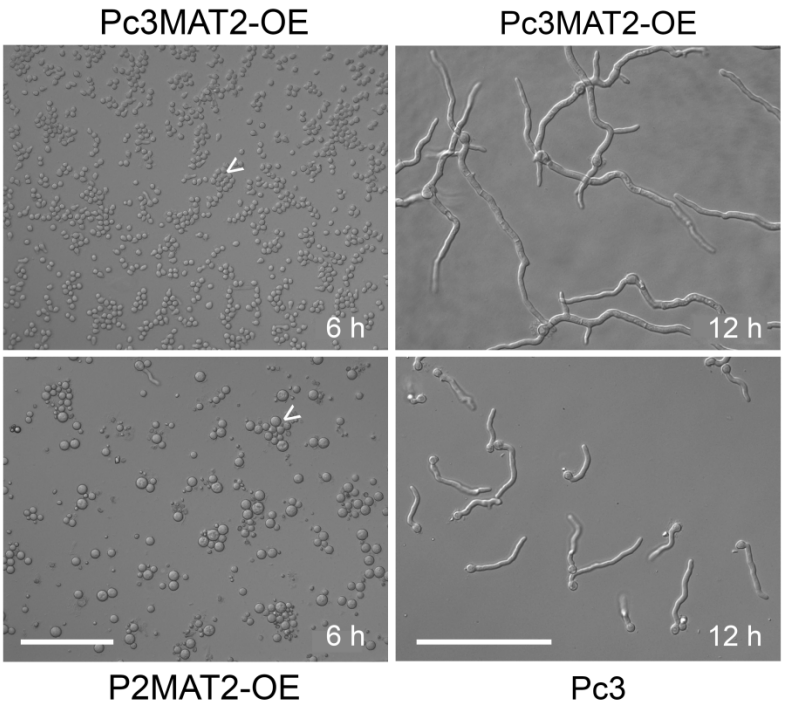


Fig. S5. Overview to demonstrate agglutination of conidiospores (Pc3MAT2-OE, P2MAT2-OE) and formation of one or two biopolar germ tubes (Pc3MAT2-OE, Pc3). Arrowheads point to agglutinated conidiospores. Scale bars correspond to 100 µm in all images.

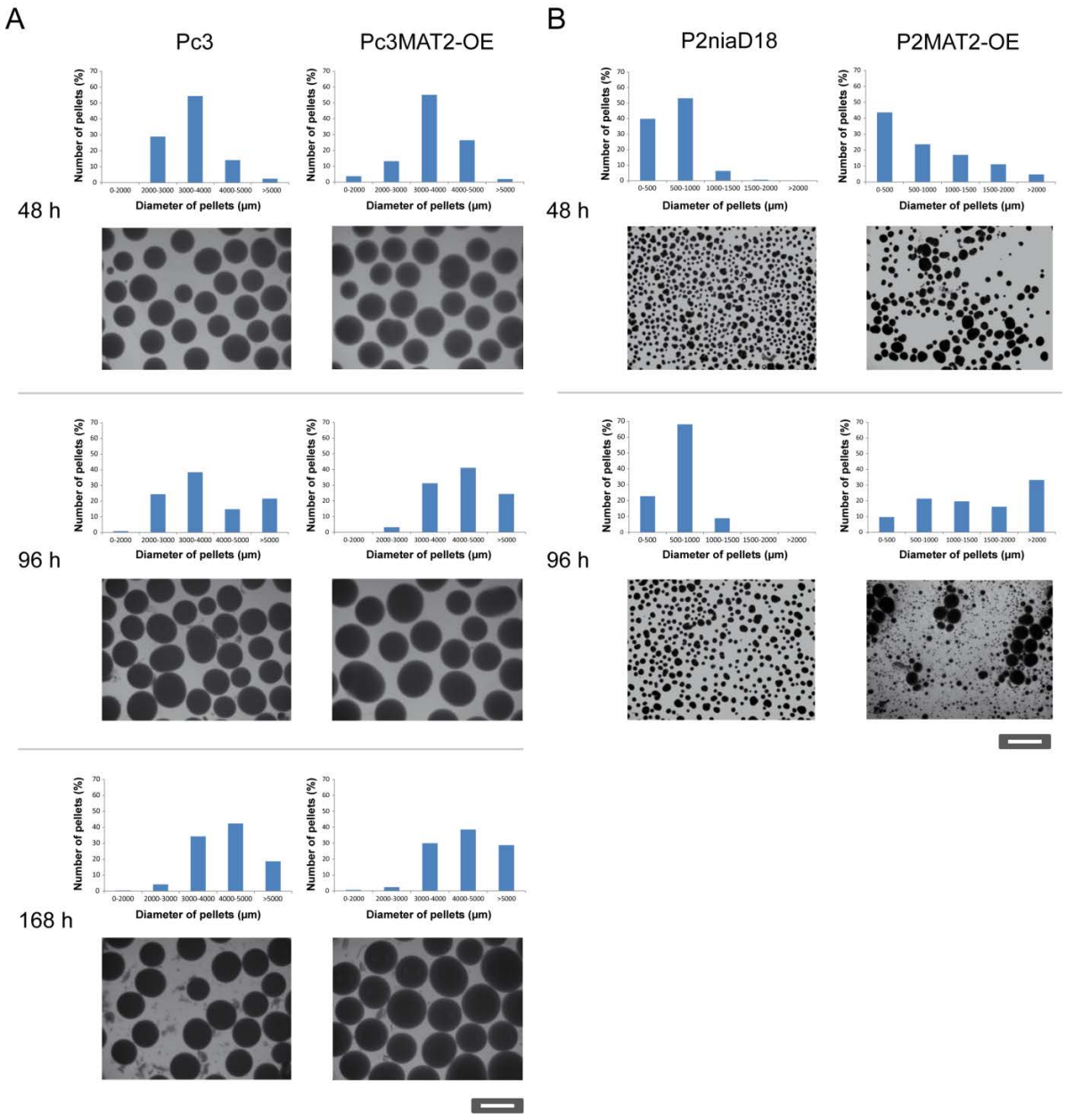


Fig. S6. Pellet formation in liquid shaking cultures formed by reference and recombinant *MAT1-2* and recombinant *MAT1-1* strains. **A**. Quantification of pellet diameters formed from *MAT1-2-1* overexpression strains in the Pc3 background in liquid shaking CCM cultures. Measurements were taken at three different time points, as indicated. Error bars represent the mean ± SD of 240 random pellets from three independent strains. **B**. Quantification of pellet diameters formed from *MAT1-2-1* overexpression strains in the P2niaD18 background in liquid shaking CCM cultures. Measurements were taken at two different time points, as indicated. Error bars represent the mean ± SD of 240 random pellets from three independent strains.

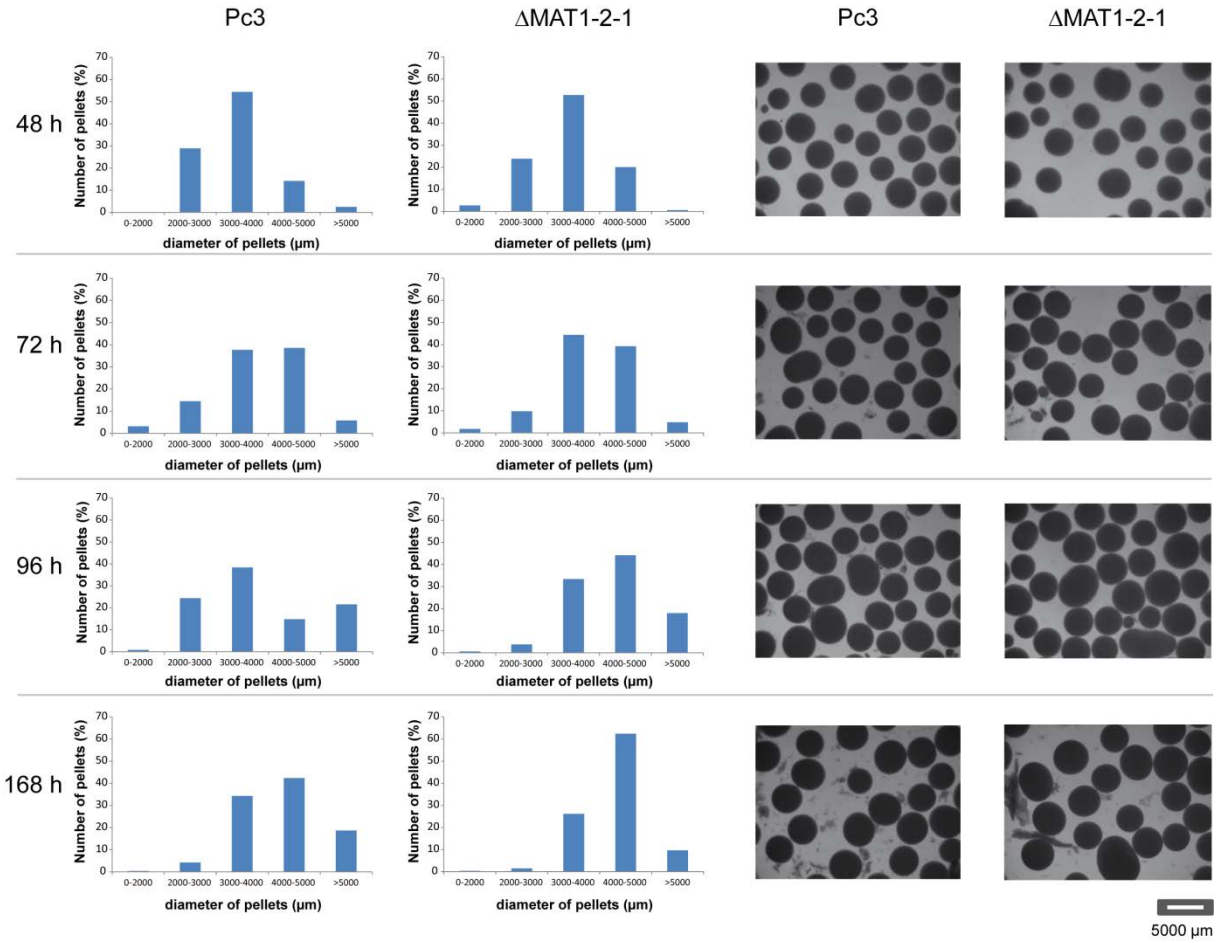


Fig. S7. Pellet formation in liquid shaking cultures. (Left) Quantification of pellet diameters formed from *MAT1-2-1* deletion strains in liquid shaking CCM cultures. Measurements were taken at four different time points, as indicated. Error bars represent the mean ± SD of 240 random pellets from three independent strains. (Right) The pellet phenotypes are illustrated with representative micrographs.

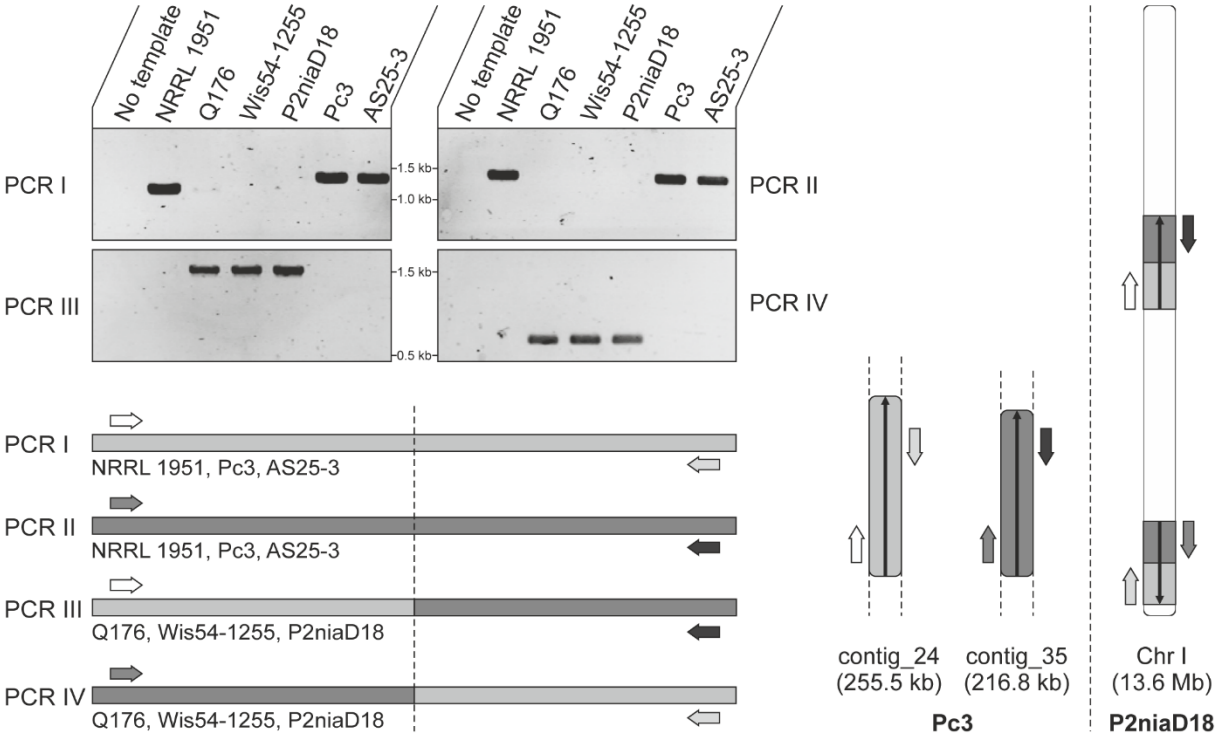


Fig. S8. Verification of a chromosomal rearrangement in wild-type and penicillin-production strains by PCR analysis. The wild-type strains, NRRL 1951 and Pc3, and ascospore isolate, AS25-3, show the same chromosomal structure; in contrast, all low- and high-penicillin producers of the Wisconsin lineage (Q176, Wisconsin54-1255, and P2niaD18) have a different chromosomal organization at this locus. Primers for PCR amplification are marked by different shading, and binding sites are shown on chromosome I of P2niaD18 (Specht *et al.* 2014) and contig_24 and contig_35 of wild type Pc3.

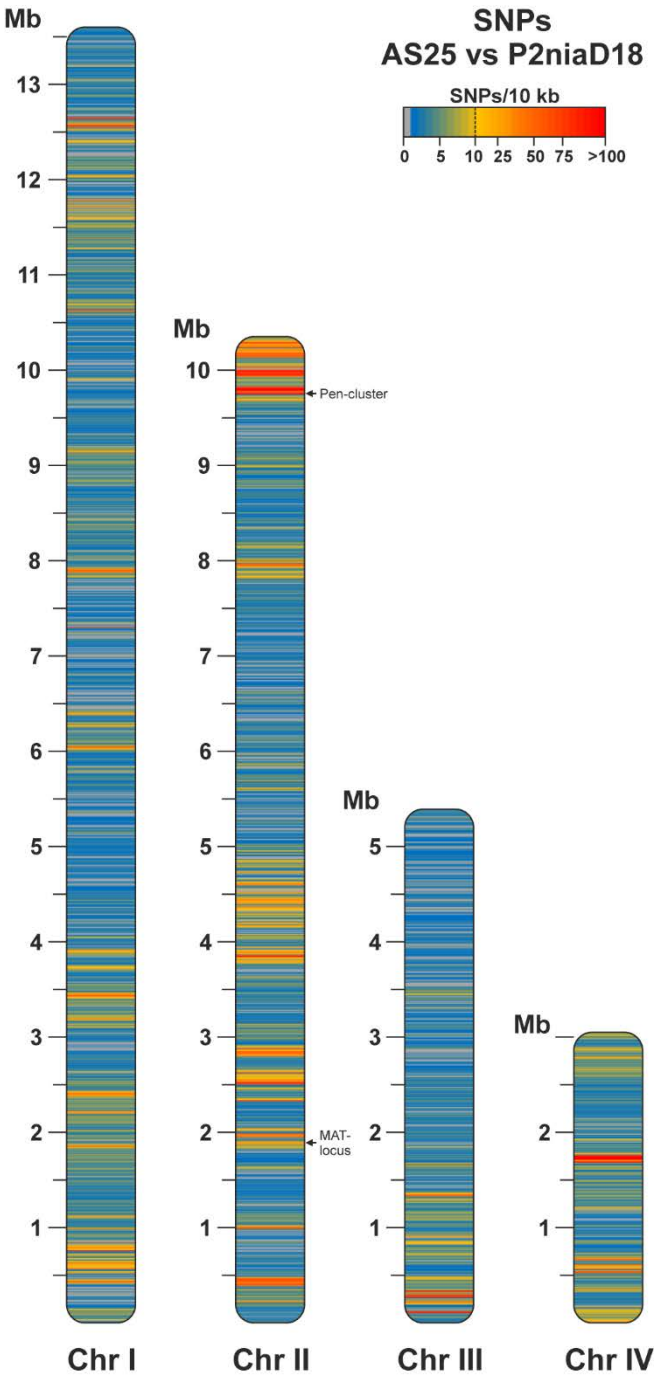


Fig. S9. Comparison of AS25 sequence to the published sequence of P2niaD18 (T. Specht *et al.* 2014). Genome sequencing of AS25 reveals a high number of single nucleotide polymorphisms (SNPs) compared to the industrial strain. The number of SNPs within 10 kb windows is indicated with color. Locations of genes of particular interest are highlighted.

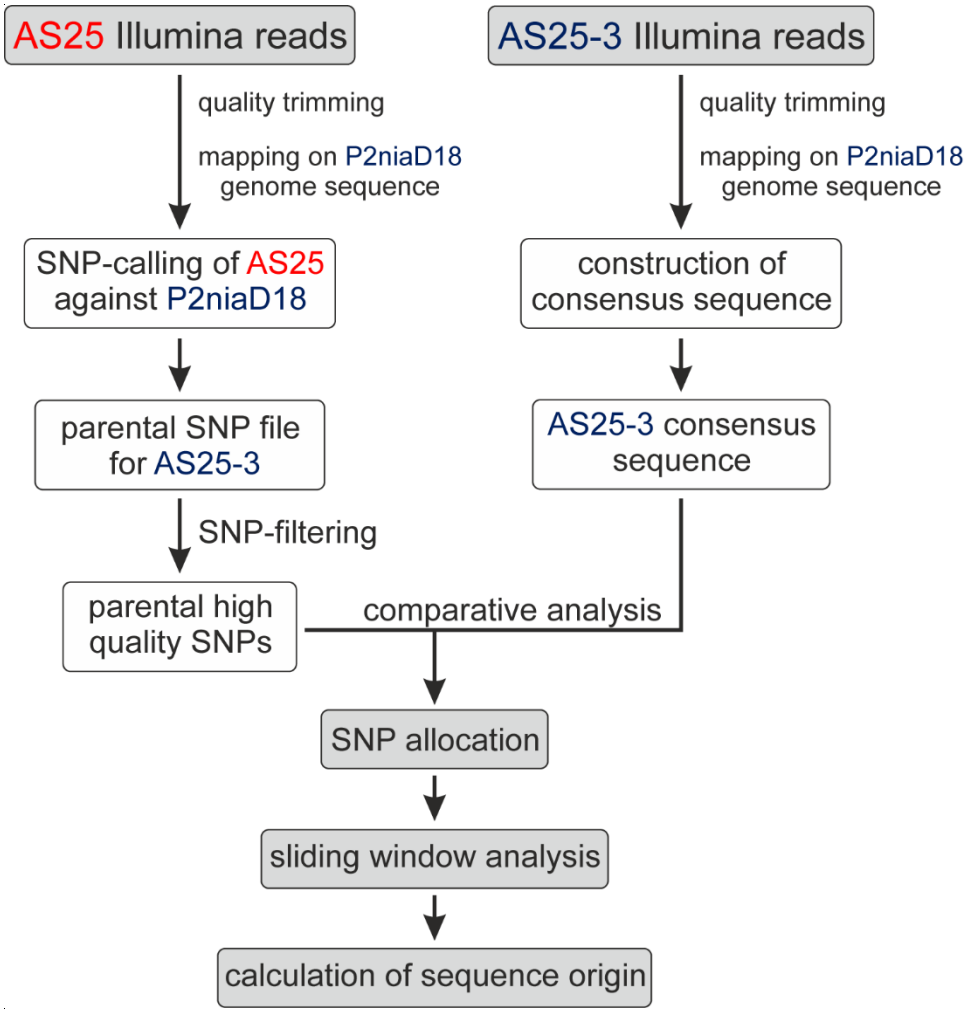


Fig. S10. Flow chart of bioinformatic analysis to identify differently inherited genomic regions in AS25-3. Illumina reads of ascospore isolates AS25 and AS25-3 were trimmed and mapped on the P2niaD18 reference genome sequence (*MAT* loci are indicated by colors: blue for *MAT1-1*, and red for *MAT1-2*). SNPs in the parental strains P2niaD18 and AS25, and the AS25-3 consensus sequence were predicted. To clearly assign SNPs on the AS25-3 genome, comparative analysis were performed using parental high-quality SNPs (quality score >50, and allele frequency = 1.0). After SNP allocation, sliding-window analysis provides the basis to calculate differently inherited genomic regions inside AS25-3.

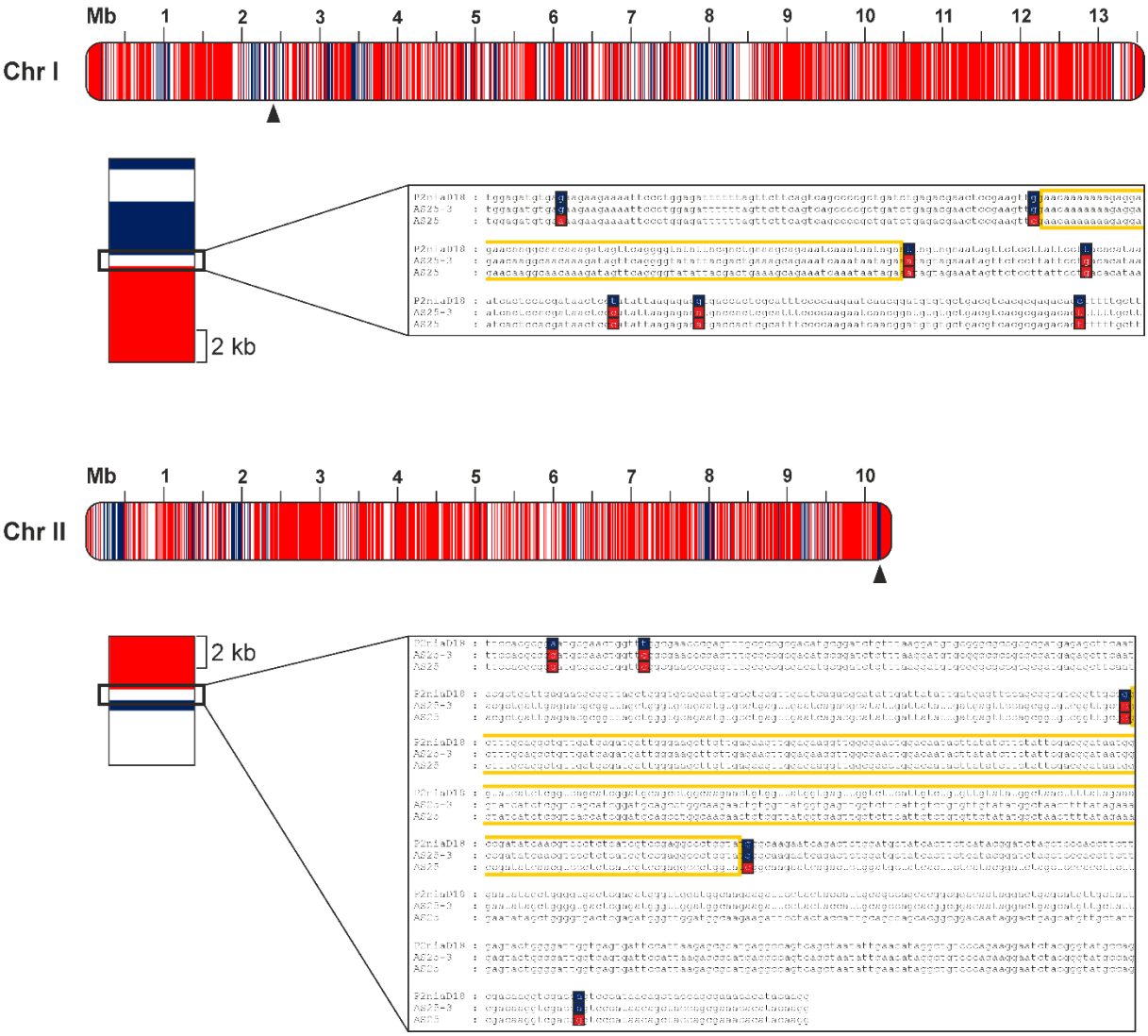


Fig. S11. Sequence validation of recombinant sites in the ascospore progeny AS25-3. Appropriate PCR amplicons from AS25-3, AS25, and P2niaD18 were used for Sanger sequencing and further comparison (positions are indicated by triangles). Differently inherited regions are indicated by colors: blue for P2niaD18, red for AS25. Yellow frames mark the recombination sites between two differently inherited SNPs.