Cell Research (2010) 20:408-420.

www.nature.com/cr

ORIGINAL ARTICLE

A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand

Dan Li[1, 2, *], Yang Dong[1, 2, *], Yu Jiang[1, 2, *], Huifeng Jiang[1, 3, *], Jing Cai[1, 2], Wen Wang[1]

[1]CAS-Max Planck Junior Research Group, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences (CAS), Kunming, Yunnan 650223, China; [2]Graduate School of Chinese Academy Sciences, Beijing 100039, China

Recent transcription profiling studies have revealed an unexpectedly large proportion of antisense transcripts in eukaryotic genomes. These antisense genes seem to regulate gene expression by interacting with sense genes. Previous studies have focused on the non-coding antisense genes, but the possible regulatory role of the antisense protein is poorly understood. In this study, we found that a protein encoded by the antisense gene *ADF1* acts as a transcription suppressor, regulating the expression of sense gene *MDF1* in *Saccharomyces cerevisiae*. Based on the evolutionary, genetic, cytological and biochemical evidence, we show that the protein-coding sense gene *MDF1* most likely originated *de novo* from a previously non-coding sequence and can significantly suppress the mating efficiency of baker's yeast in rich medium by binding MATα2 and thus promote vegetative growth. These results shed new light on several important issues, including a new sense-antisense interaction mechanism, the *de novo* origination of a functional gene, and the regulation of yeast mating pathway.

Keywords**:** *de novo*, sense-antisense interaction, mating pathway, *Saccharomyces cerevisiae*
Cell Research (2010) **20**:408-420. doi: 10.1038/cr.2010.31; published online 2 March 2010

Introduction

Antisense transcription (transcripts from the opposite strand of a sense gene) is widespread in eukaryotes, from yeast to mammals [1]. Studies in various organisms revealed that antisense transcripts are involved in degradation of the corresponding sense transcripts (RNA interference) [2]. However, in *Saccharomyces cerevisiae*, components of the RNAi machinery are absent [3], and antisense repression can be mediated by transcription interference (TI) or histone deacetylation. TI was thought to be an unavoidable suppressive consequence of two convergent promoters directing transcripts that overlap for at least part of their sequences [4]. But in the case of PHO84, the sense gene is regulated by accumulation of antisense RNAs, which leads to targeted histone deacetylation and the silencing of sense transcription [5]. However, all the reported mechanisms rely on the non-coding antisense RNAs. Whether the protein-coding antisense gene can serve a regulatory role remains an open question. In this study, we identified a pair of functionally linked protein-coding sense and antisense genes, YCL058C (*MDF1*) (previously named as a dubious gene *FYV5*, whose function was thought to be required for yeast viability 5 [6]) and YCL058W-A (*ADF1*) in *S. cerevisiae*. Through extensive genetic, cytological, and biochemical experiments, we demonstrate that the regulation that YCL058W-A confers to YCL058C is not due to previously known mechanisms, but results from binding of YCL058W-A protein as a transcription repressor to the promoter region of YCL058C. Thus, our results reveal a new molecular mechanism of interaction between the sense and antisense pair.

In addition to the regulatory connection between this gene pair, the unusual property of the sense gene

*These four authors contributed equally to this work.
Correspondence: Wen Wang
Tel: +86-871-5192979
E-mail: wwang@mail.kiz.ac.cn
[3]Present address: Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA
Received 23 January 2010; revised 26 January 2010; accepted 28 January 2010; published online 2 March 2010

YCL058C itself also caught our attention. In our sequence comparative analysis, we did not find any significantly homologous open reading frame (ORF) of YCL058C in any other yeast species. Therefore, YCL058C probably originated *de novo* from a previously non-coding sequence in *S. cerevisiae*.

The origination of new genes, a fundamental process for all organisms, has been extensively studied in the past few years [7]. The majority of newly evolved genes are derived from pre-existing genes, and their origination mechanisms include duplication divergence, retrotransposition, exon shuffling, and lateral gene transfer [7]. Completely *de novo* origination of a protein-coding gene from a non-coding sequence has been thought to be an almost impossible event, as stated by Susumo Ohno that "Each new gene must have arisen from an already existing gene" [8], and by François Jacob that "The probability that a functional protein would appear *de novo* by random association of amino acids is practically zero" [9]. However, a number of *de novo* genes have recently been identified mainly by Begun's and our group [10, 11]. These putative *de novo* genes have already generated intensive debate and discussion (e.g. Casci [12]; http://richarddawkins.net/forum/viewtopic.php?f=4&t=45460). These controversial examples are not supported by direct evidence of their protein-coding capacity, but only by the existence of putative ORFs and expression sequences. The direct evidence of their protein-coding capacity still remains to be provided. Moreover, a concrete molecular mechanism or pathway has not been demonstrated for any young duplicated genes, let alone *de novo* genes. The discovery of a concrete molecular mechanism/pathway for a newly evolved gene would convincingly show the biological significance of origin of new genes and significantly contribute to our mechanistic understanding of functional evolution in general.

Here we performed comprehensive evolutionary and experimental analyses on YCL058C and showed that this new gene is not only capable of encoding a protein but also takes essential cellular tasks in the mating pathway of *S. cerevisiae*. By binding the MATα2 protein, one of the determinants of yeast mating types, YCL058C suppresses yeast mating behavior and allows quick vegetative growth. As the previous name for YCL058C, *FYV5* [6], was not functionally distinguished from that of its antisense gene, YCL058W-A, which nests on the antisense strand of YCL058C, we propose to name YCL058C as *MDF1* (Mating Depressing Factor 1) and its anti-sense partner YCL058W-A as *ADF1* (Antisense of Depressing Factor 1) to reflect the newly uncovered properties of YCL058C and the functional relationship between this gene pair.

Results

Both MDF1 and ADF1 are subject to selection and encode proteins

MDF1 with an ORF of 152 amino acids is located in chromosome III of *S. cerevisiae*, while *ADF1* with an ORF of 113 amino acids completely nests on the opposite strand of *MDF1*. To initially test if *MDF1* and *ADF1* are functional protein-coding genes in *S. cerevisiae*, we conducted an evolutionary analysis to look at whether they have been subject to functional constraint by estimating their nucleotide substitutions within and between yeast species [13]. For *MDF1*, we conducted an intraspecies analysis and found that *MDF1* is fixed in all 39 sequenced *S. cerevisiae* strains from geographically and ecologically diverse sources, and there are no frame-shift or nonsense polymorphisms, suggesting that the gene may be under functional constraint. The seven polymorphic sites in these 39 *S. cerevisiae* strains are all non-synonymous substitutions. It is significantly different from neutral expectation ($P = 0.038$) by Z test [13], implying positive selection on *MDF1* and thus suggesting the functionality of *MDF1*. For *ADF1*, the evolutionary rates of non-synonymous and synonymous substitution among species in the *sensu stricto* group are significantly smaller than 1 (Supplementary information, Table S1), suggesting strong functional constraints on *ADF1*.

We next examined the functionality of *MDF1* and *ADF1* by testing whether transcription and further translation in *S. cerevisiae* are possible. Strand-specific RT-PCR showed that both *MDF1* and *ADF1* expressed in normal condition in *S. cerevisiae*, but not in other yeast species (Figure 1A). In an effort to obtain the final proof of the protein-coding capabilities of *MDF1* and *ADF1*, 3HA and 13Myc-tags were annealed to the 3′-ends of *MDF1* and *ADF1*, respectively. The western-blot analyses detected positive signals (Figure 1B), which states clearly that *MDF1* and *ADF1* can encode proteins.

MDF1 and ADF1 have antagonistic effects on growth in rich medium

Previous preliminary phenotypic screening analyses indicated that the *MDF1*Δ mutant appeared to show reduced growth in rich medium [14]. This encouraging hint suggests that *MDF1* or *ADF1* may influence growth. To discriminate the functional effects of *MDF1* and *ADF1*, we cloned *MDF1* (*M* for short) and *ADF1* (*A* for short) separately into the whole locus deletion (M^-A^-) strain in the background of α cells of *S. cerevisiae* using pRS316 vector. For the relatively short *ADF1*, the coding sequence plus upstream flanking sequence of *ADF1* can simply be used to construct the M^-A^+ strain. For the

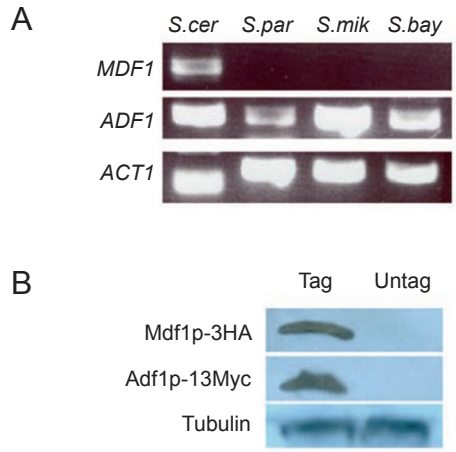


Figure 1 Both *MDF1* and *ADF1* are protein-coding genes. **(A)** The strand-specific RT-PCR experiments showed that *MDF1* only expressed in *S. cerevisiae* in the YPD medium, while *ADF1* expressed constantly in the *sensu stricto* group species. *ACT1* was used as the internal control. *S.cer*, *S. cerevisiae*; *S.par*, *S. paradoxus*; *S.mik*, *S. mikatae*; *S.bay*, *S. bayanus*. **(B)** Endogenous Mdf1p and Adf1p tagged with 3HA and 13Myc respectively were detected by western blotting, untagged yeast was used as the negative control, and tubulin as a positive control.

relatively long *MDF1*, a stop codon was introduced into the 5′-end of *ADF1* without changing the coding ability of *MDF1* by site-directed mutagenesis to construct the M^+A^- strain. After genetically separating *MDF1* and *ADF1*, we measured the influence of *MDF1* and *ADF1* on proliferation by both competition experiments and growth rate analyses at 30 °C in the rich medium. The competition experiments showed that the M^+A^- strain grew more quickly than the wild-type strain, whereas growth defects were observed in both M^-A^+ and M^-A^- strains (Figure 2A). In agreement with this finding, the M^+A^- strain enjoyed faster growth in growth rate analyses, but M^-A^+ and M^-A^- strains proliferated more slowly than the reference wild-type strain (Supplementary information, Figure S1A). The growth defects of M^-A^- strain could be remedied by re-introducing both *MDF1* and *ADF1* (Figure 2A and Supplementary information, Figure S1A). In addition, the growth superiority of M^+A^- strain was repeatedly supported by our two-dimensional gel electrophoresis data, which showed that some essential genes involved in the energy and substance metabolism, such as *ATP1*, *PGK1*, *MDH1*, *SAM1*, were distinctly increased in the M^+A^- strain compared with the wild type (Supplementary information, Figure S1B). The antagonistic effects of *MDF1* and *ADF1* on growth raise the possibility of sense-antisense interaction. Therefore, we seek further evidence for this interaction phenotypically and mechanistically in our next experiments.

Adf1p negatively regulates the expression of MDF1 by binding to its promoter

To examine if *ADF1* has an effect on *MDF1* expression, we overexpressed *ADF1* using the inducible pYES3/CT vector in the wild-type *S. cerevisiae*. Strikingly, the sense (i.e. *MDF1*) transcripts could be completely abolished by overexpressed *ADF1* (Figure 2B). Because the overexpressed *ADF1* on the plasmid does not physically overlap with the chromosomal *MDF1*, transcription interference [4] is not a probable cause. This transcriptional suppression is instead probably due to the RNA or protein of *ADF1* present in the cells. In view of the absence of RNAi machinery in *S. cerevisiae* [3], it is more likely that the repression occurred at the protein level. Subcellular localization of the Adf1p provides further support for a role as a transcription factor. By constructing a GFP-fusion plasmid to localize Adf1p within yeast cells, we observed that Adf1p resided in the nucleus (Figure 2C), representing a major characteristic of a transcription factor.

As it was of interest for us to explore whether the Adf1p could actually regulate the transcriptional activity of *MDF1* as a transcription repressor, we subsequently performed chromatin immunoprecipitation (ChIP) assay to investigate the direct association of Adf1p with the *MDF1* promoter in a yeast strain overexpressing His-tagged Adf1p. The ChIP results show that Adf1p does bind to the upstream region of *MDF1* (Figure 2D). Taken together, these results strongly support a novel mechanism of sense-antisense interaction, in which the antisense-encoded protein negatively regulates the expression of the sense gene by binding to the promoter of the sense gene (Figure 2E).

Mdf1p significantly decreases the mating efficiency of α cells

In an attempt to uncover the underlying mechanism for the rapid growth in M^+A^- strain, we conducted global microarray analyses among strains M^+A^-, M^-A^+ and wild type. Unexpectedly, our microarray data indicated that most of the down-regulated genes in M^+A^- strain are enriched in the yeast mating pathway in comparison with the wild-type and M^-A^+ strains (Figure 3A and Supplementary information, Table S2). Our quantitative mating assays further confirmed that the M^+A^- strain was substantially less successful than the wild-type α strain ($P < 0.01$) in mating, whereas the mating efficiencies of M^-A^+ and M^-A^- strains were comparable to those of wild type (Figure 3B). Therefore, it is intuitively appealing to assume that *MDF1* fulfills a role in the mating pathway.

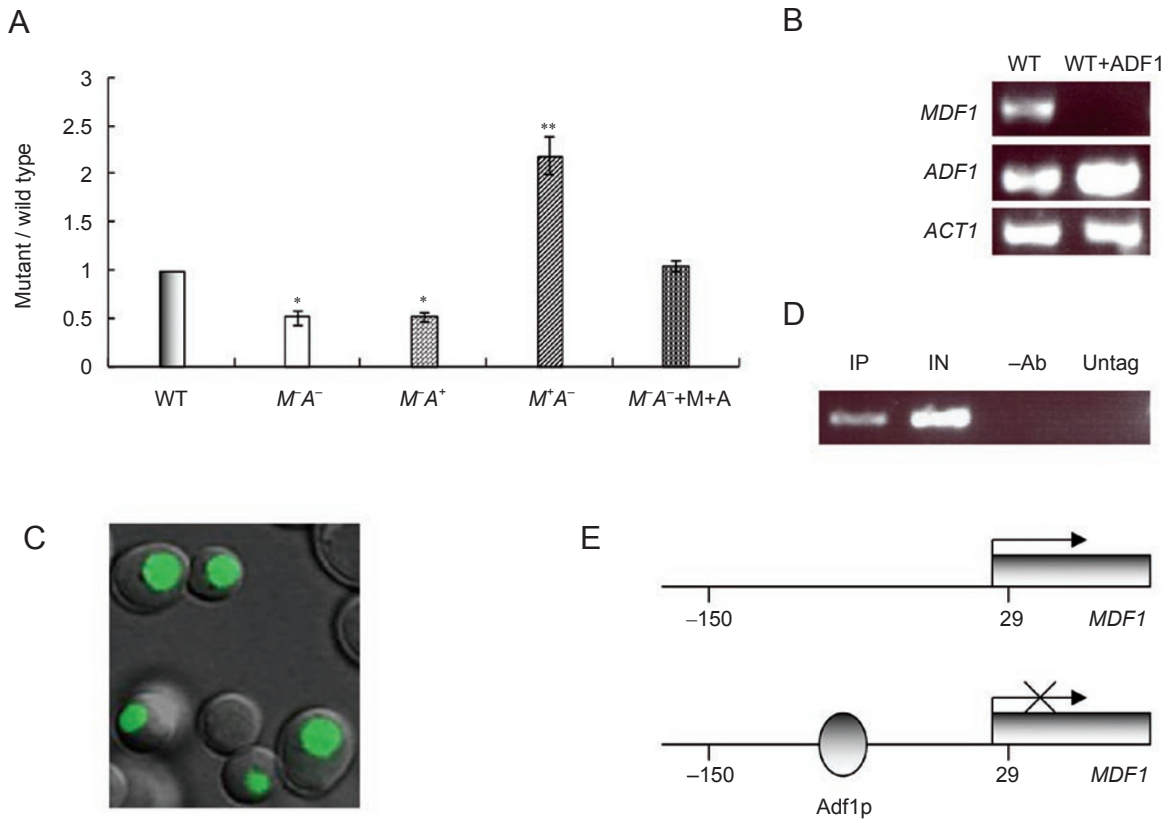
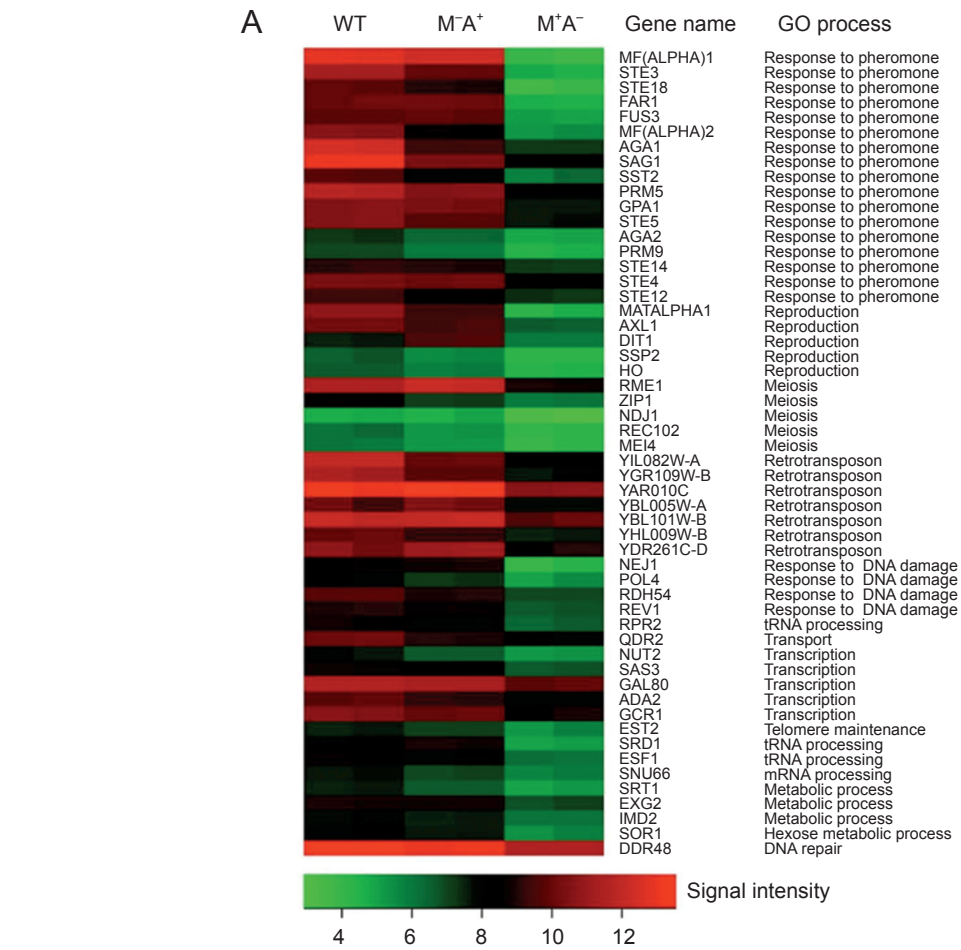
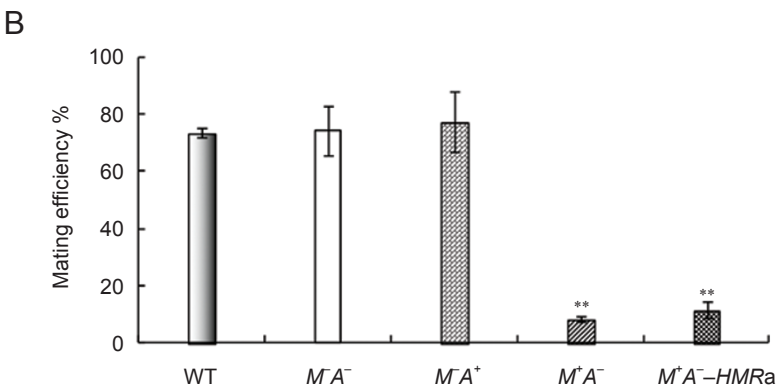


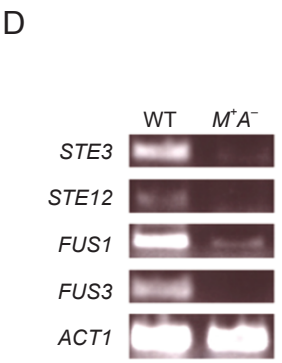
Figure 2 Adf1p negatively regulates the expression of *MDF1* by binding the promoter region of *MDF1* **(A)** Competition experiments indicate that *MDF1* and *ADF1* have antagonistic effects on yeast growth, i.e., M^+A^- ($MDF1^+ADF1^-$) strain grew much faster than the wild type strain (**P < 0.01), whereas M^-A^+ and M^-A^- strains grew worse than the wild type strain (*P < 0.05). Histograms represent the clone numbers of mutants divided by the clone numbers of wild type. The values are average of three independent experiments (with standard deviations). WT, wild type strain; M^-A^-, strain with both *MDF1* and *ADF1* deleted; M^-A^+, strain with *MDF1* deleted and *ADF1* left; M^+A^-, strain with *ADF1* deleted and *MDF1* left; M^-A^- + M + A, strain with *MDF1* and *ADF1* simultaneously transformed back to M^-A^- strain. **(B)** Overexpressed Adf1p inhibits the expression of *MDF1* completely. WT, wild type; WT + ADF1, *ADF1* was overexpressed in the background of wild type; *ACT1*, house-keeping gene as internal control. **(C)** Nuclear localization of Adf1p is visualized by Adf1p -GFP fusion protein. **(D)** ChIP assays shows that Adf1p binds the promoter region of *MDF1*. The final DNA extracts were amplified using a pair of primers that cover the promoter region (between −150 to +29 bp) of *MDF1*. IN, input; IP, immunoprecipitation; −Ab, control for non-specific binding in the absence of antibody; untag, untagged yeast as a negative control. **(E)** The model of a new sense-antisense interaction mechanism, in which the antisense-encoded protein (Adf1p) negatively regulates the expression of the sense gene (*MDF1*) by binding the promoter of the sense gene. The promoter region (between −150 to +29 bp) of *MDF1* used for ChIP assays was indicated.

The mating pathway (mitogen-activated protein kinase (MAPK) pathway) is currently one of the best-characterized pathways in yeast [15]. Three distinct cell types exist in *S. cerevisiae*: haploid cell types a and α, and diploid cell type a/α. The *MAT* loci encode master regulators of cell type: MATal is encoded by the *MAT*a locus, present in a cells and diploids, while MATαl and MATα2 are encoded by the *MAT*α locus, present in α cells and diploids. In a cells, a-specific and haploid-specific genes function by default, and so MATa1 does not contribute anything. In α cells, the MATα1 protein turns on α-specific genes, including *STE3*, the entrance of the MAPK pathway; the MATα2 protein turns off a-specific genes, while haploid-specific genes function normally. In response to mutual pheromone stimulation, the mating pathway is triggered, and thus a and α cells can fuse to form diploids. In diploid cells, MATα2 protein still turns off a-specific genes, while MATa1 and MATα2 dimerize to suppress haploid-specific genes, including *MAT*α1. The diploid cells can undergo meiosis and transform into a or α haploids in the scarcity of fermentative carbon and nitrogen sources.

When scrutinizing our array data, we found that


A
WT
M⁻A⁺
M⁺A⁻
Gene name
GO process
MF(ALPHA)1 Response to pheromone
STE3 Response to pheromone
STE18 Response to pheromone
FAR1 Response to pheromone
FUS3 Response to pheromone
MF(ALPHA)2 Response to pheromone
AGA1 Response to pheromone
SAG1 Response to pheromone
SST2 Response to pheromone
PRM5 Response to pheromone
GPA1 Response to pheromone
STE5 Response to pheromone
AGA2 Response to pheromone
PRM9 Response to pheromone
STE14 Response to pheromone
STE4 Response to pheromone
STE12 Response to pheromone
MATALPHA1 Reproduction
AXL1 Reproduction
DIT1 Reproduction
SSP2 Reproduction
HO Reproduction
RME1 Meiosis
ZIP1 Meiosis
NDJ1 Meiosis
REC102 Meiosis
MEI4 Meiosis
YIL082W-A Retrotransposon
YGR109W-B Retrotransposon
YAR010C Retrotransposon
YBL005W-A Retrotransposon
YBL101W-B Retrotransposon
YHL009W-B Retrotransposon
YDR261C-D Retrotransposon
NEJ1 Response to DNA damage
POL4 Response to DNA damage
RDH54 Response to DNA damage
REV1 Response to DNA damage
RPR2 tRNA processing
QDR2 Transport
NUT2 Transcription
SAS3 Transcription
GAL80 Transcription
ADA2 Transcription
GCR1 Transcription
EST2 Telomere maintenance
SRD1 tRNA processing
ESF1 tRNA processing
SNU66 mRNA processing
SRT1 Metabolic process
EXG2 Metabolic process
IMD2 Metabolic process
SOR1 Hexose metabolic process
DDR48 DNA repair
Signal intensity
4
6
8
10
12


B
Mating efficiency %
100
80
60
40
20
0
**
**
WT
M⁻A⁻
M⁻A⁺
M⁺A⁻
M⁺A⁻–HMRa


D
WT
M⁺A⁻
STE3
STE12
FUS1
FUS3
ACT1

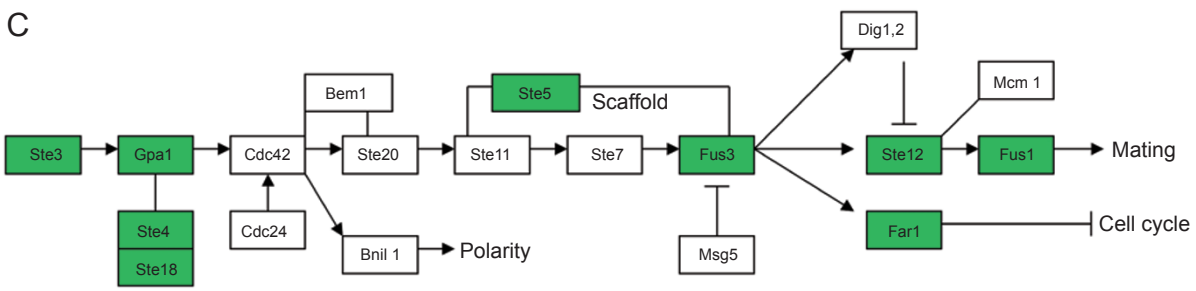

C
Ste3
Gpa1
Ste4
Ste18
Cdc42
Cdc24
Bem1
Ste20
Bnil 1
Polarity
Ste5
Scaffold
Ste11
Ste7
Fus3
Msg5
Dig1,2
Mcm 1
Ste12
Fus1
Mating
Far1
Cell cycle

Figure 3 Mdf1p significantly decreases the mating efficiency of α cells. **(A)** Most of the down-regulated genes in M^+A^- strain via array analyses were associated with yeast mating pathway. WT, wild type strain; M^-A^+, strain with *MDF1* deleted and *ADF1* left; M^+A^-, strain with *ADF1* deleted and *MDF1* left. **(B)** The mating efficiency tests demonstrate that the M^+A^- strain mated far worse than the wild type α strain (**P < 0.01), while the mating efficiencies of M^-A^+ and M^-A^- strains were comparable to the wild type α cells. The mating defect of M^+A^- strain cannot be rescued by deleting *HMRa* in M^+A^- strain (M^+A^--*HMRa*) (**P < 0.01). **(C)** Key components of MAPK pathway were significantly down-regulated. More than three times down-regulated genes were marked with green. The MAPK pathway information for *S. cerevisiae* was downloaded from Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/). **(D)** The significantly down-regulated genes involved in the MAPK pathway were confirmed by semi-quantitative RT-PCR.

*MAT*α1 was almost completely suppressed and all the known haploid-specific genes, some of which are key components of the MAPK signaling pathway (Figure 3C), were among those significantly down-regulated in the M^+A^- strain. The a-specific genes were off, consistent with the observation that the transcription level of *MAT*α2 remained relatively normal. Meanwhile, the α-specific genes were off as well, which was an unavoidable consequence of suppressed *MAT*α1 gene expression. The down-regulation of these genes revealed by microarray results was validated by semi-quantitative RT-PCR (Figure 3D). Therefore, the M^+A^- cells, which are physically α haploids, behave more like diploids, in which the a-, α- and haploid-specific genes are all shut down or down-regulated.

In addition to the *MAT* locus, *S. cerevisiae* carries two unexpressed, but complete copies of mating-type genes *HMLα* and *HMRa*, which are usually transcriptionally silenced [16]. One explanation for the above unusual expression pattern is the abnormal activation of the cryptic mating-type loci *HMLα* and *HMRa*, which can lead to the coexpression of a and α information in haploid cells [17]. However, our array data showed that *HMLα* and *HMRa* remained silenced in the M^+A^- strain. The unchanged mating inhibition phenotype observed when *HMRa* was deleted in the M^+A^- strain (Figure 3B) further ruled out the possibility that Mdf1p activates the silent mating cassette *HMRa* and allows a1/α2 suppressor to be formed. Therefore, the simplest mechanistic explanation for this pseudo-diploid phenotype is that Mdf1p in α cells may bind MATα2 protein, similar to what MATa1 does in diploid cells. The first piece of evidence of Mdf1p mimicking MATa1 came from the predicted secondary structure of Mdf1p by the online protein structure prediction server, PORTER (http://distill.ucd.ie/porter/) [18]. Similar to MATa1protein, Mdf1p looks like a three-helix-bearing protein, which is the foundation for binding MATα2 and the targeted DNA [19].

Mdf1p regulates the mating pathway of S. cerevisiae by binding MATα2

If the Mdf1p-MATα2 interaction hypothesis is right, we would expect that Mdf1p functions upstream of *MAT*α1 which is targeted by a1/α2 heterodimer in diploid cells and Mdf1p should function differently in a and α cells because of the absence of MATα2 in a cells. To test the first deduction, we overexpressed MATα1 in the M^+A^- strain in α cells. As we anticipated, the M^+A^- strain recovered much of the mating ability (Figure 4A). To test the second deduction, we further deleted *ADF1* alone in a cells, and found that the mating efficiency of the M^+A^- (a) strain was not as affected as the M^+A^- (α) strain (Figure 4A). Furthermore, to better understand the mechanistic aspects of Mdf1p, we examined the subcellular localization of Mdf1p by adding GFP to the C-terminus of Mdf1p. The fluorescence illustrated that Mdf1p exists in both the cytoplasm and nucleus (Figure 4B), which does not conflict with our Mdf1p-MATα2 interaction hypothesis. Hence, on the whole, the above evidence matches the proposed role for Mdf1p as a transcription suppressor for the mating pathway.

To obtain direct evidence on Mdf1p-MATα2 interaction, we employed the yeast two-hybrid assays. Mdf1p was fused with the DNA-binding domain of Gal4 (DB) and MATα2 protein was fused with the activation domain of Gal4 (AD). The yeast two-hybrid assay results suggest that Mdf1p can interact with MATα2 protein *in vivo* (Figure 4C). *In vitro* GST pull-down assays were carried out to further substantiate the results of yeast two-hybrid assays. MATα2 was expressed as a GST-fusion protein in *E. coli*, while Mdf1p was expressed as a His-fusion protein in yeast. Figure 4D shows that MATα2 and Mdf1p physically interact with each other *in vitro*. Overall, both yeast two-hybrid and GST pull-down assays support the Mdf1p-MATα2 interaction hypothesis.

Mdf1p and MATα2 cooperatively bind to the haploid-specific gene operator

Having established that Mdf1p and MATα2 can interact, we next investigated whether MATα2 and Mdf1p co-bind to the regulatory DNA elements that control haploid-specific genes. ChIP assays were carried out for 10 known haploid-specific genes (*MAT*α1, *STE4*, *STE5*, *STE18*, *FUS1*, *FUS2*, *FUS3*, *GPA1*, *SST2*, and *RME1*)

[20] using the antibody against Mdf1p -6XHis. Except *STE18*, our ChIP experiments successfully recovered the promoters of all the genes (Figure 5A), indicating that Mdf1p specifically contacted with the haploid-specific genes. We further used *in vivo* electrophoretic mobility shift assays (EMSAs) to confirm this result. In the traditional MATa1-MATα2 model, the role of al and α2 proteins is to recognize a roughly 20-bp motif, called the haploid-specific gene (hsg) operator, and suppress the expression of the cognate gene, and the recognition of the hsg operator requires both al and α2 proteins [21]. We chose the most conserved reported motif [21] labeled with biotin to test the affinity by the Mdf1p/MATα2 complex, assuming Mdf1p takes the role of MATa1. Consistently, in our EMSA experiments no detectable binding to the binding motif was observed when only Mdf1p (Figure 5B, lane 8) or MATα2 protein (Figure 5B, lane 9) was contained in the nuclear extracts, whereas Mdf1p and MATα2 cooperatively bound to the biotin-labeled binding motif using the nuclear extracts prepared from M^+A^-

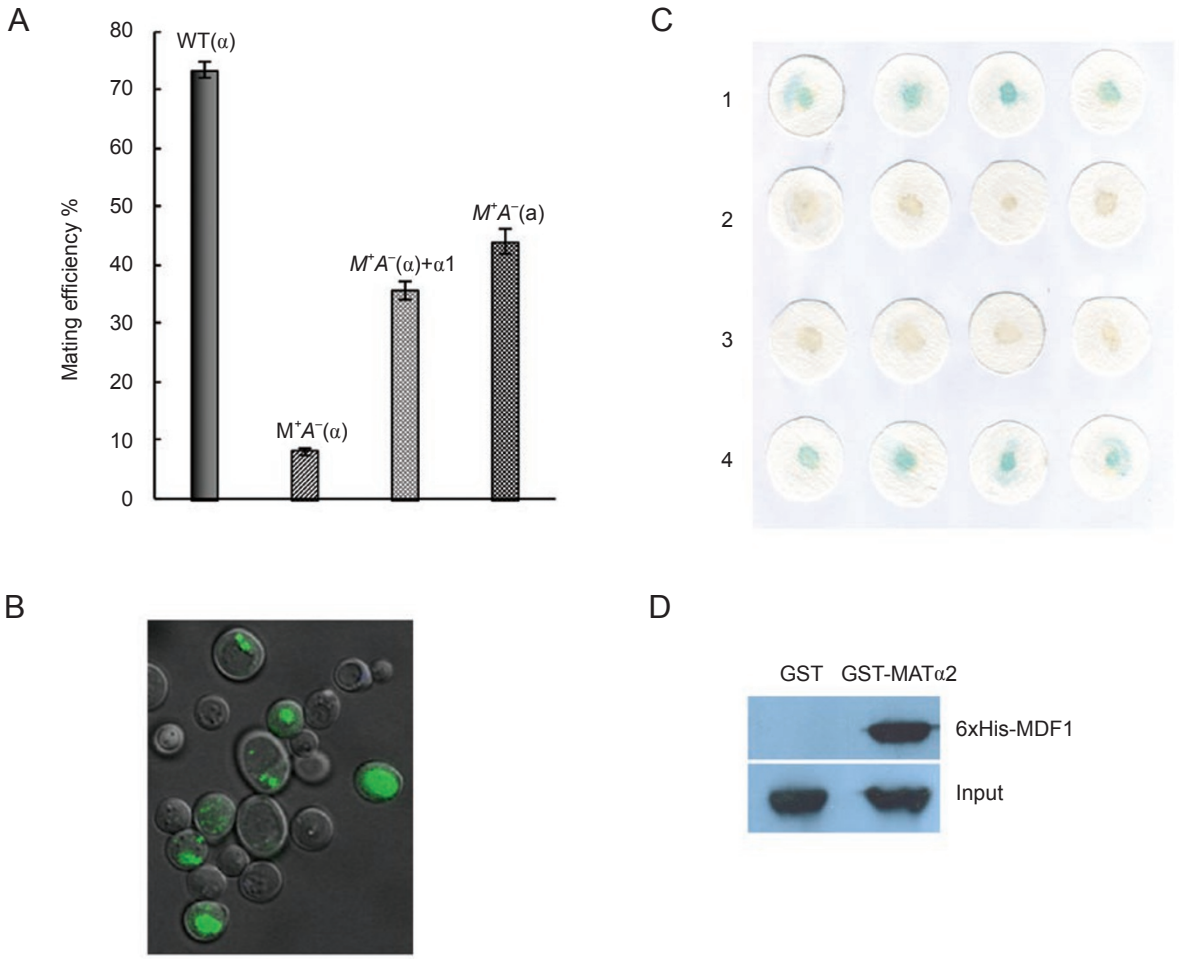


Figure 4 Mdf1p regulates the mating pathway of *S. cerevisiae* by binding MATα2 protein. **(A)** Mdf1p functions upstream of *MAT*α1. Overexpression of α1 protein can extricate M^+A^- strain from mating defect to some extent in α cells, and M^+A^- strain exhibits relatively normal mating efficiency in the background of a cells lacking *MAT*α1. WT (α), wild type (α cells); M^+A^- (α), *ADF1* was deleted in α cells; M^+A^- (α) + α1, α1 was overexpressed in M^+A^- (α); M^+A^- (a), *ADF1* was deleted in a cells. **(B)** Mdf1p is localized in both the nucleus and cytoplasm. **(C)** Yeast two-hybrid assays show that Mdf1p interacts with α2 protein *in vivo*. 1. P53-SV40 as the positive control; 2. Mdf1p fused with the DNA-binding domain of Gal4 (DB) as negative control; 3. MATα2 fused with the activation domain of Gal4 (AD) as negative control. 4. Mdf1p-MATα2 interaction. Four independent clones were patched in the selective plates. **(D)** Pull-down assays prove that Mdf1p physically binds to MATα2 protein *in vitro*. Purified His-tagged Mdf1p was incubated with MATα2 fused to GST or with GST alone, and was detected by western blotting using mouse anti-6XHis-tag monoclonal antibody. Twenty percent of purified His-tagged Mdf1p used for each pull-down reaction is shown as input.

strain (α cells) (Figure 5B, lane 3). These results indicate that Mdf1p and MATα2 also function in a mutually dependent manner. More importantly, the Mdf1p -6XHis antibody and MATα2-Flag tag antibody separately supershifted the band in an antibody concentration-dependent manner (Figure 5B, lanes 4-7), indicating that Mdf1p and MATα2 are indispensable components in the binding complex. The next and even more challenging task is to look for the precise binding site of Mdf1p within the hsg operator. We tried a series of mutated probes labeled with

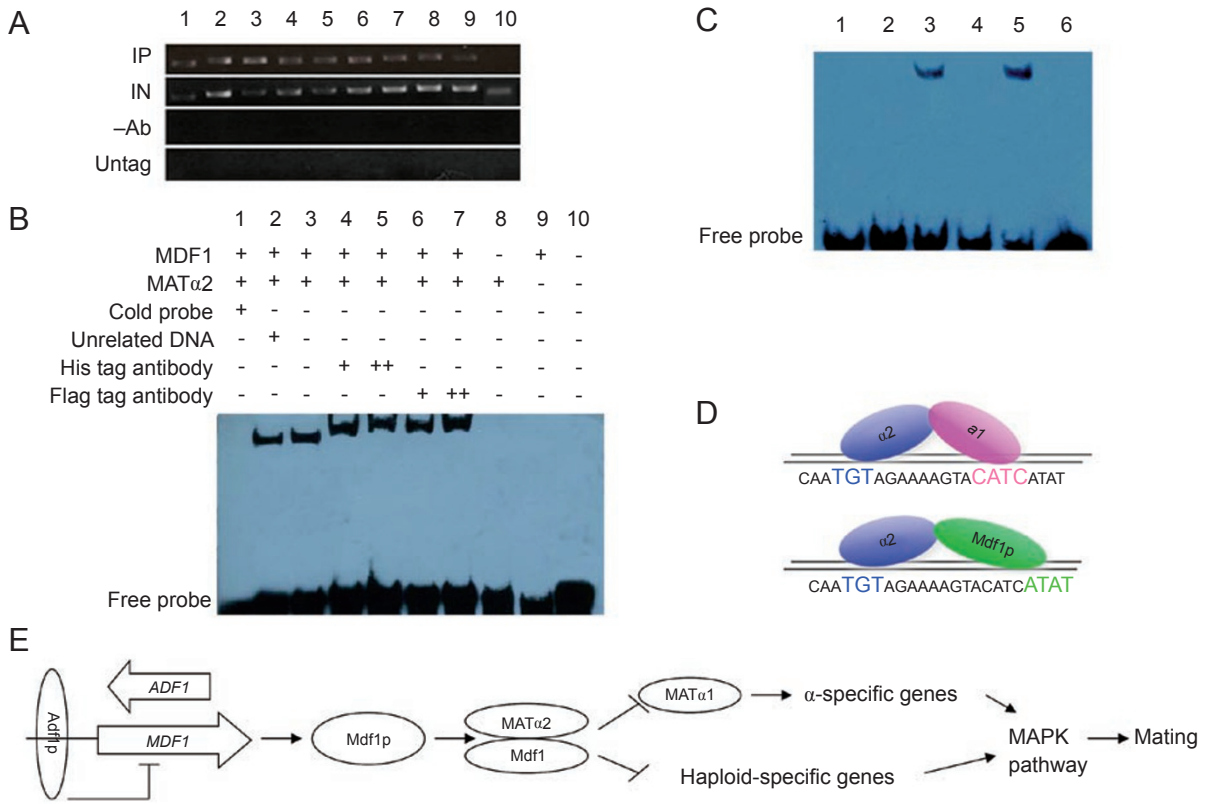


Figure 5 Mdf1p and MATα2 cooperatively bind to the haploid-specific gene operator. **(A)** Results of ChIP show that Mdf1p-MATα2 can bind to the promoters of haploid-specific genes. The final DNA extracts were amplified using a pair of primers that cover the 200-bp upstream flanking region of each haploid-specific gene. IN, input; IP, immunoprecipitation; −Ab, control for non-specific binding in the absence of antibody; untag, untagged yeast as a negative control. 1. *MAT*α1, 2. *STE4*, 3. *STE5*, 4. *FUS1*, 5. *FUS2*, 6. *FUS3*, 7. *GPA1*, 8. *SST2*, 9. *RME1*, 10. *STE18*. **(B)** Separate and cooperative DNA binding activities of Mdf1p and MATα2 to the 3′ biotin-labeled, double-stranded hsg operator probe were measured by EMSA. Nuclear extracts of M^+A^- (α type) cells with Mdf1-6XHis tag and MATα2-Flag tag (lanes 1-7) were used to analyze the DNA binding activities of the MDF1-MATα2 complex. The specificity of the binding is demonstrated by competition with a 200-fold excess of the cold probe (lane 1) and unrelated DNA (lane 2) compared with normal shift band (lane 3). Lanes 8-10 represent nuclear extracts of M^-A^+ cells (α type), M^+A^- cells (a type) and no nuclear extracts. Lanes 4 and 5 represent super-shift experiments after the addition of 0.5 and 1 μg His tag monoclonal antibody, whereas lanes 6 and 7 represent super-shift experiments after the addition of 0.5 and 1 μg Flag tag monoclonal antibody. **(C)** EMSAs using a series of mutated biotin-labeled probes show that the α2 protein in Mdf1p-α2 heterodimer still binds to α2-half sites, and the linker between the two halves is crucial in aiding the binding. The position of Mdf1p on the hsg operator is not in the original a1-half site, but slightly moves four nucleotides (Mdf1p-half site) away from the a1-half site. (1) Probe with α2-half sites mutated, (2) probe with linker mutated, (3) probe with a1-half sites mutated, (4) probe with four nucleotides flanking the a1-half sites mutated, (5) probe without mutation, (6) cold competition. **(D)** Model for the DNA binding features of a1-α2 and Mdf1p-α2 heterodimers. In a1-α2 heterodimer, a1and α2 proteins bind a1-half (pink) and α2-half (blue) sites, respectively; in Mdf1p-α2 heterodimer, Mdf1 and α2 proteins bind Mdf1p-half (green) and α2-half (blue) sites, respectively. **(E)** A model for the functions of Mdf1p and Adf1p in the mating pathway. Mdf1p and MATα2 are physically cross-linked to the promoters of haploid-specific genes and *MAT*α1 which is in charge of opening α-specific genes, thereby repressing the MAPK pathway which is responsible for triggering a series of physiological changes in preparation for mating. To prevent the concomitant side effect of Mdf1p, the expression of *MDF1* is negatively regulated by the transcriptional repressor Adf1p encoded by its antisense strand.

biotin. When we mutated the α2-half sites and the linker between a1 and α2-half sites, no shift bands could be observed (Figure 5C, lanes 1 and 2), indicating that as in a1-α2 heterodimer, the α2 protein in Mdf1p-α2 heterodimer still binds to α2-half sites, and the linker between the two halves is also crucial in aiding the binding. However, contrary to the simple expectation, the position of Mdf1p on the hsg operator is not in the original a1-half site (Figure 5C, lane 3), but slightly moves four nucleotides away from the a1-half site (Figure 5C, lane 4). These results (Figure 5D), combined with other data, strongly support that Mdf1p and MATα2 proteins are bound to each other and jointly regulate those haploid-specific genes. From the convergent evidence obtained so far, a model for the function of Mdf1p in the mating pathway can be drawn as shown in Figure 5E. Through binding MATα2, the central component of the yeast mating pathway, and cooperatively with MATα2 targeting the hsg operator, Mdf1p inhibits *MAT*α1 and other haploid-specific genes from opening the MAPK pathway which is responsible for triggering intracellular mating signal transduction, and consequently decreases the mating efficiency of *S. cerevisiae*.

Computational and experimental analyses strongly support that MDF1 is most likely a de novo gene in S. cerevisiae while ADF1 is conserved across species

The next important question deserving a close investigation is the origination process of both genes. We searched the UniRef90 protein dataset using PSI-BLAST, and found that *ADF1* is conserved in all the sequenced members of hemiascomycete subdivision of fungi except the most distant clade, *Yarrowia lipolytica* (Supplementary information, Figure S2). Undoubtedly, *ADF1* originated at least before the separation of *S. cerevisiae* with the CTG clade 300 million years ago (mya) [22] (Figure 6A). By contrast, *MDF1* does not have significantly homologous ORF in all the other organisms except two short truncated ORFs in the close relatives *S. bayanus* and *S. mikatae* (Supplementary information,

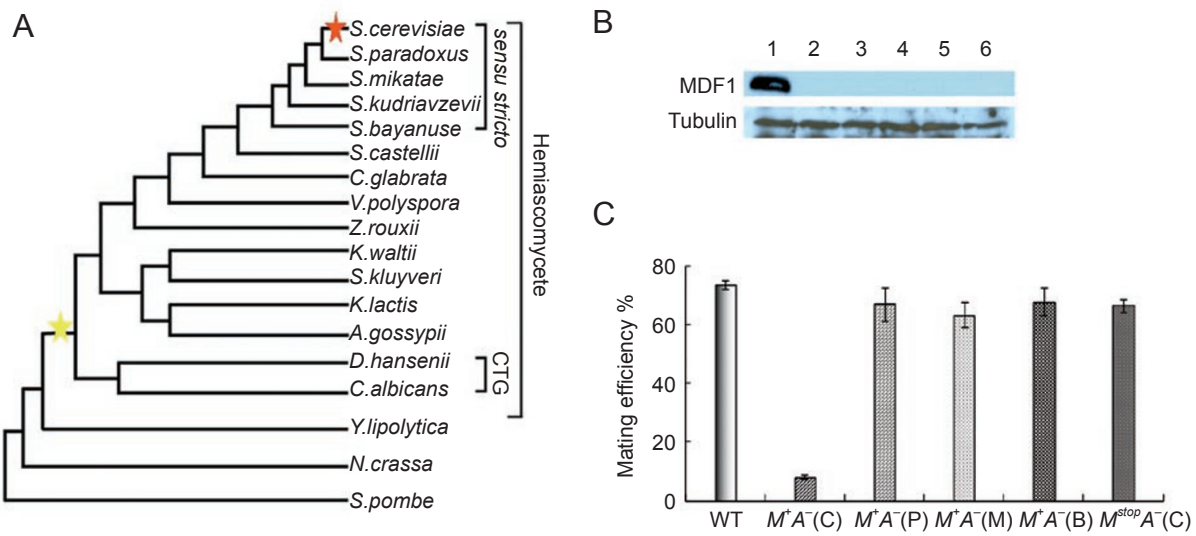


Figure 6 Comparative genomics and experimental analyses support that *MDF1* is most likely a *de novo* originated gene in *S. cerevisiae* while *ADF1* is conserved across species. **(A)** The phylogenetic tree [24] illustrates our hypothesis that *MDF1* emerged specifically in *S. cerevisiae*, while *ADF1* is conserved in all the sequenced members of hemiascomycete subdivision of fungi except the most distant clade, *Yarrowia lipolytica*, which almost completely lost synteny relation with *S. cerevisiae*. The red and yellow stars denote the generation events for *MDF1* and *ADF1*, respectively. **(B)** Western blotting results showed that there is no Mdf1p in other *sensu stricto* group species. His-tags were fused to the 3′-ends of homologous sequences of *MDF1* in *S. paradoxus*, *S. mikatae* and *S. bayanus*, and to the 3′-ends of shorter ORFs in *S. mikatae* and *S. bayanus*, respectively. Tagged Mdf1p in *S. cerevisiae* was used as positive control, tubulin as loading control. 1. *S. cerevisiae*, 2. *S. paradoxus*, 3. *S. mikatae*, 4. *S. bayanus*, 5. Shorter ORF in *S. mikatae*, 6. Shorter ORF in *S. bayanus*. **(C)** Mating assays prove that the existence of the intact *MDF1* is required for acting as a regulator of mating process. When we replaced *MDF1* in *S. cerevisiae*, respectively, with the homologous sequences of *S. bayanus*, *S. mikatae*, and *S. paradoxus*, or replaced the ACA of *MDF1* in *S. cerevisiae* with TGA (stop codon in *S. bayanus* and *S. kudriavzevii*), no mating defect was observed. WT, *S. cerevisiae* wild type; M^+A^- (C), M^+A^- in *S. cerevisiae*; M^+A^- (P), M^+A^- (M), M^+A^- (B), *MDF1* was replaced with homologous sequences of *S. paradoxus*, *S. mikatae*, and *S. bayanus*, respectively; $M^{stop}A^-$ (C), ACA of *MDF1* in *S. cerevisiae* was replaced with TGA.

Figure S3). The flanking genes of *MDF1* in *S. cerevisiae*, *KRR1* and YCL057C-A, are both conserved across fungi. This gene order is maintained in all 13 sequenced hemiascomycete species from *S. cerevisiae* to *Ashbya gossypii* (Supplementary information, Figure S4). When we manually aligned the intergenic region between these two flanking genes in other species, this region could not encode for proteins in any other species, due to the presence of multiple stop codons and frame-shifting indels (Supplementary information, Figure S5). However, it is still theoretically possible that the homologous sequences could maintain some ancestral function in other species in a way that circumvents the stop codons by nonsense suppression (read-through of stop codons), or that the truncated ORFs were functional. To test these alternative hypotheses, first we tested whether *MDF1* and *ADF1* are transcribed and further translated in other *sensu stricto* species. Our strand-specific RT-PCR experiments showed that *MDF1* only expresses in *S. cerevisiae*, while *ADF1* expressed constantly in the *sensu stricto* group (Figure 1A). When His-tags were fused to the 3′-ends of homologous sequences of *MDF1* in *S. paradoxus*, *S. mikatae* and *S. bayanus*, and to the 3′-ends of short ORFs in *S. mikatae* and *S. bayanus*, no protein could be detected under the same conditions (Figure 6B). Second, to specifically test the possibility of read-through of stop codons, we further replaced *MDF1* in *S. cerevisiae* with the homologous sequences in *S. bayanus*, *S. mikatae* and *S. paradoxus* containing stop codons and indels (Supplementary information, Figure S5); no reduction in mating efficiency was observed in all these substituted strains (Figure 6C). We also experimentally replaced the ACA in the 3′-end of *MDF1* in *S. cerevisiae* with TGA (stop codon in *S. bayanus* and *S. kudriavzevii*) (Supplementary information, Figure S5), and observed that the truncated Mdf1p was unable to cause the mating defect (Figure 6C). All the analyses fit the hypothesis that *MDF1*'s homologous sequences in other species are non-coding and the intact ORF of *MDF1* is indispensable for acting as a regulator of mating processes.

However, on account of reported widespread multiple gene losses in yeast after the whole genome duplication (WGD) event [23], more proof is still desired to distinguish between evolutionary innovation and multiple losses in evolution of *MDF1*. Logically, if *MDF1* was an old gene lost in other species, we would have to assume at least nine independent losses in 13 sequenced hemiascomycete lineages based on the phylogeny (Figure 6A). This is in sharp contrast to the fact that most gene-loss events were confined to duplicated copies after whole genome duplications, which was after the split of the lineage leading to *S. cerevisiae* from *K. lactis* about 100 mya [24], and that the most extreme and rare multiple gene-loss cases only have independent gene losses in three or four lineages [25]. In addition, we reconstructed the ancestral consensus sequence of the region that corresponds to *S. cerevisiae*'s *MDF1* gene based on the sequences from the *sensu stricto* species, and found that there were at least two stop codons and two frames-shifting indels in the common ancestor (Supplementary information, Figure S5), indicating that it is unlikely that the *MDF1* gene was lost in all the other species but remained intact only in *S. cerevisiae*. Overall, all the above comparative genomics and experimental data favor the hypothesis that *MDF1* evolved through *de novo* origination rather than multiple losses or extension of the ancestral short functional sequences.

Discussion

New mechanism of sense-antisense interaction

One of our remarkable findings is that the way in which Adf1p regulates *MDF1* fits none of the known sense-antisense interaction mechanisms, i.e., RNAi, transcription interference (TI), or antisense RNA-induced histone deacetylation. The following three pieces of evidence demonstrate that the traditional explanations for the sense-antisense interaction cannot be applied to the Adf1p case. First, when we introduced a stop codon into the N-terminus of Adf1p by site-directed mutagenesis to construct the M^+A^- strain, Adf1p was eliminated, but not the *ADF1* RNA (data not shown). If the regulation was RNA-dependent, the M^+A^- strain in which both RNAs existed should act like the wild type, but it is simply not the case. Thus, RNAi and antisense RNA-induced histone deacetylation can be ruled out. Second, the overexpressed *ADF1* on the plasmid which could completely abolish the sense (i.e. *MDF1*) transcript does not overlap with the chromosomal *ADF1*. Therefore, TI can also be dismissed. Together with our ChIP results of Adf1p, we put forward a new sense-antisense regulation mechanism, in which Adf1p represses the transcription of *MDF1* by binding to the promoter region of *MDF1*. This new finding will certainly widen our understanding of gene regulation and deepen our comprehension on how species with compact genomes use genetic materials economically.

If *MDF1* is a *de novo* gene and *ADF1* is conserved across all hemiascomycetes, what function did *ADF1* play prior to the origination of *MDF1*? In order to give some preliminary hints about the genuine function of *ADF1*, we deleted *ADF1* in *S. paradoxus*, which does not possess a functional *MDF1*, and observed defective growth (Supplementary information, Figure S6A). We

also sequenced many DNA fragments obtained through ChIP for Adf1p in *S. cerevisiae*. In addition to precipitating the *MDF1* promoter, we also obtained the promoters of a number of other old genes that are unrelated to mating (Supplementary information, Table S3). Some of these genes with multiple hits in our shotgun-clone sequencing take roles in pre-rRNA processing, cell wall formation or mitochondrial morphology. Therefore, *ADF1* should have ancestral functions as a transcription factor and was later recruited to repress *MDF1* in *S. cerevisiae*. More studies are needed to address the detailed original functions of *ADF1*, which will shed further light on the evolution of pathways.

MDF1 is an unprecedented example for the de novo origination of a protein-coding gene, leading to additional novel gene function and pathway evolution

Various mechanisms underlying gene origination have been revealed in some genes reported recently [7]. So far only a few *Drosophila* new genes have received evidence for possible functions, such as *Jingwei* and *Sphinx* [26, 27]. One of the most striking findings in this study is that *MDF1*, most likely generated *de novo* from a non-coding sequence, plays very important roles in two fundamental biological processes, namely mating and growth. To our knowledge, this is the first study to provide solid evidence that the *de novo* originated gene can truly encode a protein and play important roles in basic biological processes.

Moreover, the evolution of the intricate pathway upon which natural selection acts is a central and long-standing issue in evolutionary studies. So far, no new gene-involved pathway has ever been reported. Here we present appealing evidence that a *de novo* originated gene *MDF1* can be integrated into the yeast mating pathway at the farthest upstream position. The uniqueness of *MDF1* lies not only in a novel association with a fundamental pathway but also in the position where *MDF1* has been recruited in, i.e., *MDF1* impacts the mating pathway from the very beginning by binding the initiator of the mating process, although it seems more acceptable for a newly evolved gene to be recruited at the downstream nodes of a pathway. Our analyses on *MDF1* enriched our understanding of pathway evolution.

Roles of Mdf1p in mating and growth of S. cerevisiae and implications on evolution of the baker's yeast

Yeasts can reproduce both sexually and asexually (facultative sex); selective forces might have favored either vegetative fitness or mating ability under different conditions and a negative correlation between these two traits might exist [28]. It is one of nature's wonders to recruit a new component Mdf1p into the mating pathway to make yeast better able to balance the gain and cost of these two physiological phenomena. In benign condition, especially after the haploids' recovery from growth arrest under unfavorable conditions, vegetative proliferation is advantageous in rapid resource consumption and Mdf1p shuts down the mating pathway to limit the cost of mating, and thus *S. cerevisiae* is at a selective advantage relative to their more efficiently mating but slower-growing competitors; while in stressful condition, mating is favorable and *MDF1* is suppressed by Adf1p to gain the benefit of sexual reproduction. The new regulatory circuit involving this sense-antisense gene pair might have aided *S. cerevisiae* in exquisitely adapting to the changing environment.

When the mating pathway is stimulated by a pheromone secreted by a nearby cell of the opposite mating type, yeast cells undergo a series of physiological changes in preparation for mating [15]. These include arrest in the G1 phase of the cell cycle. Mdf1p is able to promote growth and decrease the mating efficiency of *S. cerevisiae* simultaneously. One possible connection between growth and mating is that by binding MATα2 protein and further silencing the downstream haploid-specific genes, which sends the fictitious signal of diploids, Mdf1p may push yeast cells away from cell cycle arrest and thus accelerate mitotic cell growth. This hypothesis is consistent with a recent conclusion that a growth-rate advantage can be gained by losing signaling at multiple points in the mating pathway [29]. However, it is noteworthy that the comprehensive molecular mechanism of Mdf1p promoting growth is still not clear. The GFP fusion protein assay showed that Mdf1p exists in both the cytoplasm and nucleuses (Figure 4B), while Mdf1p binding MATα2 could only explain the nuclear localization of Mdf1p. Therefore, it is plausible to assume the existence of additional MATα2-independent interacting factor(s) with Mdf1p in the cytoplasm. Our microarray and two-dimensional electrophoresis data show that many metabolic genes are influenced by Mdf1p (Figure 3A and Supplementary information, Figure S1B), but how these effects have happened remain unclear to us. Future studies are still needed to reveal the detailed pathway/network involved by Mdf1p to promote growth.

However, mad growth without orchestrating the internal and external conditions is not always beneficial to yeast cells. In fact, in contrast to the superiority of the M^+A^- strain in rich medium, Mdf1p is unable to promote growth in nonfermentative medium (Supplementary information, Figure S6B), in which sexual reproduction is advantageous [28]. Hence, the non-mating haploid M^+A^- strain may not be very good at coping with harsh nutri-

tional condition owing to the low efficiency of sporulation which is a normal strategy to resist against adverse circumstances. Besides, as revealed by microarray data, M^+A^- strain seems to exhibit some defects in DNA damage repairing due to the inhibited transcription of some DNA damage regulators (Figure 3A). Therefore, *MDF1* should be under stringent control to avoid its side effect. Opportunely, *MDF1* recruited its antisense gene *ADF1* as a negative regulator. Interestingly, as demonstrated by previous microarray analysis, the expression level of *ADF1* is fluctuating [30]. This intriguing pattern hints that the regulatory circuit should be dynamic in response to the change of physiological condition in wild type. In future, we are anticipating that more studies on both *MDF1* and *ADF1* will lead us towards an integrated understanding on how *MDF1* and *ADF1* regulate growth and other biological processes.

Materials and Methods

Competition experiment

The deletion strains including M^-A^-, M^+A^-, M^-A^- in the background of *S. cerevisiae* and A^- in *S. paradoxus* contain genetic markers with resistance to geneticine conferred by the *kan*MX4 cassette inserted into a deleted chromosomal *MDF1* locus. To obtain differently marked competitors, we introduced nourseothricine resistant to the wild types *S. cerevisiae* (α cells) and *S. paradoxus* (α cells) by inserting *nat*MX4 into the *HO* locus. Previous experiments established that these markers are neutral compared to unmarked wild-type strains [31]. The competition experiments were carried out as follows: equal volumes of overnight cultured competing pairs were mixed. After 24 h competition, the mixed cultures were printed onto two selective agar media with geneticine or nourseothricine added. The resulting surface cultures were photographed and the clone numbers were counted after 48 h of incubation at 30 °C.

Mating efficiency test

The efficiency of mating was determined as follows (modified from Hartwell [32]): cells were grown in YPD broth to a density of 3×10^7 cells per ml. The α-cell cultures to be tested were mixed 100:1 with the a-cell cultures at room temperature. The mating cultures were spread on selective medium (Met^- and Lys^-) at 30 °C to determine the number of diploids and on a different selective medium (Met^-) to determine the number of haploids. The mating efficiency is defined as the number of diploids observed on the first selective medium divided by the number of haploids observed on the second selective medium.

Yeast two-hybrid assay

All procedures essentially followed the Yeast Protocols handbook (Clontech). Briefly, the coding sequence of *MDF1* was fused with the DNA-binding domain of Gal4 (DB) and the coding sequence of *MAT*α2 was fused with the activation domain of Gal4 (AD). Then these two plasmids were co-transformed to identify interaction or transformed separately as negative controls to the host strain Y190 (*MAT*a, *gal4-542*, *gal80-538*, *his3*, *trp1-901*, *ade2-101*, *ura3-52*, *leu2-3, 112*, URA3::GAL1-LacZ, Lys2::GAL1-HIS3cyh^r). After selection on SD–Trp–Leu–His plates, 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) was added to evaluate the strength of interaction.

GST pull-down assay

The *MAT*α2 coding sequence with the stop codon was cloned into pGEX-4T-1 vector to be a GST-MATα2 plasmid. The GST-MATα2 plasmid was transformed into the *E. coli* strain BL21. The expression of the fusion protein was induced by adding IPTG with a final concentration of 1 mmol/ml and incubated at 16 °C for 4 h. After lysis of the bacterial cells by sonication, GST or GST-MATα2 fusion protein was immobilized on glutathione-Sepharose 4B beads according to the manufacturer's (GE Healthcare) instruction. The beads were washed three times with cold PBS. 100 OD 6XHis-tagged *MDF1* overexpressed yeasts were lyzed in 500 μl of lysis buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, protease inhibitor cocktail (Roche), 0.2 mM Na_3VO_4, 100 mM NaF, 0.2% NP-40) by glass bead beating and centrifuged at 12 000 × *g* for 5 min at 4 °C. 200 μl of supernatant was incubated with 20 μl of GST or GST-MATα2 immobilized glutathione-Sepharose 4B beads overnight at 4 °C. After incubation, the beads were washed with lysis buffer four times. The bound proteins were analyzed by western blotting using anti-6-XHis tag antibody (R&D Systems).

Chromatin immunoprecipitation assay

100 ml of cells overexpressing Mdf1p or Adf1p (2.0×10^7 cells/ml) was crosslinked with 2% formaldehyde for 15 min at room temperature. Glycine was added to a final concentration of 250 mM, and the incubation continued for an additional 5 min. The suspension was sonicated seven times for 10 s each, with the amplitude set at 30% using an ultrasonic processor (Sonic ultracell). Samples were incubated on ice for 2 min between sonications. The suspension was clarified by centrifugation for 5 min at 10 000 × *g* at 4 °C in a microcentrifuge. 1 μl of RNase (10 μg/μl) was added to the samples, and they were incubated for 30 min at 37 °C. Afterwards, sheared chromatin was purified using QIAquick spin columns (Qiagen). Then 250 μl of supernatant was incubated with 15 μl of anti-His monoclonal antibody (R&D Systems). The promoter primers of 10 haploid-specific genes (*MAT*α1, *STE4*, *STE5*, *STE18*, *FUS1*, *FUS2*, *FUS3*, *GPA1*, *SST2*, *RME1*) residing in the 200-bp upstream flanking region of each gene were used for the PCR analysis. The *MDF1* promoter primers used were as follows: *MDF1*-Chip-Fwd, 5′-TAG TCT TAA GCG ACG ATG CTT TAT-3′, and *MDF1*-Chip-Rev, 5′-CAG AAA AAT CAA AAA CAA ACG ACA G-3′, which flank the –150 bp to +29 region of the *MDF1* gene.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed using nuclear extracts from M^+A^- (α type) cells coexpressing Mdf1-6XHis tag and MATa2-Flag tag as described previously [33] with modified extraction buffer (HEPES, pH 8.0, 20 mM, NaCl 400 mM, EDTA 1 mM, DTT 1 mM, NP-40 1%, glycerol 10%, protease inhibitor cocktail). The oligonucleotide probes of the hsg operator labeled with 3′-biotin are listed in Supplementary information, Table S4. For antibody supershift assays, anti-6XHis tag monoclonal antibody or anti-Flag tag mono-

clonal antibody was incubated for 30 min followed by EMSA procedures [21] using North2South® Chemiluminescent Hybridization and Detection Kit (PIERCE) for detection.

(Supplemental materials and methods are depicted in the Supplementary information, Data S1)

Acknowledgments

We thank Chuck Cannon (Texas Tech University, USA); Zhenglong Gu, Roberto Flores, Shu-Bing Qian and Thomas Fox (Cornell University, USA) for their helpful comments and/or English editing on the manuscript. We are grateful to all members of CAS-Max Planck Junior Research Group for their help in the experiments and discussions during preparation of the manuscript. We also thank Jinqiu Zhou (Shanghai Institute for Biological Sciences, China) for providing yeast strains and Simon Newstead (Imperial College, United Kingdom) for the GFP-fusion plasmid. This work was supported by a CAS key grant (KSCX2-YW-N-060), two NSFC key grants (No. 30930056 and 30623007), a 973 Program (No. 2007CB815703-5), a CAS-Max Planck Society Fellowship, and the 100 Talent Program of CAS to W.W.

References

1 Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet* 2005; **21**:93-102.
2 Almeida R, Allshire RC. RNA silencing and genome regulation. *Trends Cell Biol* 2005; **15**:251-258.
3 Aravind L, Watanabe H, Lipman DJ, Koonin EV. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci USA* 2000; **97**:11319-11324.
4 Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. *Trends Genet* 2005; **21**:339-345.
5 Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* 2007; **131**:706-717.
6 Page N, Gerard-Vincent M, Menard P, *et al.* A Saccharomyces cerevisiae genome-wide mutant screen for altered sensitivity to K1 killer toxin. *Genetics* 2003; **163**:875-894.
7 Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 2003; **4**:865-875.
8 Ohno S. *Evolution by Gene Duplication*. New York: Springer-Verlag, 1970.
9 Jacob F. Evolution and tinkering. *Science* 1977; **196**:1161-1166.
10 Begun DJ, Lindfors HA, Kern AD, Jones CD. Evidence for *de novo* evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. *Genetics* 2007; **176**:1131-1137.
11 Cai J, Zhao R, Jiang H, Wang W. *De novo* origination of a new protein-coding gene in Saccharomyces cerevisiae. *Genetics* 2008; **179**:487-496.
12 Casci T. Evolution: a gene is born. *Nat Rev Genet* 2008; **9**:415.
13 Li W-H. *Molecular Evolution*. Sunderland, MA: Sinauer, 1997.
14 Winzeler EA, Shoemaker DD, Astromoff A, *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 1999; **285**:901-906.
15 Bardwell L. A walk-through of the yeast mating pheromone response pathway. *Peptides* 2005; **26**:339-350.
16 Rine J, Herskowitz I. Four genes responsible for a position effect on expression from HML and HMR in Saccharomyces cerevisiae. *Genetics* 1987; **116**:9-22.
17 Astrom SU, Okamura SM, Rine J. Yeast cell-type regulation of DNA repair. *Nature* 1999; **397**:310.
18 Pollastri G, McLysaght A. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 2005; **21**:1719-1720.
19 Ke A, Wolberger C. Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein Sci* 2003; **12**:306-312.
20 Herskowitz I. A regulatory hierarchy for cell specialization in yeast. *Nature* 1989; **342**:749-757.
21 Goutte C, Johnson AD. Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J* 1994; **13**:1434-1442.
22 Dujon B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* 2006; **22**:375-387.
23 Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* 2004; **428**:617-624.
24 Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* 2007; **449**:54-61.
25 Hittinger CT, Rokas A, Carroll SB. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci USA* 2004; **101**:14144-14149.
26 Zhang J, Dean AM, Brunet F, Long M. Evolving protein functional diversity in new genes of Drosophila. *Proc Natl Acad Sci USA* 2004; **101**:16246-16250.
27 Dai H, Chen Y, Chen S, *et al.* The evolution of courtship behaviors through the origination of a new gene in Drosophila. *Proc Natl Acad Sci USA* 2008; **105**:7478-7483.
28 Zeyl C, Bell G. The advantage of sex in evolving yeast populations. *Nature* 1997; **388**:465-468.
29 Lang GI, Murray AW, Botstein D. The cost of gene expression underlies a fitness trade-off in yeast. *Proc Natl Acad Sci USA* 2009; **106**:5755-5760.
30 Williams RM, Primig M, Washburn BK, *et al.* The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast. *Proc Natl Acad Sci USA* 2002; **99**:13431-13436.
31 Goldstein AL, McCusker JH. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* 1999; **15**:1541-1553.
32 Hartwell LH. Mutants of Saccharomyces cerevisiae unresponsive to cell division control by polypeptide mating hormone. *J Cell Biol* 1980; **85**:811-822.
33 Murphy M, Shimizu M, Roth S, Dranginis A, Simpson R. DNA-protein interactions at the S. cerevisiae {alpha} 2 operator *in vivo*. *Nucleic Acids Res* 1993; **21**:3295.

(**Supplementary information** is linked to the online version of the paper on *Cell Research* website.)