

# Comparative Genome Analysis Between *Aspergillus oryzae* Strains Reveals Close Relationship Between Sites of Mutation Localization and Regions of Highly Divergent Genes among *Aspergillus* Species

MYCO Umemura<sup>1,†</sup>, HIDEAKI Koike<sup>2,†</sup>, NORIKO Yamane<sup>2</sup>, YOSHINORI Koyama<sup>2</sup>, YUKI Satou<sup>3</sup>, IKUYA Kikuzato<sup>3</sup>, MORIMI Teruya<sup>4</sup>, MASATOSHI Tsukahara<sup>5</sup>, YUMI Imada<sup>5</sup>, YOUJI Wachi<sup>3</sup>, YUKINO Miwa<sup>5</sup>, SHUICHI Yano<sup>5</sup>, KOICHI Tamano<sup>1</sup>, YUTAKA Kawarabayasi<sup>2</sup>, KAZUHIRO E. Fujimori<sup>2</sup>, MASAYUKI Machida<sup>1,\*</sup>, and TAKASHI Hirano<sup>3</sup>

Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Higashi-Nijo 17-2-1, Tsukisamu, Sapporo, Hokkaido 062-8517, Japan<sup>1</sup>; Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Higashi 1-1-1, Tsukuba, Ibaraki 305-8566, Japan<sup>2</sup>; Okinawa Science and Technology Promotion Center, 12-2 Suzaki, Uruma, Okinawa 904-2234, Japan<sup>3</sup>; Okinawa Industrial Technology Center, 12-2 Suzaki, Uruma, Okinawa 904-2234, Japan<sup>4</sup> and Tropical Technology Center Ltd, 5-1 Suzaki, Uruma, Okinawa 904-2234, Japan<sup>5</sup>

\*To whom correspondence should be addressed. Tel. +81 11-857-8504. Fax. +81 29-861-6174.  
E-mail: m.machida@aist.go.jp

Edited by Katsumi Isono  
(Received 4 June 2012; accepted 12 July 2012)

## Abstract

***Aspergillus oryzae* has been utilized for over 1000 years in Japan for the production of various traditional foods, and a large number of *A. oryzae* strains have been isolated and/or selected for the effective fermentation of food ingredients. Characteristics of genetic alterations among the strains used are of particular interest in studies of *A. oryzae*. Here, we have sequenced the whole genome of an industrial fungal isolate, *A. oryzae* RIB326, by using a next-generation sequencing system and compared the data with those of *A. oryzae* RIB40, a wild-type strain sequenced in 2005. The aim of this study was to evaluate the mutation pressure on the non-syntenic blocks (NSBs) of the genome, which were previously identified through comparative genomic analysis of *A. oryzae*, *Aspergillus fumigatus*, and *Aspergillus nidulans*. We found that genes within the NSBs of RIB326 accumulate mutations more frequently than those within the SBs, regardless of their distance from the telomeres or of their expression level. Our findings suggest that the high mutation frequency of NSBs might contribute to maintaining the diversity of the *A. oryzae* genome.**

**Key words:** *Aspergillus oryzae*; genetic alteration; niche adaptation; non-syntenic block

## 1. Introduction

*Aspergillus oryzae* has been utilized for over 1000 years in Japan for the production of various traditional foods, including soy sauce, *sake* (rice wine), soybean paste, and vinegar. Various *A. oryzae* strains have

been isolated and/or selected for the effective degradation of target materials during the fermentation. In relation to a wide range of fermentation products produced by *A. oryzae*, genetic alteration between *A. oryzae* species is of particular interest.

As a wild-type strain of *A. oryzae*, the genome sequence of *A. oryzae* RIB40 [37 Mb; the National Research Institute of Brewing (NRIB), Japan; NITE Biological Resource Center (NBRC) stock no.

† These authors contributed equally to this work.

100959; American Type Culture Collection (ATCC) no. 42149] was completed<sup>1</sup> and compared with those of *Aspergillus nidulans* (30 Mb)<sup>2</sup> and *Aspergillus fumigatus* (29 Mb),<sup>3</sup> revealing that the *A. oryzae* genome could be characterized into distinct syntenic blocks (SBs) and non-SBs (NSBs). Because the three species are phylogenetically closely related, SBs occupy ~70–80% of the three genomes.<sup>2</sup> NSBs are considered to be unique genetic elements of individual species, and the *A. oryzae* genome has the largest total size of NSBs (Fig. 1).

Compared with NSBs found in the genomes of yeast and other filamentous fungi, NSBs of *A. oryzae* show the significant enrichment of genes involved in metabolism, including secondary metabolism, hydrolase, and transporter genes,<sup>1</sup> and contain genes that are less conserved than those of SBs.<sup>4</sup> In general, the transcriptional expression of NSB genes is markedly weaker than that of genes located in SBs; however, several NSB genes functioning as secretory hydrolases and transporters are highly induced in solid-state cultivation and reach absolute expression levels comparable with those of SB genes.<sup>5</sup> Another characteristic feature of NSB genes is that they are often repressed by heat shock.<sup>5</sup> Together, these findings indicate that SBs and NSBs represent distinct genetic segments in the *A. oryzae* genome in terms of biological function, diversity, and transcriptional regulation, and NSBs can contribute to a wide range of fermentation capability of *A. oryzae* strains.<sup>6</sup> However, a difference of genetic alterations between SBs and NSBs remains unclear.

To reveal mutational characteristics of NSBs, we sequenced the genome of *A. oryzae* RIB326 and compared it with strain RIB40. RIB326 was isolated from an industrial *A. oryzae* strain isolated from shoyu koji (inocula for soy sauce fermentation) in a conida supplier (Osaka-Konno), independently from RIB40.<sup>7,8</sup> We considered that a genomic comparison of the two strains would help understand the development of mutations during the domestication process,

particularly relating to NSBs containing genes responsible for fermentation and secondary metabolism.<sup>4</sup>

## 2. Materials and Methods

### 2.1. Strain and medium

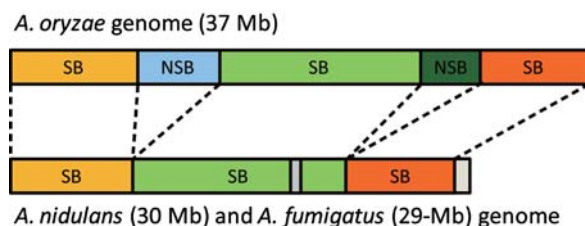
The fungal strain used in this study, *A. oryzae* RIB326, was obtained from the National Research Institute of Brewing, Japan (<http://www.nrib.go.jp/ken/asp/strain.html>). For DNA isolation, the fungus was grown in liquid yeast extract peptone dextrose medium (Difco) at 30°C for 2 days.

### 2.2. DNA preparation

Genomic DNA was isolated from RIB326 by first grinding mycelia to a fine powder in liquid nitrogen, and then mixing ~40 g of the ground mycelia with 200 ml of 10 mM Tris buffer (pH 7.0) containing 50 mM ethylenediaminetetraacetic acid (EDTA), 200 mM NaCl, 1% Triton X-100, and 10 mg each of cellulase (Yakuruto), yatalase (Takara) and lysing enzyme (Sigma). The mixture was incubated for 1 h at 37°C, followed by the addition of RNase A (Funakoshi) to a final concentration of 20 mg/ml and further incubation for 1 h. After the addition of proteinase K (Takara) to a final concentration of 0.1 mg/ml, the solution was incubated for 2 h at 50°C, followed by centrifugation at 2300 × *g* for 10 min. The resulting supernatant was subjected to phenol, phenol/chloroform and chloroform extraction, and DNA was then precipitated using ethanol. After dissolving the precipitated DNA in Tris-EDTA buffer, genomic DNA was purified using a Genomic-tip column (Qiagen).

### 2.3. Sequencing and assembly of the *A. oryzae* RIB326 genome using a next-generation sequencing system

The whole genome sequencing of RIB326 using both fragment and mate-paired libraries was performed using the SOLiD system (ABI), a next-generation sequencing system, according to the manufacturer's instruction. Fragment sequencing yielded 70 268 590 sequence reads of 50 bp, whose depth on the reference sequence of RIB40 [~37 Mb, DNA Data Bank of Japan (DDBJ): AP007150–AP007177; GenBank: AP007150–AP007177] was 67 and covered 94.6% of the reference nucleotide sequence after mapping. The mate-paired sequencing yielded 86 209 971 sequence reads of 50 bp, which were subjected to *de novo* assembly using Velvet (version 1.0.15)<sup>9</sup> on a DELL Precision T7500 (CPU: Xeon E5620, memory: 96 GB, hard disk: 2 TB × 5, OS: Ubuntu Linux 10.04). The assembled scaffolds had a total length of 36 395 679 bp; N50 length of 643 359 bp; and mean length of 7286 bp. N50 is defined as the scaffold length



**Figure 1.** Organizations of SBs and NSBs in the *A. oryzae*, *A. fumigatus*, and *A. nidulans* genomes. SBs, which were identified from syntenic analysis between *A. oryzae*, *A. fumigatus*, and *A. nidulans*,<sup>1</sup> are common between the genomes, whereas NSBs in *A. oryzae* consist of the extra genetic regions from other two smaller genome sequences.

such that using equal or longer scaffolds produces half the bases of all scaffolds. The longest scaffold had a length of 2 218 714 bp. The sequences of the longest 200 scaffolds are available in the DDBJ (BAEZ01000001–BAEZ01000200).

#### 2.4. Nucleotide mutation analysis between strains RIB40 and RIB326

Conserved sequences and nucleotide alterations in strain RIB326 were analyzed using Corona-Lite software (Version 4.2) with the RIB40 genome sequence as a reference. Genome sequence rearrangements between strains RIB40 and RIB326 were analyzed using the longest 200 scaffolds of RIB326 determined from *de novo* assembly. The 200 scaffolds represent 35 425 415 bp (97.3%) of the total length of the entire RIB326 assembly. Homology searches of the 200 RIB326 scaffolds against 22 RIB40 scaffolds using the LAST algorithm<sup>10–12</sup> generated 5521 pairs of nucleotide sequences having scores and identities of >500 and >98%, respectively. Deletions, insertions and inversions of longer than 500 bp in the RIB326 scaffolds were counted. For the display of the RIB40 genome, sequences of the 22 scaffolds were concatenated as one continuous sequence with an artificial N sequence of 10 kb between the neighbouring scaffolds. The final length of the RIB40 genome sequence was determined to be 37 268 201 bp.

#### 2.5. Analysis of mutation frequency in coding sequences between RIB40 and RIB326

To identify coding sequences (CDSs) in RIB326, each amino acid sequence of RIB40 genes was subjected to tBlastn searches<sup>13,14</sup> against the nucleotide sequence of the RIB326 genome. Each codon of RIB40 CDS was compared with the corresponding RIB326 codons, and codons were classified as either with or without mutation(s), or containing unidentified nucleotide(s). Codons with mutation(s) were further classified as those without (silent) or with (non-silent) amino acid substitutions.

#### 2.6. Expressed sequence tags

*Aspergillus oryzae* expressed sequence tags (ESTs) generated under eight culture conditions<sup>15</sup> were obtained from <http://riodb.ibase.aist.go.jp/ffdb/EST-DB.html>. The eight culture conditions were: (i) liquid nutrient-rich culture, (ii) liquid nutrient-rich culture at high temperature, (iii) liquid maltose-inductive culture, (iv) liquid carbon-starved culture, (v) liquid germinated conidia and conidia, (vi) alkaline pH culture on agar plates, (vii) solid-state wheat bran culture, and (viii) solid-state soybean culture (for details, see Akao *et al.*<sup>15</sup>). Gene numbers and averaged

amino acid substitution ratios were evaluated for SBs and NSBs for each EST level. The Wilcoxon signed-rank test was used to evaluate the *P*-value for differences between the amino acid substitution ratios of SBs and NSBs for each EST level when more than four genes were at the EST level in both SBs and NSBs.

#### 2.7. Eukaryotic orthologous group classification of CDSs

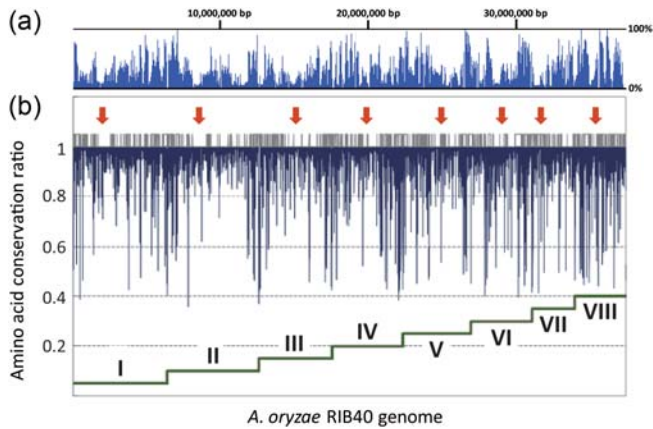
The functional categories of *A. oryzae* genes were assigned according to the eukaryotic orthologous group (KOG) classification<sup>16</sup> by searching for homology against amino acid sequences in the KOG database with a bit score of  $\geq 60$ . The genes categorized as belonging to more than one KOG category were counted for every relevant category. The 22 KOG categories were classified into three meta-categories: I, information storage and processing; II, cellular process and signalling; and III, metabolism. The three meta-categories include the following categories. Meta-category I: J, translation, ribosomal structure, and biogenesis; A, RNA processing and modification; K, transcription; L, replication, recombination, and repair; and B, chromatin structure and dynamics; meta-category II: D, Cell cycle control, cell division, and chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; Z, cytoskeleton; U, intracellular trafficking, secretion, and vesicular transport; O, post-translational modification, protein turnover, and chaperones; Y, nuclear structure; and W, extracellular structures; meta-category III: C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; and Q, secondary metabolite biosynthesis, transport, and catabolism.

### 3. Results and Discussion

#### 3.1. Are mutations localized in NSB or only in sub-telomeric region?

To grasp genetic alterations between the two *A. oryzae* strains, the localization of nucleotide alterations and amino acid substitutions between RIB326 and RIB40 were mapped on the RIB40 genome (Fig. 2). In regions proximal to centromeres, a decrease in the nucleotide substitution frequency was observed, indicating that these regions are subjected to strong selective pressure. In contrast, the nucleotide substitution ratio was higher in regions proximal to chromosomal termini (sub-telomeric regions). This result is in good agreement with the fact that the



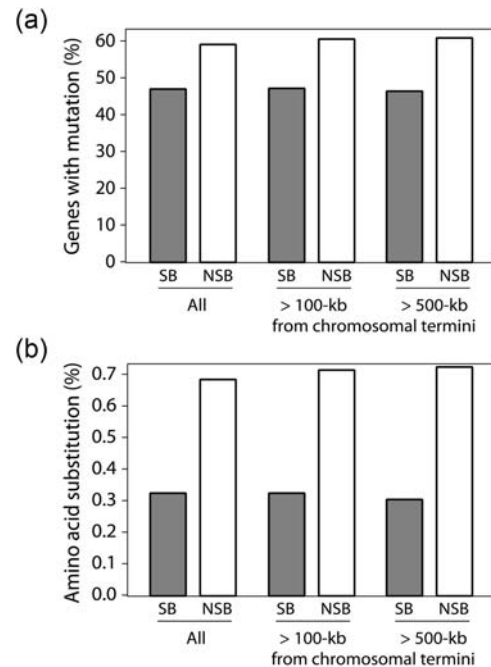


**Figure 2.** Distribution of mutations in the RIB326 genome. (a) The nucleotide substitution percentage and (b) amino acid conservation ratio in the *A. oryzae* RIB40 genome against RIB326 were plotted using a 10-kb window with a 5-kb overlap. Each chromosomal region (I–VIII) is indicated by dark green stepwise line located at the bottom of the graph, and the position of the centromere in each chromosome deduced from the terminal of scaffolds is indicated by red arrows. The short vertical gray lines located in at the middle of the figure represent the position of NSBs determined from syntenic analysis between *A. oryzae*, *A. nidulans*, and *A. fumigatus*.<sup>1</sup>

sub-telomeric regions of yeast and humans are unstable due to telomere loss and thus, accumulate mutations at a higher rate.<sup>17–19</sup> In *Caenorhabditis elegans*, the single-nucleotide polymorphism density on the chromosomal terminus was ~4.5-fold more abundant than that found on the central part of the chromosome.<sup>20</sup> The location of mutations appears to be correlated with NSBs, which are generally localized proximal to chromosomal termini. Pairwise comparison of the genes between the two genomes showed that only 1.2% of RIB326 genes had more than 20% deletions/substitutions of amino acids from corresponding RIB40 genes. Therefore, almost all SBs and NSBs in RIB326 are located at the same positions of those in RIB40 and the higher mutation frequency in NSBs is consistent using either RIB326 or RIB40 genomes.

Figure 3 shows the frequency of gene alterations for CDS located within SBs and NSBs according to their location within the RIB326 genome. The analysis revealed that genes containing mutations were ~10% more frequent within NSBs than those located in SBs and were more prevalent when excluding genes closest (100 or 500 kb) from chromosomal termini (Fig. 3a). Similarly, the amino acid substitution ratio of all NSB genes was higher than that observed in SB genes and further increased on the exclusion of genes located closest to termini (Fig. 3b). These results indicate that NSBs accumulate a greater number of mutations than SBs regardless of the chromosomal location.

The distribution of deletions, insertions, and inversions within the RIB326 genome, as mapped on the

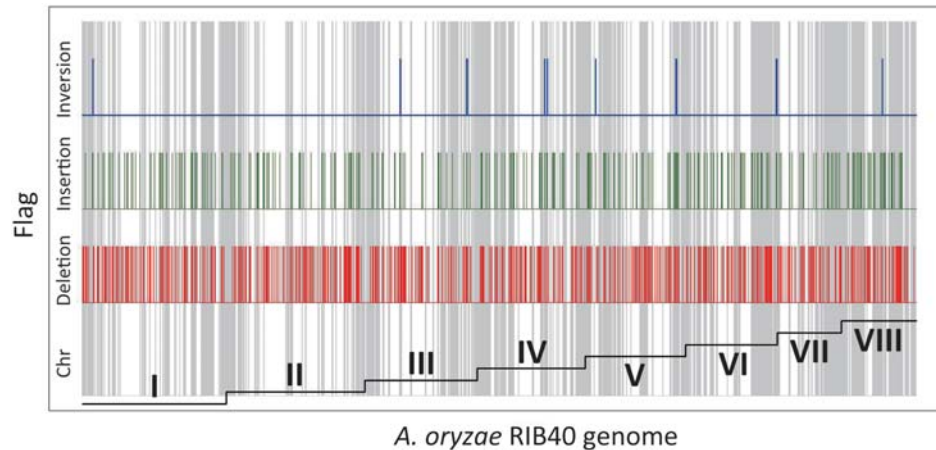


**Figure 3.** Mutation of SB and NSB genes in relation to the distance from chromosomal termini. The percentages of (a) genes with mutation and (b) averaged amino acid substitutions are shown for all SB and NSB genes located in the genome, and those located greater than 100- or 500-kb distant from the closest chromosomal termini.

RIB40 genome, was also determined, revealing that genetic rearrangements also accumulated in NSBs of strain RIB326 (Fig. 4). Table 1 summarizes the number and ratio of deletions, insertions, and inversions in the RIB40 genome sequence against that of RIB326, which were determined using the longest 200 scaffolds (>576 bp) obtained by sequencing of the mate-paired library of RIB326. Notably, 24.2 and 19.8% of the deletions and inversions, respectively, were found within NSBs on the RIB40 genome. Considering that the nucleotides defined as NSBs occupy only 12.4% of the entire genome sequence of RIB40, the results indicate that deletions and inversions are highly enriched in NSBs of *A. oryzae*.

### 3.2. Are NSB genes functionally active?

Figure 5 shows the number of genes and frequency of amino acid substitutions for genes located within SBs and NSBs against the redundancy of ESTs, which were previously obtained from eight different cultural conditions.<sup>15</sup> Our analysis revealed that the percentage of genes within SBs and NSBs decreased as the EST level increased, with the number of NSB genes decreasing at a higher rate than that of SB genes (Fig. 5a). This result is in good agreement with the finding that NSB genes typically have markedly lower expression levels than those of SB genes.<sup>4</sup> Nevertheless, a small percentage of NSB genes was



**Figure 4.** Distribution of deletions, insertions, and inversions in the RIB40 genome. The regions of deletion (red), insertion (green), and inversion (blue) were evaluated between the RIB40 reference scaffolds and RIB326 scaffolds obtained from *de novo* assembly. Each RIB40 chromosome is indicated in a step graph at the bottom of the chart. For insertions, the starting points were plotted.

**Table 1.** Number and percentage of deletion, insertion, and inversion regions in the *A. oryzae* RIB40 genome

	Deletion	Insertion	Inversion
Number	645	107	9
Ratio against the RIB40 genome sequence (%)	3.3	– <sup>a</sup>	~0
NSB ratio in the region (%)	24.2	– <sup>a</sup>	19.8

<sup>a</sup>Not determined, as the insertion length was not accurately evaluated.

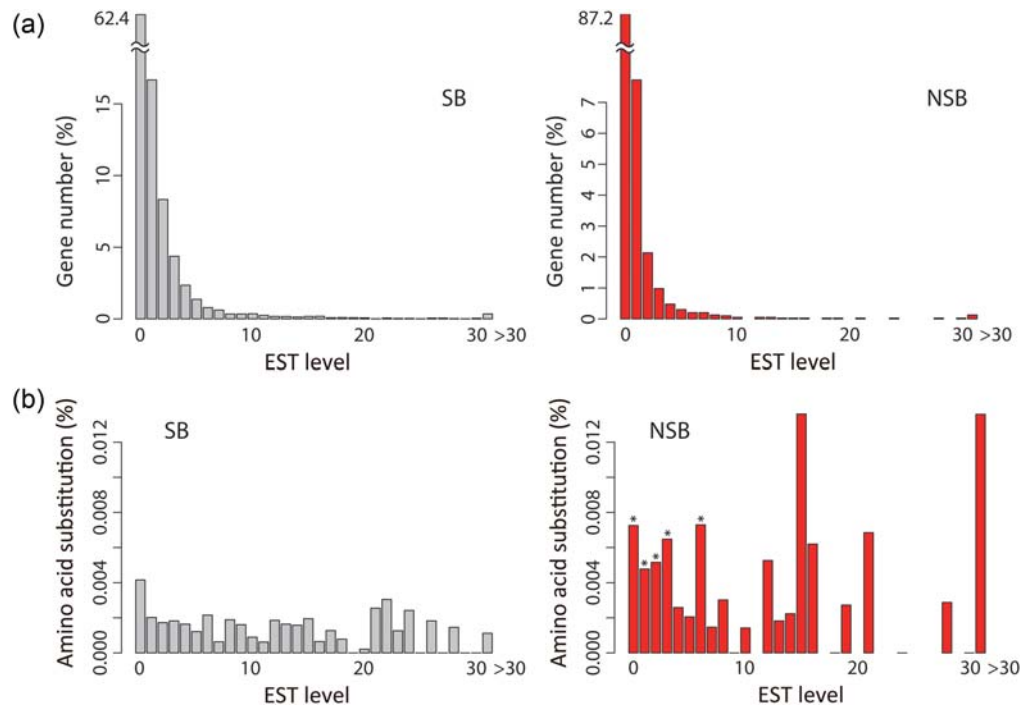
detected at EST levels of >10, indicating that several genes within NSBs are functional. As shown in Figure 5b, the average amino acid substitution ratios were significantly ( $P < 0.05$ ) higher for NSB genes than those of SB genes at any EST level, when greater than four genes were available for the analysis. This result demonstrates that the genes within NSBs accumulated more mutations than those within SBs, even those that exhibited high-expression levels.

The amino acid substitution frequency within the RIB326 genome was calculated to evaluate the activity of NSB genes. The number of genes with silent or non-silent mutations, and the ratio of mutated genes within SBs and NSBs is summarized in Table 2. Based on the amino acid frequency within *A. oryzae* genes, the probability of a synonymous mutation would only be 5.4% when a random point mutation mechanism is assumed (see the frequency within *A. oryzae* genes and probability of silent mutations for each amino acid in the Supplementary data). However, the ratio of genes with silent mutations was much more abundant (55–58%) than that of genes with non-silent mutations in both SBs and NSBs. This finding demonstrates that the NSB genes of strain RIB326 are functionally active, even though they accumulated more mutations than the SB genes.

Comparison of the sequences showed ~5.4% of RIB40 nucleotides unidentified by the reads from RIB326, among which continuous unidentified nucleotides of >500 bp occupy 3.3%. As a control, we confirmed that SOLiD short reads from RIB40 covered 99.94% of the RIB40 reference nucleotides. Further, long deletions (>30 kb) in RIB326 compared with RIB40 reached ~3.3%, which was identified by a preliminary comparative genomic hybridization experiment using a DNA microarray having single probes for each RIB40 gene (data not shown). The comparison of the sequences also showed 0.32% of the RIB40 nucleotides being substituted in the RIB326 genome. Notably, these frequencies were markedly reduced within CDS compared with the non-CDS regions. The ratio of deleted nucleotides was only 3.6% in exons, but reached 5.9 and 4.7% in the intergenic and intron regions, respectively. Similarly, the substitution ratio was 0.26% in exons, whereas intergenic and intron regions displayed ratios of 0.38 and 0.37%, respectively. These observations indicate the existence of selective pressure on CDS located within both NSBs and SBs.

### 3.3. What are the role of NSBs in regard to gene function?

Figure 6 depicts the gene numbers and mean amino acid substitution frequencies of SB and NSB genes for the major functional categories of the KOG of proteins.<sup>16</sup> The number of NSB genes belonging to meta-category I, which represents information storage and processing, was significantly smaller than that of SB genes (Fig. 6a). The number of NSB genes belonging to meta-category III, which includes secondary metabolism genes (Q), hydrolases, and transporters, was significantly greater than the other two meta-categories (Fig. 6a). This result is in



**Figure 5.** Gene number and averaged amino acid substitution ratios within SBs and NSBs of strain RIB326 against gene expression levels. (a) Gene numbers and (b) averaged amino acid substitution frequencies within SBs and NSBs are plotted against the level of EST. The *P*-values for the difference between the amino acid substitution ratios within SBs and NSBs were evaluated using the Wilcoxon signed-rank test for each EST level that included greater than four genes. Asterisks indicate a *P*-value of <0.05 in the graph of NSBs.

**Table 2.** Number and ratio of silent and non-silent mutations in SBs and NSBs of *A. oryzae* RIB326

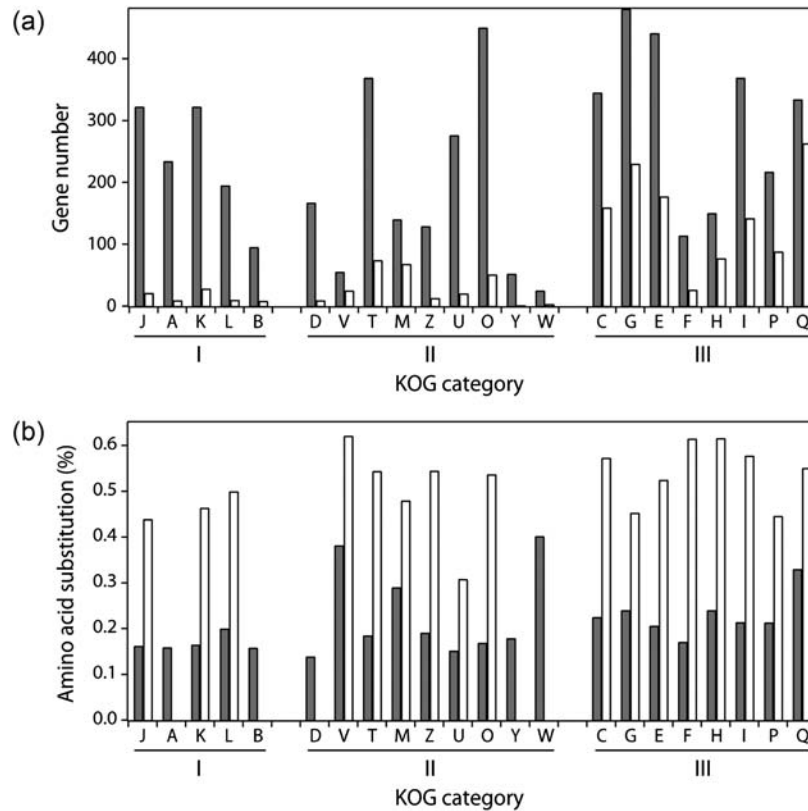
	Gene number			Total number of mutations		
	All	With silent mutation(s)	With non-silent mutation(s)	Silent	Non-silent	Silent/non-silent
SB	9345	4753	4431	14 229	10 180	1.4
NSB	3784	2114	2253	8328	7228	1.2

accordance with our previous finding that NSBs are the regions where secretory hydrolases and secondary metabolic genes unique to *A. oryzae*.<sup>5</sup> In addition, it was determined that the amino acid substitution frequency was higher within NSBs for all examined functional categories (Fig. 6b).

As secondary metabolic pathways are not directly involved in the growth, development, or reproduction of fungi, it is considered that mutations which alter or reduce the activity of these pathways do not markedly affect survival, whereas functional mutations in primary metabolic pathways are often lethal. Such mutations in secondary metabolic genes might increase survival rates through the production of new compounds that prevent the growth of other organisms. NSBs may serve as a diverse gene pool for promoting niche adaptation by accumulating genetic alterations.

It is presently unknown why NSBs are widely distributed over the chromosomes of *A. oryzae*. The

distributed regions might be suitable for secondary metabolic regulation by global regulatory factors such as LaeA and CclA, which function as histone methyltransferases.<sup>21,22</sup> Our genomic sequencing revealed that RIB326 possesses the biosynthetic genes for kojic acid (AO090113000136), containing two silent mutations, and those for penicillin (AO090038000544). However, although RIB40 is known to produce both kojic acid<sup>23</sup> and penicillin<sup>24</sup> RIB326 does not produce either compound. The inability of RIB326 to produce these two secondary metabolites is thought to be due to the lack of expression of the respective biosynthetic genes and/or lack of biosynthetic precursor(s). RIB326, which was selected from an industrial strain for fermentation, generally produces less extrolites than RIB40.<sup>7,8</sup> Mutations in RIB326 may have accumulated during the selective breeding of this strain for the reduced production of secondary metabolites during fermentation.



**Figure 6.** Mutation of SB and NSB genes based on KOG functional categories. (a) Gene numbers and (b) averaged amino acid substitution frequencies within SBs and NSBs are plotted for each KOG. The gray and white bars indicate the values for SBs and NSBs, respectively. Each KOG category was classified into three meta-categories: I, information storage and processing (J to B); II, cellular process and signalling (D to W); and III, metabolism (C to Q), as described in the Materials and methods section. In (b), the values for categories that did not include  $\leq 10$  genes were not drawn to avoid misinterpretation of biological insignificance.

Among extensive comparative studies between *A. oryzae* strains examining the productivity of hydrolases, amylases, proteases,<sup>25–27</sup> and biosynthetic homologues of cyclopiazonic acid<sup>28,29</sup> and aflatoxin,<sup>7,8,30,31</sup> this represented the first genome-wide comparison between *A. oryzae* strains. It is possible that the higher mutation frequency of NSBs, which are enriched for secretory hydrolases and secondary metabolic genes, contributes to the introduction of unique characteristics into *A. oryzae* strains for niche adaptation. From an industrial point of view, the wide diversity of *A. oryzae* strains, which may have resulted from the high-mutation frequency of NSBs, might allow for effective breeding for specific fermentation applications by introducing unique metabolism for raw material degradation and/or flavour improvement.

**Supplementary Data:** Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

## Funding

This work was supported in part by the Okinawa Cutting-Edge Genome Project.

## References

1. Machida, M., Asai, K., Sano, M., et al. 2005, Genome sequencing and analysis of *Aspergillus oryzae*, *Nature*, **438**, 1157–61.
2. Galagan, J.E., Calvo, S.E., Cuomo, C., et al. 2005, Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*, *Nature*, **438**, 1105–15.
3. Nierman, W.C., Pain, A., Anderson, M.J., et al. 2005, Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*, *Nature*, **438**, 1151–6.
4. Machida, M., Terabayashi, Y., Sano, M., et al. 2008, Genomics of industrial aspergilli and comparison with toxigenic relatives, *Food Addit. Contam.*, **25**, 1147–51.
5. Tamano, K., Sano, M., Yamane, N., et al. 2008, Transcriptional regulation of genes on the non-syntenic blocks of *Aspergillus oryzae* and its functional relationship to solid-state cultivation, *Fungal Genet. Biol.*, **45**, 139–51.
6. Machida, M., Yamada, O. and Gomi, K. 2008, Genomics of *Aspergillus oryzae*: learning from the history of Koji mold and exploration of its future, *DNA Res.*, **15**, 173–83.
7. Kiyota, T., Hamada, R., Sakamoto, K., Iwashita, K., Yamada, O. and Mikami, S. 2011, Aflatoxin non-



- productivity of *Aspergillus oryzae* caused by loss of function in the aflJ gene product, *J. Biosci. Bioeng.*, **111**, 512–7.
8. Tominaga, M., Lee, Y.H., Hayashi, R., et al. 2006, Molecular analysis of an inactive aflatoxin biosynthesis gene cluster in *Aspergillus oryzae* RIB strains, *Appl. Environ. Microbiol.*, **72**, 484–90.
  9. Zerbino, D.R. and Birney, E. 2008, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.*, **18**, 821–9.
  10. Kielbasa, S.M., Wan, R., Sato, K., Horton, P. and Frith, M.C. 2011, Adaptive seeds tame genomic sequence comparison, *Genome Res.*, **21**, 487–93.
  11. Frith, M.C., Wan, R. and Horton, P. 2010, Incorporating sequence quality data into alignment improves DNA read mapping, *Nucleic Acids Res.*, **38**, e100.
  12. Frith, M.C., Hamada, M. and Horton, P. 2010, Parameters for accurate genome alignment, *BMC Bioinformatics*, **11**, 80.
  13. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403–10.
  14. States, D.J. and Gish, W. 1994, Combined use of sequence similarity and codon bias for coding region identification, *J. Comput. Biol.*, **1**, 39–50.
  15. Akao, T., Sano, M., Yamada, O., et al. 2007, Analysis of expressed sequence tags from the fungus *Aspergillus oryzae* cultured under different conditions. *DNA Res.*, **14**, 47–57.
  16. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., et al. 2003, The COG database: an updated version includes eukaryotes, *BMC Bioinformatics*, **4**, 41.
  17. Murnane, J.P. 2006, Telomeres and chromosome instability, *DNA Repair (Amst.)*, **5**, 1082–92.
  18. Borneman, A.R., Forgan, A.H., Pretorius, I.S. and Chambers, P.J. 2008, Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain, *FEMS Yeast Res.*, **8**, 1185–95.
  19. Carreto, L., Eiriz, M.F., Gomes, A.C., Pereira, P.M., Schuller, D. and Santos, M.A. 2008, Comparative genomics of wild type yeast strains unveils important genome diversity, *BMC Genomics*, **9**, 524.
  20. Koch, R., van Luenen, H.G., van der Horst, M., Thijssen, K.L. and Plasterk, R.H. 2000, Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*, *Genome Res.*, **10**, 1690–6.
  21. Bok, J.W. and Keller, N.P. 2004, LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot. Cell*, **3**, 527–35.
  22. Bok, J.W., Chiang, Y.M., Szewczyk, E., et al. 2009, Chromatin-level regulation of biosynthetic gene clusters, *Nat. Chem. Biol.*, **5**, 462–4.
  23. Terabayashi, Y., Sano, M., Yamane, N., et al. 2010, Identification and characterization of genes responsible for biosynthesis of kojic acid, an industrially important compound from *Aspergillus oryzae*, fungal. *Genet. Biol.*, **47**, 953–61.
  24. Marui, J., Yamane, N., Ohashi-Kunihiro, S., et al. 2011, Kojic acid biosynthesis in *Aspergillus oryzae* is regulated by a Zn(II)(2)Cys(6) transcriptional activator and induced by kojic acid at the transcriptional level, *J. Biosci. Bioeng.*, **112**, 40–3.
  25. Zhu, L.Y., Nguyen, C.H., Sato, T. and Takeuchi, M. 2004, Analysis of secreted proteins during conidial germination of *Aspergillus oryzae* RIB40, *Biosci. Biotechnol. Biochem.*, **68**, 2607–12.
  26. te Biesebeke, R., Record, E., van Biezen, N., et al. 2005, Branching mutants of *Aspergillus oryzae* with improved amylase and protease production on solid substrates, *Appl. Microbiol. Biotechnol.*, **69**, 44–50.
  27. Bocking, S.P., Wiebe, M.G., Robson, G.D., Hansen, K., Christiansen, L.H. and Trinci, A.P. 1999, Effect of branch frequency in *Aspergillus oryzae* on protein secretion and culture viscosity, *Biotechnol. Bioeng.*, **65**, 638–48.
  28. Chang, P.K., Ehrlich, K.C. and Fujii, I. 2009, Cyclopiazonic acid biosynthesis of *Aspergillus flavus* and *Aspergillus oryzae*, *Toxins (Basel)*, **1**, 74–99.
  29. Kato, N., Tokuoka, M., Shinohara, Y., et al. 2011, Genetic safeguard against mycotoxin cyclopiazonic acid production in *Aspergillus oryzae*, *Chembiochem.*, **12**, 1376–82.
  30. Kusumoto, K., Nogata, Y. and Ohta, H. 2000, Directed deletions in the aflatoxin biosynthesis gene homolog cluster of *Aspergillus oryzae*, *Curr. Genet.*, **37**, 104–11.
  31. Watson, A.J., Fuller, L.J., Jeenes, D.J. and Archer, D.B. 1999, Homologs of aflatoxin biosynthesis genes and sequence of aflR in *Aspergillus oryzae* and *Aspergillus sojae*, *Appl. Environ. Microbiol.*, **65**, 307–10.