

論文 / 著書情報
Article / Book Information

題目(和文)	麹菌Aspergillus oryzae産業用近縁株の大規模比較 -家畜化とゲノム進化-
Title(English)	
著者(和文)	渡来直生
Author(English)	Naoki Watarai
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11398号, 授与年月日:2020年3月26日, 学位の種別:課程博士, 審査員:山田 拓司,伊藤 武彦,北尾 彰朗,本郷 裕一,平沢 敬
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11398号, Conferred date:2020/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

令和元年度

博士論文

麹菌 *Aspergillus oryzae* 産業用近縁株の大規模比較
—家畜化とゲノム進化—

東京工業大学

生命理工学院 生命理工学系

渡来 直生

指導教員

山田 拓司 准教授

第1章 序論.....	4
1.1 麹菌の産業利用の現状.....	4
1.2 <i>Aspergillus oryzae</i> のゲノム解析の歴史.....	4
1.3 <i>Aspergillus oryzae</i> の有性生殖.....	5
1.4 系統推定について.....	6
1.5 本研究の目的及び構成.....	7
第2章 <i>Aspergillus oryzae</i> のゲノム多様性.....	8
2.1 使用株.....	8
2.2 ドラフトゲノム作成及び遺伝子予測.....	8
方法.....	8
結果.....	9
2.3 遺伝子のオルソログクラスタリング.....	10
方法.....	10
結果.....	11
2.4 全ゲノム系統樹.....	11
方法.....	11
結果.....	11
2.5 ミトコンドリアゲノム系統樹.....	13
方法.....	13
結果.....	14
2.6 アフラトキシン生合成遺伝子クラスター系統樹.....	15
方法.....	15
結果.....	16
2.7 ゲノム交雑の可視化.....	17
方法.....	18
結果.....	18
第3章 家畜化が <i>Aspergillus oryzae</i> のゲノム進化に及ぼした影響.....	21
3.1 仮説提唱.....	21
3.2 クレード内外遺伝子変異圧解析.....	22

方法	22
結果	23
3.3 クレード内遺伝子欠損・重複解析	26
方法	26
結果	27
3.4 グループ特異的遺伝子	28
方法	28
結果	29
3.5 機能及び系統予測のための NRIB ジェノタイピングとの比較	32
方法	32
結果	33
3.6 クレード内系統樹とゲノムの変化	35
方法	35
結果	36
第4章 結論	42
参考文献	43
付録	50
謝辞	73

第1章 序論

1.1 麹菌の産業利用の現状

麹菌は、アミラーゼやプロテアーゼの分解酵素活性が非常に強く、主に東アジアで発酵食品の製造に用いられる産業的に重要なカビである(Machida et al. 2008)。中国で3,000年ほど前に真菌の産業利用が始まり、日本では室町時代の頃から全国に点在する種麹屋が特徴の異なるさまざまな種麹を管理・販売するようになった。種麹とは麹のもとになるタネのことを指し、現代では多くが単離株であるが、旧くは米にさまざまな真菌が生えた状態から木灰による集積が行われたものであった。大正時代に現株式会社秋田今野商店によって純培養が確立されるまで、種麹は純粋株ではない状態で継代されてきたと推測される。実際に本研究で扱ったサンプルの中には古くからの状態を受け継いでいるものもあり、ゲノム解析の結果2株が混合したものも存在した(解析からは取り除いた)。

現代の種麹屋は言わば菌株保存機関のような役割を果たしており、種麹から麹を作る製麹と呼ばれる過程は、麹屋(製麹業者)や、発酵食品メーカーが自社で行っている。種麹は酵素力価や胞子の色、褐色性などが多様化し、それぞれの種麹屋において識別され管理されている。実際には世間の需要に合わせて一部の株のみが商品として採用されており、また販売される商品はいくつかの純粋株を組み合わせたものである場合が多い。

麹菌として代表的な種が *Aspergillus oryzae* であり、日本では酒、醤油、味噌、甘酒、味醂などさまざまな発酵食品製造に用いられる。他にも醤油醸造などに用いられる種として *Aspergillus sojae* や *Aspergillus tamarii*、焼酎製造などに用いられる種として *Aspergillus luchuensis* などがあるが、種麹屋が保有する株としては *A. oryzae* が群を抜いて多様である。

1.2 *Aspergillus oryzae* のゲノム解析の歴史

A. oryzae は同一種内で発酵特性に多様性があることは知られているが、それぞれの株に対する全ゲノム解析はほとんど行われておらず、発酵特性と遺伝的要因との関連性はほとんど明らかになっていない。2005年、*A. oryzae* RIB40株の全ゲノムが解読され(Machida et al. 2005)、これが現状唯一の *A. oryzae* の完全ゲノムとなっている。なお、RIB40はアカデミアにおいて野生株として扱われることが多いが、1950年に京都のマルカ醤油株式会社のソラマメから採取された株で、野生に近いと判断されゲノム解読に供されたものであり(坂本 2006)、完全に野生株であるとは言い難い。当時すでに全ゲノムが解読されていた *Aspergillus nidulans*, *Aspergillus fumigatus* との比較ゲノム解析によって、*A. oryzae* はそれらと比べて7-9Mb程度ゲノムが増加していることがわかり(Galagan et al. 2005; Machida et al. 2005)、細菌からの水平伝播によって遺伝子を獲得していることが示唆された(Machida et al. 2005)。属内の共通領域を Synteny Block (SB)、非共通領域(新たに獲得された領域)を Non-Synteny Block (NSB) として区別されることがある。NSBの遺伝子は通常はほとんど発現しておらず(Kobayashi et al. 2007)、発酵に大きく関与しない遺伝子群についてはその機能のほとんどが未知のままである。

また、*A. oryzae* には近い近縁種として *Aspergillus flavus* があり、これらの間の比較解析がよく行われている。*A. flavus* は一部の株がアフラトキシンやアフラトレムといったカビ毒を生成することから(Kusumoto et al. 2000; Klich 2007)、産業的、食品衛生的に重要な種である。歴史的に *A. flavus* と *A. oryzae*

は形態的な差や毒性の有無によって区別されてきたが(Klich 2007; Jørgensen 2007)、*A. oryzae* と *A. flavus* は遺伝的に非常に近縁であり、これらのゲノムは 99.5%の相同性がある(Rokas et al. 2007)。*A. oryzae* のいくつかの株はアフラトキシン生合成遺伝子クラスターのすべてまたは一部を有しているが、すべて非毒性であることがわかっている(Kusumoto et al. 2000)。

このような背景から、*A. oryzae* は *A. flavus* を人間が家畜したことによって無毒化、分化した種であるとする説がある(Geiser et al. 1998; Machida et al. 2008; Rokas 2009)。11 個の遺伝子の系統解析(Geiser et al. 1998)、アフラトキシン生合成遺伝子クラスターの比較解析(Geiser et al. 2000; Chang et al. 2006)では、*A. flavus* の一部の系統 (L-type) から分化した単一系統の種であることが示唆された。また、全ゲノムの SNP 解析によって得られた系統樹上でも、*A. oryzae* は *A. flavus* の一つのクレードから派生した単一系統であることが示されている(Gibbons et al. 2012)。

また、これらの先行研究について、本研究ではゲノム解析の面から検証を行っており、詳しくは各論で紹介する。

1.3 *Aspergillus oryzae* の有性生殖

A. oryzae 及び *A. flavus* は長らく無性種 (有性世代をもたず、無性生殖のみによって増殖する種) であると考えられてきたが(Geiser et al. 1996)、近年、*A. flavus* において、実験室および自然環境での人為的な有性生殖 (交雑) が可能であること確認された(Horn et al. 2009; Horn et al. 2014; Horn et al. 2016)。ただし、マイクロサテライトによる集団解析の結果では、自然環境では交雑はほとんど起こっていないことが報告されている(Grubisha and Cotty 2009; Islam et al. 2018)。ゲノム解析からも有性生殖に必要な遺伝子群をほぼ完全に有していることがわかっており(Galagan et al. 2005; Machida et al. 2005)、特に *A. oryzae* 及び *A. flavus* のすべての株は、ゲノム中に一つの交配型 (MAT 型) 座位を有し、MAT1-1 または MAT1-2 のいずれか一方をもっていることが示されている(Ramirez-Prado et al. 2008; Wada et al. 2014)。

また、*A. flavus* においては、MAT 型とは別の自他認識システムとして、少なくとも 13 種の Vegetative Compatibility Group (VCG) が存在することが知られている(Bayman and Cotty 1991)。VCG は細胞融合可能な株群のことで、異なる VCG の株が細胞融合するとプログラム細胞死に至る(Glass and Dementhon 2006)。細胞融合によって細胞死が引き起こされる性質は Vegetative Incompatibility (VI) や Heterokaryon Incompatibility (HI) と呼ばれ、*Neurospora crassa*, *Podospora anserina*, *Cryphonectria parasitica* では HET domain をもつ遺伝子 (以下、HET 遺伝子) によってコントロールされている(Daskalov et al. 2016; Paoletti 2016)。しかし、*A. oryzae* RIB40 において HET 遺伝子の存在は 60 以上確認されており、*A. flavus* においても同程度の数が存在すると考えられ、VCG と HET 遺伝子の関連性は明らかになっていない。

一方で、*A. oryzae* については、まだ完全な有性生殖は確認されておらず、*A. flavus* に見られるような明確な VCG も見つかっていない。なぜなら、細胞融合が可能であることは 60 年以上前に確認されているが(Ishitani and Sakaguchi 1956)、*A. oryzae* は細胞融合が著しく乏しく(Hara et al. 2012; Wada et al. 2014; Okabe et al. 2018)、VCG を定義するための VI の観察が困難だからである。新たな技術の導入に

よって、細胞融合能をもついくつかの株間で VI を起こすことが観測されたが (Okabe et al. 2018; Mori et al. 2019)、遺伝的な要因との関連性は完全に明らかになっていない。また、*A. flavus* で有性生殖に必要とされている菌核という組織の形成が *A. oryzae* では起こりにくいため (Chang et al. 2006)、直接有性生殖を実証することも困難となっている。

ゲノム情報解析からのアプローチはさらに乏しく、MAT 型と 1 つの遺伝子系統樹の連鎖不平衡から、祖先株の間で交雑が起きていたことが示唆されているに留まる (Chang and Ehrlich 2010)。*A. oryzae* の MAT1-1 及び MAT1-2 株間の系統関係も明らかになっていない。

1.4 系統推定について

本研究では、全ゲノム系統樹を、株の真の進化系統を求めるために推定している。系統樹の推定方法は旧くよりさまざまな方法が用いられてきたが、現代ではゲノミクスの台頭により、形態学による分類から分子系統を用いる方法が最も信頼性が高いとして覇権を得ている。コウジカビにおいても、例えば、焼酎用麹である *A. luchuensis* は、旧来は *A. kawachii*, *A. awamori* や *A. saitoi*, *A. usamii* など分離源や多少の形態差によって呼称がさまざまであったが、現在では学术界においては *A. luchuensis* に統一されている。*A. oryzae* もその俎上に置かれた歴史があり、*A. flavus* との差に注目されたことがあったが、本論 1.2 で述べた通り、分子系統的アプローチより単系統であることが示されたため、これらの分類群は継続して用いられている。

分子系統からの種系統推定においては、gene tree (遺伝子系統樹) と species tree (種系統樹) の関係性についてはよく議論されている (Szöllösi et al. 2015)。種は、それがもつゲノム、すなわち遺伝子群とともに進化しているため、単一の遺伝子の系統樹と種の進化系統は強く関連する。一方で、異なる種からの遺伝子水平伝播や、同一ゲノム内での遺伝子重複、対立遺伝子のような集団中に系統の異なる遺伝子が存在することによって生じる Incomplete Lineage Sorting (ILS, deep coalescence, 種の系統と遺伝子の系統の不整合) の問題があることから、複数の遺伝子系統樹をうまく組み合わせて種の系統を推定する必要があった。

また、遺伝子配列のアラインメント情報を用いた系統推定に関しても、近隣結合法、最節約法、非荷重結合法、最尤法などさまざまな方法が提案されてきた。複数の遺伝子系統情報を統括する方法として、近年では、比較したい系統群に存在する共通のシングルコピー遺伝子または Single Nucleotide Polymorphism (SNP) を結合して最尤法で推定するという手法が、統計的に不確実であるという理論的な指摘はありつつも (Roch and Steel 2015)、最も簡略で妥当性のある系統樹を返すために、近縁ゲノムの比較において用いられている (Bennett et al. 2007; Merker et al. 2015)。原核生物ではこの方法で、かつ進化速度を考慮した系統分類の再評価が行われ、データベース整備が進んでいる (Chaumeil et al. 2019)。

上述の手法は、基本的には種、すなわち特に自然に有性生殖の起こらない個体群の分類に対して提案されてきたものであるが、本研究ではこれを種内の近縁な株間の分類にも用いている。種内で交雑が起きているとすれば、この手法で推定される系統樹は正確には系統樹とは言えない (ノードが分化を意味していない)。一方で、遺伝的な多様性を俯瞰するには非常に効果的な手法であるため、株群の真の進化系統を推定するための土台として用いている。

1.5 本研究の目的及び構成

本研究では日本の種麴屋から産業用株の全ゲノム解析を通じて、先行研究で提唱された *A. oryzae* の進化にまつわる諸説をゲノム解析の面から再検証し、特に人間による家畜化と進化の関係を解明することを目的とする。

第2章では、新たに取得された5社82株の産業用株の全ゲノムシーケンスからドラフトゲノムを作成し、*A. oryzae* 及び *A. flavus* の公開ゲノムデータを含めた系統樹を構築することで、産業用途とゲノムレベルの多様性の関係を明らかにした。また、得られた結果の考察から、*A. oryzae* の祖先株間で交雑が起こっていたと仮定し、ゲノム座位ごとの相同性の可視化を行い、交雑株の組み合わせ推定を試みた。

第3章では、先行研究及び第2章の解析結果を踏まえて、「人間による家畜化は *A. oryzae* の交雑を引き起こさず、発酵特性を多様化させるゲノム変化を引き起こす」という新たな仮説を提唱した。仮説検証として、*A. oryzae* のクレード内外の変異蓄積解析及びゲノム重複・欠損の解析を行ったところ、家畜化がゲノム全体における非同義変異やフレームシフト変異の集中を引き起こしていることを示唆する結果を得た。また、各クレード内において産業用途における特性と具体的な遺伝子の関係について個別の解析を行った。

第2章 *Aspergillus oryzae* のゲノム多様性

2.1 使用株

本研究で使用した *A. oryzae*, *A. sojae* の計 85 株は東京工業大学ぐるなび食の価値創成共同研究講座において種麹屋から収集され (Table 1, S1)、シーケンシングは国立遺伝学研究所によって HiSeq2000 を用いて行われた。

Table 1 本研究における種麹屋ごとの全ゲノム解読株数

種麹屋	都道府県	ID	<i>A. oryzae</i>	<i>A. sojae</i>
(株)菱六	京都府	P1	11	0
(株)ビオック	愛知県	P2	19	1
(株)樋口松之助商店	大阪府	P3	20	0
(株)秋田今野商店	秋田県	P4	15	0
日本醸造工業(株)	東京都/茨城県	P5	17	2

2.2 ドラフトゲノム作成及び遺伝子予測

方法

HiSeq Paired-end リードを Platanus_trim v1.0.7(Kajitani et al. 2014)によってトリムし、Platanus v1.2.4 を用いて assemble から gap-closure までを行った。細菌ゲノムの混入を取り除くため、1,000 bp 未満の contig または scaffold 取り除き、LAST v869(Kielbasa et al. 2011)によって RIB40 のコンプリートゲノム (GCA_000184455.3) に対し残りの配列のアラインメントを行った。そのうち RIB40 ゲノムに対して 1 箇所もヒットがなかった配列について、さらに RefSeq DB release 89(Pruitt et al. 2007)に対して blastn v2.2.8(Altschul et al. 1990)を用いてアラインメントを行い(-max_target_seqs 1 -evalue 0.000001)、細菌ゲノムにヒットしたものを取り除いた。

その後、85 株のドラフトゲノムに NCBI Assembly (Kitts et al. 2016)からダウンロードした、*A. oryzae*, *A. flavus*, *A. sojae*, *Aspergillus parasiticus* の計 67 株を加え、これら 152 のドラフトゲノム配列に対し、以下の遺伝子予測を行った。比較ゲノムに際し、同じ遺伝子予測パイプラインを用いる必要があると考えたため、公開されているゲノムに対するそれぞれのアノテーションは用いなかった。なお、*A. parasiticus* は *A. sojae* に非常に近縁な毒素産生種であり、本研究では *A. parasiticus* 及び *A. sojae* を、*A. oryzae* 及び *A. flavus* に対するアウトグループとして主に扱った。また、“*A. oryzae* ATCC12892” (GCA_002894705.1, (Deng et al. 2018)) は本研究の結果 *A. sojae* に非常に近縁であることが分かったため、*A. sojae* として扱った (Table S2)。

遺伝子領域は、以下に示すように *ab initio* 予測と参照配列のマッピングによる予測の 2 つを行い、結果をマージして得た。*ab initio* 予測は GeneMark-ES v4.32 (Lukashin and Borodovsky 1998) (gmes_petap.pl --ES --fungus) と AUGUSTUS v3.3 (Stanke et al. 2006) (--species=aspergillus_oryzae) の 2 つを用いた。参照配列は *A. oryzae* RIB40 s01-m08-r29, *A. niger* CBS_513.88 s01-m07-r04, *A. flavus* NRRL_3357 のアミノ酸配列を *Aspergillus* Genome Database (AspGD, <http://www.aspgd.org> (Arnaud et al. 2012)) よりダウンロードし、GMAP v2017-11-15 (Wu and Watanabe 2005) を用いてマッピングによる予測を行った。これらの遺伝子予測結果を EVIDENCEModeler v1.1.1 (Stanke et al. 2006) (優先度は、“*A. oryzae* RIB40 s01-m08-r29”: 10, “*A. niger* CBS 513.88 s01-m07-r04” and “*A. flavus* NRRL_3357”: 5, “AUGUSTUS”: 2, “GeneMark-ES”: 1) によってマージした。遺伝子領域予測結果から、開始コドンが欠損しているかフレームシフトしている配列を取り除き、各サンプルに対する遺伝子配列を得た。これらの遺伝子配列を用い、BUSCO v3.0.1 (Simão et al. 2015) (eurotiomycetes_odb9 の 4,046 コア遺伝子) によって、ゲノムの完全度を評価した。

結果

アセンブルによって得られた scaffold (>1,000bp) の統計値に関して、*A. oryzae* のコンティグの数は 36–256 で、合計長は 35.9–38.66 Mb であった。*A. sojae* のコンティグ数は 45–71 で、total length は 40.0–40.1 Mb であった。2 種のドラフトゲノムサイズはこれまでの報告と大きく異なることはなかった (Machida et al. 2005; Sato et al. 2011)。全結果の詳細を Table S2 に示す。すべてのサンプルについて depth が 100 以上になるようにシーケンスしたが、run3 の一部及び run4, run5 のサンプルについては、ゲノム抽出にヤタラーゼを使用した影響で、ヤタラーゼの生産に用いられた Actinobacteria (*Corynebacterium*) のゲノムが混入していたため、これらを取り除いた結果、最終的な contig の平均 coverage が比較的少なくなった (一部の株は細胞壁が固く、DNA 抽出効率が悪かったため細胞壁分解酵素であるヤタラーゼを使用した結果として収量が悪くなってしまったため使用は控えるべきである)。なお、クオリティの悪いサンプルについては、scaffold に対しリードをマッピングすることによって単一ゲノム由来かどうかを Integrated Genome Viewer (Robinson et al. 2011) を用いて目視にて確認した。その結果、1 つのサンプルで異なる 2 株以上の混合が認められたため、本研究のすべての記述から除外した (Table 1, S1, S2 からもこのサンプルを除外してある)。

遺伝子予測の結果、*A. oryzae* 及び *A. sojae* に予測された遺伝子数はそれぞれ 11,196–11,716 及び 13,309–13,317 であり、先行研究との大きな乖離はなかった。また、BUSCO の評価スコアは 94.6–98.8% だった。RIB40 のコンプライートゲノムに対して、本研究のパイプラインを用いたスコアは 97.6% であった (なお、参照配列 s01-m09-r06 のスコアは 92.7% であり、本研究の遺伝子予測パイプラインを RIB40 ゲノム配列に適用した結果は BUSCO のスコア上では精度が高いことを確認した)。BUSCO は近縁のゲノムに存在するコア遺伝子 (共通のシングルコピー遺伝子) がどれだけ存在するかによってゲノムの完全度を測るツールであり、本研究において参照配列は eurotiomycetes_odb9 を用いたが、*A. oryzae* においてはゲノムが完全であっても 100% になることはないようである。

本研究で用いたサンプルの一つである *A. oryzae* TK-54 株は、アセンブルのクオリティも高く異種混入も認められなかったものの、BUSCO スコアが 94.6% であり、すなわち特異的に遺伝子欠損が多かった。

このような数%に及ぶ遺伝子欠損は他の 84 株には認められなかった。種麴屋に対する聞き取り調査によると TK-54 は最も古い単離株であり、現代まで保管する間に遺伝子欠損が蓄積したと考えられる。

2.3 遺伝子のオルソログクラスタリング

方法

152 サンプルの遺伝子予測の結果及び参照配列 2 種 (AspGD よりダウンロードした *A. oryzae* RIB40 s01-m08-r29, s01-m09-r06) すべてに対して、OrthoFinder v2.2.6 (Emms and Kelly 2015) (diamond v0.9.22 (Buchfink et al. 2015)をアラインメントソフトウェアとして指定) を用いてオルソログクラスタリングを行った。ここで、シングルコピー-OG (以下、SCG。各サンプルについて多くとも 1 個が存在するオルソログ遺伝子グループ) について、遺伝子領域及びその上流・下流領域の DNA 配列が完全に同じであっても、遺伝子予測ソフトウェアのランダムネスによってエクソン予測部位が変化することを確認した。また、本来のエクソン内にフレームシフト変異やナンセンス変異が入っていた場合、それを避けて誤ったエクソン領域が予測されることがあったため、BUSCO スコアが 96%以上だった 101 サンプルを用いて、各オルソロググループ (OG) に対して以下の方法でコンセンサス coding region (Fig. 1 (左), コンセンサス CDR)を推定し、代表配列を抽出した。

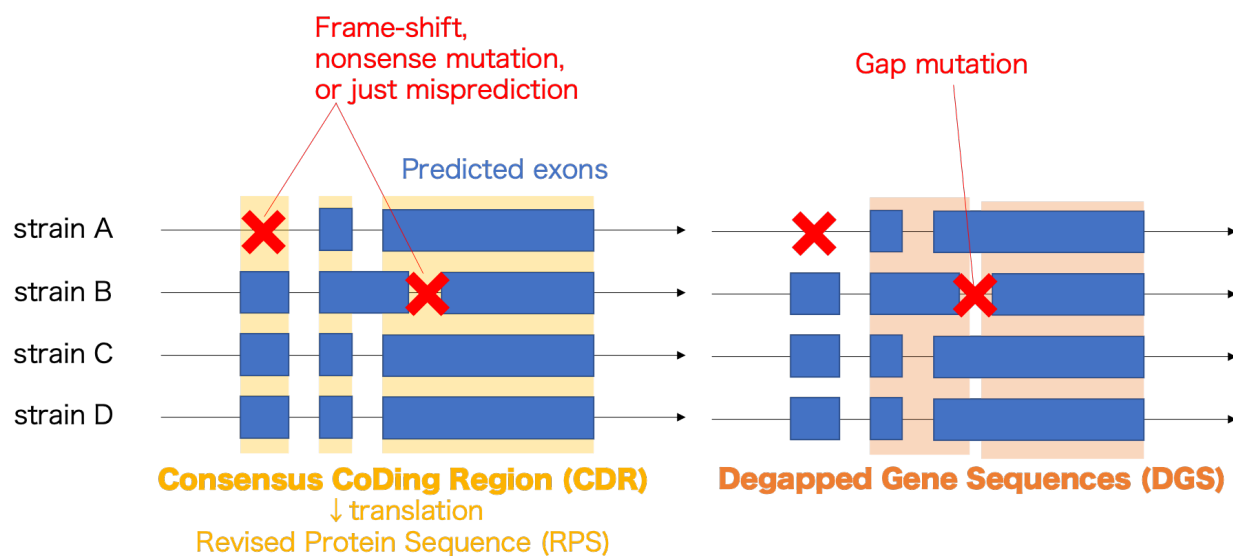


Fig. 1 本研究で定義したコンセンサス CDR, RPS 及び DGS の概略図

まず、各 OG に含まれる各遺伝子について、元のゲノムの遺伝子アノテーションファイルを参照してエクソンとイントロンを結合した配列を再構成し、これらの配列のマルチプルアラインメントを MAFFT v7.2.2.2 (Katoh and Standley 2013) を用いて作成した。次に、その中で最も多く出現する予測エクソンの組み合わせをコンセンサス CDR として、すべての配列について適用し、補正翻訳アミノ酸配列 (Revised Protein Sequence, RPS) を得た (Fig. 1 (左))。

各 OG のアミノ酸配列から、代表配列に選ばれた CDR をもつ配列のうちの 1 つを代表配列として抽出し、InterProScan v5.28-67.0(Quevillon et al. 2005)及び GhostKOALA v2.2(Kanehisa et al. 2016)を用いて機能アノテーションを行った。

結果

全 1,882,788 遺伝子のアミノ酸配列に対する OrthoFinder の結果から 15,614 個の OG が形成され、クラスタリングされずに残った遺伝子は 1,395 個あった。また、152 サンプルかつ参照配列 2 セットに共通して存在する SCG は 3,951 個あった。しかし、これはクオリティの低いサンプルを含んでいるため、これらを除くため BUSCO スコアが 96%以上のサンプルに限ったところ、SCG は 9,012 個あった。よって、*A. oryzae* における本来のコア遺伝子は約 12,000 の全遺伝子の 3/4 程度に相当すると考えられる。

15,614 OG のうち、BUSCO のスコアが 96%以上のサンプルに存在していた 14,710 個の OG について、それぞれ RPS を取得した。そのうち InterProScan によってドメイン情報が付与できたのは 90.3%にあたる 13,289 OG であり、GhostKOALA によって KO がアサインできたのは 30.4%にあたる 4,466 OG だった。また、13,061/15,614 OG に対して、先行研究(Machida et al. 2005)で定義された“AO”で始まる RIB40 の遺伝子座位 ID を付与することができた。

2.4 全ゲノム系統樹

方法

全ゲノム系統樹は、152 サンプルかつ参照配列 2 セットに共通して存在する SCG を用いて作成した。それぞれの OG のエクソン及びイントロンを結合した配列に対してマルチプルアラインメントを作成し、gap 領域を取り除いた (Fig. 1(右), Degapped Gene Sequence (DGS))。各 OG の DGS をそれぞれ MAFFT v7.2.2.2 でアラインメントした後、直列結合し、RAxML v8.2.11(Stamatakis 2014) (-m GTRGAMMAI -N 100) を用いて系統樹推定を行った。生成された Newick ファイルは FigTree v1.4.3 (<https://github.com/rambaut/figtree>)を用いて可視化し、マニュアルで reroot した後、アウトグループの *A. sojiae*, *A. parasiticus* を取り除いた。

また、各ドラフトゲノムに対する MAT 型を、先行研究で提案されている PCR プライマー配列(Wada et al. 2012)を Bowtie2 v2.2.6(Langmead and Salzberg 2012)を用いて直接マッピングすることで決定した(--sensitive-local --no-discordant --no-mixed -X 10000 --end-to-end)。

結果

共通して存在する 3,951 SCG の DGS を直列につなげた結果、5,677,852 bp (ゲノム全長の約 15%)であった。推定した系統樹を Fig. 2、アウトグループ(*A. sojiae* 及び *A. parasiticus*)を含めた系統樹を Fig. S1 に示す。

先行研究に示されているように(Geiser et al. 1998; Geiser et al. 2000; Chang et al. 2006; Gibbons et al.

2012)、*A. oryzae* は *A. flavus* にネストした単系統となった。また、日本の産業用株はいくつかのクレードを形成した (A-H, Table S2)。以下、「クレード」とは、単一の祖先株の変異またはゲノム内 rearrangement の蓄積のみで分化したことが明確である clonal な株群のことを指す。BUSCO スコア 96% の株では、クレード内の DGS の類似性は 0.01% に収まっていた。クレードごとに「酒・味噌用」または「醤油用」と用途が分類された。

いくつかの *A. flavus* とされていた株 (WRRL1519, NRRL35739, IFM54693, IFM57535, IFM59975, IFM60655, 2017 Washington T4) は、全ゲノム系統樹上で *A. oryzae* のグループに属しており、アフラトキシン生合成遺伝子クラスター(本論 2.6)や、先行研究で記述されている毒性などの形態から、本研究では *A. oryzae* であると定義した。このうち、例えば NRRL35739 や WRRL1519 は、*A. flavus* よりも生育が貧弱ながら非アフラトキシン産生の *A. flavus* として注目されていた株である (Pennerman et al. 2019)。これらは自然環境から単離した野生株や臨床分離株であるがゆえに *A. flavus* とされていたようであり、逆に言えば、産業用途でない *A. oryzae* は自然環境に少なからず存在することを意味している。*A. oryzae* の名は稲 (*Oryza*) から単離されたことに由来するが、本研究で扱ったサンプルの中には、日本の医療現場で喀痰から得られた株 (IFM 株 (Toyotome et al. 2019)) や、アメリカのミシシッピ州の落花生畑から得られた株 (2017 Washington T4 (Weaver and Abbas 2019)) も存在した。しかも、例えば IFM60655 は産業用株と同じクレード E に属する結果となり、野生/非野生は、非常に細かな遺伝子の違いでないと区別できないか、あるいは全く区別できない可能性があると考えられる。中国の産業用株 (AS3.951, 100.8, 3.042 (Zhao et al. 2012; Zhao et al. 2014)) や韓国の産業用株 (BP-2, SRCM 株) は、日本の産業用株 (TK-29 以外) とは系統的に離れたグループを形成し、ここには野生株も多く含まれていた。

MAT 型は、152 株のうち NCBI Assembly よりダウンロードした 6 株を除く 146 株についてアサインでき、クレードごとに MAT1-1 及び MAT1-2 型に分類された。MAT 型にユニークな遺伝子は、MAT1-1 で OG0012491/AO090020000089 (Pfam; PF04769; Mating-type protein MAT alpha 1 HMG-box)、MAT1-2 で OG0012281 (Pfam; PF00505; HMG box) + OG0012282 (no domain assigned) のみだった。OG0012281 と OG0012282 は隣接しており、OG0012282 にドメインが見つからなかったことから、これら 2 つが本来 1 つの遺伝子であるか、あるいは OG0012282 が misprediction である可能性が高い。なお、RIB40 は MAT1-1 株であるため、RIB40 において定義された AO で始まる ID は MAT1-2 遺伝子には付与されていない。

例えば、クレード A と B、H と G など、全ゲノム系統樹上で近縁であるにも関わらず、MAT 型が異なるものが多くあった。すなわち、全ゲノム系統樹上での系統的關係と MAT 型に不整合が見られることから、本研究で定義したクレードの分化はある一つの祖先株内での変異蓄積ではなく、複数の祖先株間の交雑によって生じていることが明らかである。さらに、*A. oryzae* と *A. flavus* は全ゲノム系統樹上で分別できていることから、これらの 2 種間での交雑は起こらず、各種内の祖先株間での交雑が起きたと考えられる。これは他の糸状菌で見られる VCG と同じく、異なる遺伝子の混入を防ぐために、系統的に近い株同士でしか交雑が起こらないように制御されていることと矛盾しない。

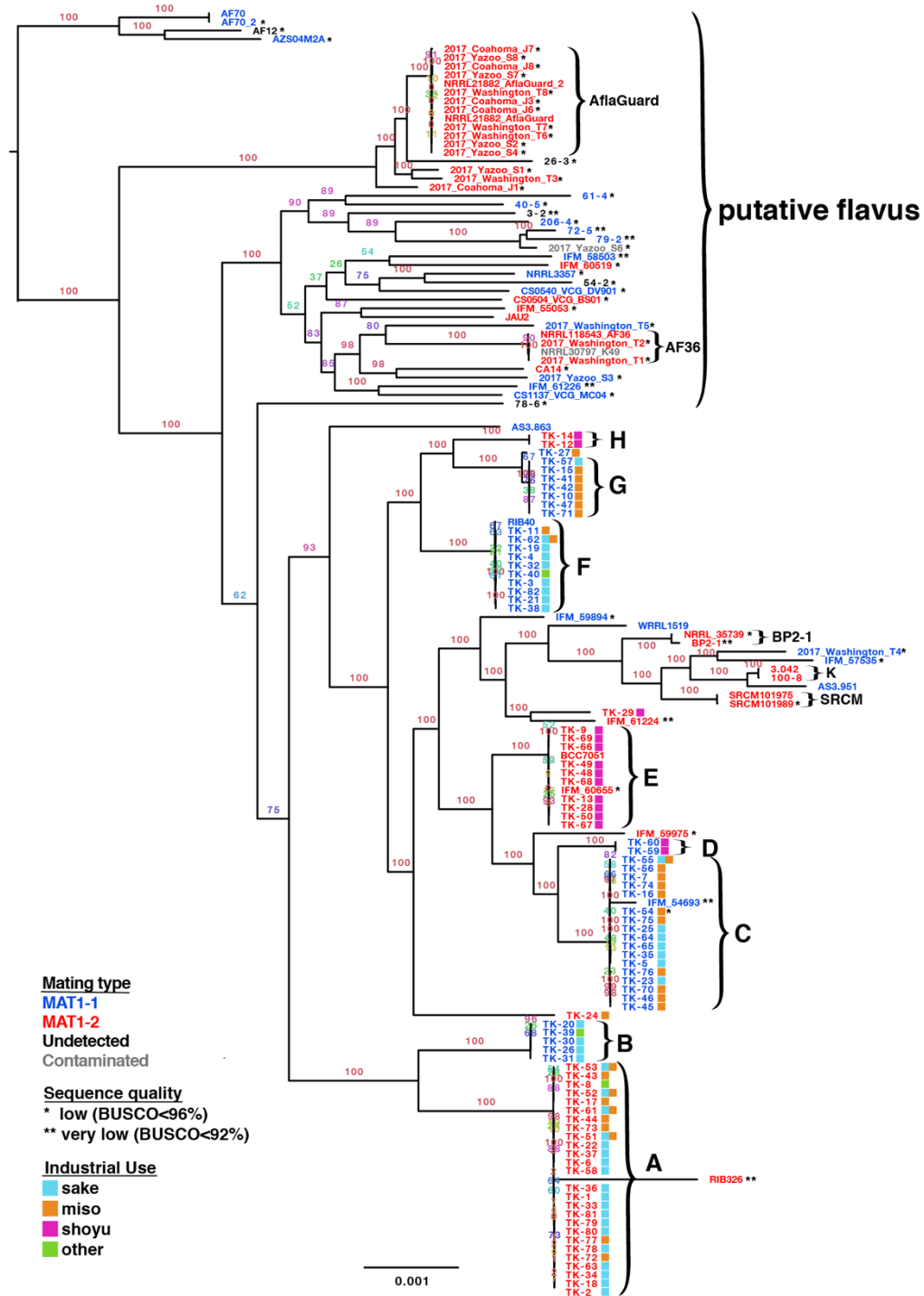


Fig. 2 *A. oryzae*, *A. flavus* の全ゲノム系統樹 (3,951 遺伝子)

A-H は日本の産業用株が属するクレード。AflaGuard, AF36, BP2-1, K, SRCM はその他のクレード。

2.5 ミトコンドリアゲノム系統樹

方法

ミトコンドリアは1細胞中に数十個ほど存在するため、DNA抽出時にミトコンドリアを取り除いていなければ、そのゲノムもシーケンスリード中に染色体ゲノムの数十倍の depth で存在する。ミトコンド

リアゲノムは数万 bp の AT-rich な環状 DNA で、染色体ゲノムとミトコンドリアゲノムが混合した状態のリードからはアセンブルが困難である。前章における方法でアセンブルした場合にミトコンドリアの完全ゲノムは得られなかった。そこで、以下の手順に沿って、それぞれの株についてミトコンドリアゲノムを構築した。また、Table S2 に示すシーケンスリード (8 *A. oryzae* + 30 *A. flavus*) を NCBI Sequence Read Archive (Kodama et al. 2012) よりダウンロードし、本研究サンプル (82 *A. oryzae* + 3 *A. sojae*) と共に解析に供した。

- 1) *A. oryzae* RIB40, *A. flavus* TCM2014, *A. niger* CBS 513.88 のミトコンドリアのコンプリートゲノム (AP007176.1, NC_026920.1, NC_007445.1) を NCBI Assembly よりダウンロードした。
- 2) 環状を考慮し、それぞれの配列を直列に 2 つつなげた配列を作成し、参照配列とした。
- 3) Bowtie2 v2.2.6 を用いて各シーケンスリードを参照配列にマッピングし(--sensitive --end-to-end)、マップされたものだけを抽出した。
- 4) 抽出されたリードを Platanus_trim で処理後、Platanus でアセンブル及び scaffolding, gap-closure した。
- 5) 最長の配列が 28,780 bp (RIB40 参照配列の 98%) 以上の FASTA を、RIB40 参照配列 (AP007176.1) に対して blastn v2.2.8 を用いてアラインメントし、結果ファイルを用いて開始位置を揃えた。

なお、(5)において元の配列の両端を結合させる際、RIB40 参照配列に対して長さが足りないものは曖昧塩基 N で埋めた。このとき、偶然 contig の両端に真の挿入・欠失が入ってしまった場合は再構成できていないが、可能性は低く、系統樹推定にはほとんど影響しないため無視した。

そして、得られた配列群にコンプリートゲノム(*A. oryzae* RIB40 (AP007176.1)及び *A. oryzae* 3.042 (JX129489.1))を加え、曖昧塩基 N を gap に変換し RAxML v8.2.11 (-m GTRGAMMAI -N 100) を用いて系統樹推定を行った。生成された Newick ファイルは FigTree v1.4.3 を用いて可視化し、マニュアルで reroot した。

結果

125 のリードデータから、(5)の手順で 8 サンプルが除外され、117 のミトコンドリアゲノムが構築できた。うち 89 は曖昧塩基を含まない完全な配列だった。完全に構築できた配列の長さは *A. oryzae* で 29,188–29,210 bp、*A. flavus* で 29,188–29,470 bp、*A. sojae* で 29,136 bp だった。系統樹推定に用いた 117 の配列のアラインメントは gap を含めて 29,465 列 (bp) だった。

ミトコンドリア系統樹 (Fig. 3) は、*A. oryzae* は *A. sojae* とは分別できたものの、*A. flavus* とははっきりと分別できなかつた。全ゲノム系統樹上で分離可能であることを考えると、*A. oryzae* と *A. flavus* はそれぞれ集団的に多様なミトコンドリアゲノムを種分化以前から有してきたと言える。特に、*A. oryzae* の中でクレード B は他の株群と比較的離れた系統に位置した。また、全ゲノム系統樹と比較して解像度が低いながらも、おおまかなクレードの分別はできており、日本の産業用株の属するクレード (A–H) 内の配列変異は確認できなかつた。クレード G と TK-24, TK-27, TK-29 はアラインメント領域の変異は見られなかつたため、gap を無視した最尤推定では分別できなかつたが、それぞれ挿入・欠失による変異が

あった。

なお、ミトコンドリアは *A. flavus* において母性遺伝 (MAT 型によらず、菌核を作った方の株が遺伝) することが知られている (Horn et al. 2016)。これが *A. oryzae* においても同様と仮定すると、クレード分別が可能な程度の変異や挿入・欠失による多様性が見られる中で完全に同一な配列がないということは、本研究で用いたサンプル群から確認されたクレードの分化 (交雑) は、ミトコンドリアゲノムに変異が入る程度の過去に起こったと言える。

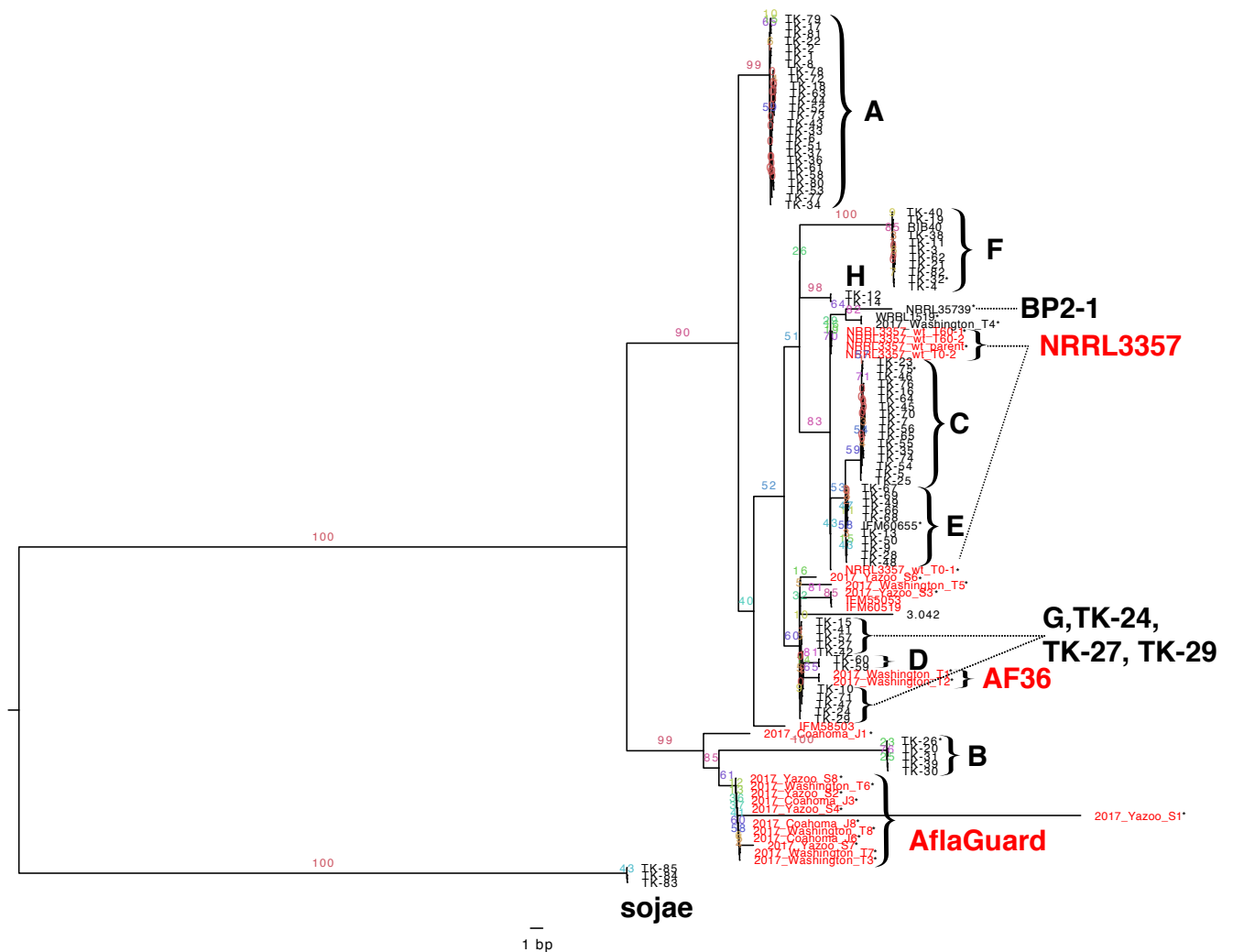


Fig. 3 ミトコンドリア系統樹

全ゲノム系統樹で定義したクレードを太字、かつ *A. flavus* の株名およびクレードを赤字で示している。

なお、1 bp に満たないずれは、本来同一の位置にあるべき配列が可視化の際にずれたものである。

アスタリスク(*)は曖昧塩基 N を含む不完全な配列を示す。

2.6 アフラトキシン生合成遺伝子クラスター系統樹

方法

アフラトキシン生合成遺伝子クラスター領域は、*A. oryzae* RIB40 ゲノム (GCA_000184455.3) に対し

て *A. flavus* AF36 の周辺領域配列 (AY510455.1) を blastn v2.2.8 によってアラインメントすることによって同定した。そこに含まれる 19 個の SCG (AO0900260000-08-13, 15, 20-22, 26, 28-30, 32-36) のエキソン+イントロン配列をそれぞれ MAFFT v7.2.2.2 でアラインメントした後、遺伝子欠損 (アSEMBル不良も含む) を gap で埋めながら直列に結合し、RAxML v8.2.11 によって系統樹を推定した (GTRGAMMAI -N 100)。生成された Newick ファイルは FigTree v1.4.3 を用いて可視化、マニュアルで reroot し、アウトグループを取り除いた。

結果

アフラトキシン生合成遺伝子クラスターのアラインメントは 45,432 列 (bp) となり、ここから推定した系統樹上では (Fig. 4)、これまで報告されているのと同様 (Machida et al. 2005; Chang and Ehrlich 2010)、*A. oryzae* の配列はすべて 3 つのグループのいずれかに分類できた。Group 1 はすべての遺伝子を有し (ただし無毒性)、Group 2 は半分程度が欠落し、Group 3 は 8 割程度が欠落しているグループである。また、これまでこれらのアフラトキシングループの系統的關係性は諸説あったが、本研究の結果から、アフラトキシン生合成遺伝子クラスター内で交雑による部分的な相同組み換えが起きていないと仮定すれば、Group 3 は Group 2 から派生したものであり、かつ Group 1 と Group 2, 3 は別の起源であることが言える。

アフラトキシン生合成遺伝子クラスター系統樹上で *A. oryzae* が単系統でないことから、少なくとも Group 1 及び Group 2, 3 は *A. oryzae* 種分化当時から別の祖先型配列であったと考えられ、すなわち Group 1 と Group 2, 3 は独立に無毒化したと言える。よって、「*A. oryzae* は *A. flavus* の家畜化の結果、アフラトキシンの無毒化によって生まれた種である」という説は一般的に正しくない。また、本研究で定義した全ゲノム系統樹上の *A. oryzae* であれば、産業用かどうかに関わらずアフラトキシン非産生性が実証されており、系統的にも非毒性のアフラトキシン生合成遺伝子クラスターをもっていた。

現在の地球環境におけるアフラトキシンの不要性については先行研究でも議論されており (Chang and Ehrlich 2010)、家畜化の外にある *A. flavus* においても無毒化が進んでいると考えられる。例えば、AflaGuard は無毒性の *A. flavus* としてよく知られている株であるが、産業用途ではないにもかかわらずアフラトキシン生合成遺伝子クラスターのほぼ全て (*A. oryzae* Group 1 よりも多く) を欠落しており、その名の通りアフラトキシン汚染を予防するために農業の場で使用されている株である。また、*A. sojae* と *A. parasiticus* の関係は *A. oryzae* と *A. flavus* の関係と相似しており、無毒化が系統と独立に進んでいるという説が支持される。よって、アフラトキシン生合成遺伝子クラスターの進化及び無毒化に対して家畜化はほとんど関与しておらず、蒸米に生えやすく分解酵素活性が高いといった、醸造に適した性質の株が偶然アフラトキシン非産生性であったという方が妥当であると考えられる。

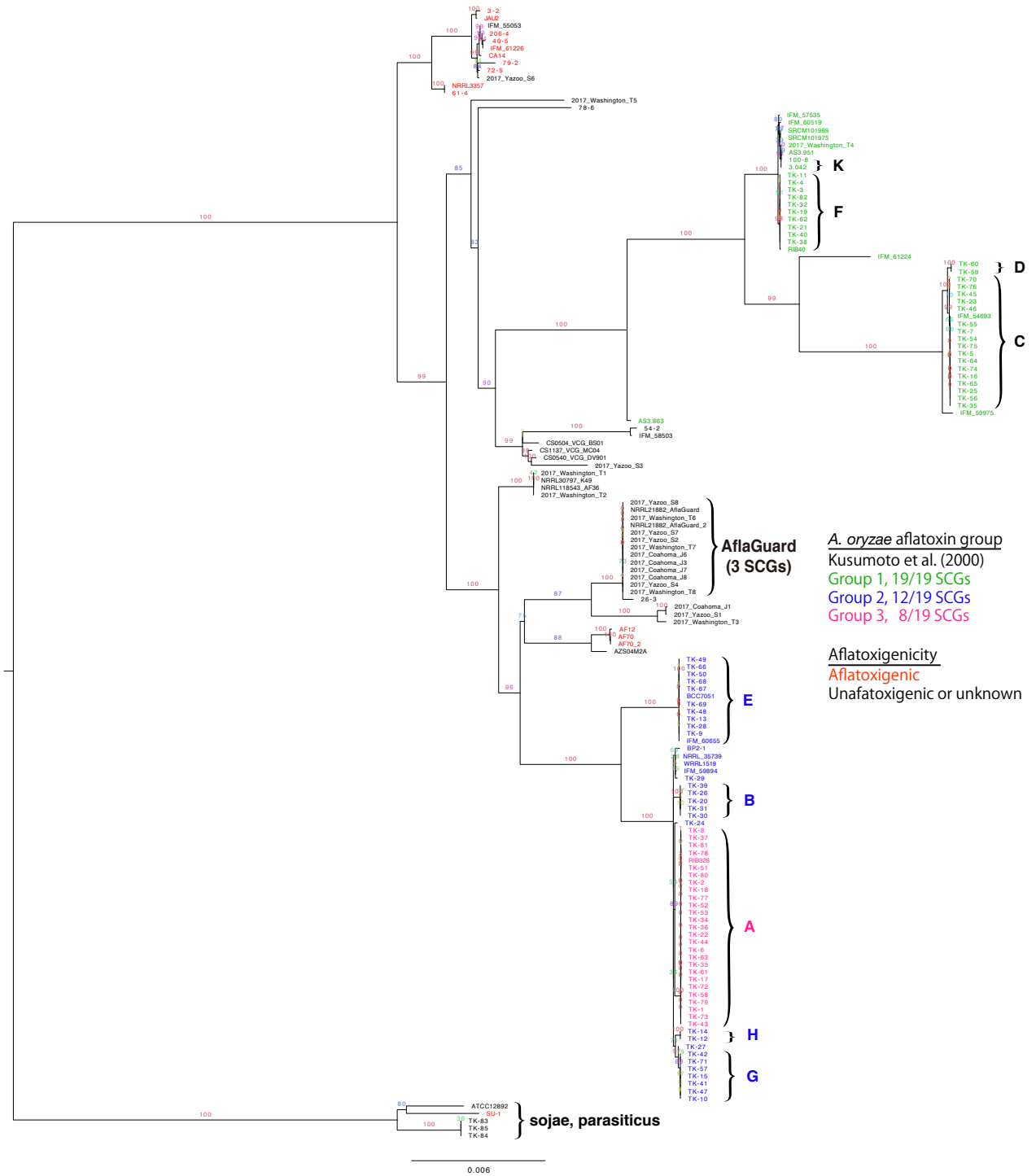


Fig. 4 アフラトキシン合成遺伝子クラスター系統樹

2.7 ゲノム交雑の可視化

本章の結果を踏まえると、全ゲノム系統樹のクレードと MAT 型の不整合から、*A. oryzae* の祖先株間において交雑が行われていた可能性が示唆され、またミトコンドリア系統樹のクレード分解能から、比較的最近の交雑によって生まれた株は存在しない可能性が示された。そこで、交雑の様子（例えば、ある

クレードが別の 2 つのクレードの交雑種であるかどうか) を調べるため、ゲノム座位ごとの遺伝的近縁度の可視化を試みた。

方法

- 1) RIB40 ゲノムに存在する 12,203 個の SCG に対し、それぞれの OG 内で DGS を完全一致でクラスタリングした。
- 2) 各 OG に対して RIB40 におけるゲノム座位をアサインした。
- 3) 各株の各 OG について、同一のクラスタに属する配列を含むクレードとの近縁度を 1、それ以外を 0 として計算した。各株自身の所属するクレードは計算から除いた。
- 4) スムージングとして、ゲノム座位上を動く window を設定し、window_size=100,000 bp, interval_size 50,000 bp として、window 内に含まれる OG が 2 つ以上含まれる場合、その近縁度の平均をとった。1 つの window に含まれる OG が 2 個未満であった場合は、欠損として扱った。
- 5) 各株内で、window ごとに最大の近縁度をもつクレードを決定した。
- 6) 最大近縁度の window 数が多い上位 12 クレードとそれ以外に色分けし、最大近縁度を棒グラフの長さとして、Circos v0.69-6(Krzywinski et al. 2009)によって可視化した。また、スムージングをしない可視化として、各 OG について同一の配列をもつクレードを積み上げ棒として表したグラフを作成した。

結果

A. oryzae, *A. flavus* のすべてのクレードのゲノムが複雑なモザイク構造を呈した。例えば、クレード A に属する TK-22 株 (Fig. 5) では、クレード B との完全に同一な DGS が遺伝子の個数で 46%あり、それがゲノム上に偏って存在していた。完全の同一な DGS が固まっている領域が、クレード A とクレード B との SB と考えられる。一方で、クレード B との NSB では、局所的にクレード F と近縁な領域も確認できた (Fig. S2)。このように、どのクレードの株であっても、染色体の位置によって最も近縁なクレードが異なっていた。系統樹上で近いクレード間では、ゲノム上で相同な配列の割合が大きかった。しかし一方で、系統樹上で離れたクレード間でも局所的に排他的に相同な領域を有する例があった (Fig. 6)。これらの結果は、*A. oryzae* 及び *A. flavus* のクレード分化が、共通の祖先株の突然変異ではなく、複数株の複数回の交雑によってもたらされたことを示唆する。交雑は、少なくとも観測されたクレードの数より多く起きたはずである。

特に、*A. oryzae* クレードでは、ゲノムのほとんどの領域が他の *A. oryzae* クレードとの混合として表された。これは、*A. oryzae* 及び *A. flavus* は主に種内での交雑が起きており、種をまたいで交雑することは稀であることを示唆している。一方で、*A. oryzae* と *A. flavus* が一部の領域で高い相同性を示すことがあった。これについては、種分化後に 2 種間での交雑があった可能性は否定できないが、相同な領域の大きさ及び相同性が非常に小さいことから、本研究で観測できなかったクレードの影響だと考えられる。例えば、アフラトキシン生合成遺伝子クラスターと同様、種分化以前から集団的に 2 つ以上の異なるタイプの遺伝子配列を有しており、それが半ば偶然に、観測した株の中で相同な領域として検出された可能性などが考えられる。さらに多くの株を解析に追加することで、モザイク構造が単純化されてクレー

ドの交雑の経緯が明確になる可能性があるが、親株が絶滅していた場合、交雑の様子を明らかにすることは困難になると考えられる。

本研究で収集したサンプルは人間が家畜化し管理してきた株であるため、その中で交雑が起こっていたのならば、その過程を追跡して推測できる可能性があった。しかし、本研究で用いた株群では、明確な別の2つゲノムの混合として表されるものは一つも観測できなかった (Fig. S2)。このことから、人間の家畜化が *A. oryzae* の進化にもたらす影響は、ほとんど個別のゲノム内への変化に対してであり、交雑を引き起こすことは稀であると推測できる。本解析からは、少なくとも近年に2株の交雑が起きた例はないと言える。交雑による育種が成功していないという事実と一致する (Hara et al. 2012; Wada et al. 2012; Wada et al. 2014; Okabe et al. 2018)。

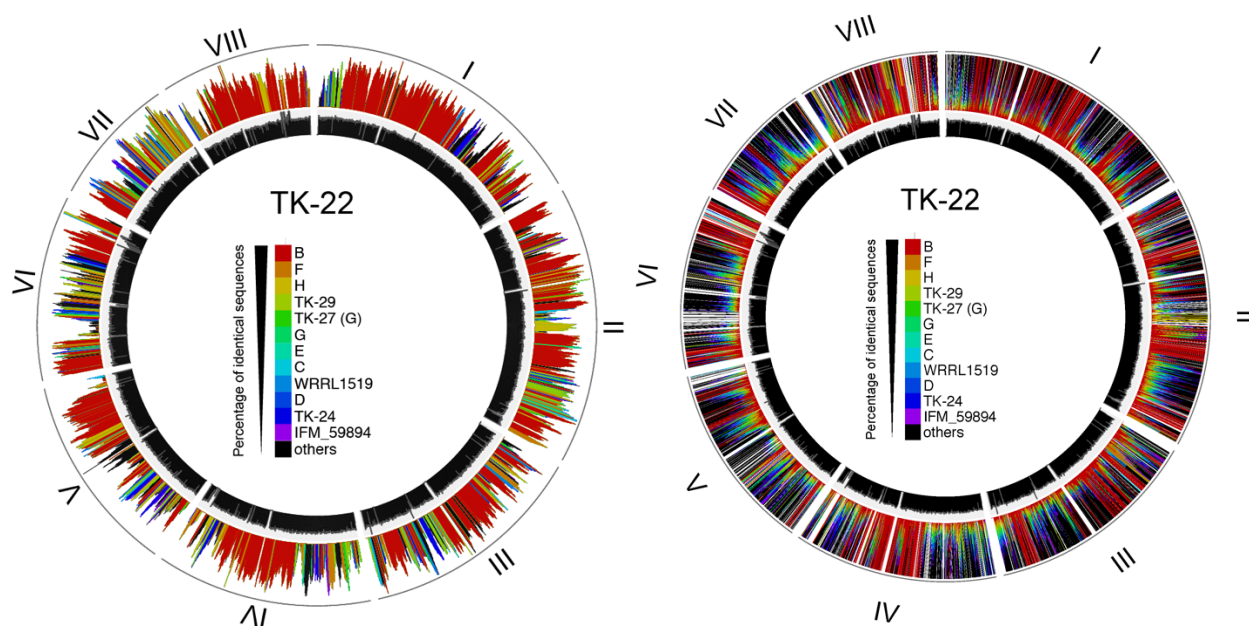


Fig. 5 TK-22 株(クレード A)のゲノム座位ごとの対他クレード近縁度
I-VIII は RIB40 で定義された染色体番号を示す。左はスムージングあり、右はスムージングなし。

OG0002327/AO090023000363
Phytanoyl-CoA dioxygenase

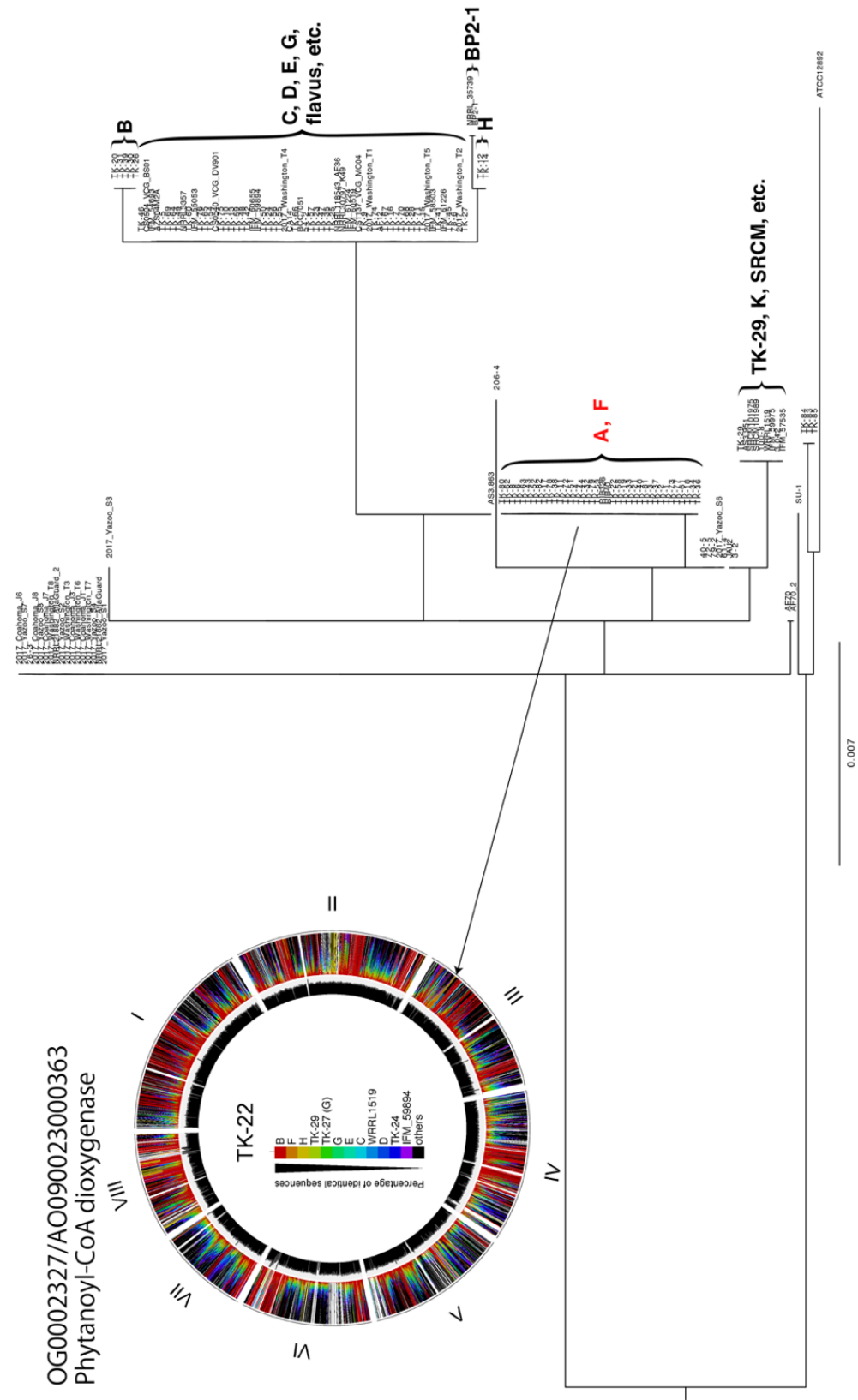


Fig. 6 クレイド A, F の相同領域に属する遺伝子の系統樹

第3章 家畜化が *Aspergillus oryzae* のゲノム進化に及ぼした影響

3.1 仮説提唱

一連の結果から、*A. oryzae* の祖先株間で、複数の交雑が起きていたことが浮き彫りとなった。一方で、現在種麹屋が所有している株間での単純な交雑は見られず、異なる出自でありながらほとんど同様のゲノム構造をもった株も多く存在することから、種麹屋は自然界に存在する醸造に適した菌株を選択し、交雑させることなく家畜化してきたという実情が想像できる。そこで、人間の家畜化が麹菌の進化に及ぼした影響は、クレード分化後のゲノム内変異や遺伝子重複等に顕著に現れていると考えられる (Fig. 7)。

先行研究に置いて、*A. flavus* における VCG 分化は 50,000–189,000 年前に起きたとされている (Chang and Ehrlich 2010)。*A. oryzae* が系統樹上で *A. flavus* 系統樹にネストした単系統であることを考慮すると、*A. oryzae* と *A. flavus* の種分化も同時期に起きたと考えられる。また、麹菌の家畜化及び産業利用は 3,000–2,000 年前に中国で始まり、種麹としての販売・利用は 700–500 年前に日本で始まった (Machida et al. 2008)。よって、家畜化が進化に及ぼす影響は広く見て最近の 3,000 年間、交雑後に限れば 700 年間に蓄積していると考えられる。すなわち、単純なタイムスケールで言えば、*A. oryzae* の進化史において、家畜化による影響は非常に限定的である。特に種麹屋は、その名の通り親となる「たね」を保存し、継代による発酵機能の減退を防いできた側面がある。そのため、家畜化の選択圧がかかっていると仮定しても、継代回数は時間スケールに比較してもかなり少なく、結果としてゲノムに対する変化はさらに限定的になっていると考えられる。

実際には、種麹屋によっては、もとを辿れば江戸の頃より受け継がれてきた株もあれば、大正時代に単離技術が確立されて以降に醸造元から採取された株も多く所有している (太古のように稲穂から起こすことは稀である)。そのため、ゲノムに対する変化は、種麹屋のストックとしての環境下で起こったものと、醸造の実用環境下で起こったものの少なくとも 2 種類があると考えられる。前者は酵素力価や生育などの形質によって選択圧がかかっていると考えられ、後者は醤油などの実際の製品としての環境下での淘汰圧がかかっていると考えられる。

先行研究においては、家畜化を議論するための解析として、*A. oryzae* 及び *A. flavus* の比較 (Gibbons 2012) や、RIB40 及び RIB326 の比較 (Toyotome et al. 2019) が行われてきたが、テロメア及びセントロメア周辺領域に変異が蓄積していると述べている。一方で、これらの解析は自然界と実用下での選択圧の区別が曖昧であり、家畜化の影響を正しく評価しているとは言い難い。本研究においては、クレード間では自然界、クレード内では家畜化の影響下でのゲノムに対する変化が濃縮しているという仮説の下、比較ゲノム解析を行った。

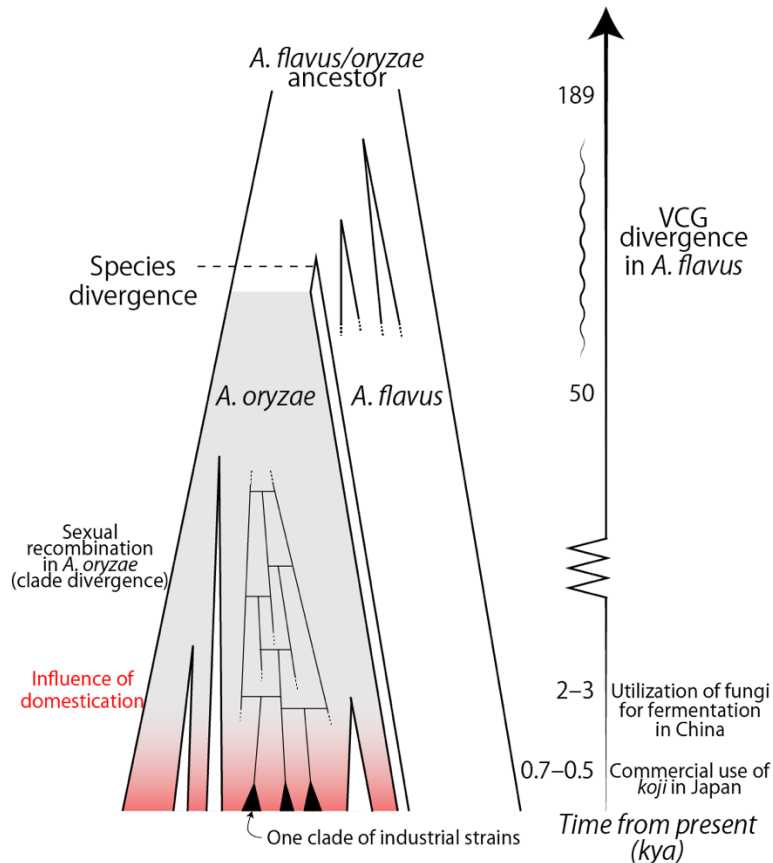


Fig. 7 *A. oryzae* の進化、及び家畜化の影響の仮説

3.2 クレード内外遺伝子変異圧解析

方法

本解析では、偽陽性を少なくするため、BUSCO スコアが 96%以上のサンプルを用いた。以下の方法で、*A. flavus*, *A. oryzae* の種内、及びクレード A-H のクレード内・クレード間変異解析を行った。なお、コンセンサス CDR は本論 2.3 において、すべてのサンプルを用いて各 OG 内で作成したものを利用した。

- 1) 各 OG に対して各クレード/各種内でコンセンサス CDR の配列 (gap 含む) を抽出した。
- 2) フレームシフト変異箇所をカウントした。なお、gap が連続している場合は 1 個のフレームシフト変異としてカウントした。また、クレード内のすべてのサンプルに共通している gap はクレード分化後のフレームシフト変異かどうか判別がつかないため、カウントしなかった。
- 3) gap を含まない配列からコンセンサス配列を作成した。クレード内の比較においてコンセンサス配列からハミング距離で CDR 長の 2%以上離れている配列はパラログとして取り除いた。gap およびパラログを取り除いて残った配列が 2 本以上ある場合、同義置換と非同義置換の数をクレード内/種内でそれぞれカウントした。また、このとき、同義置換と非同義置換の数の和を遺伝子帳で除した値 (以下、進化速度) を、クレード内/種内で計算した。

- 4) *A. oryzae* クレード内変異は、クレード A-H の和とした。*A. oryzae* クレード間変異は、*A. oryzae* の種内の変異の数から *A. oryzae* クレード内変異の数を差し引くことで得た。
- 5) *A. oryzae* クレード内変異、*A. oryzae* クレード間変異、*A. flavus* 内変異における同義置換、非同義置換、フレームシフト変異の間の有意差を、それぞれ Fisher の正確確率検定によって出力した。
- 6) AspGD からダウンロードしたオルソログ情報 (A_oryzae_RIB40_A_fumigatus_Af293_orthologs.txt, A_oryzae_RIB40_A_nidulans_FGSC_A4_orthologs.txt)を用いて、*A. oryzae*RIB40 の遺伝子のうち、*A. fumigatus* 及び *A. nidulans* の少なくとも一方と共有しているものを SB 遺伝子、その他を NSB 遺伝子として、(5)と同様の比較解析を行った。

なお、(4)において計算上クレード間変異が負になるものは、以降の解析では除外した。これらはクレードをまたいで同一の変異が入っているか、または(3)で取り除けなかった非常に近いパラログであると考えられるため、個別に判断する必要がある。

Table 2 *A. oryzae* のクレード内間比較 オッズは *A. flavus* の同義置換の数を基準としている。

		同義置換	非同義置換	frameshift		
<i>A. oryzae</i>	クレード内	count	265	528	93	
		odds		2.9*	10.3*	
	クレード間	count	90610	77396	5546	
		odds		1.2*	1.8*	
	<i>A. flavus</i>		count	93564	64845	3190
			odds		1	1

* p-value < 1.0E-10 (Fisher's exact test)

結果

BUSCO スコアが 96%以上だった 101 サンプル中の 14,711 個の SCG のうち、クレード内変異は、201 OG 中に同義置換が 265 個、453 OG 中に非同義置換が 528 個、70 OG 中にフレームシフト変異が 93 個見つかった。クレード内変異は上記のように非常に少なく、1 遺伝子に 1 個あるかないかの程度であったため、遺伝子ごとにクレード内クレード間での変異の偏りを検定するには不十分であった。一方で、ゲノム全体では、同義置換に対する非同義置換及びフレームシフト変異のオッズが、*A. oryzae* のクレード間変異及び *A. flavus* の species 内変異に比べて、クレード内変異で非常に大きくなっていた (Table 2)。一方で、*A. nidulans* 及び *A. fumigatus* との SB/NSB の間で変異の傾向の差はほとんど見られなかった (Table S4, *A. oryzae* クレード内の比較では SB において僅かに非同義置換が多かったが有意差はなく、クレード間及び *A. flavus* 種内においても 10%以上の差はなかった)。

クレード内で非同義置換及びフレームシフト変異の見られた遺伝子で、遺伝子名がアノテーションされているものでは、*wA*/OG0003175/AO090102000545 と *laeA*/OG0001970/AO0900030000489 の 2 つがあった。*wA* は *A. nidulans* において緑色の色素産生に関連するポリケチド合成酵素であり、*A. oryzae* においても、この遺伝子機能が喪失すると白色変異を引き起こすことが知られている (Mayorga and Timberlake 1990)。また、*laeA* は *Aspergillus* 属において global regulator として知られており (Bok and Keller 2004)、*A. oryzae* についてはコウジ酸産生にも関わっていることが報告されている (Oda et al. 2011)。特に *wA* 変異は味噌用株に特異的に存在しており、色素のコントロールが重要である味噌製造における選択圧が如実に現れていると考えられる。変異箇所も系統によってさまざまであり、変異が並行に起こっていたことがわかる。

興味深いことに、産業的によく知られている (20 個以上アノテーションされている) アミラーゼやプロテアーゼといった重要な分解酵素遺伝子には、非同義置換やフレームシフト変異は一切見られなかった。変異が見られた遺伝子のうち、*AspGD* において遺伝子シンボルが付与されていた遺伝子の機能は、シグナル伝達、二次代謝、分泌/トランスポーター、細胞形質に関わるもののみだった (Table S5)。発酵に関わる遺伝子そのものというよりもその周辺の一部の形質を変化させるような変異を集中させている可能性が高い。そもそも種麴とは、その名の通り、継代による発酵機能の衰退を防ぐための、保存的役割をもつ存在である。その中で、目に見えてわかる変異体がピックアップされ、別の種麴として保存されてきたために、このような変異に多様性が見られるのではないかと推測される。特にフレームシフト変異が多いことから、これらの変異は Loss Of Function (LOF) 変異である可能性が非常に高い。すなわち、特定の遺伝子に対する変異を以て機能を向上させることは稀でありこれが分解酵素遺伝子の保存性にも表れていると考えられる。

方法 (4) において検出された、クレード間で同一の変異が入っていた遺伝子は、OG0000313/AO090003000933/PF01048; Phosphorylase superfamily, OG0009541/AO090009000369/SSF48371; ARM repeat superfamily, OG0011005/AO090206000126/PF13374; Tetratricopeptide repeat, OG0012755/consensus disorder prediction, OG0012821/Coils coil, consensus disorder prediction で、これら 5 つはすべてフレームシフト変異として検出された。しかし、これらの遺伝子はコーディング領域に (3 個はタンパク質ドメインとしても) リピート配列を含んでおり、リピート数の変化によって配列の多様性が生まれるため、クレード間で共通の変異が生じた (またはミスアセンブルのよってそのように検出された) と考えられる (Fig. 9)。そのため、これらは家畜化の特異的な選択圧によって共通の変異が生まれたとは言い難い。

また、遺伝子進化速度について、種内及びクレード内にて計算した結果、14,711 個の SCG のうち約 15% の遺伝子は *A. oryzae* または *A. flavus* の種内で完全に同一の配列であった (Fig. 8)。種内の進化速度は、遺伝子全長の 0-0.4 で、99.6% の遺伝子の進化速度は 0-0.1 に収まっていた。また、いずれかのクレード内において遺伝子進化速度でパラログと判定された遺伝子は 3 個 (OG0012138: 291 bp, OG0012489: 1,629 bp, OG0013032: 297 bp) が存在した。クレード内 gap の入っていた遺伝子は上記に含んでいないが、いくつかクレード内系統樹との不整合がみられたため (本論 3.6)、この中にも真のパラログがいくつか含まれていると考えられる。

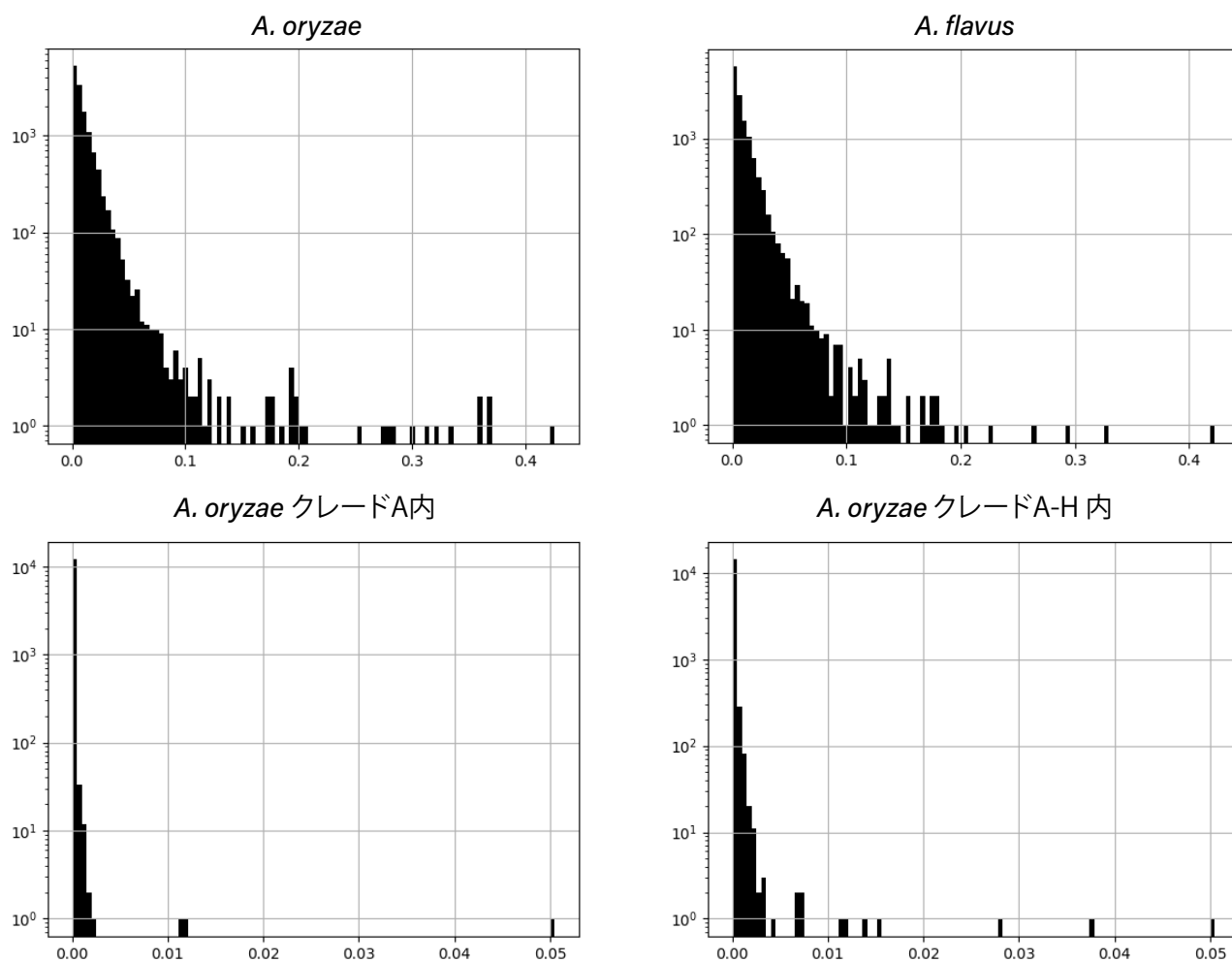


Fig. 8 遺伝子進化速度のヒストグラム

横軸は遺伝子進化速度、比較群内でのコンセンサス配列に対する変異の種数を遺伝子長で除した値。

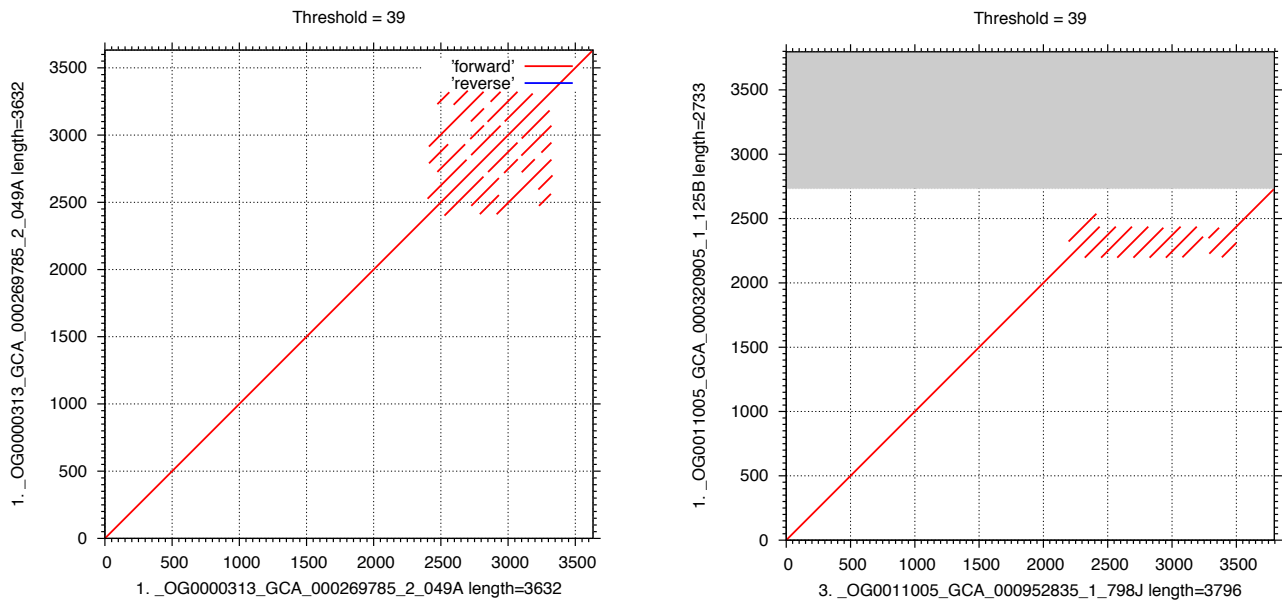


Fig. 9 OG0000313 の自己配列ドットプロット及び OG0011005 の異なる 2 つの配列のドットプロット
LAST プログラムによる。40 塩基以上連続したアラインメント領域が赤線で描画されている。

3.3 クレード内遺伝子欠損・重複解析

方法

遺伝子重複及び欠損は、本研究で新たに扱った 82 株に対して推定した。アセンブルによって得られた contig や scaffold では、遺伝子重複が depth の変化として吸収されることがあり、また、アセンブルの不足による遺伝子予測の失敗と真の遺伝子欠損とを区別することが難しいため、直接シーケンスリードをマッピングすることによって推定した。ここでは、点変異やフレームシフト変異による LOF は検出していない。

- 1) *A. oryzae* RIB40 ゲノム (GCA_000184455.3) に対して Bowtie2 v2.26 を用いて、シーケンスリードをマッピングした (--very-sensitive --end-to-end)。
- 2) SAMtools v1.3.1(Li et al. 2009) を用いて、RIB40 ゲノム座位に沿って depth を計算した (samtools depth)。その後、depth を 0 以外の最頻値で除することにより 正規化し normalized depth とした。
- 3) AspGD からダウンロードした“A_oryzae_RIB40_version_s01-m09-r06_chromosomal_feature.tab”の遺伝子領域ごとに平均 normalized depth を求めた。これらの遺伝子は AO で始まる ID でアノテーションされているため、OG に変換した。重複する遺伝子が存在する OG は、OG ごとに平均 normalized depth を加算した。
- 4) SCG の平均 normalized depth が 0.2 以下のものを欠損、1.6 以上のものを重複としてカウントした。2 個以上存在する OG はクレード内で平均 normalized depth 0.6 以上の差があったものをカウントした (以下、推定遺伝子数と言う)。

結果

221 個の OG (うち 117 個は SCG) で、クレード内推定遺伝子数に変化が見られた。そのうち 179 個の OG (うち 20 個は SCG) は、2 つ以上のクレードにまたがって変化が検出された。推定遺伝子数が異なっていた遺伝子の機能 (KEGG BRITE) の偏りを調べるため、全ゲノム中の遺伝子を母数として Fisher の正確確率検定に供したが、 $p=0.05$ を閾値とした有意な偏りは見られなかった。遺伝子重複は、個別の遺伝子に対してというよりも、ある程度連続した領域に起きていたため検出されなかった可能性も考えられる。

推定遺伝子数が顕著に変化していた遺伝子として、トランスポゾン遺伝子である OG0000041/PF03221/Tc5 transposase 及び OG0000321/PF14529/endonuclease-reverse transcriptase があり、クレード B と G では 15-20 のバリエーションが見られた。また、rRNA 遺伝子もかなり広い範囲で推定遺伝子数に変化しており、クレード C では 13-48 の変動が見られた。ただし、*A. fumigatus* においては自然環境で容易にコピー数が増加することが示されており (Herrera et al. 2009)、これらの遺伝子数の変化はクレード内の変化であるとはいえ、家畜化の影響が強いとは言い難い。同様に RIB40 に 276 個あるうちの 102 個の tRNA 遺伝子領域において、推定遺伝子数が増加していた。

産業上重要である α -アミラーゼ (ジアスターゼ) の遺伝子についてはよく研究されており、RIB40 においては 3 コピー存在することが知られている (*amyA/amyB/amyC/OG0011956*) (Machida et al. 2005)。 α -アミラーゼ遺伝子は *A. flavus* においては 1 コピーしか存在しないため、産業利用の中で (*A. oryzae* へと分化する過程で) コピー数が増加したとする説も存在したが、本研究では *A. oryzae* 株の中においても以下のような多様性が確認できた。推定遺伝子数は、クレード A では 3-4 コピー、クレード B では 2 コピー、クレード D 及び E では 3 コピー、クレード G (TK-27) では 4 コピー、TK24, TK-29 では 1 コピーの多様性が見られた。また、同様に *A. sojae* 株では 1 コピー存在し、先行研究と一致した (Sato et al. 2011; Yoshino-Yasuda et al. 2013)。クレード内での α -アミラーゼ推定遺伝子数の増加はクレード A の TK-61 株にしか見られず、産業的な選択圧の中でこそ遺伝子数が増加したとは言い切れない。クレード分化以前に自然環境でもコピー数が増加していたと考えるほうが妥当である。

また、一般的な傾向として、テロメア周辺領域において depth の変化が多く見られた (Fig. 10, 染色体 I, Fig. S3)。一部のクレードでは、RIB40 (クレード F) に存在する領域が大規模に欠損していた。また、染色体 VI と VIII でも染色体内部に大規模な欠損領域があり、かつ depth が大きく変動する領域が存在した。これらの領域の遺伝子には他の *Aspergillus* 属とのオルソログも多数含まれており、特にクレード F において新たに獲得された領域というわけではないようである。なお、アフラトキシン生合成遺伝子クラスターは染色体 III のテロメア付近に存在しており、欠損領域近くにテロメア領域のコピーが侵入していることが知られている (Lee et al. 2006)。染色体 VI 及び VIII の領域はそれよりもセントロメア寄りの領域であり、ゲノム内 rearrangement については、本研究で用いたショートリードでは検出しきれなかった。これらを実際に検証するには、ロングリードや Mate-pair リードによって染色体構造を解析する必要があると考えられる。また、本研究では RIB40 株のみを参照配列としているため、ここで欠損している領域については議論できていない。OG 数比で考えても単一株のゲノムではパンゲノムサイズに対して 10-20% 割程度の不足があり、パンゲノム参照配列をうまく作成できれば、より広範囲の遺伝子重複・欠損解析が可能になると考えられる。

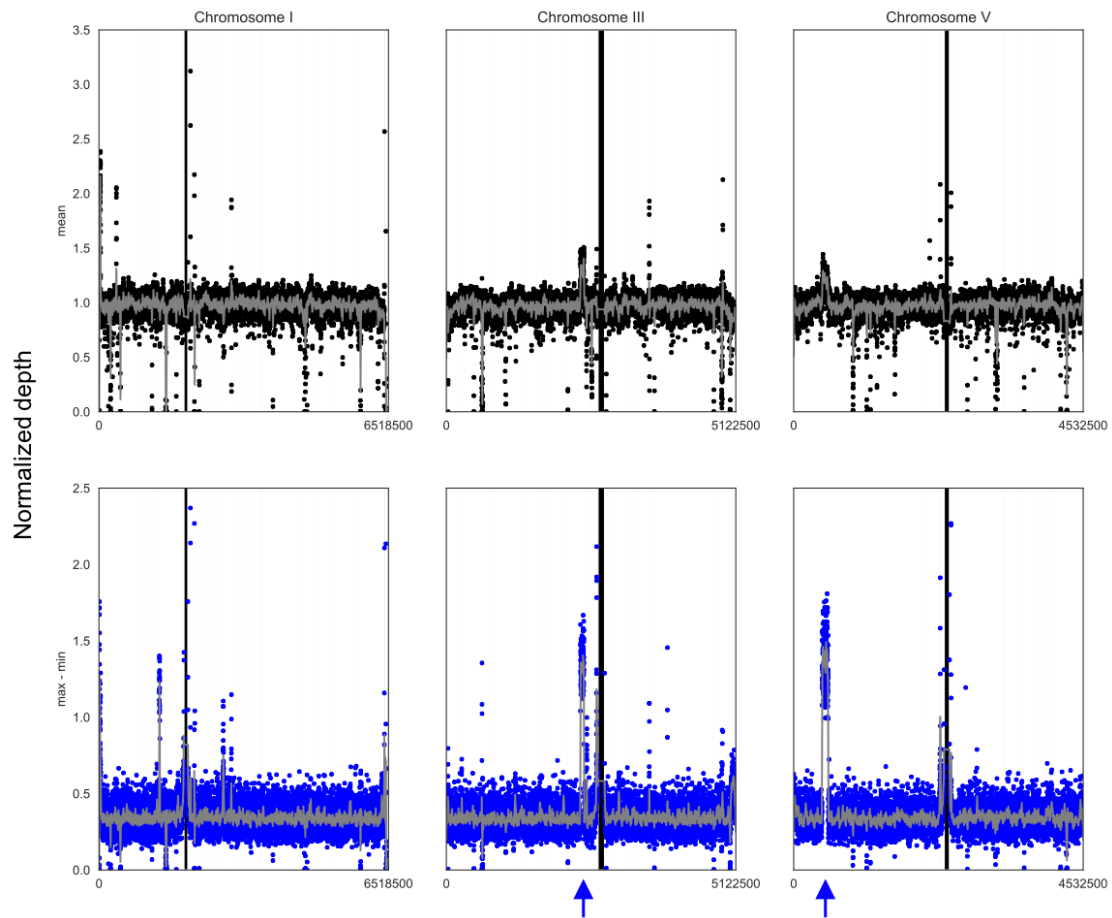


Fig. 10 例：クレード C の染色体 I, III, V における depth の変化

上段(黒)が平均 normalized depth で、下段(青)がクレード内の最大値と最小値の差。N で示される領域(多くはセントロメア領域)は黒く塗りつぶしてある。灰色は 10,000 bp window の平均移動線を示している。青い矢印は約 50,000 bp の重複領域を示している。

3.4 グループ特異的遺伝子

方法

本研究で用いた株は、種やクレードなど系統的な分類群と、産業用途や生育、酵素力価などさまざまなメタデータで分類されている。これらの特徴は遺伝子の有無や Single Amino Acid Polymorphism (SAP) などによって決定されている可能性がある。そこで、いくつかのカテゴリデータに特異的な遺伝子または SAP を抽出した。なお、BUSCO スコア 96%以上の 101 サンプルを使用し、比較するカテゴリデータが付随していないサンプルはその比較から取り除いた。遺伝子の有無は本論 2.3 で得た OrthoFinder の

出力を、アミノ酸配列は本論 2.3 で得た 14,713 OG に対する RPS 配列を用いた。

結果

種特異性

本研究では、*A. oryzae* や *A. flavus* に特異的な遺伝子は見つからなかった。先行研究における RIB40 ゲノムへのマッピングベースの解析では *A. oryzae* にのみ存在する遺伝子が 5 個あると推定されていたが (Gibbons et al. 2012)、本研究では先行研究より多様な系統の株を扱ったため結果が異なっていた。本研究の結果から、種分化以前から *A. oryzae* と *A. flavus* は集団的に遺伝子プールを共有しており、特定の一部の遺伝子の有無で種及び種の特徴が決定されることはないと考えられる。

一方で、種で特異的な SAP については、RIB40 でゲノム座位が付された 69 個の遺伝子中で見つかった。これらの遺伝子のゲノム座位は偏っており (Table S6)、特定の周辺ゲノム領域が *A. oryzae/flavus* で種特異的に保存されていることを示唆している。また、特に OG0003458/AO090701000370, OG0003884/AO090020000391 は HET 遺伝子であった。特異的変異箇所は HET domain ではなかったものの、*N. crassa* など HI が観測されている他の真菌に保存されている領域であり (PTHR24148, PTHR33112)、*A. oryzae* と *A. flavus* の種内特異的交雑に寄与している可能性がある。HET 遺伝子はそれぞれ細胞死を誘導するための相手となる遺伝子が存在し、一部の遺伝子ではそれらの相互作用が対立遺伝子の配列差によってもたらされることが知られているため (Bastiaans et al. 2014; Paoletti 2016)、これらと相互作用する遺伝子を実験的に見つけることができれば、*A. oryzae* における有性生殖のメカニズムを解明する手がかりになるかもしれない。

また、AspGD において遺伝子シンボルがアノテーションされている遺伝子は、OG0011554/AO090011000434/*CYP620K1*, OG0007388/AO090010000047/*aoiF*, OG0007389/AO090010000046/*aoiE* があった。これらは 2 次代謝に関わる遺伝子で、前者は *A. nidulans* において、遺伝子クラスター解析から、ステリグマトシスチンやアフラトキシンなど真菌毒の生成に関わっていると推定された P450 酵素のオルソログであるが (David et al. 2008)、*A. oryzae* において機能は明らかになっていない。*aoiF/aoiE* はポリケチド合成遺伝子クラスター中の遺伝子であり、特に *aoiF* は、強制発現実験の結果、色素前駆体である hexaketide isocoumarin 誘導体の水酸基をメチル化するメチルトランスフェラーゼであると推定されている (Nakazawa et al. 2012)。これらの二次代謝遺伝子クラスターの特異性は、*A. oryzae* に特徴的な形質や無毒性を保証する一つの理由になっている可能性がある。

用途特異性

本研究で用いた株は清酒用、味噌用、醤油用、その他(清酒/味噌用、みりん用等)に分類される。十分なサンプル数が確保できた清酒用、味噌用、醤油用の 3 つの用途について、特異的な遺伝子及び SAP を探索した。その結果、醤油用にのみ、いくつかの特異的な SAP をもつ OG が 13 存在した。また、醤油用に使われる *A. sojae* 株を含めても共通の変異が見つかった遺伝子はそのうち 3 個あった (Table 3)。特に、系統的に遠い *A. sojae* 株を含めて共通の SAP が見つかった IV 染色体上の領域(4,274,751–4,278,832)

周辺は、遺伝子間領域を含めて機能に関連している可能性が高い。

醤油用株は他の用途の株に比べて緑色が濃く、短毛で、プロテアーゼ活性が高いといった特徴があり、また、(実用上は蒸した大豆に生やして製麴するものだが)蒸米に植菌しても醤油のような色や香りを漂わせるように、容易に見分けがつくほど形態が異なる。さらに興味深いことに、*A. sojae* と *A. oryzae* は系統樹上で大きく離れているにも関わらず、醤油用株は形態が酷似している(*A. sojae* は醤油用しか存在しないが、*A. oryzae* の醤油用株と形態による分類がしばしば誤っていることがある)。

よって、ゲノム上にその特徴を決定づける領域が保存されていると考えられ、本研究で同定した OG 及びその周辺のゲノム領域が醤油用株の特徴に関与している可能性がある。例えば OG0002019/AO090003000691/K11771 は酵母におけるクロマチン制御タンパク質のオルソログであり、遺伝子の発現量を変化させることが知られている(Monahan et al. 2008)。また、OG0009397 は転写調節因子ドメインをもっており、これらの遺伝子が形態変化に関係している可能性がある。他にも、OG0006944/AO090701000985 は、コードするタンパク質が有する WD repeats が HET 遺伝子の認識部位として知られるドメインで、また NACHT ドメインがアポトーシスに関連することから、HI 関連遺伝子である可能性が高い。醤油株特異的な SAP をもつ遺伝子の中に HET 遺伝子は検出できなかったものの、このような自他認識関連遺伝子の存在は、例えば豆に生えやすい株同士での交雑を優先するような働きをすることで、自然界におけるニッチを区別し、結果として醤油株の特徴を決定するゲノム上の領域の保持に貢献してきた可能性があると考えられる。

清酒用・味噌用については、兼用の株の存在から分かる通り、同一クレード内 (A, C, G, F) の比較でさえ、特異的な遺伝子及び変異は見つけることができなかった。種麴屋 に対する聞き取り調査においても、清酒用・味噌用を分けるような特別な指標があるわけではなく、形態として区別もできないとのことであったため、結果は妥当であったと考えられる。本論 3.2 の通り、味噌用株は白色変異株であることが多いこと (あるいは白色変異株は味噌用株として扱われやすいこと) は言える。

Table 3 醤油株特異的な SAP が見られた遺伝子

“sojae”は *A. oryzae* 及び *A. sojae* の 2 種にまたがる SAP が存在することを示している。

	sojae	染色体	positon	アノテーション (KEGG, InterPro)
OG0005278		I	5,284,010	Pfam;PF00004;ATPase family associated with various cellular activities (AAA)
OG0002019	○	II	3,959,715	K11771;SWI/SNF chromatin-remodeling complex subunit SWI1
OG0003032	○	IV	1,728,787	Pfam;PF07690;Major Facilitator Superfamily
OG0011390		IV	4,271,722	K17922;Sorting nexin 8, regulating endosome-to-Golgi transport

OG0006405	○	IV	4,274,751	Pfam;PF00004;ATPase family associated with various cellular activities (AAA)
OG0009414	○	IV	4,278,832	Pfam;PF10395;Utp8 family
OG0009426	○	V	1,023,142	consensus disorder prediction
OG0008374		VI	618,069	MobiDBLite;mobidb-lite;consensus disorder prediction
OG0009397		VI	1,980,253	Pfam;PF04082;Fungal specific transcription factor domain
OG0006943		VI	2,777,764	Pfam;PF07971;Glycosyl hydrolase family 92
OG0006944		VI	2,787,600	ProSiteProfiles;PS50294;Trp-Asp (WD) repeats circular profile. Pfam;PF05729;NACHT domain
OG0004156		VI	3,175,028	Pfam;PF08386;TAP-like protein Pfam;PF00561;alpha/beta hydrolase fold
OG0009121		VI	3,178,702	SUPERFAMILY;SSF56801;Acetyl-CoA synthetase-like superfamily

3.5 機能及び系統予測のための NRIB ジェノタイピングとの比較

方法

酒類総合研究所 (NRIB)において、ゲノムアレイの結果から *A. oryzae* の種内で株の系統及び発酵特性を分類するためのプライマーセットが提案されている(岩下 2013)(Table 4)。これらを Bowtie2 v2.26 によって *A. oryzae* RIB40 株のゲノム(GCF_000184455.2)にマッピングすることにより、インサート領域を同定し、これを参照配列として blastn v2.2.8 によって、各サンプルの contig に対しローカルアラインメントを行った。ヒットした配列を擬似的な PCR の結果として、インサートパターンを導出した。

Table 4 NRIB によって提案されたジェノタイピング用のプライマー配列

ID	Fwd	Rev
192	CCACAACCTCAGAGAAGTCC	AGGCTGTTACATCACCAAGT
089	GCATCAAGGGAATACCATCT	TATTGTAAGAGCCCTTCCTG
027	CATATCTGCATCGCGACTTG	TTAGCATTGGGTGGAAGGAC
025	GGCCATTCTAACTCGGTTCA	AACTCTGCCACTTGTGCTAC
225	CCCCATCCTCGATTTCTCTG	TTGTGGAGTTGTTGCTAGGT
367	ACTTTACGCGTGTTCGTTGT	ACCTGACGGCAGTACTCATC
217	GCATGCTATCCAACAGAAGG	CTGAACGGCTTTGATGAGAG
239	TATCAACCCAGCCATCTGTC	TCATATGACCCTTGCGAGTT
091	CGCGAGGTAACCTACCTTCTG	CCCTGTATACGACGTTAGCC
140	TCTTCCTCATCGGTGACTCT	CCTGTGGCATATTCACCGTT
161	CAAGAGATTTCGTGGATGCCT	TTTAGTGCGTTACCAGCACA
187	GCCAGCAAGAATGTCTCCTT	GTTCTGACCCACGACGTTTG
232	GATACAAAAGCCACCTCCAA	TCTGAAGTTCCCAATGTCCA

結果

本研究で用いた *A. oryzae* 実用株において、10 個の PCR インサート領域の検出パターンが見つかった。このうち、NRIB が特許本文中で示した増幅パターンの 13 クラスターとは 3 つが同一となった。本研究で用いたサンプルにおいて、プライマー-225, 1410, 161, 187, 232 は分類に寄与しなかったため、これらを除いてパターンを調べたが、それでも 5 クラスターしか一致しなかった。すなわち、NRIB の示す分類群のうち 10 クラスターは観測されず、逆に、本研究では新たに 8 ないし 7 個のパターンが見つかった。

同一のクレードの株は同一のパターンを示したが、クレード A と F など、いくつか分類できないクレードが存在した (Table S7)。既知のクラスターとの積集合が少なかった理由として、プライマーのインサート領域のアセンブルが困難で検出できなかったことが考えられるが (false negative)、クレード内ですべて同一の結果に落ち着いているため、この可能性は低い。一方で、PCR プライマー領域だけではなく PCR 増幅領域を含めてアラインメントしているためプライマー領域の変異による affinity の変化を考慮しなかった可能性は高い (false positive)。ただし、このような誤謬を考慮しても、本研究で見出したパターンと NRIB のパターンは大きく異なっており、NRIB の保有株および自然界には本研究で観測しなかった *A. oryzae* クレードが存在する可能性が高いと考えられる。

本研究における検証では、このようなゲノムアレイをベースとした PCR プライマーによる株の系統分類はある程度有効ではあるが、本論 2.7 において示したように *A. oryzae* はゲノムの構造がクレードごとに異なっており、また未知のクレードの存在があることも加味すると、完全にクレードを分類することは困難であると考えられる。例えば、プライマー-091 は機能未知遺伝子 AO090701001243 と AO090038000501 の遺伝子間領域を検出しているが、当該領域は *Aspergillus bombycis* acyl-CoA oxidase (ABOM_009898) の mRNA 配列 (XM_022537026) の一部 (563 bp) に identity 86% でヒットした。一方で、この配列は *A. oryzae* のみならず *A. sojae*, *A. flavus* において偽遺伝子になっており、この領域自体の欠損による機能変化は大きくないと考えられる。PCR 増幅バンドによる分類は、発酵機能の大まかな予測のために開発されたものであるが、実用的には全ゲノム系統樹上のクレードを分類するためのプライマーと、クレード内での発酵機能に関わる遺伝子領域変化を検出するためのプライマーを区別して設計することで、機能予測の向上が望めるだろう。

Table 5 RIB プライマーパターンでの分類(完全一致)

パターン	株名	クレード	NRIB
d_0	3.042, 100-8, TK-29	K, TK-29	SO1
d_1	AS3.951	AS3.951	
d_2	AS3.863, RIB326, RIB40, TK-3, TK-8, TK-4, TK-1, TK-2, TK-6, TK-11, TK-21, TK-22, TK-12, TK-32, TK-33, TK-17, TK-18, TK-19, TK-34, TK-36, TK-43, TK-44, TK-37, TK-38, TK-40, TK-14, TK-61, TK-51, TK-62, TK-52, TK-53, TK-63, TK-58, TK-77, TK-78, TK-79, TK-80, TK-72, TK-73, TK-81, TK-82	A, F, AS3.863	RIB40
d_3	BCC7051, TK-9, TK-13, TK-49, TK-50, TK-48, TK-28, TK-68, TK-69, TK-66, TK-67	E	
d_4	SRCM101975, SRCM101989, BP2-1	SRCM, BP2-1	
d_5	TK-7, TK-5, TK-25, TK-16, TK-35, TK-45, TK-46, TK-23, TK-54, TK-64, TK-65, TK-55, TK-56, TK-75, TK-76, TK-70, TK-74	C	(Sake/Miso)
d_6	TK-10, TK-15, TK-42, TK-41, TK-47, TK-57, TK-71	G	
d_7	TK-24	TK-24	
d_8	TK-27	TK-27	
d_9	TK-26, TK-30, TK-31, TK-20, TK-39	B	(NS1)
d_10	TK-83, TK-85, TK-84	sojae	
d_11	TK-59, TK-60	D	A

3.6 クレード内系統樹とゲノムの変化

方法

本論 2.3 で得られた全ゲノム (DGS) 系統樹からクレード内だけに焦点を当て、それぞれのブランチにおけるゲノムの変化 (遺伝子のアミノ酸配列変化、遺伝子欠損・重複) 及び各株のメタデータをマニュアルでマップした。クレード内のブランチの長さはアウトグループを含めたそれに比べて非常に小さな変化であり、尤度が極小値で描画されることがあったため、枝の位置や長さは適宜マニュアルで調整した (枝の長さは変異頻度を反映していない)。なお、マニュアル操作とは、具体的に近接する複数株間の比較によってブランチごとのゲノム上の変化を埋めていく方法で行った (酒、味噌用が混在しているクレード A, C, F, G を対象にした)。

アミノ酸配列変化は対象の 3 株以上の RPS の比較によって検出し、このとき片方に遺伝子欠損がある場合は遺伝子予測の失敗と区別ができないため無視し、また、すでにフレームシフト変異が入った後の変異は無視した。また、本論 3.2 で得た配列中にリポートをもつ 5 個の OG は解析から取り除いた。不整合が見られるゲノム上の変化については DGS の系統関係を優先した。また、系統樹と不整合が見られた遺伝子はミスアラインメント及びミスアセンブルの可能性が高いため、個別にパラログの存在やリポート配列を確認して除外した。遺伝子シンボルが与えられている OG にはそれを追記した。

遺伝子への遺伝子欠損・重複の検出は、本論 3.3 で得た遺伝子領域の depth を対象の 3 株以上で比較することによって検出した。比較対象株間で normalized depth の最大値と最小値の差が 0.9 以上あるもので、明らかに連続的な変化の延長でないものを遺伝子重複または欠損とした。染色体の最端の遺伝子領域は depth の誤差が大きいため無視し、RNA 遺伝子も depth の変化量が大きいため局所的な増減は無視した。また、大規模に変化したゲノム領域に含まれる遺伝子のオルソログはそれに付随して depth が変化する傾向にあったが、これも無視した。大規模に重複した領域の中での局所的な欠損は遺伝子減少として区別した。

最後に、それぞれの種麴屋から提供されたメタデータのうち、用途と、すべての株に共通して存在した「菌糸の長さ」および「生育速度」について「普通」以外を書き加えた。

クレード F

クレード F は MAT1-1 交配型であり、本研究で用いた株の中では唯一 Potato Dextrose Agar (PDA) 培地上で菌核と呼ばれる組織を形成するクレードである。菌核は *A. flavus* において有性生殖に必要な組織であり (Horn 2009)、このことからクレード F は *A. oryzae* における有性生殖の可能性を探す上で重要な存在だと考えられる。このような背景があったかどうかは不明だが、クレード F の RIB40 は最も野生株に近いとされ、完全ゲノムが解読されている。実用的には酒や味噌用株が属し、(麴の褐変の原因となる) チロシナーゼ活性を有するといった特徴がある。

クレード内系統樹を見ても (Fig. 11)、RIB40 と実用菌株は比較的離れた系統に位置し、その間には 100 個以上の非同義置換変異が観測された。これらの変異蓄積が、RIB40 株の採取された醸造蔵(ソラマメ)におけるものか、菌株保存機関の管理下におけるものかは判別できない。マッピングベースの解析において遺伝子をまたぐ規模のゲノム領域の重複や欠損は一切検出されなかった。遺伝子コピー数を変化させるような rearrangement は起こっていないものの、いくつかの株においては短毛化や早熟・晩熟化が起きているため、これらは遺伝子変異が原因である可能性が高い。しかしながら、これらの形態変化はほとんどすべての遺伝子が関係しており、少なくとも系統樹上にマップした遺伝子変異から妥当なものを見つけることはできなかった。RNA の推定遺伝子数は誤差が多いため無視したが、これが真に関係している可能性も考えられる。

クレード A

クレード A は、アフラトキシン生合成遺伝子クラスターのほとんどが欠失していることが特徴である。アフラトキシン毒性の再発性が限りなく低いことから、海外への輸出用としても注目されている。実用上の特性としては、酵素力価は他のクレードと比較してそれほど高くないため、清酒用としても弱い味になると言われている。そのため、古くはあまり使用されていなかったが、近年では趣向の変化により甘口の酒が好まれるようになり、使用頻度が増加しているようである (種麴屋の間では新しいタイプという認識があるらしい)。本研究で用いたサンプルの中では、クレード A に属する株が最も多かったことからそれが伺える。チロシナーゼ活性がないため、色合いとしても好まれる。

クレード内系統樹 (Fig. 12) でも、各所で独立の白色変異 (*wA* の LOF 変異) が起きており、色合いが重視されていることがわかる。これらは味噌用として使われることが多いようで、TK-61 および TK-72 のみ白色変異 (*wA* 変異) が起きていない味噌用株であった (TK-77 の *wA* の非同義置換は LOF 変異であった)。また、他のクレードと比較して同一ノードに位置する株が非常に多く、すなわち非同義置換やフレームシフト変異が全く見られない株群が多く得られた。その中には、種麴屋が異なるものもあり、有用性が認識された単離株が流通し各所で保存されている経緯が反映されていると言える。これらの株はメタデータ上では増殖速度や菌糸の長さが異なっているが、管理される種麴屋の培養条件や用途が異なれば相対的な評価も異なるため、一概に遺伝子外変異が形質変化を及ぼしているとは言い難い。同一クレード内での定量的生育実験を行って、ゲノム変異との関連性を考察する必要がある。また、クレード A は参照配列である RIB40 (クレード F) と全ゲノム系統樹上で遠くに位置しており、RIB40 とは異なるアクセサリ遺伝子を多くもつため、これらの重複数が変化している可能性も考えられる。パンゲノム解析を行うことで、より詳細な株間比較ができる可能性がある。

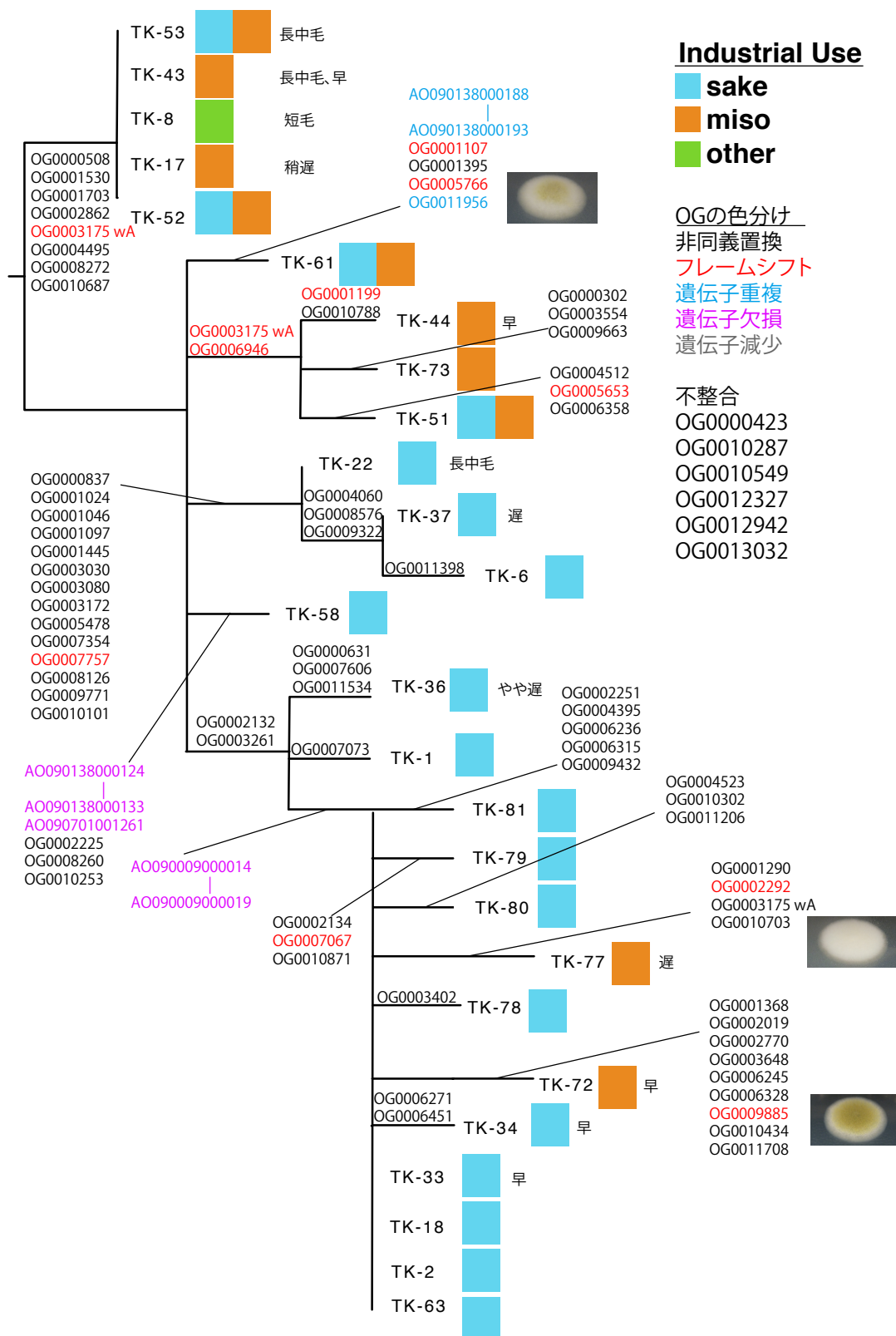


Fig. 12 クレード A 内系統樹

クレード C

クレード C はチロシナーゼ活性を有し、酵素力価が非常に高く、旧くから重宝されてきた株が多く含まれている（クレード A に比較して古いタイプと認識されている）。本研究でドラフトゲノムを構築したサンプル中では唯一 BUSCO 96%以下でコア遺伝子欠損が多かった TK-54 株もここに含まれている（比較解析からは除外された）。その酵素力価の高さから、味噌用や強い味の酒の製造に使われることが多い。TK-55 株のような、焼酎用黄麹としても使われることのある珍しい株も存在した。

クレード内系統樹を見ると（Fig. 13）、Fig. 10 でも示したとおり、クレード内の遺伝子領域重複や欠損が多く見られた。しかも、これらは系統樹中段のブランチで起きており、種麹屋の保有株中でその過程が追えているということと、旧くから使われている菌株が多いことを考えると、家畜化の過程で起こった可能性が高い。面白いことに、TK-75 の近縁株で起きたと考えられる AO090023000907-AO090023000925 の大規模なゲノム領域の重複後の株は清酒用として利用され、それ以外はほとんど味噌用である。これらの事実を踏まえると、どこかの種麹屋が新奇な変異株を取得し、その有用性が認識され、別用途（味噌用→清酒用）として別の種麹屋に流通した経緯が想像できる（TK-23 の周辺でも同様の関係が見られる）。個別の株に注目すると、TK-56 は TK-55 の変異株であるとされ、短毛化と OG0002895/AO090012000255 のアミノ酸変異が起きていた。本研究で用いた株の中で実際の分離元が明確な株は少なく、この遺伝子欠損が短毛化に関連している可能性は高い。この遺伝子は K22149 mevalonyl-CoA ligase にアサインされ、*A. fumigatus* において 2 次代謝に関わっていることが報告されている（Gründlinger et al. 2013）。一方で、この遺伝子の変異はこれらの株間のみでしか観測されず、他の株の短毛化はまた別の遺伝的变化によってもたらされていることが示唆される。同様に、TK-65 は分離元が明確な TK-64 の変異株であり、形状のみが異なるとされている。しかしながら、遺伝子領域の変化は RNA 遺伝子の depth 以外に全く検出されず、真に RNA 遺伝子のコピー数変化が効いているか、遺伝子間領域の影響あるいはエピジェネティックな制御が異なる可能性も考えられる。形状の変化は同一コロニー内でも頻繁に見られる現象であり、これらも同一の仕組みによる可能性がある。

クレード G

クレード G はほとんどが味噌用株である。チロシナーゼ活性を有し、酵素力価は中の上であるが特徴は薄い。TK-57 は唯一清酒用であったが、その理由は定かでない。

系統樹上では非常に興味深い側系統の TK-27 が存在する（Fig. 14）。本研究では家畜化の影響を考察するために TK-27 とクレード G とを区別して解析を行ったが、MAT 型（MAT1-2）やミトコンドリアの系統などを総合的に考えてもクレード G と同一の祖先株をもつ（すなわちクレード G の株）と考える方が妥当である。TK-27 と他の株の間には 1,000 個以上の OG で非同義置換変異が蓄積しており、ゲノム領域の重複や欠損も非常に広い範囲で観測された。TK-27 とそれ以外の間の株が観測されなかったことから、これらのゲノム上の変化が家畜化によるものとは考えにくい。逆に言えば、自然界においても、遺伝子重複や欠損が頻繁に起きることを示唆しており、家畜化の影響を正しく推量するためには、自然環境と家畜化の過程を考慮して比較することが重要であると再認識できる。

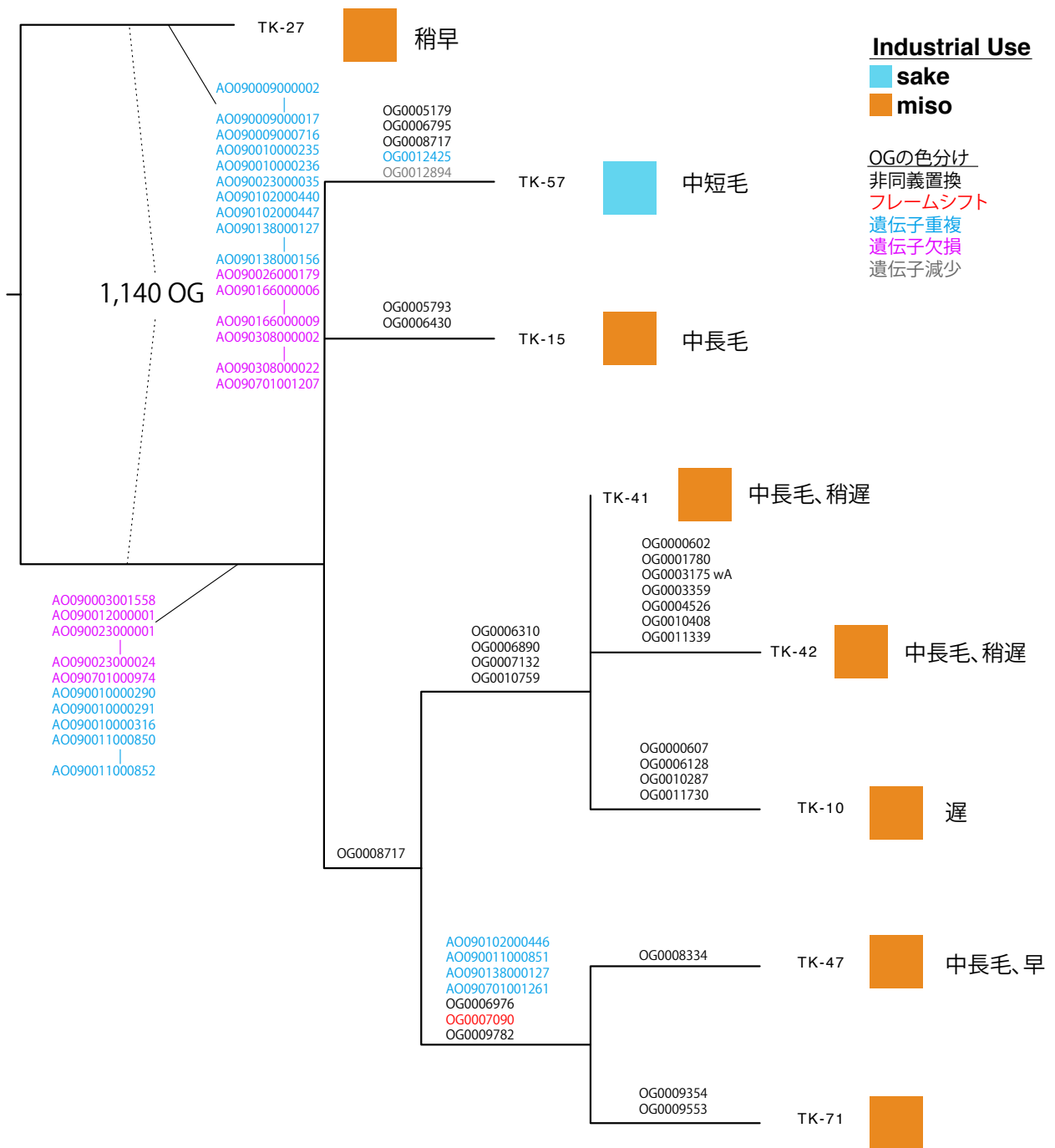


Fig. 14 クレード G 及び TK-27 の系統樹

第4章 結論

本研究では、*A. oryzae* の進化における有性生殖の経緯及びそれに関わる家畜化の影響を推定するため、産業用株の大規模ゲノム比較を行った。*A. oryzae* の全ゲノムデータは、これまで完全ゲノム及びドラフトゲノム含めて十数しか公開されていなかったが、本研究では新たに種麴屋から 82 株という大規模かつハイクオリティなドラフトゲノムの作成を行い、サンプル数を大幅に増加させたことで、より詳細な近縁株の比較ゲノム解析を行うことができた。まず、*A. flavus* を含めた全ゲノム系統樹上において、*A. oryzae* が *A. flavus* にネストした単系統であることを示し、さらに同一の近縁な祖先株をもつと推定される「クレード」を定義することができた。そして、クレード間のゲノム配列の近縁度と MAT 型の間には不整合を見いだせたことから、これらのクレード分化が有性生殖によってもたらされていたことを明らかにした。ついで、アフラトキシン生合成遺伝子クラスターの系統関係を明らかにし、*A. oryzae* には祖先の異なる 2 種類が存在することを見出し、*A. oryzae* と *A. flavus* の種分化がアフラトキシン整合性遺伝子クラスターの無毒化と無関係であり、かつ家畜化とは無関係であることを示した。また、ミトコンドリアゲノムの構築及び系統樹を行うことにより、ミトコンドリアゲノムでは *A. flavus* と *A. oryzae* が分離できないこと、またクレード間でそれぞれミトコンドリア配列が異なることを示した。このことから、*A. oryzae* と *A. flavus* は集団的に多様なミトコンドリアゲノムを有しており、かつ最近に（ミトコンドリアゲノムを共有するイベントである）有性生殖が起きていないことを示した。有性生殖の組み合わせを調べるために、ゲノム座位ごとの近縁度比較を行ったところ、すべてのクレードで複雑なモザイク構造が確認され、本研究で用いた種麴屋が保有する株間では有性生殖が起きていないことを示した。

これらの大局的なゲノム解析の結果から、家畜化は *A. oryzae* の有性生殖を引き起こさず、個別のゲノム内の変化にのみ影響を及ぼしているという仮説を提唱した。すなわち、家畜化の影響はクレード間よりもクレード内の比較によって推定されると考え、これらを区別した比較ゲノム解析を行った。その結果、ゲノム全体における同義置換に対する非同義置換およびフレームシフト変異の頻度が、クレード内で大幅に大きくなっていることを示した。一方で、変異が見られた遺伝子の特徴を調べたところ、産業上有用であると同定されている分解酵素遺伝子のクレード内変異は全く検出されず、すなわち家畜化の過程はこのような重要な遺伝子の変異に対しては保存的に働いていることが明らかになった。このことから、家畜化の影響は、見た目の形質に関わる遺伝子の機能変化、特に LOF を伴う変異を選択していると推察した。産業利用上の特徴に関して、同一クレードの株では大まかな特徴が一致していた。一方で、クレード内での遺伝子変異を細かく調べたところ、白色変異と *wA* 遺伝子への変異には完全な関連性が確認できたが、それ以外の、増殖速度や菌糸の長さと言った形態的と具体的な遺伝子の関連は明らかにすることはできなかった。これらの形質変化は遺伝子間領域やエピジェネティックな制御による可能性も考えられ、具体的な原因を明らかにするためには他の実験的手法を取り入れることが必要となると考える。

本研究で得られた考察は、人間が麴を利用してきた文化的背景とも合致する。種麴屋の存在は、菌株を何度も継代することによって有用な形質を増強させるというよりも、むしろ偶然見つかった有用な株を保存し、安定的な醸造を助けるという役割が大きい。「家畜化」というと、特に哺乳類に対し生殖及び形質変化を促進する意味合いが強いが、コウジカビにおいては、形質の異なる株を別に単離することはあれど、形質変化を抑制する側面が強いと言えるだろう。

参考文献

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410 . [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Arnaud MB, Cerqueira GC, Inglis DO, Skrzypek MS, Binkley J, Chibucos MC, Crabtree J, Howarth C, Orvis J, Shah P, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G, Wortman JR (2012) The *Aspergillus* Genome Database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic Acids Res* 40:D653–D659 . <https://doi.org/10.1093/nar/gkr875>
- Bastiaans E, Debets AJM, Aanen DK, van Diepeningen AD, Saupe SJ, Paoletti M (2014) Natural Variation of Heterokaryon Incompatibility Gene *het-c* in *Podospora anserina* Reveals Diversifying Selection. *Mol Biol Evol* 31:962–974 . <https://doi.org/10.1093/molbev/msu047>
- Bayman P, Cotty PJ (1991) Vegetative compatibility and genetic diversity in the *Aspergillus flavus* population of a single field. *Can J Bot* 69:1707–1711 . <https://doi.org/10.1139/b91-216>
- Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) Species status of *Neisseria gonorrhoeae*: evolutionary and epidemiological inferences from multilocus sequence typing. *BMC Biol* 5:35 . <https://doi.org/10.1186/1741-7007-5-35>
- Bok JW, Keller NP (2004) *LaeA*, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot Cell* 3:527–35 . <https://doi.org/10.1128/ec.3.2.527-535.2004>
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60 . <https://doi.org/10.1038/nmeth.3176>
- Chang P-K, Ehrlich KC (2010) What does genetic diversity of *Aspergillus flavus* tell us about *Aspergillus oryzae*? *Int J Food Microbiol* 138:189–199 . <https://doi.org/10.1016/J.IJFOODMICRO.2010.01.033>
- Chang P-K, Ehrlich KC, Hua S-ST (2006) Cladal relatedness among *Aspergillus oryzae* isolates and *Aspergillus flavus* S and L morphotype isolates. *Int J Food Microbiol* 108:172–177 . <https://doi.org/10.1016/J.IJFOODMICRO.2005.11.008>
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz848>
- Daskalov A, Habenstein B, Sabaté R, Berbon M, Martinez D, Chaignepain S, Couлары-Salin B, Hofmann K, Loquet A, Saupe SJ (2016) Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proc Natl Acad Sci U S A* 113:2720–2725 . <https://doi.org/10.1073/pnas.1522361113>
- David H, Özçelik İŞ, Hofmann G, Nielsen J (2008) Analysis of *Aspergillus nidulans* metabolism at the genome-scale. *BMC Genomics* 9:163 . <https://doi.org/10.1186/1471-2164-9-163>

Deng S, Pomraning KR, Bohutskyi P, Magnuson JK (2018) Draft Genome Sequence of *Aspergillus oryzae* ATCC 12892. Genome Announc 6:e00251-18 . <https://doi.org/10.1128/genomeA.00251-18>

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157 . <https://doi.org/10.1186/s13059-015-0721-2>

Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, Batzoglou S, Lee S-I, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Sczzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MÁ, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438:1105–1115 . <https://doi.org/10.1038/nature04341>

Geiser DM, Dorner JW, Horn BW, Taylor JW (2000) The Phylogenetics of Mycotoxin and Sclerotium Production in *Aspergillus flavus* and *Aspergillus oryzae*. Fungal Genet Biol 31:169–179 . <https://doi.org/10.1006/FGBI.2000.1215>

Geiser DM, Pitt JI, Taylor JW (1998) Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. Proc Natl Acad Sci U S A 95:388–93 . <https://doi.org/10.1073/PNAS.95.1.388>

Geiser DM, Timberlake WE, Arnold ML (1996) Loss of meiosis in *Aspergillus*. Mol Biol Evol 13:809–817 . <https://doi.org/10.1093/oxfordjournals.molbev.a025641>

Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A (2012) The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus *Aspergillus oryzae*. Curr Biol 22:1403–1409 . <https://doi.org/10.1016/j.cub.2012.05.033>

Glass NL, Dementhon K (2006) Non-self recognition and programmed cell death in filamentous fungi. Curr. Opin. Microbiol. 9:553–558

Grubisha LC, Cotty PJ (2009) Twenty-four microsatellite markers for the aflatoxin-producing fungus *Aspergillus flavus*. Mol Ecol Resour 9:264–267 . <https://doi.org/10.1111/j.1755-0998.2008.02378.x>

Gründlinger M, Yasmin S, Lechner BE, Geley S, Schrettl M, Hynes M, Haas H (2013) Fungal siderophore biosynthesis is partially localized in peroxisomes. Mol Microbiol 88:862–875 . <https://doi.org/10.1111/mmi.12225>

Hara S, Jin FJ, Takahashi T, Koyama Y (2012) A further study on chromosome minimization by protoplast fusion in *Aspergillus oryzae*. Mol Genet Genomics 287:177–187 . <https://doi.org/10.1007/s00438-011-0669-1>

Herrera ML, Vallor AC, Gelfond JA, Patterson TF, Wickes BL (2009) Strain-dependent variation in

18S ribosomal DNA Copy numbers in *Aspergillus fumigatus*. J Clin Microbiol 47:1325–32 .
<https://doi.org/10.1128/JCM.02073-08>

Horn BW, Gell RM, Singh R, Sorensen RB, Carbone I (2016) Sexual Reproduction in *Aspergillus flavus* Sclerotia: Acquisition of Novel Alleles from Soil Populations and Uniparental Mitochondrial Inheritance. PLoS One 11:e0146169 . <https://doi.org/10.1371/journal.pone.0146169>

Horn BW, Moore GG, Carbone I (2009) Sexual reproduction in *Aspergillus flavus*. Mycologia 101:423–429 . <https://doi.org/10.3852/09-011>

Horn BW, Sorensen RB, Lamb MC, Sobolev VS, Olarte RA, Worthington CJ, Carbone I (2014) Sexual Reproduction in *Aspergillus flavus* Sclerotia Naturally Produced in Corn. Phytopathology 104:75–85 .
<https://doi.org/10.1094/PHYTO-05-13-0129-R>

Ishitani C, Sakaguchi K (1956) HEREDITARY VARIATION AND RECOMBINATION IN KOJI-MOLDS (*ASPERGILLUS ORYZAE* AND *ASP. SOJAE*). J Gen Appl Microbiol 2:345–400 .
<https://doi.org/10.2323/jgam.2.345>

Islam M-S, Callicott KA, Mutegi C, Bandyopadhyay R, Cotty PJ (2018) *Aspergillus flavus* resident in Kenya: High genetic diversity in an ancient population primarily shaped by clonal reproduction and mutation-driven evolution. Fungal Ecol 35:20–33 . <https://doi.org/10.1016/J.FUNECO.2018.05.012>

Jørgensen TR (2007) Identification and Toxigenic Potential of the Industrially Important Fungi, *Aspergillus oryzae* and *Aspergillus sojae*. J Food Prot 70:2916–2934 . <https://doi.org/10.4315/0362-028X-70.12.2916>

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res 24:1384–95 .
<https://doi.org/10.1101/gr.170720.113>

Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–731 .
<https://doi.org/10.1016/J.JMB.2015.11.006>

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772–780 .
<https://doi.org/10.1093/molbev/mst010>

Kielbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Res 21:487–93 . <https://doi.org/10.1101/gr.113985.110>

Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, Smith RG, Tatusova T, Xiang C, Zherikov A, DiCuccio M, Murphy TD, Pruitt KD, Kimchi A (2016) Assembly: A resource for assembled genomes at NCBI. Nucleic Acids Res 44:D73–D80 . <https://doi.org/10.1093/nar/gkv1226>

- Klich MA (2007) *Aspergillus flavus*: the major producer of aflatoxin. *Mol Plant Pathol* 8:713–722 .
<https://doi.org/10.1111/j.1364-3703.2007.00436.x>
- Kobayashi T, Abe K, Aasi K, Gomi K, Juvvadi PR, Kato M, Kitamoto K, Takeuchi M, Machida M (2007) Genomics of *Aspergillus oryzae*. *Biosci Biotechnol Biochem* 71:646–670 .
<https://doi.org/10.1271/bbb.60550>
- Kodama Y, Shumway M, Leinonen R (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res* 40:D54–D56 . <https://doi.org/10.1093/nar/gkr854>
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19:1639–1645 .
<https://doi.org/10.1101/gr.092759.109>
- Kusumoto KI, Nogata Y, Ohta H (2000) Directed deletions in the aflatoxin biosynthesis gene homolog cluster of *Aspergillus oryzae*. *Curr Genet* 37:104–111 . <https://doi.org/10.1007/s002940050016>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359 .
<https://doi.org/10.1038/nmeth.1923>
- Lee Y-H, Tominaga M, Hayashi R, Sakamoto K, Yamada O, Akita O (2006) *Aspergillus oryzae* strains with a large deletion of the aflatoxin biosynthetic homologous gene cluster differentiated by chromosomal breakage. *Appl Microbiol Biotechnol* 72:339–345 . <https://doi.org/10.1007/s00253-005-0282-5>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079 .
<https://doi.org/10.1093/bioinformatics/btp352>
- Lukashin A, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26:1107–1115 . <https://doi.org/10.1093/nar/26.4.1107>
- Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K-I, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nierman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R, Iwashita K, Juvvadi PR, Kato M, Kato Y, Kin T, Kokubun A, Maeda H, Maeyama N, Maruyama J, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, Yamada O, Yamagata Y, Anazawa H, Hata Y, Koide Y, Komori T, Koyama Y, Minetoki T, Suharnan S, Tanaka A, Isono K, Kuhara S, Ogasawara N, Kikuchi H (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438:1157–1161 . <https://doi.org/10.1038/nature04300>
- Machida M, Yamada O, Gomi K (2008) Genomics of *Aspergillus oryzae*: learning from the history of Koji mold and exploration of its future. *DNA Res* 15:173–83 . <https://doi.org/10.1093/dnares/dsn020>
- Mayorga ME, Timberlake WE (1990) Isolation and molecular characterization of the *Aspergillus nidulans wA* gene. *Genetics* 126:73–79

Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MGB, Rüscher-Gerdes S, Mokrousov I, Aleksic E, Allix-Béguec C, Antierens A, Augustynowicz-Kopeć E, Ballif M, Barletta F, Beck HP, Barry CE, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebæk T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T (2015) Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 47:242–249 .

<https://doi.org/10.1038/ng.3195>

Monahan BJ, Villén J, Marguerat S, Bähler J, Gygi SP, Winston F (2008) Fission yeast SWI/SNF and RSC complexes show compositional and functional differences from budding yeast. *Nat Struct Mol Biol* 15:873–880 . <https://doi.org/10.1038/nsmb.1452>

Mori N, Katayama T, Saito R, Iwashita K, Maruyama J ichi (2019) Inter-strain expression of sequence-diverse HET domain genes severely inhibits growth of *Aspergillus oryzae*. *Biosci Biotechnol Biochem* 83:1557–1569 . <https://doi.org/10.1080/09168451.2019.1580138>

Nakazawa T, Ishiuchi K, Praseuth A, Noguchi H, Hotta K, Watanabe K (2012) Overexpressing Transcriptional Regulator in *Aspergillus oryzae* Activates a Silent Biosynthetic Pathway to Produce a Novel Polyketide. *ChemBioChem* 13:855–861 . <https://doi.org/10.1002/cbic.201200107>

Oda K, Kobayashi A, Ohashi S, Sano M (2011) *Aspergillus oryzae laeA* Regulates Kojic Acid Synthesis Genes. *Biosci Biotechnol Biochem* 75:1832–1834 . <https://doi.org/10.1271/bbb.110235>

Okabe T, Katayama T, Mo T, Mori N, Jin FJ, Fujii I, Iwashita K, Kitamoto K, Maruyama J (2018) BiFC-based visualisation system reveals cell fusion morphology and heterokaryon incompatibility in the filamentous fungus *Aspergillus oryzae*. *Sci Rep* 8:2922 . <https://doi.org/10.1038/s41598-018-21323-y>

Paoletti M (2016) Vegetative incompatibility in fungi: From recognition to cell death, whatever does the trick. *Fungal Biol. Rev.* 30:152–162

Pennerman KK, Yin G, Bennett JW, Hua SST (2019) *Aspergillus flavus* NRRL 35739, a poor biocontrol agent, may have increased relative expression of stress response genes. *J Fungi* 5: . <https://doi.org/10.3390/jof5020053>

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61–D65 . <https://doi.org/10.1093/nar/gkl842>

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. *Nucleic Acids Res* 33:W116–W120 . <https://doi.org/10.1093/nar/gki442>

Ramirez-Prado JH, Moore GG, Horn BW, Carbone I (2008) Characterization and population analysis of the mating-type genes in *Aspergillus flavus* and *Aspergillus parasiticus*. *Fungal Genet Biol* 45:1292–

1299 . <https://doi.org/10.1016/J.FGB.2008.06.007>

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26

Roch S, Steel M (2015) Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor Popul Biol* 100:56–62 .
<https://doi.org/10.1016/j.tpb.2014.12.005>

Rokas A (2009) The effect of domestication on the fungal proteome. *Trends Genet* 25:60–63 .
<https://doi.org/10.1016/J.TIG.2008.11.003>

Rokas A, Payne G, Fedorova ND, Baker SE, Machida M, Yu J, Georgianna DR, Dean RA, Bhatnagar D, Cleveland TE, Wortman JR, Maiti R, Joardar V, Amedeo P, Denning DW, Nierman WC (2007) What can comparative genomics tell us about species concepts in the genus *Aspergillus*? *Stud Mycol* 59:11–17 .
<https://doi.org/10.3114/SIM.2007.59.02>

Sato A, Oshima K, Noguchi H, Ogawa M, Takahashi T, Oguma T, Koyama Y, Itoh T, Hattori M, Hanya Y (2011) Draft Genome Sequencing and Comparative Analysis of *Aspergillus sojae* NBRC4239. *DNA Res* 18:165–176 . <https://doi.org/10.1093/dnares/dsr009>

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212 . <https://doi.org/10.1093/bioinformatics/btv351>

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313 . <https://doi.org/10.1093/bioinformatics/btu033>

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62 .
<https://doi.org/10.1186/1471-2105-7-62>

Szöllősi GJ, Tannier E, Daubin V, Boussau B (2015) The inference of gene trees with species trees. *Syst Biol* 64:e42–e62 . <https://doi.org/10.1093/sysbio/syu048>

Toyotome T, Hamada S, Yamaguchi S, Takahashi H, Kondoh D, Takino M, Kanesaki Y, Kamei K (2019) Comparative genome analysis of *Aspergillus flavus* clinically isolated in Japan. *DNA Res* 26:95–103 . <https://doi.org/10.1093/dnares/dsy041>

Wada R, Jin FJ, Koyama Y, Maruyama J, Kitamoto K (2014) Efficient formation of heterokaryotic sclerotia in the filamentous fungus *Aspergillus oryzae*. *Appl Microbiol Biotechnol* 98:325–334 .
<https://doi.org/10.1007/s00253-013-5314-y>

Wada R, Maruyama J-I, Yamaguchi H, Yamamoto N, Wagu Y, Paoletti M, Archer DB, Dyer PS, Kitamoto K (2012) Presence and functionality of mating type genes in the supposedly asexual filamentous fungus *Aspergillus oryzae*. *Appl Environ Microbiol* 78:2819–29 .

<https://doi.org/10.1128/AEM.07034-11>

Weaver MA, Abbas HK (2019) Field Displacement of Aflatoxigenic *Aspergillus flavus* Strains Through Repeated Biological Control Applications. *Front Microbiol* 10: .

<https://doi.org/10.3389/fmicb.2019.01788>

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875 . <https://doi.org/10.1093/bioinformatics/bti310>

Yoshino-Yasuda S, Fujino E, Matsui J, Kato M, Kitamoto N (2013) Molecular Analysis of the α -Amylase Gene, *AstaaG1*, from Shoyu Koji Mold, *Aspergillus sojae* KBN1340. *Food Sci Technol Res* 19:255–261 . <https://doi.org/10.3136/fstr.19.255>

Zhao G, Yao Y, Hou L, Wang C, Cao X (2014) Draft genome sequence of *Aspergillus oryzae* 100-8, an increased acid protease production strain. *Genome Announc* 2: .

<https://doi.org/10.1128/genomeA.00548-14>

Zhao G, Yao Y, Qi W, Wang C, Hou L, Zeng B, Cao X (2012) Draft Genome Sequence of *Aspergillus oryzae* Strain 3.042. *Eukaryot Cell* 11:1178–1178 . <https://doi.org/10.1128/EC.00160-12>

坂本和俊 (2006) 麹菌ゲノム解析のおしまいは、はじまりなのだ。NRIB(エヌリブ). 3

岩下和裕 (2013) JP6179936B2 - 麹菌株の系統判定法及び醸造特性予測法

付録

遺伝子アノテーションは紙面の都合上省略した。各種データは <http://matsu.bio.titech.ac.jp/datadir> に置いた。

Table S1 本研究で収集した *A. oryzae* 及び *A. sojae* 株
Run はシーケンスのラン番号。Yatalase は DNA 抽出の際に Yatarase を使用し、結果として
Actinobacteria ゲノムの混入が認められたことを示す。

Run	株名	種	種廻屋	用途	Yatalase
1	TK-1	oryzae	P1	sake	
1	TK-2	oryzae	P1	sake	
1	TK-3	oryzae	P1	sake	
1	TK-4	oryzae	P1	sake	
1	TK-5	oryzae	P1	sake	
1	TK-6	oryzae	P1	sake	
1	TK-7	oryzae	P1	miso	
1	TK-8	oryzae	P1	other	
1	TK-9	oryzae	P1	shoyu	
1	TK-10	oryzae	P1	miso	
1	TK-11	oryzae	P1	miso	
2	TK-12	oryzae	P2	shoyu	
2	TK-13	oryzae	P2	shoyu	
2	TK-15	oryzae	P2	miso	
2	TK-16	oryzae	P2	miso	
2	TK-17	oryzae	P2	miso	
2	TK-18	oryzae	P2	sake	
2	TK-19	oryzae	P2	sake	
2	TK-20	oryzae	P2	sake	
2	TK-21	oryzae	P2	sake	
2	TK-22	oryzae	P2	sake	
2	TK-24	oryzae	P2	miso	
2	TK-25	oryzae	P2	sake	
2	TK-26	oryzae	P2	sake	
2	TK-27	oryzae	P2	miso	
2	TK-29	oryzae	P2	shoyu	

2	TK-30	oryzae	P2	sake	
2	TK-31	oryzae	P3	sake	
2	TK-32	oryzae	P3	sake	
2	TK-33	oryzae	P3	sake	
3	TK-14	oryzae	P2	shoyu	
3	TK-23	oryzae	P2	sake	
3	TK-28	oryzae	P2	shoyu	
3	TK-34	oryzae	P3	sake	
3	TK-35	oryzae	P3	sake	
3	TK-36	oryzae	P3	sake	
3	TK-37	oryzae	P3	sake	
3	TK-38	oryzae	P3	sake	
3	TK-39	oryzae	P3	other	
3	TK-40	oryzae	P3	other	
3	TK-41	oryzae	P3	miso	
3	TK-42	oryzae	P3	miso	
3	TK-43	oryzae	P3	miso	
3	TK-44	oryzae	P3	miso	
3	TK-45	oryzae	P3	miso	
3	TK-46	oryzae	P3	miso	○
3	TK-47	oryzae	P3	miso	
3	TK-48	oryzae	P3	shoyu	○
3	TK-49	oryzae	P3	shoyu	
3	TK-50	oryzae	P3	shoyu	○
4	TK-51	oryzae	P4	sake/miso	○
4	TK-52	oryzae	P4	sake/miso	○
4	TK-53	oryzae	P4	sake/miso	○
4	TK-54	oryzae	P4	miso	○
4	TK-55	oryzae	P4	sake/miso	○
4	TK-56	oryzae	P4	miso	○
4	TK-57	oryzae	P4	sake	○
4	TK-58	oryzae	P4	sake	○
4	TK-59	oryzae	P4	shoyu	○
4	TK-60	oryzae	P4	shoyu	○

4	TK-61	oryzae	P4	sake/miso	○
4	TK-62	oryzae	P4	sake/miso	○
4	TK-63	oryzae	P4	sake	○
4	TK-64	oryzae	P4	sake	○
4	TK-65	oryzae	P4	sake	○
5	TK-66	oryzae	P5	shoyu	○
5	TK-67	oryzae	P5	shoyu	○
5	TK-68	oryzae	P5	shoyu	○
5	TK-69	oryzae	P5	shoyu	○
5	TK-70	oryzae	P5	miso	○
5	TK-71	oryzae	P5	miso	○
5	TK-72	oryzae	P5	miso	○
5	TK-73	oryzae	P5	miso	○
5	TK-74	oryzae	P5	miso	○
5	TK-75	oryzae	P5	miso	○
5	TK-76	oryzae	P5	miso	○
5	TK-77	oryzae	P5	miso	○
5	TK-78	oryzae	P5	sake	○
5	TK-79	oryzae	P5	sake	○
5	TK-80	oryzae	P5	sake	○
5	TK-81	oryzae	P5	sake	○
5	TK-82	oryzae	P5	sake	○
2	TK-83	sojae	P2	shoyu	
5	TK-84	sojae	P5	shoyu	○
5	TK-85	sojae	P5	shoyu	○

Table S2 アセンブルの結果及びクレード

クオリティの指標である BUSCO のスコアが 92%以下のものを赤色、96%以下のものを黄色としている。accession には、NCBI Assembly からダウンロードした配列には NCBI Assembly accession number、本研究でシーケンスしたサンプルには BioSample accession number を記載してある。

accession	株名	種	クレード	MAT 型	BUSCO
GCA_000269785.2	3.042	oryzae	K	MAT1-2	98.00%
GCA_000278405.1	AS3.951	oryzae	AS3.951	MAT1-1	97.70%
GCA_000278425.1	AS3.863	oryzae	AS3.863	MAT1-1	97.70%
GCA_000320905.1	RIB326	oryzae	A	MAT1-2	88.80%
GCA_000691885.1	100-8	oryzae	K	MAT1-2	98.00%
GCA_000952835.1	AF70	flavus	AF70	MAT1-1	97.50%
GCA_000956085.1	SU-1	parasiticus	parasiticus	MAT1-1	98.00%
GCA_001576635.1	79-2	flavus	79-2	MAT1-1	90.60%
GCA_001576645.1	61-4	flavus	61-4	MAT1-1	95.20%
GCA_001576655.1	40-5	flavus	40-5	MAT1-1	95.60%
GCA_001576715.1	26-3	flavus	26-3	N.A.	92.50%
GCA_001576725.1	78-6	flavus	78-6	N.A.	94.60%
GCA_001576735.1	72-5	flavus	72-5	MAT1-1	90.80%
GCA_001576745.1	206-4	flavus	206-4	MAT1-1	94.40%
GCA_001576795.1	54-2	flavus	54-2	N.A.	95.70%
GCA_001695535.1	3-2	flavus	3-2	N.A.	72.90%
GCA_002007945.1	BCC7051	oryzae	E	MAT1-2	98.70%
GCA_002214955.1	SRCM101975	oryzae	SRCM	MAT1-2	98.40%
GCA_002214965.1	SRCM101989	oryzae	SRCM	MAT1-2	95.70%
GCA_002217615.1	JAU2	flavus	JAU2	MAT1-2	97.70%
GCA_002217635.1	NRRL21882_AflaGuard	flavus	AflaGuard	MAT1-2	96.90%
GCA_002443195.2	NRRL21882_AflaGuard	flavus	AflaGuard	MAT1-2	98.50%
GCA_002443215.2	NRRL30797_K49	flavus	AF36	contaminated	98.60%
GCA_002456175.2	NRRL118543_AF36	flavus	AF36	MAT1-2	98.80%
GCA_002864195.1	WRRL1519	oryzae	WRRL1519	MAT1-1	98.60%
GCA_002894705.1	ATCC12892	sojae	ATCC12892	MAT1-2	98.40%
GCA_003709025.1	CA14	flavus	CA14	MAT1-2	95.30%
GCA_003711285.1	CS1137_(VCG_MC04)	flavus	CS1137_(VCG_MC04)	MAT1-1	95.00%
GCA_003711305.1	CS0504_(VCG_BS01)	flavus	CS0504_(VCG_BS01)	MAT1-2	94.90%

GCA_003711315.1	CS0540_(VCG_DV901)	flavus	CS0540_(VCG_DV901)	MAT1-1	95.10%
GCA_003711345.1	AF12	flavus	AF12	MAT1-2	95.00%
GCA_003711355.1	AZS04M2A	flavus	AZS04M2A	MAT1-1	95.40%
GCA_003711385.1	AF70	flavus	AF70	MAT1-1	95.20%
GCA_003953485.1	2017_Yazoo_S8	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953495.1	2017_Yazoo_S4	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953505.1	2017_Yazoo_S3	flavus	2017_Yazoo_S3	MAT1-1	95.10%
GCA_003953515.1	2017_Yazoo_S6	flavus	2017_Yazoo_S6	contaminated	95.40%
GCA_003953525.1	2017_Yazoo_S7	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953585.1	2017_Yazoo_S1	flavus	2017_Yazoo_S1	MAT1-1	95.20%
GCA_003953595.1	2017_Washington_T8	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953605.1	2017_Washington_T7	flavus	AflaGuard	MAT1-2	95.10%
GCA_003953615.1	2017_Washington_T4	oryzae	2017_Washington_T4	MAT1-2	95.20%
GCA_003953625.1	2017_Washington_T6	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953685.1	2017_Washington_T3	flavus	2017_Washington_T3	MAT1-2	95.40%
GCA_003953695.1	2017_Coahoma_J8	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953705.1	2017_Washington_T1	flavus	AF36	MAT1-2	95.00%
GCA_003953715.1	2017_Washington_T2	flavus	AF36	MAT1-2	95.00%
GCA_003953735.1	2017_Coahoma_J7	flavus	AflaGuard	MAT1-2	93.80%
GCA_003953785.1	2017_Coahoma_J6	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953795.1	2017_Coahoma_J3	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953805.1	2017_Yazoo_S2	flavus	AflaGuard	MAT1-2	95.30%
GCA_003953825.1	2017_Washington_T5	flavus	2017_Washington_T5	MAT1-1	95.30%
GCA_003953865.1	2017_Coahoma_J1	flavus	2017_Coahoma_J1	MAT1-2	95.40%
GCA_003967615.1	IFM_54693	oryzae	C	MAT1-1	82.10%
GCA_003967635.1	IFM_55053	flavus	IFM_55053	MAT1-2	93.10%
GCA_003967655.1	IFM_57535	oryzae	IFM_57535	MAT1-1	92.60%
GCA_003967675.1	IFM_58503	flavus	IFM_58503	MAT1-1	91.90%
GCA_003967695.1	IFM_59894	oryzae	IFM_59894	MAT1-1	92.50%
GCA_003967715.1	IFM_59975	oryzae	IFM_59975	MAT1-1	93.50%
GCA_003967735.1	IFM_60519	flavus	IFM_60519	MAT1-2	94.00%
GCA_003967755.1	IFM_60655	oryzae	E	MAT1-2	93.40%
GCA_003967775.1	IFM_61224	oryzae	IFM_61224	MAT1-2	84.30%
GCA_003967795.1	IFM_61226	flavus	IFM_61226	MAT1-1	87.90%

GCA_004329145.1	NRRL_35739	oryzae	BP2-1	MAT1-2	93.70%
GCA_004353305.1	BP2-1	oryzae	BP2-1	MAT1-2	91.10%
GCF_000006275.2	NRRL3357	flavus	NRRL3357	MAT1-1	96.90%
GCF_000184455.2	RIB40	oryzae	F	MAT1-1	98.60%
SAMD00154426	TK-7	oryzae	C	MAT1-1	98.60%
SAMD00154422	TK-3	oryzae	F	MAT1-1	98.70%
SAMD00154427	TK-8	oryzae	A	MAT1-2	98.70%
SAMD00154423	TK-4	oryzae	F	MAT1-1	98.70%
SAMD00154420	TK-1	oryzae	A	MAT1-2	98.70%
SAMD00154421	TK-2	oryzae	A	MAT1-2	98.70%
SAMD00154424	TK-5	oryzae	C	MAT1-1	98.60%
SAMD00154428	TK-9	oryzae	E	MAT1-2	98.70%
SAMD00154425	TK-6	oryzae	A	MAT1-2	98.70%
SAMD00154429	TK-10	oryzae	G	MAT1-1	98.60%
SAMD00154430	TK-11	oryzae	F	MAT1-1	98.60%
SAMD00154440	TK-21	oryzae	F	MAT1-1	98.70%
SAMD00154441	TK-22	oryzae	A	MAT1-2	98.70%
SAMD00154443	TK-24	oryzae	TK-24	MAT1-2	98.70%
SAMD00154446	TK-27	oryzae	TK-27 (G)	MAT1-1	98.70%
SAMD00154445	TK-26	oryzae	B	MAT1-1	98.60%
SAMD00154502	TK-83	sojae	sojae	MAT1-1	98.70%
SAMD00154432	TK-13	oryzae	E	MAT1-2	98.60%
SAMD00154431	TK-12	oryzae	H	MAT1-2	98.50%
SAMD00154444	TK-25	oryzae	C	MAT1-1	98.60%
SAMD00154448	TK-29	oryzae	TK-29	MAT1-2	98.70%
SAMD00154449	TK-30	oryzae	B	MAT1-1	98.60%
SAMD00154450	TK-31	oryzae	B	MAT1-1	98.50%
SAMD00154451	TK-32	oryzae	F	MAT1-1	98.70%
SAMD00154452	TK-33	oryzae	A	MAT1-2	98.70%
SAMD00154434	TK-15	oryzae	G	MAT1-1	98.60%
SAMD00154435	TK-16	oryzae	C	MAT1-1	98.60%
SAMD00154436	TK-17	oryzae	A	MAT1-2	98.60%
SAMD00154437	TK-18	oryzae	A	MAT1-2	98.70%
SAMD00154438	TK-19	oryzae	F	MAT1-1	98.70%

SAMD00154439	TK-20	oryzae	B	MAT1-1	98.50%
SAMD00154453	TK-34	oryzae	A	MAT1-2	98.80%
SAMD00154454	TK-35	oryzae	C	MAT1-1	98.60%
SAMD00154455	TK-36	oryzae	A	MAT1-2	98.70%
SAMD00154461	TK-42	oryzae	G	MAT1-1	98.50%
SAMD00154462	TK-43	oryzae	A	MAT1-2	98.80%
SAMD00154463	TK-44	oryzae	A	MAT1-2	98.70%
SAMD00154464	TK-45	oryzae	C	MAT1-1	98.40%
SAMD00154456	TK-37	oryzae	A	MAT1-2	98.80%
SAMD00154457	TK-38	oryzae	F	MAT1-1	98.70%
SAMD00154458	TK-39	oryzae	B	MAT1-1	98.60%
SAMD00154459	TK-40	oryzae	F	MAT1-1	98.70%
SAMD00154460	TK-41	oryzae	G	MAT1-1	98.50%
SAMD00154468	TK-49	oryzae	E	MAT1-2	98.60%
SAMD00154469	TK-50	oryzae	E	MAT1-2	98.70%
SAMD00154465	TK-46	oryzae	C	MAT1-1	98.50%
SAMD00154466	TK-47	oryzae	G	MAT1-1	98.50%
SAMD00154467	TK-48	oryzae	E	MAT1-2	98.60%
SAMD00154433	TK-14	oryzae	H	MAT1-2	98.50%
SAMD00154442	TK-23	oryzae	C	MAT1-1	98.50%
SAMD00154447	TK-28	oryzae	E	MAT1-2	98.60%
SAMD00154480	TK-61	oryzae	A	MAT1-2	98.70%
SAMD00154470	TK-51	oryzae	A	MAT1-2	98.80%
SAMD00154481	TK-62	oryzae	F	MAT1-1	98.60%
SAMD00154471	TK-52	oryzae	A	MAT1-2	98.70%
SAMD00154472	TK-53	oryzae	A	MAT1-2	98.70%
SAMD00154473	TK-54	oryzae	C	MAT1-1	94.60%
SAMD00154482	TK-63	oryzae	A	MAT1-2	98.60%
SAMD00154483	TK-64	oryzae	C	MAT1-1	98.60%
SAMD00154484	TK-65	oryzae	C	MAT1-1	98.60%
SAMD00154474	TK-55	oryzae	C	MAT1-1	98.50%
SAMD00154475	TK-56	oryzae	C	MAT1-1	98.60%
SAMD00154476	TK-57	oryzae	G	MAT1-1	98.60%
SAMD00154477	TK-58	oryzae	A	MAT1-2	98.70%

SAMD00154478	TK-59	oryzae	D	MAT1-1	98.70%
SAMD00154479	TK-60	oryzae	D	MAT1-1	98.10%
SAMD00154494	TK-75	oryzae	C	MAT1-1	98.60%
SAMD00154487	TK-68	oryzae	E	MAT1-2	98.70%
SAMD00154495	TK-76	oryzae	C	MAT1-1	98.60%
SAMD00154488	TK-69	oryzae	E	MAT1-2	98.70%
SAMD00154504	TK-85	sojae	sojae	MAT1-1	98.80%
SAMD00154489	TK-70	oryzae	C	MAT1-1	98.50%
SAMD00154485	TK-66	oryzae	E	MAT1-2	98.70%
SAMD00154496	TK-77	oryzae	A	MAT1-2	98.70%
SAMD00154497	TK-78	oryzae	A	MAT1-2	98.70%
SAMD00154498	TK-79	oryzae	A	MAT1-2	98.80%
SAMD00154499	TK-80	oryzae	A	MAT1-2	98.60%
SAMD00154491	TK-72	oryzae	A	MAT1-2	98.70%
SAMD00154492	TK-73	oryzae	A	MAT1-2	98.60%
SAMD00154493	TK-74	oryzae	C	MAT1-1	98.50%
SAMD00154490	TK-71	oryzae	G	MAT1-1	98.60%
SAMD00154486	TK-67	oryzae	E	MAT1-2	98.60%
SAMD00154500	TK-81	oryzae	A	MAT1-2	98.70%
SAMD00154501	TK-82	oryzae	F	MAT1-1	98.70%
SAMD00154503	TK-84	sojae	sojae	MAT1-1	98.80%

Table S3 ミトコンドリア系統樹解析に用いたリードデータ とアセンブル結果
「長さ」で黄色がけになっているものはアセンブルの失敗として解析には用いなかった。

株名	SRA ID	種	曖昧塩基 N の有無	長さ	%GC
IFM54693	DRR107963	oryzae	—	17,348	27.00%
IFM55053	DRR107964	flavus		29,470	26.20%
IFM57535	DRR107965	oryzae	—	25,863	26.60%
IFM58503	DRR107966	flavus		29,188	26.20%
IFM59894	DRR107967	oryzae	—	4,165	23.80%
IFM59975	DRR107968	oryzae	—	12,439	27.60%
IFM60519	DRR107969	flavus		29,469	26.20%
IFM60655	DRR107970	oryzae		29,207	26.20%
IFM61224	DRR107971	oryzae	—	25,867	26.60%
IFM61226	DRR107972	flavus	—	27,469	26.50%
NRRL3357_wt_T60-1	SRR4142424	flavus		29,206	26.20%
NRRL3357_wt_parent	SRR4142427	flavus	○	29,234	26.10%
NRRL3357_wt_T60-2	SRR4142438	flavus	○	29,206	26.20%
NRRL3357_wt_T0-2	SRR4142439	flavus		29,206	26.20%
NRRL3357_wt_T0-1	SRR4142447	flavus	○	28,925	26.20%
WRRL1519	SRR5906257	oryzae	○	29,240	26.10%
Afla-Guard	SRR7615260	flavus	—	2,000	33.20%
NRRL35739	SRR8526600	oryzae		29,209	26.10%
2017_Yazoo_S7	SRR8554025	flavus	○	28,941	26.20%
2017_Yazoo_S6	SRR8554026	flavus	○	28,823	25.80%
2017_Yazoo_S8	SRR8554027	flavus	○	29,017	26.10%
2017_Yazoo_S3	SRR8554742	flavus	○	28,950	26.20%
2017_Yazoo_S4	SRR8554743	flavus	○	29,078	26.20%
2017_Yazoo_S1	SRR8554744	flavus	○	29,121	26.10%
2017_Yazoo_S2	SRR8554745	flavus	○	28,878	26.30%
2017_Washington_T8	SRR8556558	flavus	○	28,936	26.20%
2017_Washington_T1	SRR8556559	flavus	○	28,780	26.30%
2017_Coahoma_J8	SRR8556560	flavus	○	29,120	26.20%
2017_Coahoma_J7	SRR8556561	flavus	—	17,308	27.70%

2017_Coahoma_J6	SRR8556562	flavus	○	29,023	26.20%
2017_Washington_T5	SRR8556563	flavus	○	28,946	26.20%
2017_Washington_T4	SRR8556564	flavus	○	28,939	26.20%
2017_Washington_T3	SRR8556565	flavus	○	29,127	26.10%
2017_Washington_T2	SRR8556566	flavus	○	29,132	26.20%
2017_Washington_T7	SRR8556567	flavus	○	29,019	26.20%
2017_Washington_T6	SRR8556568	flavus	○	29,181	26.10%
2017_Coahoma_J3	SRR8556698	flavus	○	28,929	26.20%
2017_Coahoma_J1	SRR8556699	flavus	○	29,182	26.10%
3.042 (JX129489.1)		oryzae		29,192	26.20%
RIB40 (AP007176.1)		oryzae		29,202	26.20%
TK-1		oryzae		29,193	26.20%
TK-2		oryzae		29,193	26.20%
TK-3		oryzae		29,202	26.20%
TK-4		oryzae		29,202	26.20%
TK-5		oryzae		29,205	26.20%
TK-6		oryzae		29,193	26.20%
TK-7		oryzae		29,205	26.20%
TK-8		oryzae		29,193	26.20%
TK-9		oryzae		29,207	26.20%
TK-10		oryzae		29,201	26.20%
TK-11		oryzae		29,202	26.20%
TK-12		oryzae		29,205	26.20%
TK-13		oryzae		29,206	26.20%
TK-14		oryzae		29,210	26.20%
TK-15		oryzae		29,202	26.20%
TK-16		oryzae		29,202	26.20%
TK-17		oryzae		29,192	26.20%
TK-18		oryzae		29,192	26.20%
TK-19		oryzae		29,202	26.20%
TK-20		oryzae		29,188	26.20%
TK-21		oryzae		29,202	26.20%

TK-22	oryzae		29,192	26.20%
TK-23	oryzae		29,204	26.20%
TK-24	oryzae		29,203	26.20%
TK-25	oryzae		29,208	26.20%
TK-26	oryzae	○	29,188	26.20%
TK-27	oryzae		29,197	26.20%
TK-28	oryzae		29,206	26.20%
TK-29	oryzae		29,201	26.20%
TK-30	oryzae		29,188	26.20%
TK-31	oryzae		29,188	26.20%
TK-32	oryzae	○	29,202	26.10%
TK-33	oryzae		29,192	26.20%
TK-34	oryzae		29,192	26.20%
TK-35	oryzae		29,205	26.20%
TK-36	oryzae		29,192	26.20%
TK-37	oryzae		29,192	26.20%
TK-38	oryzae		29,202	26.20%
TK-39	oryzae		29,188	26.20%
TK-40	oryzae	○	29,202	26.20%
TK-41	oryzae		29,202	26.20%
TK-42	oryzae		29,202	26.20%
TK-43	oryzae		29,192	26.20%
TK-44	oryzae		29,192	26.20%
TK-45	oryzae		29,205	26.20%
TK-46	oryzae	○	29,205	26.20%
TK-47	oryzae		29,202	26.20%
TK-48	oryzae		29,206	26.20%
TK-49	oryzae		29,206	26.20%
TK-50	oryzae		29,206	26.20%
TK-51	oryzae		29,192	26.20%
TK-52	oryzae		29,192	26.20%
TK-53	oryzae		29,192	26.20%
TK-54	oryzae		29,205	26.20%
TK-55	oryzae		29,205	26.20%

TK-56	oryzae		29,205	26.20%
TK-57	oryzae		29,202	26.20%
TK-58	oryzae		29,192	26.20%
TK-59	oryzae		29,201	26.20%
TK-60	oryzae		29,201	26.20%
TK-61	oryzae		29,192	26.20%
TK-62	oryzae		29,202	26.20%
TK-63	oryzae		29,192	26.20%
TK-64	oryzae		29,205	26.20%
TK-65	oryzae		29,205	26.20%
TK-66	oryzae		29,206	26.20%
TK-67	oryzae		29,206	26.20%
TK-68	oryzae		29,206	26.20%
TK-69	oryzae		29,206	26.20%
TK-70	oryzae		29,205	26.20%
TK-71	oryzae		29,201	26.20%
TK-72	oryzae		29,192	26.20%
TK-73	oryzae		29,192	26.20%
TK-74	oryzae		29,205	26.20%
TK-75	oryzae	○	29,205	26.20%
TK-76	oryzae		29,202	26.20%
TK-77	oryzae		29,192	26.20%
TK-78	oryzae		29,192	26.20%
TK-79	oryzae		29,192	26.20%
TK-80	oryzae		29,192	26.20%
TK-81	oryzae		29,192	26.20%
TK-82	oryzae		29,202	26.20%
TK-83	sojae		29,136	26.20%
TK-84	sojae		29,136	26.20%
TK-85	sojae		29,136	26.20%

Table S4 *A. nidulans* 及び *A. fumigatus* との SB/NSB における変異頻度比較

		同義置換	非同義置換	frameshift	
<i>A. oryzae</i>	クレード内	SB	156	325	65
		NSB	108	193	28
	クレード間	SB	48,844	40,626	1,080
		NSB	27,587	25,211	708
<i>A. flavus</i>		SB	50,371	32,959	853
		NSB	28,571	20,809	475

Table S5 クレード内非同義置換またはフレームシフト変異があった、AspGD において遺伝子シンボルが与えられていた遺伝子

RIB40_s01-m09-r06	symbol	アノテーション
クレード内非同義置換のあった遺伝子		
AO090003001208	<i>amyR</i>	Regulatory protein that indirectly affects production of hemicellulolytic and cellulolytic enzymes, likely through a carbon catabolite repression-mediated control
AO090003000489	<i>laeA</i>	Methyltransferase; global transcriptional regulator of secondary metabolic gene clusters; required for kojic acid gene regulation and biosynthesis;
AO090009000638	<i>steA</i>	Ortholog of <i>Ste12p</i> with a predicted role in regulation of transcription
AO090001000237	<i>veA</i>	Ortholog of <i>A. nidulans</i> VeA, a global gene regulator involved in light-sensitive control of differentiation and secondary metabolism; positively regulates penicillin production of <i>A. oryzae</i>
AO090003000491	<i>rpbA</i>	Predicted RNA polymerase II largest subunit; has 25 repeats in its C-terminal domain
AO090026000360	<i>gprD</i>	Family A G-protein coupled receptor (GPCR)-like
AO090001000439	<i>schA</i>	Ser/Thr protein kinase related to the PKA catalytic subunit
AO090001000512	<i>cyaA</i>	Adenylate cyclase
AO090003001305	<i>aglB</i>	Putative alpha-galactosidase; expression altered by <i>manR</i> disruption but not by ManR overexpression
AO090003001507	<i>tgLA</i>	Triacylglycerol lipase with role in degradation of triglycerides
AO090001000445	<i>CYP505A3</i>	Cytochrome P450 monooxygenase

AO090005000070	<i>CYP620H3</i>	Cytochrome P450 monooxygenase; involved in 7-hydroxycoumarin production
AO090012000465	<i>CYP620H9</i>	Cytochrome P450 monooxygenase; involved in 7-hydroxycoumarin production
AO090038000488	<i>csyC</i>	Putative type III polyketide synthase
AO090011000926	<i>dffA</i>	L-ornithine N5-oxygenase; enzyme required for biosynthesis of an iron-chelating compound deferriferrichrysin; siderophore biosynthesis
AO090001000009	<i>wykN</i>	Non-ribosomal peptide synthase (NRPS) involved in the synthesis of a dipeptidyl peptidase IV 2 inhibitor
AO090102000632	<i>can1</i>	Has domain(s) with predicted role in amino acid transport, transmembrane transport and integral component of membrane, membrane localization
AO090103000127	<i>bg1F</i>	Secretory aryl beta-glucosidase
AO090003001144	<i>vti1</i>	T-SNARE
AO090701000589	<i>chsB</i>	Chitin synthase; required for normal hyphal growth and conidiation
AO090026000337	<i>sec31</i>	Vesicle coat complex COPII, subunit; expression increased in MAT1-2 strain compared to MAT1-1 strain
AO090102000545	<i>wA</i>	Hydroquinone:oxygen oxidoreductase; orthologs have a role in asexual spore wall assembly, melanin biosynthesis, pathogenesis and pigment metabolism

クレード内フレームシフト変異のあった遺伝子

AO090102000545	<i>wA</i>	(Above)
AO090003000489	<i>laeA</i>	(Above)
AO090009000612	<i>amdA</i>	Sequence-specific DNA binding transcription factor

Table S6 *A. oryzae* / *A. flavus* に特異的な SNP が存在する遺伝子。

	染色体	position	symbol
OG0004956	I	375,711	
OG0000839	I	593,814	
OG0000952	I	1,309,622	
OG0007593	I	5,426,647	
OG0008742	I	5,620,192	
OG0001480	I	5,621,886	
OG0005320	I	5,907,389	
OG0007616	I	5,971,264	
OG0007617	I	5,983,241	
OG0000484	I	6,220,001	
OG0000448	II	9,327	
OG0001753	II	1,542,138	
OG0008841	II	5,174,860	
OG0007845	II	5,349,000	
OG0005751	II	5,350,682	
OG0000624	II	5,356,625	
OG0007846	II	5,364,168	
OG0002167	II	5,710,874	
OG0007854	II	5,805,683	
OG0002184	II	5,819,576	
OG0005840	III	180,363	
OG0002439	III	2,006,296	
OG0002443	III	2,021,303	
OG0002792	III	4,811,629	
OG0006137	III	4,813,620	
OG0008932	III	4,819,374	
OG0002797	III	4,896,327	
OG0002798	III	4,901,505	
OG0002803	III	4,933,687	
OG0006471	V	442,923	
OG0003449	V	1,207,583	

OG0003458	V	1,359,950	
OG0003604	V	2,696,550	
OG0006613	V	2,698,635	
OG0003619	V	2,771,228	
OG0006802	VI	586,318	
OG0003856	VI	587,676	
OG0008373	VI	590,680	
OG0003859	VI	603,678	
OG0003860	VI	606,620	
OG0003861	VI	620,795	
OG0003884	VI	843,749	
OG0006864	VI	1,491,302	
OG0008486	VII	325,124	
OG0009142	VII	517,571	
OG0009497	VII	642,936	
OG0009925	VII	661,912	
OG0004310	VII	684,545	
OG0008498	VII	686,416	
OG0007069	VII	691,008	
OG0004365	VII	1,100,451	
OG0011554	VII	1,103,282	<i>CYP620K1</i>
OG0008522	VII	1,495,111	
OG0011646	VII	1,506,504	
OG0004423	VII	1,526,515	
OG0007124	VII	1,532,825	
OG0007128	VII	1,548,275	
OG0004429	VII	1,579,029	
OG0004432	VII	1,605,437	
OG0004433	VII	1,609,896	
OG0004437	VII	1,623,370	
OG0009161	VII	1,627,525	
OG0007145	VII	1,788,589	
OG0000779	VII	1,825,029	
OG0008580	VIII	578,596	

OG0004622	VIII	646,220	
OG0007299	VIII	1,744,868	
OG0007388	VIII	3,283,215	<i>aoiF</i>
OG0007389	VIII	3,285,802	<i>aoiE</i>

Table S7 NRIB プライマーのインサート領域の検出パターン(bp)

	192	089	027	025	225	367	217	239	091	140	161	187	232
d_0	○	○			○		○	○	○	○	○	○	○
d_1	○	○			○	○	○	○	○	○	○	○	○
d_2	○	○	○	○	○	○	○	○	○	○	○	○	○
d_3	○			○	○		○		○	○	○	○	○
d_4	○	○	○	○	○		○	○	○	○	○	○	○
d_5	○	○			○		○		○	○	○	○	○
d_6	○	○	○	○	○	○		○	○	○	○	○	○
d_7	○	○	○		○		○	○	○	○	○	○	○
d_8	○	○	○	○	○	○		○	○	○	○	○	○
d_9	○	○	○	○	○	○	○	○	○	○	○	○	○
d_10		○	○	○	○		○	○	○	○	○	○	○
d_11	○	○			○	○	○		○	○	○	○	○

NRIB 分類	192	089	027	025	225	367	217	239	091	140	161	187	232
A	○	○			○	○	○		○	○	○	○	○
B	○	○	○		○		○			○	○	○	○
Sake/Miso	○	○			○		○			○		○	○
SO1	○	○			○		○	○	○	○	○	○	○
SO2	○				○		○		○	○	○	○	○
SO3		○	○		○		○	○	○		○		○
SO4		○	○		○		○	○	○	○			
C	○	○	○	○	○	○		○	○	○	○	○	○
RIB40	○	○	○	○	○	○	○	○	○	○	○	○	○
RIB40-2		○	○	○	○	○	○	○	○	○	○	○	
NS1	○	○	○	○		○	○	○		○	○	○	○
NS2	○	○	○		○	○	○	○	○	○	○	○	○
NS3	○	○	○		○	○		○		○	○	○	○

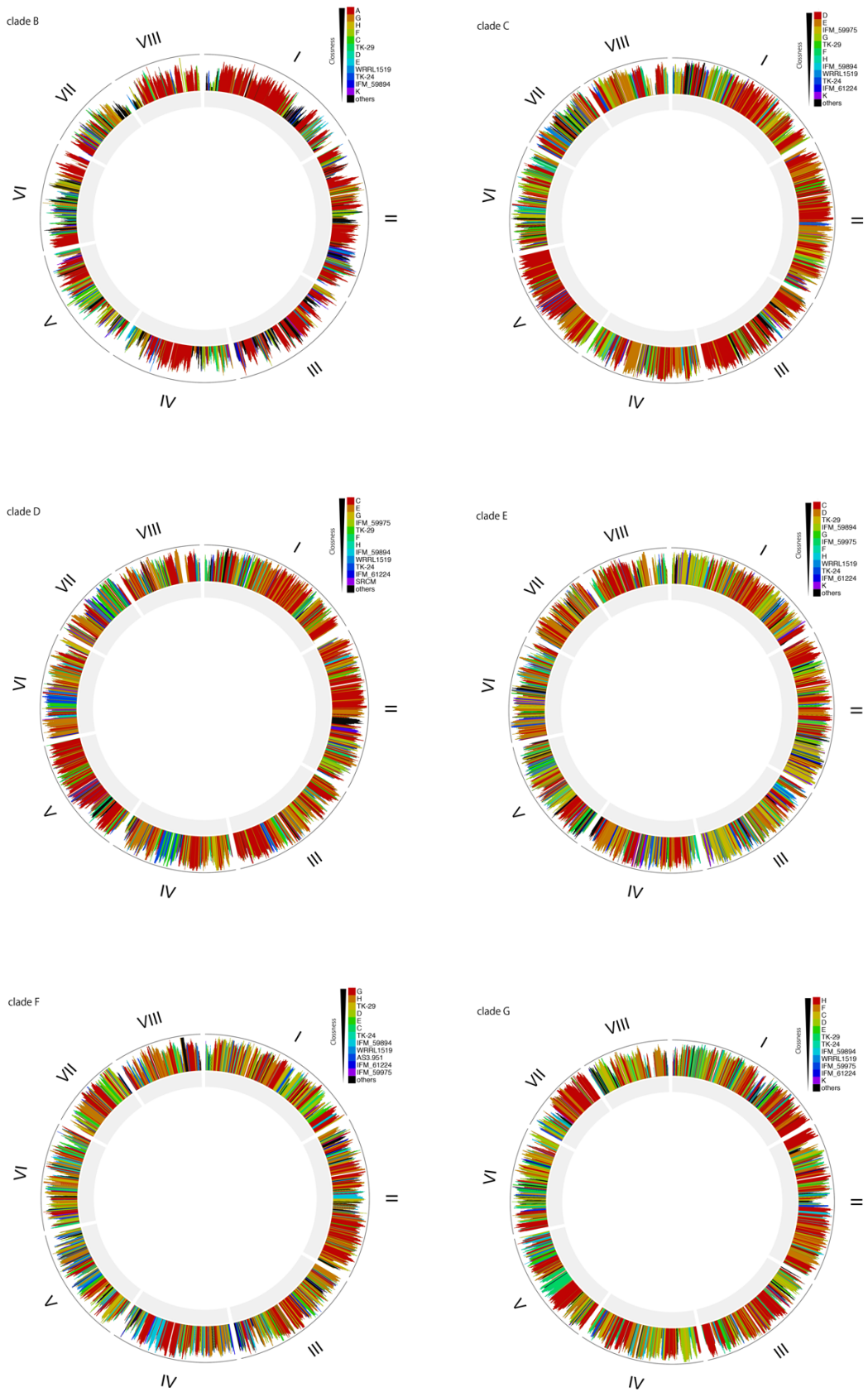


Fig. S2 クレード B-G の混合図 (TK-27 はクレード G としてある)

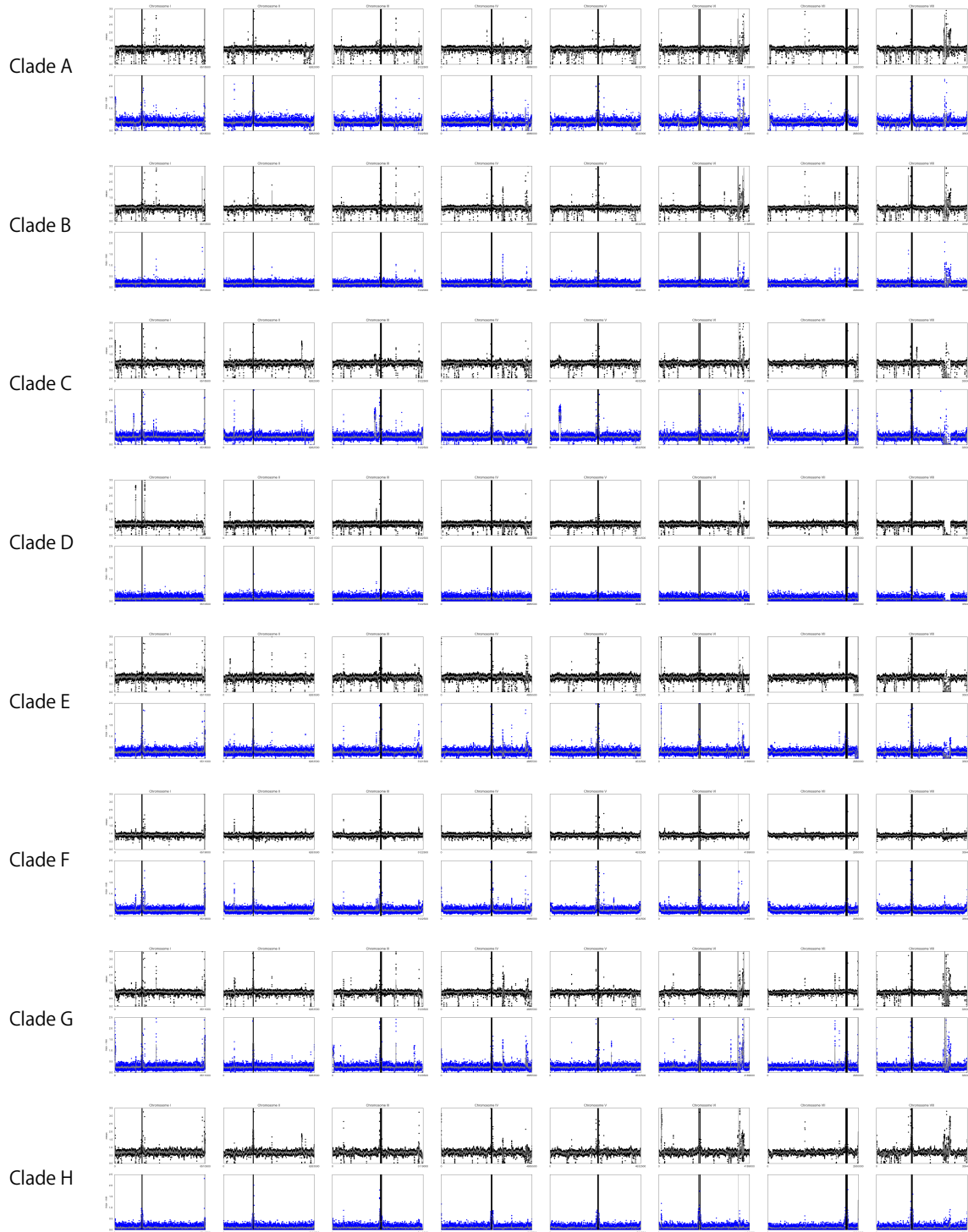


Fig. S3 クレード内 normalized depth の変化
 上段(青)が平均値、下段(黒)が最大値と最小値の差。左から染色体 I-VIII の順。
 曖昧塩基 N の領域は縦に黒く塗りつぶしてある。

謝辞

本研究を実施するに当たって、まずこのような大規模なデータ提供を下された「ぐるなび食の価値創成 共同研究講座」の皆様、資金的援助を下されたぐるなび株式会社様に深く感謝いたします。麹菌株の比較ゲノムは RIB40 株の完全ゲノム解読以来いくつか挑戦されてきましたが、予算的な都合もあってか、既知の株の多様性を網羅できるようなデータを扱うものではありませんでした。これを可能にしたのは、共同研究講座の「食文化と発酵微生物とを科学で繋げる」という独特な目標と、大規模な予算とが合致した結果だと考えます。

また、研究に関する的確なアドバイスを下さった国立遺伝学研究所黒川顕先生、森宙史先生、指導教員である東京工業大学生命理工学院山田拓司先生、また博士論文審査員の先生方、研究室の皆様へ感謝申し上げます。私自身 *A. oryzae* のゲノムを研究対象としたのは博士課程からであり、真菌に対する実験的な知識もありませんでした。まずゲノムを解析するというところから、データドリブンに仮説を積み上げていき、本論のような形にまとめることができた次第です。データを扱い出した当初は、シングルゲノムのアSEMBルも初めてで、見様見真似で先行研究のツールを叩き、先生方にご指摘いただいてコンタミの存在にも徐々に気づいていくような状態でした。株の多様性も把握できるようなものではなく、目星すらもない状態で、系統推定もさまざまな方法を用いて行っていました。最終的にきれいな形でいろいろな要素と整合性がつくような手法にたどり着きましたが、本論を書きながら思い返せばもう少しいい形にできたという部分もあり、扱うサンプルそれぞれの解析が必要だということを再認識させられているところです。

ならびに、麹菌株の提供はもとより、研究結果に対してさまざまなご意見を下さった、株式会社菱六様、株式会社ビオック様、株式会社樋口松之助商店様、株式会社秋田今野商店様、日本醸造工業株式会社様には多大なる感謝を申し上げます。解析を進めていく中で、種麹屋の皆様とのお話は非常に刺激的で貴重なものでした。特に、麹菌のゲノム解析から得た考察が、実際の使用の場での知見と一致するたびに、興奮し知的好奇心が掻き立てられました。自分自身、このテーマは基礎研究であるとして進めてきましたが、応用の場が近いということで推論の幅が広がったことは確かです。麹菌の進化と家畜化の関係というのは、ある種の歴史学であり、それそのものが産業に活かされるという類のものではありませんが、大規模な解析データを世に出したという責任と、種麹屋の皆様方への感謝を込めて、情報学的には自分にできる限りの解析を行ったと自負しています。

最後に、麹菌の研究に携ることができた幸運に感謝いたします。*Aspergillus oryzae* の名を最初に知ったのは石川雅之先生作の漫画『もやしもん』でした。思い返せば、高校生だった私はこれを読んで醸造学の深さを知り、生命系への興味を持ちました。学部時代に受講した黒川先生の講義をきっかけに生命情報の道へと舵を切りましたが、長らく頭の片隅にあった高校時代の経験が、本研究へと導いてくれたのではないかと考えます。株式会社秋田今野商店様にお邪魔した際、石川先生のサインを見つけたときに、このような思いがめぐりました。

麹菌は我々日本人には欠かせない存在で、和食の食卓を見れば必ずどこかに関わっています。日々生活の中で博士課程における研究を思い返す機会があるという幸運に感謝し、この経験に誇りを持って今後の糧としていきたいと考えています。