

麴菌の総合ゲノムデータベースを10年ぶりに大規模リニューアル

～麴菌群総合ゲノムデータベースの公開～

(Comprehensive *Aspergillus oryzae* genome database across the clades: CAoGDX)

麴菌総合ゲノムデータベース (CAoGD) を麴菌群総合ゲノムデータベース (CAoGDX) として大幅リニューアルしました。CAoGD 同様に、CAoGDX は我が国の国菌である麴菌の学術研究を支える基盤的なデータベースとして広く利用されるとともに、麴菌研究全体を活性化させることが期待されています。



CAoGDX ロゴ

CAoGDX アドレス: <https://nribf21.nrib.go.jp/CAoGDX/>



2次元コード

🧬 アップデートのポイント

- 麴菌 (*Aspergillus oryzae*) 研究の基盤となる RIB40 株の完全長ゲノムシーケンスを掲載。
(Telomer to Telomer genome sequence of *A. oryzae* RIB40)
- 遺伝子発現情報や遺伝子機能解析などを完全長 RIB40 ゲノムシーケンスに統合
- 218 株の麴菌群のゲノムシーケンス解析から作成した麴菌群の系統樹を掲載
- 麴菌群の系統樹を元に選抜した 23 株の完全長ゲノムシーケンスを掲載
- GO ターム、各遺伝子の文献情報、各遺伝子の麴菌群での多型情報等を新たに掲載
- 産官学の機関で実施した EST (Expression sequence tag) 解析データを統合
- 多くの麴菌研究者の意見を反映し、データの表示方法など細やかな使用方法の改善

🧬 麴菌群総合ゲノムデータベースについて

< 麴菌総合ゲノムデータベース (CAoGD) について >

麴菌 (*A. oryzae*) のゲノムシーケンスは当研究所を含む産官学の連携プロジェクトとして、2005年にドラフトゲノムシーケンスが発表されました。当研究所では、ゲノムシーケンスを使用した麴菌の分子生物学的な研究をプロモートするために、麴菌と他の糸状菌のゲノムを比較解析可能な、麴菌比較ゲノムデータベース (CFGD) を作成いたしました (2011年)。その後、我々は麴菌のゲノム情報を基に麴菌 DNACHIP の開発を行い、系統解析や遺伝子発現解析を行いました。これらのゲノム研

究の成果は、順次 CFGD に掲載し、2014 年には麴菌総合ゲノムデータベース (CAoGD) と名称を改めリニューアルしました。

<CAoGD の課題と麴菌ゲノム研究の進展>

CAoGD となっても、麴菌 RIB40 株のゲノムシーケンスはアップデートされ、遺伝子予測も新しいバージョンとなりました。その結果、CAoGD 中には、RIB40 株ゲノムシーケンスが 3 つのバージョン存在することとなりました。同じ株のゲノムシーケンスが 3 つあることは、奇異な事かもしれませんが、技術の発展途上の段階ではよくあることです。これらのバージョンでは、予測された遺伝子数やその領域 (配列) がわずかに異なり、各バージョンで行なったゲノム研究の結果はそのバージョンに紐付けされてしまいます。つまり、各バージョンで独自の解析結果が掲載されており、ある遺伝子について、全ての情報を網羅するためには、全 RIB40 ゲノムシーケンスのバージョンを確認する必要があります。

また、RIB40 株のゲノムシーケンス自体が不完全なもので多くのギャップ (シーケンスが繋がっていないところ) を含んでおりました。そこで、当研究所では株式会社ジーンペイと協力し RIB40 株の完全長ゲノムシーケンス (テロメアからテロメアまでギャップがないゲノムシーケンス) を取得しました。また、多くの RNA-seq データを使用して遺伝子予測の改善を行いました。完全長のゲノムシーケンスが得られたことから、複数バージョンにわたっていた情報についても、完全長ゲノムシーケンスのものとの関連性を解析し統合しました。

この完全長 RIB40 株ゲノムシーケンスを参照配列として、清酒用、味噌用、醤油用種もやしなどから単離され、当研究所に保存されている 136 株の麴菌について、新たにドラフトゲノムシーケンスを取得しました。さらに、公的なゲノムデータベースに登録されている全麴菌株のゲノムシーケンス情報を加えて、これらの菌株すべてに保存されている遺伝子 (コア遺伝子) の配列を利用して、218 株の麴菌株の系統解析を行いました。その系統樹を参考に、合計 23 株について完全長ゲノムシーケンスを取得いたしました。

<麴菌群総合ゲノムデータベース (CAoGDX) の開発>

麴菌群のゲノム研究の進展により多くの情報が得られましたので、CAoGD を 10 年ぶりに大幅リニューアルすることとし、麴菌群総合ゲノムデータベース (CAoGDX) を開発しました。CAoGDX には、完全長 RIB40 株のゲノムシーケンスをベースに、各遺伝子について配列情報やアノテーション、モチーフ情報などの基本的な情報、EST やマイクロアレイ、RNA-seq 等の遺伝子発現情報を統合し掲載しました。さらに、麴菌 23 株間での各遺伝子の多型情報、麴菌群の系統解析の情報、GO 情報や、文献情報など、新たに取得した多くの情報を掲載いたしました。また、これまでの CAoGD に対する改善の希望や新たな要望について、多くの麴菌研究者に対して調査を行い、データの表示方法やデータベースの使用方法等、細かな改善を行っています。

CAoGDX は β 版を 2024 年 12 月に公開し、バグの洗い出しを行うとともに、日本の麴菌研究者から様々なフィードバックをいただきました。 β 版の公開により明らかになったバグの修正や改善を行い、この度、CAoGDX を正式に公開いたします。

本データベースが世界で唯一の麴菌群に特化したゲノムデータベースとして、麴菌研究者に基盤

情報を提供するとともに、麴菌研究の発展に大きく貢献することを願っています。

令和7年3月31日

CAoGDX 開発責任者

独立行政法人酒類総合研究所 業務統括部門長 岩下和裕

CAoGDX についての連絡先

醸造微生物研究部門 主任研究員 織田健

成分解析研究部門 研究員 片岡涼輔

問合せ e-mail: caogdx@nrib.go.jp