

課題名：塩基配列アーカイブのデータベース構築と統合への貢献




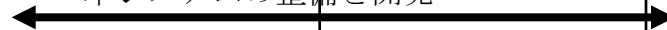
担当：国立遺伝学研究所 五條堀孝

事業の目標・概要

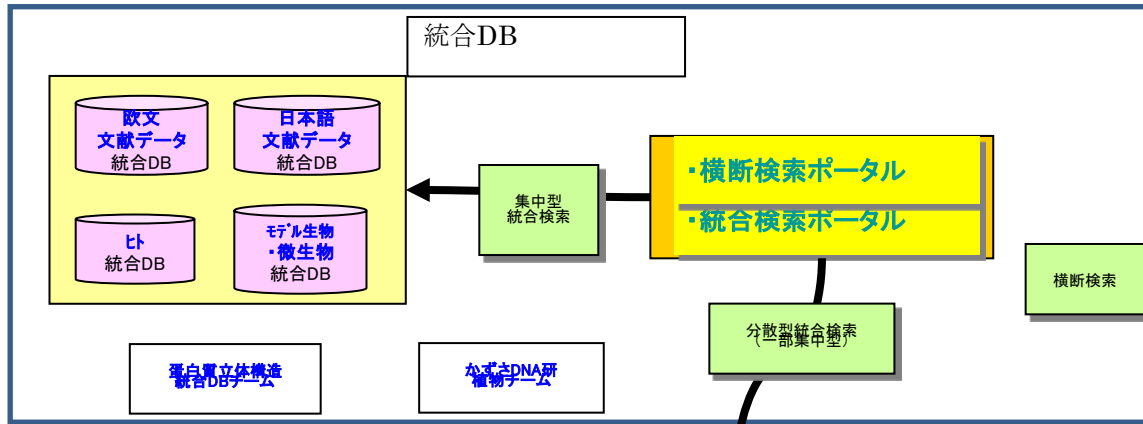
各種生物の遺伝子やゲノムの塩基配列を決定するいわゆるシーケンシングセンターにおいてはその配列決定の原データになる波形データなどのいわゆるアーカイブ(Trace archive; 以下Traceデータという)を有している。このTraceデータは、品質管理やそれを基にして行う配列決定アルゴリズムの改良ならびに配列断片を連結するアセンブリにおいて大変重要で貴重な情報である。そして、これらのTraceデータは、シーケンシングセンターの活動がその支持母体のプロジェクトの完了とともに終了するとすると、原則的には完全に消滅してしまう可能性が極めて高い状況にある。また、454やSolexaあるいはABI-SOLiDといった次世代の超高速の塩基配列決定装置の登場により、そのTraceデータの量は飛躍的に巨大化してきており、シーケンシングセンター自身でもそのデータハンドリングを含めて保存はもちろんのこと対処が非常に困難な状況になっている。

この状況の理解の下、わが国における塩基配列決定におけるTraceデータの保存と有効利用を目的として、当機関である大学共同利用機関法人 情報・システム研究機構の国立遺伝学研究所生命情報・DDBJ研究センターのDDBJが、Traceデータのデータベース構築事業とデータ提供の事業を実施する。

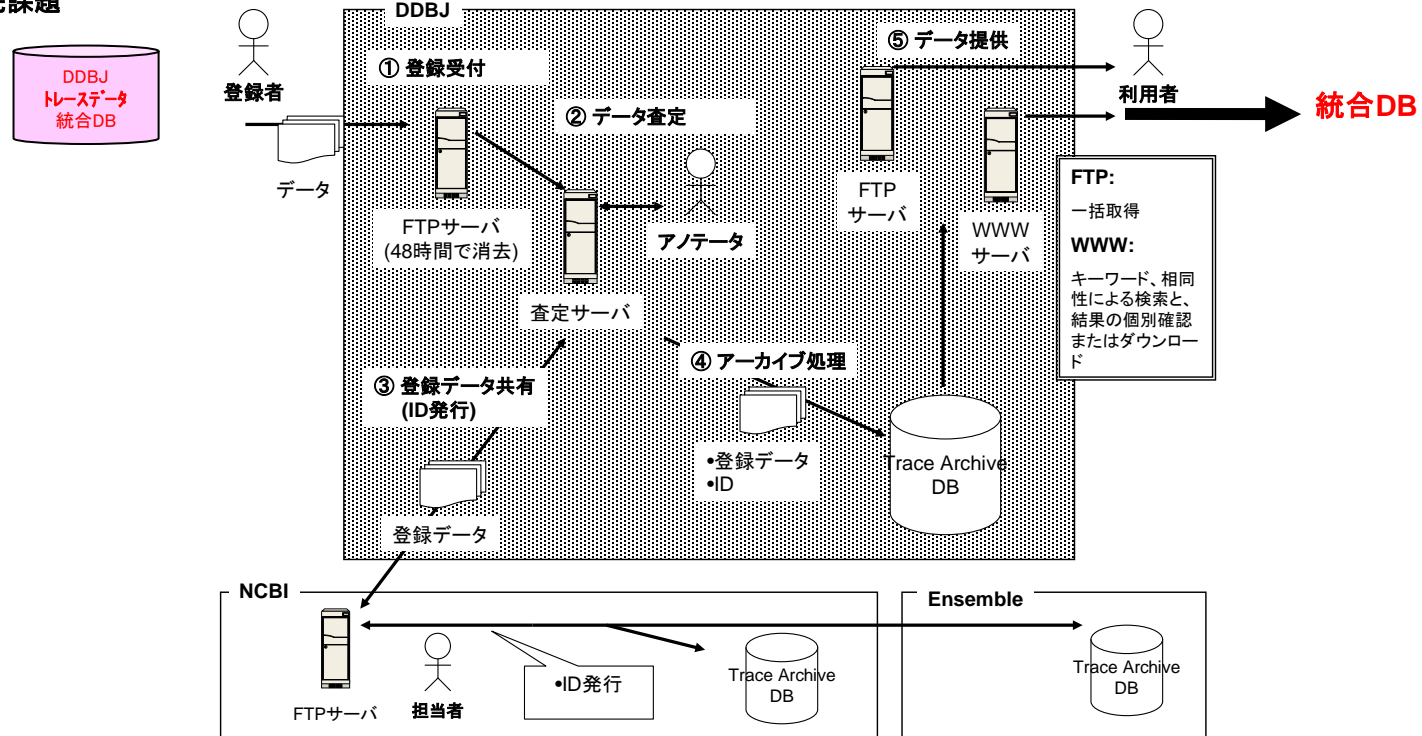
実施計画

実施項目	1年度目	2年度目	3年度目	4年度目
(1) 公開FTPサイトとWWWサイトに関する開発 (参画研究機関) なし	基盤技術整備 	公開サイトシステムの整備と開発 		取りまとめ 
(2) 登録処理および波形表示システムに関する開発 (参画研究機関) なし	試験システムの開発 	本システムの整備と開発 		

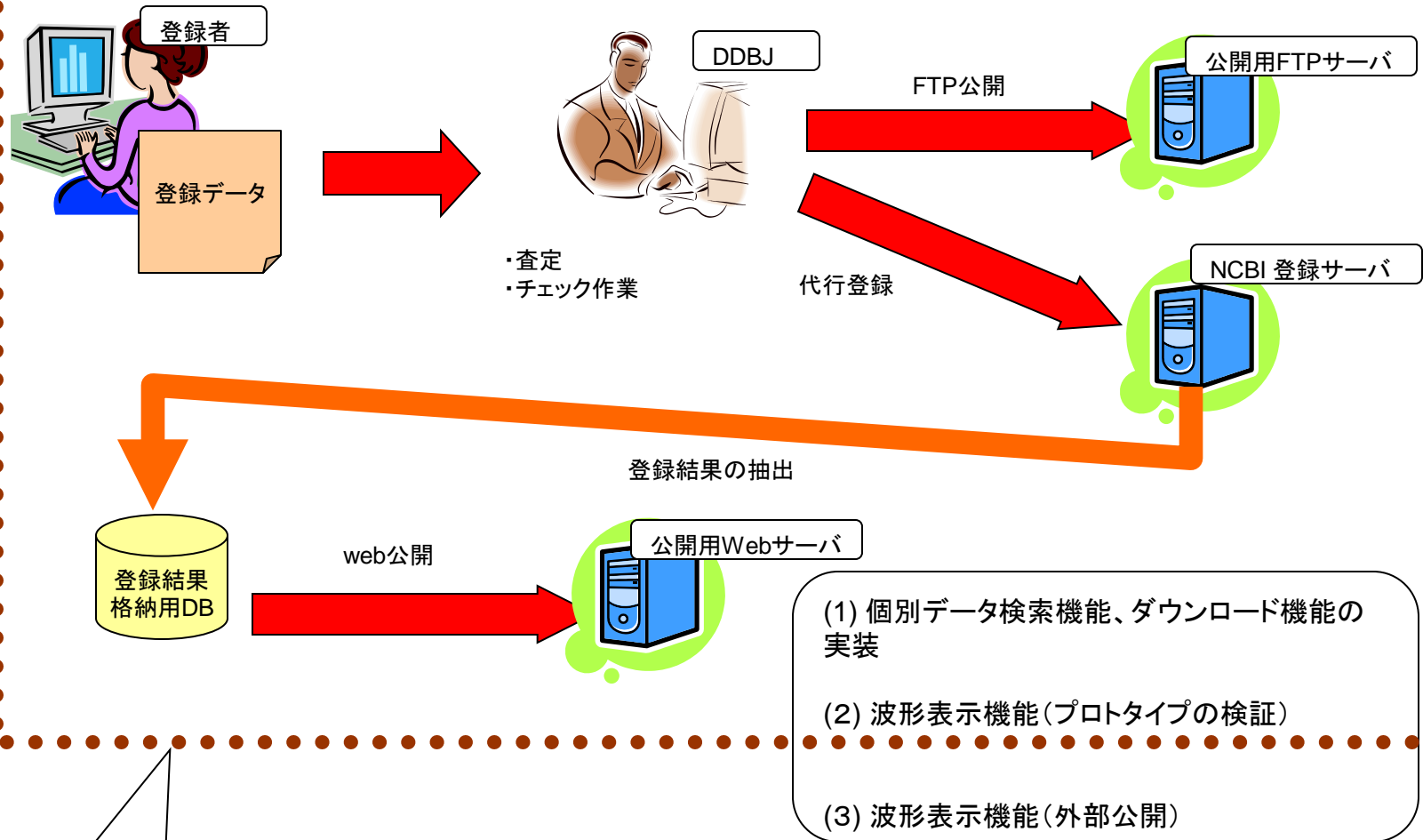
補完課題における連携の概要



補完課題



Trace Archive 処理概要



今年度達成目標範囲

Trace Archive 登録・公開実績概要

項番	機関名	対象の生物種	件数	圧縮時の容量 (bytes)	データ受付時期	現在のステータス
1	国立遺伝学研究所(NIG) 小原研究室	メダカ	約148万 件	圧縮後で約50G B	2008年3月上旬	登録・公開完了(7月上旬)
2	東京大学新領域創成科学 研究科(UTCOP) 服部研究室	ヒトメタゲノム	約106万 件	圧縮後で約40G B	2008年3月下旬	登録・公開完了(7月上旬)

The screenshot shows the DDBJ (DNA Data Bank of Japan) website interface. At the top, there are navigation links for 'Accession', 'DNA', 'Protein', 'Taxonomy', and 'Site Search'. Below this is a search bar for 'Accession numbers' with a 'Go' button. A menu bar includes 'HOME', '塩基配列の登録', '利用の手引き', '検索・解析', 'FTP・WebAPI', 'レポート・統計', and 'お問い合わせ'. The main content area features a search result for 'Trace Archive' with a '統合' (Integration) button and a text box stating '受付・公開を始めました' (Started receiving and publishing). Below the text is a DNA sequence chromatogram with the sequence: TTGC TGC CTG CA GTCG ACCTA GAGGAT CCCCAT CAT TCAACTATT GAAAA GC GCT CT. The DDBJ logo and 'DNA Data Bank of Japan' are visible at the bottom of the chromatogram.

DDBJは長年の懸案であった Trace Archive に、昨年度下期から [ライフサイエンス統合データベースプロジェクト](#) の補完課題を受けて取組を始め、2008年7月に2件を公開するに至りました。

公開ページ

国立遺伝学研究所、DNA配列解読で得られる一次データの保全システムを構築

大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJ 研究センター(五條堀 孝センター長)は、シーケンサーの一次出力データ(トレースデータ)の保全と公開を行う事業、通称 [トレースアーカイブ](#) の試験運用を平成20年8月19日より開始致しました。本サイトからは、日本国内の少数の協力サイトから提供を受けた合計約254万トレースのデータ(平成20年8月19日現在)がFTP(ftp://ftp.ddbj.nig.ac.jp/ddbj_database/trace/)よりダウンロードできます。本事業は、大学共同利用機関法人 情報・システム研究機構 ライフサイエンス統合データベースセンター(高木利久センター長)を中核とした文部科学省委託研究開発事業「[統合データベースプロジェクト](#)」の一環として行われています。

ニュースリリース