

質問	回答
前半の講義のデーターベースの関係と登録等の手順を一つの図に示したものはありますか？ マニュアルとしてまとめられていますか？ AIに頼めば作図してくれるでしょうか？	DDBJのデータ種類は、ゲノム、転写産物、1つの遺伝子など、様々なデータベースがあるため、1つのマニュアルにまとめるということは難しいのですが、DDBJのサイト (https://www.ddbj.nig.ac.jp/index.html) では、登録のナビゲーションを提供しています。トップページの「登録」の先に「登録ナビゲーション」があります。提供予定のデータがどのようなものかを選んでいくと、何をどういう順番で登録したらいいのかを提示してくれるナビゲーションです。こちらをご利用いただくとわかりやすいと思います。
SRAに登録されたリードはどれくらいでアセンブルになるのでしょうか？	これは本当にまちまちです。 例えば、ゲノムのシークエンスが登録され、そのDRAの番号が論文には書いてあるが、アセンブルされた配列が登録されてないというケースもありますし、生リードもアセンブル配列も両方登録するという場合もあり、これは研究者によってまちまちです。最近では、アセンブルがあつたりなかったりになってきています。 特にバクテリアやウイルスのゲノムといったアセンブルが比較的容易なものに関しては、SRAだけで済ませてしまうというケースが最近は増えてきています。
また、GenbankとRefSeqの違いは何でしょうか？	大変いい質問です。私が説明した INSDC のデータベースは、すべて一次データベースと呼ばれるものです。 一次データベースというのは、研究者あるいは登録者の方々が、解析して決定した配列を直接受け付けているデータベースです。 一方、米国国立バイオテクノロジー情報センター (NCBI) のRefSeqというのは、独自の基準でもう1回アノテーションした二次データベースと呼ばれるものになります。 オリジナルデータで構成される一次データベースがあって、それを整理した二次データベースというのがあります。RefSeqに関しては、その二次データベースになります。 DDBJには二次データベースがないというのが弱みなのですが、そのような違いになります。