

動物ゲノムデータのリンクドデータ化の試みについて An attempt to create the animal genomic data as a Linked Data

川島 武士^{1*}, 小柳亮¹, 行者路¹

KAWASHIMA, TAKESHI^{1*}, KOYANAGI, Ryo¹, GYOJYA, Fuki¹

¹ 沖縄科学技術大学院大学

¹Okinawa Institute of Science and Technology

DNA 配列解読機の性能の向上により、生物のゲノムの配列解読速度は、昨今になり急激すすんでいる。これは、ゲノム DNA のサイズが大きき動物においても例外ではない。ゲノム解読にともなって予測される遺伝子の塩基配列、タンパク質のアミノ酸配列、さらに遺伝子発現解析データとの対応、遺伝学的各種情報との対応関係を整備する事は、ゲノム配列解読後のデータの解析において、必須の作業であるが、そのようなデータの整備作業は、配列データの解読スピードにおいっていないというのが現状である。

とくに動物ゲノムについては、ほんの数年前までは、比較的大型の研究プロジェクトという位置づけであったが、現在では配列解読自体はかなり簡単にできるようになってきたため、一つのゲノムデータについて少人数の研究者のみが解析する事が多く、公開された段階での予測遺伝子モデルの精度の低下が懸念される。

動物の遺伝子モデルの予測は、他の動物の遺伝子モデルとの比較から作成される事が多く、この場合、精度の低い遺伝子予測が次の予測においてさらに精度の低い遺伝子予測を誘発する、いわゆる「Junk Makes Junk」がしばしば見いだされるのが現状である。

複数の動物間のゲノム比較は、過去の動物のゲノムがどのようなものであったか、つまり現生動物の進化について重要な手がかりをあたえてくれるが、そもそもの遺伝子予測やその機能予測に大きな間違いがあることは、解析結果に致命的なエラーをもたらしかねない。

このようなことをできるだけ排除するために、我々は進化的な視点から見てできるだけ多様な動物のゲノム配列に注目し、遺伝子予測の修正や、精度の高いアノテーションをつけることを行っている。またそれらのデータ間の比較解析を簡便にするために、Linked Data として利用できる形で公開する試みを行っているので、紹介する。

キーワード: 動物, ゲノム, アノテーション, オントロジー, リンクドデータ, セマンティックウェブ

Keywords: animals, genome, annotation, ontology, linked data, semantic web