

コンテストによるバイオインフォマティクス研究加速

2010年11月22日

理化学研究所(兼)産総研、JST
東京医科歯科大、慶應義塾大
八尾 徹 yao@riken.jp

今回は、各種のコンテストによるバイオインフォマティクスの技術進歩の動きをまとめてみましょう。以前に(2005/7/8)、

「遊び心のバイオインフォマティクス」

—各種のコンペティション・ジャンボリーが花ざかり—

と題して、10種ほどのイベントを紹介しましたが、そこでは、ヒト遺伝子数の賭け(2002 CSHL) や、機能同定アノテーションジャンボリー大会(2000— FANTOM, 2006— H-InvDB)や、技術開発競争(2002, ボストン 1000ドルゲノム大会)のことも含めて述べました。

その後も同様の動きは次々と企画されております。今回はそれらの中から以下のような「予測」又は「設計」に関するコンテストに限っていくつかの動きをご紹介します。

この中で、最も有名であり成果も大きいのはタンパク質立体構造予測コンペ CASP (1994—)であり、最も古いのは人工タンパク質設計コンペ (1986—) であり、そして最も新しいのは GenoCon (2010—) でしょう。

いずれも発足当初は、「そんなことが出来るはずがない」というほどチャレンジングなテーマ設定でした。しかしそれから5年以上経つ頃からそのようなイベントの効果が現れ始めた例が多いのです。明らかに科学技術の進歩を加速してきたと言えます。

どの分野でも、スポーツでも科学でも、チャレンジ精神が必要でしょう。

新しい方(2010年スタート)から取り上げていきます。

1. GenoCon(ゲノム設計国際コンテスト) <http://genocon.org>

理化学研究所生命情報基盤研究部門(BASE—豊田部門長)が中心になって、理研植物科学研究センター・京都大・北海道大・近畿大、島根大、慶應大の研究者と共同で、ゲノム設計国際コンテスト GenoCon を立ち上げました。その第1回の募集を2010年5月に開始し、9月末に締め切り、現在審査中です。

このコンテストには2つの特徴があります。一つはゲノム設計技術(Dry)だけを競うこととし、優秀設計のもの実際の作成(Wet)は専門機関ですること限定したことです。合成生物学

の結果の分散流通に歯止めをかける新しい試みです。もう一つは、高校生の部を設定したことです。将来のゲノム設計士を育てようとする遠大な計画です。

今回の募集テーマはグリーンイノベーション課題「環境浄化作用を高める植物のゲノム設計」で、具体的には「空気中からホルムアルデヒドを除去して効果的に無毒化する機能を、実験植物のシロイヌナズナに付与するための塩基配列の設計をせよ」として、ハウスシック症候群を防ぐためという身近なことを目標にしています。

募集内容の説明会は、横浜サイエンスフロンティア高校で7月24日に行われ、豊田リーダーの概要説明に続き、北大山崎教授(iGEM参加経験者)、近畿大泉山教授(植物機能専門家)及び慶大板谷教授(ゲノム設計専門家、理研兼務)のお話のあと、理研蒔田博士から具体的な応募要領が説明されました。高校生を含む若手研究者が参加しました。この説明会の模様はゲノコンのユーチューブに載っています。

日本から発信したこの国際コンテストが健全に成長することを見守りたいものです。

2. 表現型予測法コンテスト CAGI (Critical Assessment of Genome Interpretation)

<http://genomeinterpretation.org>

2010年9月にCAGIの募集が始まりました。このコンテストは、UC BerkeleyのSteve Brenner教授が中心となって、Maryland大と共同で始めたもので、ゲノムの変異が表現型に及ぼす影響を予測するコンピュータ手法を競うものです。

具体的には、6ヶ所の実験・治験研究室から集められたデータに基づき、アミノ酸の変異が、酵素機能・がんの発生・がんの治療にどう影響するかを予測することや、乳がんの細胞数十種類に対する薬物の応答予測などが課題として挙げられています。

Brenner教授は、単に既製の方法論の優劣を競うのではなく、今後のデータ急増に対し、バイオインフォマティクス社会として共同で有用なソフトウェアを開発していくきっかけを作りたいと述べています。現在どのようなソフトウェアがあるか、それらがどのような問題に適しているか、今後の課題は何かを、このようなイベントを通して関係者が共通認識を持つことが重要であると強調しています。

3. バイオ回路のリバースエンジニアリング DREAM (Dialogue on Reverse Engineering Assessment and Methods).

<http://wiki.c2b2.columbia.edu/dream>

IBMのシステムバイオロジーリーダー Dr.G.Stolovitzkyが中心となって2007年から進めてきたこのコンテストでは、今年DREAM 5として4つの挑戦テーマが提示されました。

- ① エピトープの抗体認識部位予測
- ② 転写因子の DNA モチーフ認識部位予測
- ③ 遺伝子データからの疾患表現型及び遺伝子ネットワークの予測
- ④ 遺伝子発現データからの遺伝子制御ネットワークの予測

合計で 73 チームが競って、その評価結果が今週 コロンビア大で開かれた 第 3 回の “Annual Joint Conference on Systems Biology, Regulatory Genomics and Reverse Engineering Challenges” で発表されました。

その際、A. Califano 教授は、データからの結合サイトやネットワークの予測はかなり良くなってきたが、次の課題はそのネットワークから「行動 Behavior」を予測することにあると述べ、更に高い目標を設定しました。

正に参加者の挑戦心をゆさぶるような呼びかけです。このような動きが科学の進歩を促す大きなインパクトであることは確かでしょう。

4. 合成バイオロジー設計コンペ (2004--) 前編参照 <http://igem.org>

バイオ標準回路群 BioBricks を用いて新しい機能の生物を創ることに挑戦しようという合成バイオロジー設計コンペ iGEM は、2004 年にボストンで小規模(Harvard, MIT, Stanford の大学生十数人)で始められましたが、年々拡大し 2009 年には 1000 人規模に達し、次は国際的に地域予選をしなければならないという勢いです。

その中で、2007 年には千葉大の梅野チーム、東工大の木賀チームが優秀な成績を収め、その後 大阪大・東京大・東京農工大・北海道大・首都大・京都大・京都工繊大などが好成績をとって来ました。

最近の入賞校の名前を見ると、ケンブリッジ大、UCサンフランシスコ、カリフォルニア工科、プリンストン大、ICL、ブリストール大などの有名校に加えて、上位にリュブリャナ大(スロベニア)や北京大学が名を連ねているのが注目されます。

それらの作品の中には、単に興味あるものが作られたというだけでなく、新しい血液や光る微生物や有用医薬品の合成が大きな成果として出ています。

5. タンパク質立体構造予測コンテスト CASP (1994--) 前編参照

<http://predictioncenter.org>

このコンテスト CASP については、以前にもご紹介しましたが、その後も2年毎に開かれ現在も継続中です。第 3 回に登場した D. Baker グループのフラグメントアセンブリー法はその後も圧倒

的優位を誇り、立体構造予測法に大きな一石を投じました。当初 35 チームの参加であったものが、現在ではチームとサーバーを合わせて 250 以上のグループが参加しております。日本からのチームも度々上位入賞を果たしています(初期の頃は蛋白工学研究所の西川・松尾チーム、最近では生命情報工学研究センターの野口チーム、富井チーム、北里大の梅山チーム、神戸大の高田チームほか)。今年は第 9 回目が進行中で、まもなく結果が発表されます。

尚、この発展形として、タンパク質相互作用の予測コンテスト CAPRI も進んでいます。こちらは、2001 年以來すでに 23 課題が提示され、昨年 12 月にバルセロナで第 4 回ミーティングを開いています。

以上のほかにも、類似のものとしては、

CAMDA (Critical Assessment of Massive Data Analysis)

CAFA (Critical Assessment of Function Annotations)

CASD-NMR (Critical Assessment of Automated Structure Determination by NMR)

BioCreative (Critical Assessment of Information Extraction Systems in Biology)

などがあります。

6. 人工タンパク質設計コンペ (1986—) 前編参照

7. ゲノム遺伝子予測法コンペ (2005—) 前編参照

以上すべてに共通していることは、インターネットの活用です。問題が次々とインターネット上に提示され、それに対して世界中からインターネットを介して回答が送られ、それらを評価委員が評価して、最後に全体集合で発表・討議・表彰が行われる——このスタイルは正に上記のCASPが始めたことです。

遊び心から始まったこのスタイルが、バイオインフォマティクス手法の開発・評価の大きな手段となっております。

日本からもそのような発信が始まったことを喜んでいます。

2010. 11. 22 八尾 徹