操作に必要なユーザーにも使えるよう工夫されている。

さて、結局、「EMBOSSを使えば?」という私の意見は以下の2点で要約された。1)EMBOSSのドキュメントは非常に充実しているが、英語であるためにテクニカルな修士の学生が見づらい。2)コンマンドベースのプログラム集は使いにくく、インストールや設定ができない。ウェブ上で対しては教育コストが高い。

つっこまることはたくさんある。しかし、私はその問題を議論するのではなく、2つの行動を起こすことになった。ひとつは、すぐれた英語のドキュメント(特にEMBOSSのドキュメント)を翻訳すること。そしてもうひとつは、簡便にバイオインフォマティクスのツールを利用できる環境(特にバイオインフォマティクスのツールが設定済みのCD Linux)を作ることである。前者は当社、ちょうど、日本新薬株式会社の多賀谷さんがEMBOSSの翻訳プロジェクトJAMBOを立ち上げたところであったのでそこから始めることにした。現在、「チュートリアル」「管理者の手引き」などのドキュメントの翻訳を完了しており、その成果はWebで公開されている(http://jambo.sourceforge.jp/)。後者に関しては、KNOB Project(http://knob.sourceforge.jp/)という活動を開始した。

KNOB(ノブ、Knoppix for Bio)とはバイオインフォマティクスのソフトウェアが設定されたLinuxのディストリビューションである。KNOBはKnoppix JP(http://junix.nao.ac.jp/~itri/knoppix/)というCD Bootable Linuxベースをベースとして作っている。そのため、あらゆるWindowsコンピュータのCD-ROMドライブにKNOBが入ったCD-ROMをいれて再起動だけで、Linuxが起動し、簡単にバイオインフォマティクスの解析や学習が利用できる。もちろん、もう一つはWindows環境における一切の影響を及ぼさない。KNOBは無償で配布しており、特典や利用可能の企業での利用も自由である。

しかし、CDで提供されているデモソフトもある。バイオインフォマティクス研究に欠かせないデータベースや収録できないソフトウェアを収録できないのである。KNOBではこの問題を解決するためにWebサービスを積極的に利用し、必要に応じてインターネット上のデータベースを自動的に参照する仕組みをもっている。それぞれのプログラムからユーザーが意図することなくDNA/タンパク質配列を自在に利用できる。またバイオインフォマティクスではしばしば大量の計算が必要になる。理化学研究所の西谷史一さんがKNOBを改良しKNOB High Throughput Computing edition(http://big.gac.nitek.ips/index.html/Members/umakazi/hc)を作成しており、ジョブ管理システムや共有ファイルシステムが設定された分散コンピューティングが簡単に利用できる。

KNOBの機能でなく利用場面に目を当ててみると、バイオインフォマティクス初学者や実験生物学者向けの教育教材として、KNOBが使われて始めている。例えば、慶応、医科薬科大での講義で使われており、私もKSBUが協賛している財団法人神奈川科学技術アカデミー開発のセミナーかくずさDNA研究所で行われた「生物学者による情報処理技術研究会」のハンズオンセミナーなどで、実験生物学者向けの講習をおこなってきた。KNOBはバイオインフォマティクスへの興を喚起、その専門を広げるために貢献していると言えよう。

このように、研究者が実際現場で使っている解析環境が簡単な皆さんでも手に入れるように、研究者以外の人たちは我々の研究を追出し、新しいアイデアを試すことができる。大学や研究機関などではアマチュア研究者が新しい星や新しい種を見発見するといったことは当たり前のことで、同じように、新しい発見を発見するアマチュア研究者ができてきそうではないか。ITの世界では10代の若者がすぐれたソフトウェアを作って成功している例もある。

アソリーチが呼ばれる今日、教育講義やポータルサイトのような活動の重要性は疑いない。さらに、興味を持った人がバイオインフォマティクスを体験できる原があれば、あるいはこの世界に飛び込み、あるいは人材バンクとしてこの分野の有用性を周知に説明し、ある人は科学の清潔さを保つために追試に従わぬさんがいないわけではない。KNOBのようなオープンソース活動はバイオインフォマティクスの視野を広げ、結果的には人材の交流を促進し、業界を盛り上げ、未来を投入するふさわしい学問であることを一般の皆様に理解して頂くために、いくらか貢献できるのではと、深い期待を抱いている。

■次世代バイオインフォマティクスツールを目指して——G-language Project——

荒川 和晴（慶應義塾大学先端生命科学研究科）

2001年、コンピュータゲノムが爆発的勢いで次々と読み取られる様子とまるで戦うように、様々なアルゴリズムが実装されバイオインフォマティクスが開発・公開されていた。それらソフトウェアの中にはバイオインフォマティクスという分野の成熟を見せる優れたものも多数存在した。だが、「ドライ」な科学者「ウェット」な実験主体の科学と同様、様々な手法を目的に応じて組み合わせることによって始めて研究になるのであり、実行環境もここは数種のソフトウェアに囲まれている実験プロトコル(解析バイオライン)を構築するの容易ではない。そんな中、Perlという言語の上での様々なソフトウェアを統一的に組み合わせるBioPerlプロジェクトは有望な解決策に見えた。しかしながら、実際に使ってみるとBioPerlはプログラムの知識があっても複雑で面倒な作業が多い。データベースに応じて利用するAPIを選ぶ必要があり、結果もまた用意されたオブジェクトから取り出すべきしなければならない。「バイオ」と「インフォマティクス」の間には依然として深淵が横わたっていた。

私が慶應義塾大学先端生命科学研究科においてG-language Projectを立ち上げたのはまさにこのような背景に基づいている。バイオインフォマティクスの統合環境として開発しているG-language Projectではデータベースの形式やデータ型を問わずに一つのインタフェースで扱い、その他の部分はシステム間の自動判別及び解析を行う、実装されている200以上の解析のほとんどはプログラミングを必要とせずGUIから実行可能であり、結果は一目で解析できるようなウィンドウに出力される。当然GUIで出来ることはまだまだ限られているがG-language Systemの真価を発揮する上でにはプログラミングは不