

TIN SINH HỌC MỘT CÔNG NGHỆ MỚI ĐANG PHÁT TRIỂN Ở VIỆT NAM

✍ ThS. NGUYỄN THANH BÌNH*

Khoa học nói chung và sinh học nói riêng đều dựa trên nền tảng tri thức, trên sự kết nối tri thức của nhiều lĩnh vực khác nhau. Điều này càng đúng khi các công nghệ kết hợp với nhau sản sinh dữ liệu với tốc độ chóng mặt. Theo thống kê trên các trang web: Uniprot Knowledgebase¹ - kho thông tin protein có hơn 13,5 triệu mục, tính đến tháng 1 năm 2011, GenBank² - kho dữ liệu di truyền có đến gần 164 triệu hồ sơ gen và 110 triệu bộ mã di truyền, tính đến tháng 6 năm 2013. Với lượng dữ liệu lớn như vậy, cần phải nhờ đến máy tính để “thông dịch” và sàng lọc. Vì vậy, việc ứng dụng khoa học máy tính và thống kê để phân tích dữ liệu sinh học tạo nên lĩnh vực nghiên cứu mới có tên là tin sinh học. Chính ngành khoa học mới Tin sinh học sẽ giúp thúc đẩy, cải thiện và đa dạng hóa các nghiên cứu về sinh học.

Tóm lại, **Tin sinh học** (*bioinformatics*) là một lĩnh vực khoa học sử dụng các công nghệ của các ngành toán học ứng dụng, tin học, thống kê, khoa học máy tính, trí tuệ nhân tạo, hóa học và hóa sinh (*biochemistry*) để giải quyết các vấn đề sinh học³. Một thuật ngữ thường được dùng thay thế cho tin sinh học là **sinh học tính toán** (*computational biology*). Tuy nhiên, tin sinh học thiên về việc phát triển các giải thuật, lý thuyết và các kỹ thuật thống kê và tính toán để giải quyết các bài toán bắt nguồn từ nhu cầu quản lý và phân tích dữ liệu sinh học.

Tin sinh học là lĩnh vực khoa học mới có tính ứng dụng cao trong cuộc sống, đặc biệt là trong lĩnh vực công nghệ sinh học, nông nghiệp, y dược và trong lĩnh vực di truyền. Về cơ bản, tin sinh học tập trung vào nghiên cứu và áp dụng các phương pháp cũng như các kỹ thuật trong tin học để giải quyết các bài toán trong sinh học phân tử.

Tin sinh học đã đem lại những thay đổi trong các lĩnh vực y học phân tử, di truyền học so sánh, tiến

hóa phân tử, ứng dụng di truyền vi sinh vật, nghiên cứu dược phẩm... và đặc biệt là lĩnh vực thực phẩm và dinh dưỡng.

Tin sinh học hiện đang và sẽ tiếp tục đóng góp nhiều thành tựu khoa học mới và quan trọng như thúc đẩy nhanh quá trình chẩn đoán bệnh và tìm ra các loại thuốc chữa bệnh mới, tìm ra các giống cây trồng và vật nuôi mới cho năng suất cao, xây dựng quá trình tiến hóa.

Các lĩnh vực nghiên cứu của tin sinh học

Phân tích trình tự của axit nucleic và axit amin trong protein: Vào năm 1977, lần đầu tiên Sanger và cộng sự đã xác định được trình tự ADN của virus jx-174 và từ đó đến nay, trình tự ADN của rất nhiều loài sinh vật đã được phân tích và lưu giữ trong các ngân hàng cơ sở dữ liệu gene.

Bản đồ gen và đánh dấu gen: Nghiên cứu bộ gen người giúp các nhà khoa học có cơ sở nghiên cứu về sự tiến hóa của loài người, xác định nguyên nhân một số bệnh di truyền, bệnh truyền nhiễm, từ đó có cơ sở để phát triển các loại thuốc mới, các liệu pháp chữa trị hiệu quả.

Đò tìm đột biến: Những nghiên cứu hiện nay về xác định trình tự ADN, kỹ thuật nhân bản gen (polymerase chain reaction-PCR) nhằm tìm ra các đột biến, chủ yếu là các đột biến điểm (point mutation) xảy ra trên các gen khác nhau trong ung thư.

Phân loại học phân tử: Phân tích dữ liệu gen, bộ gen và xây dựng cây phân loại không chỉ cho thấy chiều hướng tiến hóa của các loài trong sinh giới mà còn chỉ rõ mức độ quan hệ trong loài và sự tiến hóa trong nội bộ loài. Bởi thế, việc xây dựng cây phân loại phân tử có vị trí hết sức quan trọng.

Bảo tồn đa dạng sinh học: Tin sinh học thường được áp dụng trong lĩnh vực bảo tồn đa dạng sinh học (biodiversity). Thông tin quan trọng nhất được thu thập chính là tên, các đặc điểm hình thái, sinh lý, sinh hóa, di truyền, sự phân bố, tập tính... được truy xuất bởi các chương trình phần

¹ <http://www.uniprot.org/uniprot/Q9SXZ2>

² <http://www.ncbi.nlm.nih.gov/genbank/>

³ https://vi.wikipedia.org/wiki/Tin_sinh_hoc

* Trường Đại học Công đoàn

mềm để tìm kiếm, hiển thị, phân tích các thông tin một cách tự động.

Mạng điều khiển sinh học: Mạng điều khiển sinh học (biological regulatory network) là các mạng tương tác giữa gene, protein để điều khiển một quá trình sinh học nào đó trong tế bào.

Một số bài toán và ứng dụng cơ bản của tin sinh học:

Cơ bản, tin sinh học có thể định nghĩa là ứng dụng nguồn lực điện toán cho dữ liệu sinh học, chủ yếu liên quan tới các loại dữ liệu như chuỗi ADN, thành phần và cấu trúc protein. Tin sinh học đã được áp dụng khá rộng rãi trong nhiều lĩnh vực khoa học đời sống, dưới đây là một số ứng dụng đáng chú ý:

Quản lý dữ liệu: Một trong các bài toán quan trọng đầu tiên của tin sinh học là xây dựng các hệ thống dữ liệu để quản lý và khai thác một cách hiệu quả lượng dữ liệu sinh học khổng lồ. Dữ liệu sinh học bao gồm nhiều loại khác nhau, trong đó chủ yếu là các trình tự nucleotit, và các trình tự axit amin. Lượng dữ liệu sinh học có thể lên đến hàng triệu megabyte và được cập nhật liên tục theo thời gian.

Nghiên cứu tiến hóa: Nghiên cứu quá trình tiến hóa và mối quan hệ giữa các loài sinh vật là bài toán quan trọng trong sinh học. Tin sinh học tập trung vào phát triển các mô hình và phương pháp nhằm xây dựng mối quan hệ tiến hóa giữa các loài sinh vật dựa vào phân tích mức độ giống nhau giữa các trình tự sinh học của chúng.

Di truyền học so sánh

Nhiều tiến bộ quan trọng trong công nghệ di truyền cùng với sự cải thiện đáng kể về năng lực tính toán đã góp phần rất lớn cho những tiến bộ trong lĩnh vực di truyền học so sánh (DTHSS). Như tên gọi, DTHSS phân tích các điểm tương đồng và khác biệt giữa các hệ gen, cách tiếp cận này có thể tiết lộ thông tin về cấu trúc và kích thước hệ gen của từng chủng loài, làm sáng tỏ các yếu tố di truyền, phương thức tiến hóa, thành phần gen liên quan đến mầm bệnh và khả năng thích nghi môi trường. Thực tế hơn, DTHSS được áp dụng để làm suy yếu các dòng vi khuẩn trong các vắc-xin hiện có nhằm phát triển loại vắc-xin tốt hơn. Ngoài ra, DTHSS còn giúp xác định các gen liên quan đến việc tạo ra nhiên liệu sinh học, khả năng này đang được nhiều nơi nghiên cứu để đưa vào sản xuất.

Nghiên cứu dược phẩm

Tin sinh học được xem là một nhân tố quan trọng

trong quá trình nghiên cứu dược phẩm nhiều năm qua, giúp rút ngắn đáng kể thời gian của nhiều công đoạn bào chế thuốc. Nhờ dữ liệu di truyền dồi dào, người ta có thể khám phá các mục tiêu “công kích” mới cho thuốc, ước tính có khoảng 3.000 mục tiêu tiềm năng trong bộ gen người. Ngoài ra, tin sinh học còn được sử dụng để nhận diện các hợp chất chỉ khi bào chế thuốc, có thể sàng lọc hàng ngàn hợp chất một cách nhanh chóng.

Phát triển các giống cây trồng, vật nuôi: Một trong các ứng dụng nổi bật của tin sinh học là phát triển các phương pháp phân tích nhằm tìm ra các gen quý hiếm ở các loại vật nuôi, cây trồng. Việc này sẽ giúp đẩy nhanh quá trình tạo ra các giống vật nuôi cây trồng cho năng suất và hiệu quả kinh tế cao.

Tin Sinh học ở Việt Nam

Tin sinh học là một lĩnh vực mới nhưng không xa lạ. Trên thế giới đã có rất nhiều các nhà khoa học nghiên cứu và thành công trong lĩnh vực này. Bắt nhịp xu thế phát triển này của khoa học thế giới, các nhà Tin sinh học Việt Nam đã xác định được những hướng đi phù hợp và đạt được thành công bước đầu với một số công trình nghiên cứu mang tính ứng dụng cao⁴.

Việc đào tạo một đội ngũ chuyên gia về tin sinh học đương nhiên là điều quyết định cho sự thành công của sự phát triển tin sinh học. Đội ngũ này không những phải có trình độ tư duy toán học xuất sắc nhưng còn phải thông hiểu những vấn đề hiện nay của sinh học. Việc đào tạo kỹ năng nghiên cứu độc lập và khả năng cộng tác với các đồng nghiệp trên thế giới của các chuyên gia này sẽ là điều kiện thiết yếu trong quá trình phát triển của ngành tin sinh học Việt Nam.

Đội ngũ Tin sinh học Việt Nam bao gồm cả các nhà nghiên cứu tiếp cận từ hai nhánh tin học và sinh học:

· Nhánh Tin học: Nhóm của GS. Hồ Tú Bảo, PGS. TS. Lê Sỹ Vinh (ĐH Công nghệ, ĐH Quốc gia Hà Nội), GS. TS. Từ Minh Phương (ĐH Bưu chính viễn thông), TS. Trần Đăng Hưng (ĐH Sư phạm, ĐH Quốc gia Hà Nội)... với thế mạnh làm chủ kỹ thuật tính toán, thống kê, khoa học máy tính, trí tuệ nhân tạo, khai thác dữ liệu (data mining)...

⁴ Nhận xét của GS.TS. Trương Nam Hải (Viện Công nghệ sinh học, Viện Hàn lâm KH&CN Việt Nam).

· Nhánh sinh học: các nhóm của GS.TS. Trương Nam Hải, TS. Nguyễn Cường (Viện Công nghệ sinh học), PGS.TS. Lê Thị Lý (ĐH Quốc tế, ĐH Quốc gia TP. HCM), TS. Dương Quốc Chính (Viện Huyết học truyền máu TU), TS. Trịnh Thanh Bình (trường Cao đẳng Thủy sản), TS. Khuất Hữu Trung (Viện Di truyền nông nghiệp)... vốn nắm rất chắc kiến thức về hệ gene học (genomics), hệ phiên mã (transcriptomics), hệ protein (proteomics)... Trong số này, nhóm ở Viện Công nghệ Sinh học là một trong những nhóm đầu tiên ở Việt Nam có sự kết hợp chặt chẽ giữa sinh học và tin học.

Tuy lực lượng chưa nhiều nhưng Tin sinh học Việt Nam đã chọn được hướng đi phù hợp để bắt nhịp với thế giới. Hiện nhóm của GS. Hồ Tú Bảo đang tập trung vào hướng nghiên cứu và ứng dụng Bệnh án điện tử, EMR - Electronic Medical Record) nhằm thiết lập bệnh án của người bệnh ở dạng điện tử dựa trên hệ thống thông tin bệnh viện (HIS - Hospital Information System). Cũng thiên về y sinh song nhóm của PGS. Lê Thị Lý đi sâu vào nghiên cứu và phát triển dược phẩm; TS. Dương Quốc Chính tìm đột biến đơn điểm SNP (Single Nucleotide Polymorphism) gây bệnh ung thư máu, SNP gây kháng thuốc. Sau những công bố về hệ gene người Việt, PGS. Lê Sỹ Vinh tiếp tục con đường này và tìm cách cộng tác với các bệnh viện để đưa kết quả nghiên cứu vào thực tế điều trị. Riêng nhóm của GS. Trương Nam Hải và TS. Nguyễn Cường nghiên cứu xây dựng bản đồ hệ gene vi tảo, lắp ráp và chú giải hệ phiên mã tôm sú, phân tích hệ gene lúa, phân tích hệ gene metagenomics, đều là những bài toán có nhiều ý nghĩa ứng dụng.

Trong quá trình nghiên cứu, các nhà khoa học của lĩnh vực Tin sinh học ở Việt Nam cũng gặp nhiều thuận lợi và không ít những khó khăn sau:

Thuận lợi:

- Có nhiều chuyên gia đầu ngành trong các lĩnh vực Tin học, sinh học
- Vào cuối năm 2014, Tin sinh học được bổ sung vào danh mục 58 công nghệ cao được nhà nước ưu tiên phát triển.

Khó khăn:

- Sự quan tâm của nhà nước cho lĩnh vực này vẫn còn ở mức khiêm tốn nên trong quá trình nghiên cứu gặp phải nhiều khó khăn, nhất là thiếu thốn về cơ sở vật chất, đặc biệt là thiếu hệ thống máy chủ có hiệu năng cao (HPC - High Performance Computing) chạy được những phần mềm chuyên dụng.

· Tin sinh học Việt Nam còn quá ít nhiệm vụ, mặc dù tính ứng dụng từ các đề tài nghiên cứu trong lĩnh vực này rất lớn. Vì vậy, nhiều nhà nghiên cứu trẻ đã ra nước ngoài tìm đất “dụng võ” như Phan Quang Huy (Mỹ), Tạ Xuân Hưng (Phần Lan), Lê Sỹ Quang (Anh), Cao Xuân Hiếu (Đức), Thắng Phạm (New Zealand)..., dẫn đến tình trạng trong nước hiếm hoi những người có chuyên môn sâu đủ khả năng xây dựng nhiệm vụ nghiên cứu, tạo ra cái vòng luẩn quẩn của việc thiếu nhân lực và nhiệm vụ trong lĩnh vực Tin sinh học Việt Nam hiện nay.

· Các nhóm nghiên cứu Tin sinh học Việt Nam vẫn hoạt động đơn lẻ, chưa có nhiều liên kết và phối hợp với nhau trong nghiên cứu nên hạn chế khả năng giải quyết những bài toán sinh học lớn. Chính sự rời rạc của mạng lưới này cũng khiến nhiều nhà quản lý khoa học chưa đánh giá hết tiềm năng của tin sinh học Việt Nam. Với mong muốn thiết lập mạng lưới liên kết của Tin sinh học Việt Nam, TS. Nguyễn Cường đã xây dựng trang web www.tinsinhhoc.org và trang facebook fanpage với gần 1.000 thành viên đăng ký theo dõi, trên đó phổ biến những kiến thức cơ bản của Tin sinh học, cập nhật kết quả nghiên cứu trong nước và thế giới cũng như luôn đón nhận các sinh viên chuyên ngành Tin học và Sinh học quan tâm tới lĩnh vực Tin sinh học.

Những khám phá mới của nghiên cứu Tin sinh học như một cuộc cách mạng đang xảy ra ở Việt Nam cũng như ở trên thế giới, có những tác động quan trọng đến sự phát triển của nền kinh tế và xã hội. Trong tương lai, ngành khoa học Tin sinh học của Việt Nam sẽ phát triển, chói sáng và sẽ có những thành tựu đỉnh cao nếu Việt Nam có những chính sách ưu tiên và nỗ lực đặc biệt đến sự phát triển công nghệ mới này. □

TÀI LIỆU THAM KHẢO

1. TS. Lê Sỹ Vinh, *Giáo trình Nhập môn Tin sinh học, Trường Đại học Công nghệ, Đại học quốc gia Hà Nội, năm 2013*
2. *Tin sinh học: Chưa khai thác hết tiềm năng*, <http://www.tiasang.com.vn/Default.aspx?tabid=111&News=8573&CategoryID=2>, cập nhật ngày 20-4-2015
3. *Diễn đàn sinh học Việt Nam* <http://www.sinhhocvietnam.com/forum/showthread.php?t=59>
4. *Tin sinh học - Wikipedia tiếng Việt*, https://vi.wikipedia.org/wiki/Tin_sinh_hoc
5. *Kho thông tin protein - Uniprot Knowledgebase* :// www.uniprot.org/uniprot/
6. *Ngân hàng Gen*: <http://www.ncbi.nlm.nih.gov/genbank/>