

## 初版まえがき

本書は著者の農業バイオテクノロジー企業での数年間の経験の積み重ねの結果である。ゲノムデータベースのキュレーターとしてスタッフの研究者に対しバイオインフォマティクスが必要な広範な範囲をサポートした。彼らの何人かは、すでに手作業で行われていた同様の処理を自動化したいと欲したり、生物学の問題をバイオインフォマティクスで解決できるかを尋ねた。ほとんどの場合プログラミングの知識が問題を解くのに必要という共通点があった。本書の主目的は生物学の問題と解かんとするこれら研究者を助け、プログラミングの初歩を学びきることにある。この目的のため、プログラミングの概念を当然のことと考えないように努めた。プログラミング言語にはPythonを使用した。

Pythonは習得が容易なコンピュータ言語で、研究者の注目を集めている。使うのが簡単なうえにほとんどの問題を解くのに十分に強力である。Pythonを使えば、読者はすぐにプログラミングを始めることができる。*Computing in Science and Engineering*, *Briefings in Bioinformatics* や *PLOS Computational Biology*などの論文誌でもPythonについての入門記事を公開している。研究者はPythonを使って分子の可視化、ゲノムのアノテーション、データ操作他無数のアプリケーションに利用できる。

とりわけライフサイエンスの場合、Python、特にBioPythonを使った開発は非常に重要である。このため、第II部はBiopythonを取り上げた。安直にBiopythonが世界のあらゆる生物学的問題の解決策であるとは主張しない。簡単なカスタムメイドのソリューションが最先の問題にうまく適合する場合もある。ここでは読者が試してみたいと思うBioNEBやCoreBioのような他のパッケージも記されている。

本書は基本からスタートし、第I部（“プログラミング”）で読者にプログラミングの原理を教える。プログラミングの習得は、実践が一番なので初歩のうちから練習重点を置く。そのため本書のあちらこちらにプログラムの断片を記してあり、読者がそのコードを実際に動かして習得することを期待している。啓発のため他の言語による予備的な比較もするが、新しい言語を教える際、ほとんどの言語比較は良いことよりも害になる。それらは理解できない情報だし、ほとんどの読者にとっては無関係だ。

読者の関心を得るために、ほとんどの例は何らかの形で生物学に関連している。読者が当該知識をもっていなくても理解に支障はない。

本書の実用性を強化し参考資料として使用するために、第IV部は“コメント付きPythonソースコードレシピ”と名付けた。これらのプログラムはそのまま使用できるが、他のプロジェクトで再利用できる。読者は、いくつかの例が非常に単純で、特に何もしなくてもプログラムができるに気づくだろう。これが第IV部の狙いだ。よけいな機能で読者の注意をそらすことなくアプリケーションの特定の側面を解説し、複雑なプログラムで読者を落胆させないこと。基本さえしっかり学べば機能とカスタマイズの追加は後でどうにでもなる。

第III部（“高度なトピック”）は威圧的に感じるかもしれないが、本書の場合は決して難しいという意味ではない。実際に、誰もが第III部の章で「特にリレーションナルデータベース管理システム—RDBMS—～およびXML」を使用する。バイオインフォマティクス研究の重要な部分は、データベースの構築とクエリ構築であり、これがMySQLのようなRDBMSがバイオインフォマティクスのスキルセットの関連部分であると考える理由である。様々なソースからのデータ統合は、バイオインフォマティクスで最も頻繁に実行されるタスクの1つである。これに最適なツールはXMLで、アプリケーション間のデータ交換に広く使用されるプラットフォームになりつつある。PythonにはいくつかのXMLパーサがあり、それらのほとんどは本書で解説する。

付録B「選書」では、Pythonに関する入門を提供する。内容は重複するが同じ内容を違った角度で解説してくれる。

本書は研究者に有益なだけではなく、大学の教科書としても役立つように編集した。学生は本書をプログラミングを、特に新しいバイオインフォマティクスに利用できる。

## 第2版まえがき

*Python for Bioinformatics* の初版は 2008 年と 2009 年に上梓した。8 年を経た今でも同書の内容は色褪せていない。このような進歩の激しい領域においては驚異である。それにもかかわらず、本書は時代にそっており第 2 版から得られるものは大きいと思われる。

Python 2.7 はまだまだ生産システムで使われているが、Python の主バージョンは 3.6 である。それゆえバージョン間の非互換性があり、本書の全コードを Python 3 との互換性に対応させるのに苦労した。

8 年間の歳月はソフトウェアの進歩のみならず、オープンソースソフトウェアに対する企業の姿勢がずいぶん変ったし、一部の Python の扱いが劇的に変化した。共同開発やクラウドコンピューティングでは無視できない新しいコンピューティングパラダイムであった。

初版では第 14 章 **Bazaar** をベースとした共同開発：バージョンコントロールで、当時使われていた分散開発ワークフローを取り上げたが、現在は多くの開発者は **GitHub** 上で **Git** を使っているので、現在の主流に書き換えた。

Web 開発は大きく変化したが、これは Web 開発のための書籍ではない。Web アプリケーションの章では、CGI/WSGI やミドルウェアベースのアプリケーションに代え、長時間実行されるプロセスやフレームワークの現状での使い方を反映させた。フレームワークについては本章の補足として記述したが、本章では **Bottle** フレームワークの周辺について取り上げ、従来の方法は歴史的な脚注とした。

データベースでは、**NoSQL** は初版の箇条書きからとても大きく変化し、**MongoDB** として 1 つの節になり、Python レシピはこの NoSQL データベースを使うように変更された。

グラフィクスライブラリは 2009 年以降向上し、極めて高品質のグラフィクスを Python で利用可能になった。本章では **Bokeh** と呼ばれる無料の対話的な可視化ライブラリを解説する。

本書に反映されているもう 1 つの変更は、**Anaconda** と **Jupyter Notebooks** である（および Microsoft Azure<sup>1)</sup> <sup>2)</sup> で提供されるクラウド notebook 上のすべてのコード）。

<sup>1)</sup> 参照：<https://notebooks.azure.com/py4bio/libraries/py3.us>

<sup>2)</sup> 訳注：原著では、Jupyter Notebook の実行環境を Azure Notebooks で提供と記載しているが、翻訳時点では、原著者の実行環境は Google Colab に変更されている。実行は <https://github.com/serulab/Py4Bio/> にある Google Colab へのリンクをクリックすることで可能となっている。なお、Google Colab の利用方法は <https://colab.research.google.com>

ソースコードは GitHub のリポジトリ <https://github.com/Serulab/Py4Bio> に置いてあるので、本書のすべてのサンプルコードがダウンロードできる。

各々の章で訂正がある。実際に誤っているものもあるが、Python 3へのアップグレードや現在のよい習慣に基づくものがほとんどである。本書は訂正が必要と思われる所以で読者が最新情報を取得可能 Web ページを作成した <http://py3.us>。またメーリングリストの購読も実施されたい。

ソフトウェアの進化やパラダイムシフトを別にすれば、著者も開発経験を積み、教育学的事項への認識が変化した。数年間国際コンソーシアムでゲノム配列プロジェクトに従事し、NYSE 上場企業 (Globant) でシニアソフトウェア開発者として仕事をしてきた。最後の 5 年間は Salesforce や National Geographic などいくつかのよく知られたクライアントと仕事をした。現在は PLOS (Public Library of Science) で仕事をしている。

MATLABへの連絡先は以下のとおり。

MATLAB® は The MathWorks, Inc. の登録商標。製品情報についての問い合わせは以下を参照：The MathWorks, Inc. 3 Apple Hill Drive Natick, MA, 01760-2098 USA. Tel: 508-647-7000. Fax: 508-647-7001. E-mail: [info@mathworks.com](mailto:info@mathworks.com). Web: [www.mathworks.com](http://www.mathworks.com).

表紙に使用している Biopython のロゴについてライセンスを記す（ロゴを含めたすべてのファイル）。

Biopython は現在 "Biopython License Agreement" の下で公開されている（詳細は以下）。特に明記のない限り、個々のファイルヘッダ、すべての Biopython ファイルは "Biopython License Agreement" のライセンス下にある。

いくつかのファイルについては "Biopython License Agreement" もしくは "BSD 3-Clause License"（両方とも以下に詳細）の排他的な 2 つのライセンスから選択できる。これは、後にこのデュアルライセンスの下で Biopython のすべてが提供されることを意図する。

## Biopython License Agreement

Permission to use, copy, modify, and distribute this software and its documentation with or without modifications and for any purpose and without fee is hereby granted, provided that any copyright notices appear in all copies and that both those copyright notices and this permission notice appear in supporting documentation, and that the names of the contributors or copyright holders not be used in advertising or publicity pertaining to distribution of the software without specific prior permission.

THE CONTRIBUTORS AND COPYRIGHT HOLDERS OF THIS SOFTWARE DISCLAIM ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL THE CONTRIBUTORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY

SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

### BSD 3-Clause License

Copyright (c) 1999-2017, The Biopython Contributors All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

## 謝 辞

本書のようなプロジェクトはたった一人ではできない。このため多くの人々に感謝を表したい。一般の読者にはどうでもいいことだし、感謝すべき誰かの名前をおとしてしまうかもしれないが、次の人々の名前を挙げたい：私の妻 Virginia Gonzalez (Vicky) と息子の Maximo Bassi は、1年以上にわたり私が実質的に家庭にいないことを容認してくれた。また Vicky は原稿の準備に数え切れないほどの助けをしてくれた。私の両親と教授は私に重要な教訓を教えてくれた。私の家族 (Oscar, Graciela と Ramiro) は、Hugo と Lucas Bejar とともに、英文の出版作業を手伝ってくれた。Vicky, Griselda, Eugenio はライターと開発者に必要な開発抽象化レイヤーの提供を支援した。

私は地元の Python コミュニティ (<http://www.python.org.ar>) の人々にも感謝を表したい： Facundo Batista, Lucio Torre, Gabriel Genellina, John Lenton, Alejandro J. Cura, Manuel Kaufmann, Gabriel Patiño, Alejandro Weil, Marcelo Fernandez, Ariel Rossanigo, Mariano Draghi, and Buanzo.

この素晴らしいコミュニティのためだけに、再び Python を選択する。Biopython の人々にも感謝する：Jeffrey Chang, Brad Chapman, Peter Cock, Michiel de Hoon, and Iddo Friedberg。特に Peter Cock には Biopython の章に関するコメントに感謝する。 $\text{\LaTeX2e}$  の課題を手伝ってくれた Shashi Kumar と Pablo Di Napoli、そして当初から私をサポートしてくれた Sunil Nair にも感謝する。Guido Barosio, Josefina Chausovsky, Lucas Campos, Pablo Brenner や Guibert Englebienne のように、私を信頼してくれた Globant の人々、Pedro Mourelle, Chris DeBlois, Rodrigo Obi-Wan Iloro, Carlos Del Rio や Alejandro Valle のように Globant の同僚、PLOS の Jeffrey Gray や Nick Peterson にも感謝する。

## 訳者まえがき

データサイエンスという産業および研究領域の浸透と昨今の深層学習の発展を背景にプログラミング言語 Python の重要性は非常に高く認知されている。古くから自然科学研究に携わっている諸氏には説明は不要であるが、Python は化学、生物学の研究で利用されてきたし、システム記述言語としても利用されてきた。

Python については、昨今の AI ブームもあいまって、数多くの関連書籍があるし、インターネットから得られる情報も数多い。そのため本書の翻訳を出版社から打診されたとき、訳者の数名で検討した。

これまでバイオインフォマティクスのためのプログラミング言語という書籍はいくつか出版されているけれども、バイオインフォマティクスにそれらの言語が有効であるとして、それぞれの言語の解説に終始するという内容の書籍が大半であった。その点本書は筆者も初版まえがきで言及しているように、極力バイオ研究の話題を扱っている。一方でメモリ管理など、Python 利用をつきつめると問題になる項目までカバーしている。

いずれの専門領域もそうかもしれないが、バイオインフォマティクスの場合は非常に尖った知識やスキルが必要になると感じている。必ずしも自身がその知識やスキルを有する必要はなく、それを第三者に委ねられればいいのであるが、その場合でも知識やスキルを評価するスキルは必要になる。

したがって本書は研究をサポートする技術者だけではなく、実験主体の研究者やこれからバイオ系の研究を始める若い方々への有益な情報となりうると確信した。Jupyter notebook や Google Colab を用いてスクリプトを対話的に実行することで読者の理解も深まるだろう。

今回の作業では、著者である Sebastian Bassi と頻繁にやりとりし、可能な限り正確な訳出を心がけた。そのため単なる原書第2版の邦訳ではなく、現在では古くなっている情報は該当箇所を書き換えるか訳注として記した。

本書は、バイオインフォマティクスと Python という複数のトピックを扱っている。そのため内容の精査を複数の査読者に依頼した。中央大学理工学部田口善弘教授、慶應義塾大学環境情報学部鈴木治夫准教授、NEC ソリューションイノベータ株式会社渡邊亮氏そして株式会社マインドインフォ代表柴田淳氏には面倒な作業を快諾いただき感謝している。

なお、著者が謝辞で言及しているのと同様に、翻訳者の家族の協力なしには今回の翻訳作業は

なしえなかった。それぞれの家族と、共立出版株式会社の本書のご担当山内千尋氏、編集制作担当の島田誠氏、本書翻訳のきっかけを作ってくれた福島真太朗氏に深く感謝の意を表したい。そして何より Sebastian Bassi に対しては、彼の献身的な協力なくして本書の上梓はあり得ず、最大限の感謝の意を表したい。

今、世界は COVID-19 の脅威に晒されている。本書の内容がその脅威を打破する一助となりえたならば訳者としてこれ以上の喜びはない。

2021年 訳者一同