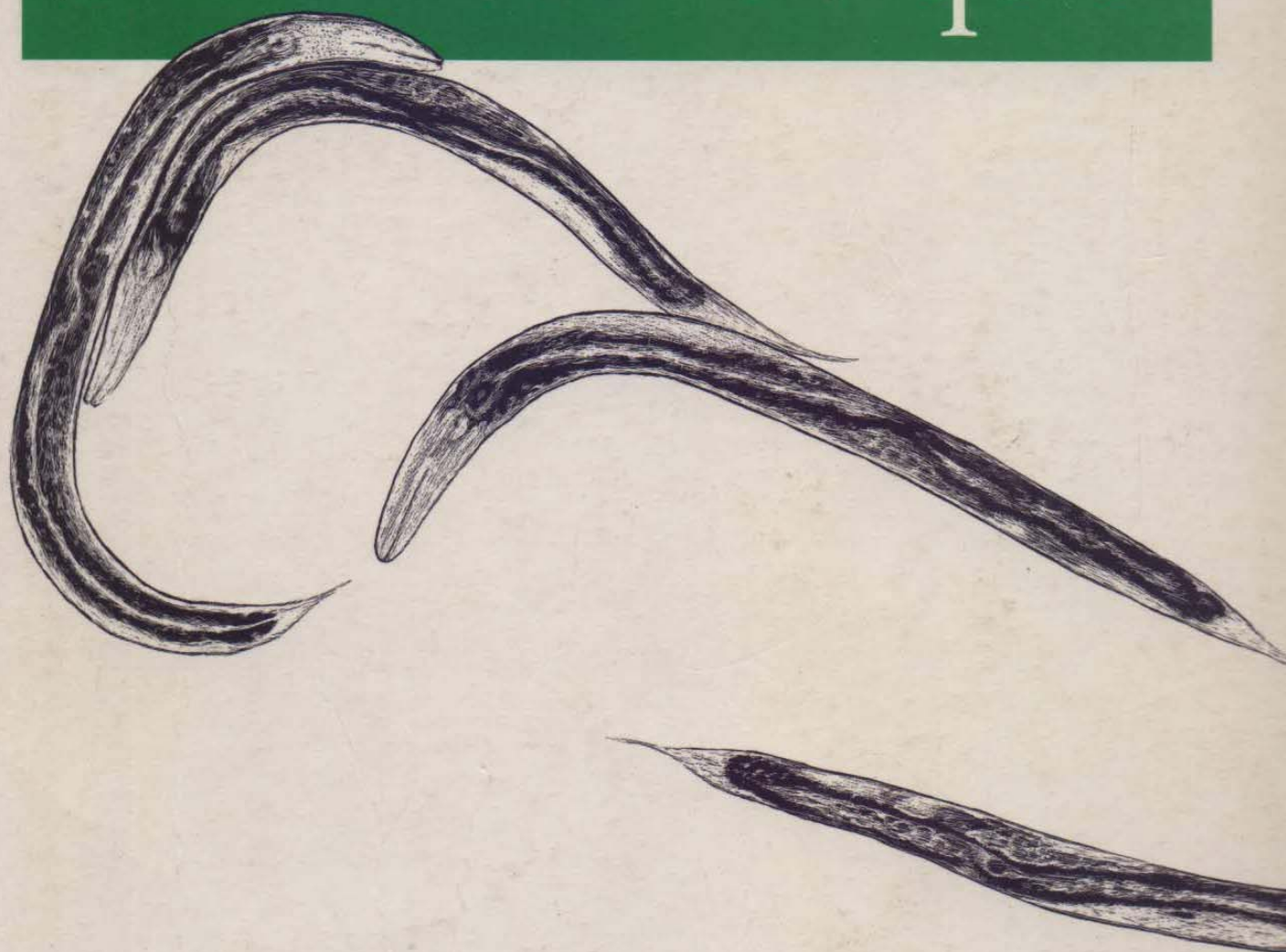


Concepts fondamentaux et outils logiciels

**Édition
française**

Introduction à la bioinformatique



Cynthia Gibas & Per Jambeck

*Traduction d'Hélène Dauchel,
Isabelle Milazzo et Laurent Mouchard*

O'REILLY®

Table des Matières

Préface	ix
À qui s'adresse ce livre ?	x
Structure de ce livre	xi
Notre approche de la bioinformatique	xiii
Conventions typographiques	xiii
Pour nous contacter	xiii
Pour être tenu au courant de nos publications	xiii
Remerciements	xiv
Édition française	xv
 I Introduction	 1
 1. La biologie à l'ère de l'informatique	 1
Comment l'informatique transforme-t-elle la biologie ?	4
Le but de la bioinformatique serait-il seulement de créer des bases de données ?	8
Que représente l'informatique pour les biologistes ?	12
Quels défis la biologie offre-t-elle aux informaticiens ?	12
Comment configurer son PC pour faire de la recherche en bioinformatique ? ..	14
Quels sont les informations et les logiciels disponibles ?	15
Est-il possible d'apprendre un langage de programmation sans enseignement ? ..	16
Quels sont les informations et les logiciels disponibles ?	17
Comment ranger, visualiser et interpréter ses données ?	17
Comment écrire un programme pour aligner deux séquences biologiques ?	18
À quelles questions la bioinformatique permet-elle de répondre ?	18
 2. Approches informatiques de quelques questions biologiques	 21
Mécanismes fondamentaux en génétique moléculaire	21
Quelles modélisations en biologie ?	25
Une expérience en bioinformatique	34

II.	<i>Station de travail bioinformatique</i>	39
3.	<i>Installation d'une station de travail</i>	41
	Travailler sur un système Unix	41
	Installer une station de travail sous Linux	44
	Comment faire fonctionner les logiciels ?	52
4.	<i>Fichiers et répertoires sous Unix</i>	59
	Les notions de base d'un système de fichiers	59
	Commandes opérant sur les répertoires et les fichiers	64
	Travailler dans un environnement multi-utilisateurs	71
5.	<i>Travailler sur un système Unix</i>	83
	Les shells Unix	83
	Émettre une commande sous Unix	84
	L'affichage et l'édition des fichiers	89
	Transformations et filtres	95
	Les statistiques sur les fichiers et les comparaisons	101
	Le langage des expressions régulières	103
	Les scripts shell Unix	106
	Communiquer avec d'autres ordinateurs	107
	Bien s'entendre avec les autres dans un environnement partagé	112
	Tableaux récapitulatifs	124
III.	<i>Outils pour la bioinformatique</i>	129
6.	<i>Recherche d'informations biologiques sur le Web</i>	131
	Retrouver des articles scientifiques	132
	Les bases publiques de données biologiques	137
	Recherche d'informations dans les banques de données biologiques	142
	Soumission de données dans les banques publiques	148
	Recherche de logiciels	149
	Juger de la qualité de l'information	149
7.	<i>Analyse et comparaison de séquences</i>	151
	Prédiction de gènes	152
	Traduction et traduction inverse	153
	Comparaison de séquences deux à deux	154
	Comparaison d'une séquence contre toutes les séquences d'une banque de données biologiques	163
	Outils multifonctions pour l'analyse de séquence	168
8.	<i>Alignements multiples de séquences, arbres et profils</i>	171
	Alignement multiple de séquences	171
	Analyse phylogénétique	175
	Profils et motifs	181
9.	<i>Visualisation de structures protéiques et calcul de propriétés structurales</i>	189
	Les formats des données de structures protéiques	190
	La structure chimique des protéines	190
	Outils web pour l'analyse de la structure des protéines	202

Visualisation des structures	204
Classification des structures	210
Alignement structural	215
CE et CL	217
VAST (Vector Alignment Search Tool)	218
Analyse Structurale	218
Accessibilité au solvant et interactions	221
Calcul des propriétés physicochimiques	224
Optimisation d'une structure	226
Bases de données de protéines	230
Presage	231
En résumé	231
10. Prédiction de la structure et de la fonction d'une protéine à partir de sa séquence	235
Détection de caractéristiques fonctionnelles dans les séquences protéiques	235
Détermination expérimentale de structures protéiques	236
Prédiction de structures protéiques	240
Prédiction de structure secondaire	242
Prédiction de la structure 3D	248
En résumé : un projet de modélisation de protéine	252
Conclusion	256
11. Outils pour la génomique et la protéomique	259
Du séquençage des gènes au séquençage des génomes	260
Accéder aux informations génomiques sur le Web	266
L'annotation et l'analyse de séquences de génomes complets	269
Génomique fonctionnelle : les nouveaux défis de l'analyse de données	272
Protéomique	277
Les bases de données de voies métaboliques	281
Modélisation de la cinétique et de la physiologie	284
Résumé	286
IV. Bases de données et visualisation	289
12. Automatisation de l'analyse des données avec Perl	291
Pourquoi Perl ?	291
Les bases de Perl	292
Recherche de motif et expressions régulières	298
Traiter une sortie de BLAST en utilisant Perl	299
Utiliser Perl en bioinformatique	304
LWP	306
13. Construire des bases de données biologiques	309
Différents types de bases de données	309
Logiciels pour les SGBD	315
Introduction à SQL	317
Installer le système de gestion de base de données MySQL	322

Conception de la base de données	326
Développer un programme web pour interagir avec des bases de données	330
14. Visualisation et fouille de données	337
Préparer les données	337
Affichage de documents graphiques	338
Visualisation de séquences	339
Visualisation des réseaux et voies métaboliques	342
Travailler avec des données numériques	343
Visualisation : récapitulation	349
Fouille de données et information biologique	350
Bibliographie	355
Unix	355
Administration Système	355
Perl	356
Références générales	356
Bioinformatique	356
Biologie moléculaire - Biologie	357
Structure des protéines et Biophysique	357
Génomique	357
Biotechnologie	357
Bases de données	358
Visualisation	358
Fouille de données	358
Index	361