

## ANALYSE BIBLIOGRAPHIQUE

Pavel A. Pevzner, *Bio-informatique moléculaire : une approche algorithmique*, traduction de Delphine Hachez, Springer, collection Iris, 2006.

La bio-informatique est un domaine de recherches interdisciplinaires, né de la collaboration de biologistes, de mathématiciens et d'informaticiens pour traiter des problèmes de biologie moléculaire. Ceux-ci sont liés à la connaissance des séquences biologiques (ADN, ARN, séquences protéiques) et à la modélisation des problèmes quand on aborde la biologie de cette façon. L'explosion de la bio-informatique date des toutes premières années 1990, parallèlement aux techniques de séquençage qui permettent d'établir de plus en plus rapidement les séquences proprement dites. La discipline évolue très vite car les progrès technologiques et méthodologiques sont fulgurants. Il est fort intéressant de souligner que cette collaboration a aussi débouché sur des travaux spécifiques en mathématiques (probabilité, espaces métriques discrets) et en informatique (algorithmique combinatoire, imagerie numérique) qui sont innovants dans ces disciplines.

Le lecteur de *Mathématiques et Sciences humaines* peut s'interroger sur les raisons pour lesquelles cet ouvrage est commenté dans la revue. La raison en est fort simple ; beaucoup d'outils de modélisation en bio-informatique sont pertinents en sciences humaines ; je ne citerai que l'étude des réseaux biologiques, si proches des réseaux sociaux, des ordres de gènes, en tous points comparables aux ordres d'événements ou de candidats à une élection, de la classification (fonctionnelle) des protéines par partitionnement de graphes, de la phylogénie qui englobe les méthodes de distances et la classification par arbre, de la comparaison des séquences qui empiète sur la lexicométrie, de la recherche de catégories sociologiques à l'aide de représentations arborées, des alignements de séquences que l'on retrouve en stratigraphie.

P. Pevzner, d'origine russe, a traversé toute cette période de développement de la bio-informatique, de 1985 à nos jours, et il a été l'un des informaticiens les plus dynamiques de cette époque, en particulier sur le séquençage par hybridation (chapitre 5 ou comment établir une séquence à partir de la liste de ses mots) et sur la *reversal distance* (chapitre 10) qui consiste à déterminer le plus petit nombre de retournements d'intervalles pour passer d'une permutation à une autre. Mais il a également abordé la quasi-totalité des problèmes combinatoires rencontrés en bio-informatique ; la cartographie physique (chapitre 2, ou comment localiser un gène), le séquençage ou l'assemblage de fragments en une séquence unique (super chaîne), la détection de zones codantes par chaîne de Markov cachées (chapitre 9), la recherche d'une séquence dans une banque ou des séquences les plus proches d'une séquence donnée, l'alignement multiple (positionnement relatif des séquences dérivées d'une unique séquence ancestrale) et bien d'autres.

Les titres des chapitres, qui sont en termes biologiques, seraient ici peu motivants. Mais derrière cette succession de sujets, sont détaillées les connaissances mathématiques nécessaires aux développements méthodologiques qui suivent. Soulignons également la qualité de la traduction qui évite d'ajouter à l'ésotérisme de la terminologie biologique les difficultés de la langue anglaise. Elle a de plus, chapitre par chapitre, été relue par des spécialistes français.

Le chercheur méthodologue en sciences humaines ou sociales trouvera dans la lecture de l'ouvrage de P. Pevzner nombre de réponses à des questions qu'il s'est déjà posées ou qui trouveraient leur pendant dans son propre domaine. Pas plus que l'auteur, qui le prouve dans son annexe de trois pages « Introduction à la biologie moléculaire »,

il n'aura besoin d'être bien savant dans ce domaine. Il pourra apprécier les solutions qui sont développées à partir des problèmes formalisés à l'aide de modèles universels et d'outils communs, comme les tables binaires, les ensembles ordonnés, les permutations, les graphes et toutes les composantes des chaînes de caractères (mots, sous-séquences, motifs).

A. Guénoche