

COMMISSION INTERDISCIPLINAIRE 44

MODÉLISATION DES SYSTÈMES BIOLOGIQUES, BIOINFORMATIQUE

Président de la CID
Bernard PRUM

Membres de la CID
Dominique CELLIER
Christian COGNARD
Gilbert DELEAGE
Érick DUFOURC
Pierre-Olivier FLAVIGNY
Christine FROIDEVAUX
Olivier GANDRILLON
Olivier GASCUEL
Thierry GRANGE
Nicolas HERMANN
Julie MAALOUM
Nathalie MAUREL
Christian MICHEL
Benoît PERTHAME
Marie-France SAGOT
Thomas SIMONSON
Éric WESTHOF

La biologie à grande échelle est à l'origine d'une masse considérable de données qui concerne tous les niveaux du vivant :

- les gènes, les protéines et leurs interactions ;
- les génomes, leur dynamique et leur évolution ;
- les cellules, leur organisation et les mécanismes moléculaires sous-jacents ;
- les organes et leur fonctionnement ;
- les organismes et leur physiologie ;
- les espèces et populations ;
- les systèmes écologiques.

L'exploitation de ces données est au cœur de la CID 44. Elle requiert à la fois des modèles mathématiques et physiques qui représentent les lois complexes du vivant, et des travaux en informatique, pour simuler ou estimer ces modèles, fouiller les données, et pour intégrer toutes ces sources d'informations hétérogènes au sein de bases de données et de connaissances. L'objectif est une meilleure compréhension du vivant, avec des enjeux dans tous les domaines, environnementaux, agronomiques, médicaux et pharmaceutiques. Les années passées ont vu ces disciplines se développer de façon extraordinaire (les articles les

plus citées aujourd'hui, toutes sciences confondues, sont liés à l'exploitation informatique des données génomiques). Le mouvement continuera très certainement. La biologie de demain sera largement faite par des biologistes « secs », modélisateurs et/ou bioinformaticiens, par opposition aux biologistes « humides » travaillant à la paillasse. L'objectif de la CID 44 est de favoriser les recherches dans ces domaines d'interface, en mettant en avant des chercheurs et des travaux innovants sur le plan méthodologique et répondant à des questions biologiques importantes. On trouvera dans la suite les principaux axes de recherche concernés, avec un regard plutôt biologique tout d'abord (quelles grandes questions biologiques?), puis plutôt méthodologique (quels modèles? quels algorithmes? quelle intégration des données?). Ces deux regards sont le plus souvent indissociables, mais ce mode de présentation facilitera la lecture par les tenants des différentes disciplines d'origine. Finalement on tracera un rapide état des lieux, avant de conclure par les recommandations essentielles.

1 – GÉNOMIQUE COMPARATIVE ET FONCTIONNELLE

On a pu croire qu'après le séquençage du génome humain la course aux génomes allait ralentir. On assiste en réalité à une forte accélération. Cette accélération, facilitée par l'apparition de nouvelles techniques de séquençage rapides et peu coûteuses, est due à l'intérêt de comparer les génomes et d'explorer les divergences évolutives à différentes échelles, depuis les études intra-spécifiques jusqu'aux analyses regroupant les grands domaines du vivant. Ainsi, 500 génomes procaryotes sont séquencés actuellement et plus de 700 projets de séquençage de génomes bactériens sont en cours. Plusieurs souches de *E. Coli* ont déjà été entièrement séquencées, tandis que 49 génomes complets d'eucaryotes sont

aujourd'hui disponibles. Et on pourrait multiplier les exemples (levures, virus, etc.), tandis que s'estompent les frontières entre espèces: plus de 70 projets de séquençage de métagénomés sont en cours, visant à caractériser le matériel génétique à partir d'échantillons environnementaux.

Cette masse considérable de données représente un nouveau changement d'échelle. La bioinformatique s'est d'abord consacrée aux gènes, à leur comparaison, leur évolution et leur fonction. Elle peut désormais s'attaquer aux génomes. Les enjeux sont considérables. Par exemple, les études sur les bactéries sont liées à la pathogénicité ou à la diversité et à l'adaptation, et ont donc un impact potentiel direct en santé et en biotechnologie. Des génomes essentiels comme celui de *Plasmodium Falciparum* (l'agent de la Malaria, responsable de deux millions de morts chaque année dans le monde) sont largement incompris et on attend beaucoup des études comparatives. Mais les difficultés sont évidentes. Se pose tout d'abord le problème de la taille: on passe typiquement de mille nucléotides pour un gène à plusieurs millions de paires de bases dans un génome de petite taille, d'où des difficultés algorithmiques importantes. Et on sait mal aujourd'hui comment évoluent les génomes (réarrangements, éléments transposables, rôle des virus, etc.), ce qui impose des travaux de modélisation et d'estimation statistique des différents modèles envisagés.

La biodiversité se développe suite à des contingences historiques et une évolution moléculaire neutre encadrée par des contraintes structurales de développement et de régulation. Alors que la microévolution (développement horizontal de l'arbre phylogénétique) est dominée par les contingences historiques, la macroévolution (développement vertical de l'arbre phylogénétique) est très fortement contrainte pour des raisons structurales et de contrôle. La génomique comparative permet de départager et de dégager les éléments moléculaires responsables de ces évolutions biologiques. Ainsi, les travaux récents de comparaisons des génomes du macaque, du singe et de l'homme ont montré

que l'allèle trouvé chez le macaque correspond à l'allèle malade chez l'homme. De telles comparaisons, généralement aussi surprenantes qu'innovantes, ont de profondes répercussions sur notre compréhension et le traitement de ces maladies humaines.

La génomique fonctionnelle qui s'attache entre autres, à comprendre la fonction des gènes, comprend notamment l'étude massive des transcriptions, des protéines et de leurs interactions au sein d'une cellule. On assiste à la multiplicité des «-omics» (genomics, transcriptomics, proteomics, metabolomics etc.). Toutes ces nouvelles approches ont transformé la biologie moléculaire moderne d'une science «pauvre en données» en une science «riche en données». Elles permettent d'envisager, dès à présent, des analyses intégrées allant des séquences complètes des génomes aux conséquences phénotypiques de mutations, en passant par les aspects structuraux et fonctionnels sur les différents acteurs cellulaires. Face à ce volume croissant de données complexes et hétérogènes, l'intégration des données couplée à des analyses bioinformatiques comparatives et prédictives est cruciale pour réaliser la description étendue de la fonction d'un gène et de la compréhension de son rôle non seulement au niveau moléculaire, mais également aux niveaux supérieurs des voies cellulaires, des complexes macromoléculaires, de la cellule ou de l'organe.

2 – BIOLOGIE DES SYSTÈMES, RÉSEAUX

La biologie des systèmes, ou biologie systémique, est souvent minimisée ou même décriée. Certes, une définition claire et acceptée n'est pas simple à trouver. La biologie moléculaire et structurale produit une quantité incroyable de données précises et fines chaque année. Leur intégration en une vision tout à la fois cohérente et synthétique reste cependant bien souvent fort lointaine. Une

des raisons est que nous ne gérons pas au niveau théorique la syntaxe des flux de l'information biologique. L'objectif de la biologie des systèmes est l'étude des réseaux dynamiques créés par les objets biologiques en interaction. La biologie des systèmes ne se borne donc pas à l'étude des réseaux métaboliques ou génétiques. Des apports théoriques importants ont été réalisés ces dernières années sur les réseaux biologiques ou non. Les similarités entre ces divers types de réseaux sont surprenantes. Toutefois, une caractéristique frappante des réseaux biologiques est leur robustesse couplée à leur évolutivité. L'espace des systèmes biologiques sur lequel nous pouvons agir à des fins d'intervention, de contrôle ou de régulation apparaît donc restreint. Les conséquences de ces observations sont lourdes pour les développements en thérapie humaine et seule une compréhension globale des interactions permettra de dégager des nouvelles voies d'attaque. Les conséquences de ces travaux en biologie (au niveau fondamental tout comme pratique) n'en sont pas encore bien comprises et peu diffusées. Les travaux de biologie systémique ouvrent la voie à la biologie synthétique dont les buts consistent à utiliser en améliorant et en simplifiant les systèmes biologiques à des fins d'ingénierie et de production. Au niveau informatique, de nombreuses difficultés inhérentes à la biologie systémique proviennent des structures mêmes de recherche et de diffusion des connaissances : nous partageons et diffusons généralement des modèles plutôt que des données brutes. Comment permettre à des ordinateurs de s'échanger ces modèles, de les vérifier sur les données brutes, de les retrouver dans les publications, de les assembler et de partager les résultats avec d'autres?

3 – BIOINFORMATIQUE STRUCTURALE

Le grand enjeu de la bioinformatique structurale est de comprendre la relation

entre structures macromoléculaires et fonctionnement biologique de la cellule. Cela passe par une compréhension large des macromolécules et de leurs complexes, en prenant en compte leurs structures, leurs mouvements, et leurs interactions. Il faut donc décrypter le vocabulaire et la grammaire d'un dialogue moléculaire très complexe à différentes échelles de taille (de l'atomique au mésoscopique), de temps (de la picoseconde à la seconde) et d'espace (compartimentation, diffusion, encombrement cellulaire). La bioinformatique structurale est fortement associée aux développements de la biologie structurale expérimentale, et donc à la fois aux programmes massifs de séquençages génomiques et aux programmes de détermination de structures tridimensionnelles. En effet, les aspects structuraux (protéiques et nucléiques, expérimentaux et théoriques) sont essentiels à la réussite des grands projets de génomique.

Dans ce contexte, plusieurs secteurs importants de la bioinformatique structurale doivent continuer à être développés ; d'autres, très récents ou nouveaux, doivent être renforcés ou créés. Un premier groupe de secteurs sont liés au problème du repliement des protéines et à la prédiction de structures : identification et classification de motifs structuraux, développement de méthodologies comparatives au niveau structural, phylogénie structurale, analyse structurale prédictive des séquences/structures (ADN, ARN, protéines), problème inverse du repliement, modélisation par homologie à grande échelle. Un deuxième groupe se situe à l'interface avec les techniques de biologie structurale expérimentale : reconstruction de gros édifices 3D en utilisant des données hétérogènes (cryoEM, RMN liquide et solide, biocristallographie, fluorescence, imagerie moléculaire, AFM) ; ces secteurs nécessitent des couplages de codes informatiques et de nouveaux développements méthodologiques. Un troisième groupe concerne la compréhension des mécanismes d'assemblage macromoléculaires (e.g., approches multi-échelles) et des forces mises en jeu (e.g., expérimentations sur molécules uniques) et de leur dynamique ; ce secteur concerne notamment les réseaux d'interaction protéine-protéine,

mais aussi l'autoassemblage de membranes lipidiques. Un quatrième secteur concerne plus directement les relations structure-fonction et les mécanismes détaillés de macromolécules d'un intérêt particulier. Il inclut la simulation des mouvements moléculaires (domaines, approches de ligands), les technologies d'ingénierie *in silico* de protéines ou de ligands (d'intérêt, par exemple), et la simulation en dynamique moléculaire de gros systèmes associant protéines, membranes, ou ARN (e.g., ribosome, facteurs de transcription, protéines membranaires, entourées d'un modèle réaliste de leur environnement).

4 - ÉVOLUTION ET ADAPTATION : DU GÈNE À L'ÉCOLOGIE

L'évolution et l'adaptation forment un autre grand pan de la biologie. Les objets biologiques sont issus d'un processus d'héritage et de mutations, et ceci à toutes les échelles, du gène aux systèmes écologiques en passant par les espèces. Comprendre et retracer l'évolution de ces objets est souvent un pas décisif dans la compréhension de la fonction (par exemple, des gènes), dans l'élucidation de la structure (par exemple, des protéines), ou de la place dans un ensemble complexe (par exemple, des espèces au sein des écosystèmes). Retracer l'évolution apparaît également essentiel dans l'études des maladies émergentes ou en évolution constante, telles que le SIDA, le SARS ou la grippe. Les études évolutives sont au centre des grands projets internationaux sur l'Arbre de la Vie, qui est la phylogénie de l'ensemble des espèces contemporaines et constituera un répertoire remarquable de la biodiversité globale.

Cette capacité à comprendre et modéliser le passé devrait prendre une nouvelle dimension dans ses applications écologiques avec les études sur le réchauffement climatique. Comment les espèces s'adapteront-elles ? Quels

nouveaux équilibres entre espèces se formeront localement, quels seront les impacts sur les sociétés humaines ? Ces questions se posent naturellement dans un contexte multidisciplinaire : la formalisation mathématique y est variée et ancienne ; les méthodes informatiques avancées sont indispensables ; le couplage de codes climatiques et de dynamique des populations est à l'ordre du jour. La modélisation biologique et géophysique, ainsi que l'intégration des données génomiques, phénotypiques, écologiques, et climatiques, sont fondamentales et nécessaires.

5 – MODÉLISATION MATHÉMATIQUE

Toute science passe depuis des siècles par des modélisations, très essentiellement de nature mathématique. La physique et la chimie sont sorties d'un mode descriptif au dix-huitième siècle avec la mise en équation de l'attraction des corps ou des réactions chimiques, passant plus tard par des calculs de potentiels. Si la biologie dans son essence a échappé à cette modélisation – il y a des contre-exemples, tels les modèles proie-prédateurs ou les dynamiques de populations –, c'est clairement du fait de sa complexité.

Ce mot, « complexité », est celui qui caractérise en premier les domaines d'interdisciplinarité reconnus (le cerveau, l'univers, les sciences sociales). Suivant un mécanisme curieusement observé tout au long de l'histoire des sciences, la disponibilité des outils abstraits va de pair, et souvent précède leur emploi dans les disciplines « concrètes ». Aujourd'hui les outils conceptuels, au premier rang desquels les outils mathématiques, ont acquis depuis un siècle, disons, la capacité de traiter de tels problèmes complexes : les fonctions dérivables ont cédé le pas au mouvement brownien et à ses avatars, les espaces euclidiens à des espaces de Hilbert de dimension infinie, les équations différentielles (réduites à être « ordi-

naires »!) partagent le terrain avec des systèmes dynamiques de plus en plus complexes.

Cette complexité se reflète d'abord par une complexité accrue des modèles mathématiques, faisant souvent appel – ce qui est symptomatique de l'ampleur des problématiques – à des domaines relevant des maths dites « pures » : c'est le cas des systèmes dynamiques et des graphes, par exemple pour les processus biochimiques, leur fonctionnement et leur évolution. C'est également le cas de la théorie des jeux, pour ce qui concerne les systèmes écologiques et leurs fragiles équilibres. C'est aussi le cas de la géométrie appelée à jouer un rôle central, non seulement pour modéliser dans l'espace usuel les positions relatives des molécules pour mieux comprendre leurs interactions (par exemple, notion de site actif présenté à un substrat), mais surtout pour rendre compte dans des espaces de très haute dimension (espaces de lacets, par exemple) de la topologie des molécules biologiques (brins d'ADN, par exemple) et de leurs déplacements possibles (repliement, ouverture, etc.) sans doute le long de géodésiques dans ces espaces complexes. Cette complexité devrait aussi conduire à des modélisations multi-échelles pour lesquelles des événements de natures différentes (événements moléculaires de nature stochastique, événements cellulaires et tissulaires plus déterministes) seront intégrés dans un même modèle utilisant des formalismes adéquats et différents selon les niveaux.

Mais les mathématiques plus traditionnellement tournées vers les applications conservent un grand rôle dans l'interaction math-biologie. Toutes les méthodes de maximisation sont mises à contribution pour tâcher de modéliser cette extraordinaire optimisation qui caractérise le Vivant. Les équations différentielles modélisent aussi bien la dynamique moléculaire que les déplacements de populations. La modélisation probabiliste est largement employée, par exemple pour représenter les événements évolutifs (mutation, spéciation, réarrangements génomiques, etc.). Et bien sûr, tous les modèles devant être choisis sur des critères d'ajustement aux observations, tous leurs paramètres devant être estimés, toute

hypothèse devant être testée, les statistiques sont centrales dans cette interdisciplinarité.

6 – MÉTHODES ET OUTILS INFORMATIQUES

Un axe essentiel de recherche en bioinformatique est l'algorithmique du texte. Exploiter les données génomiques nécessite des algorithmes toujours plus fins et rapides pour explorer les banques de données, trouver par alignement les séquences homologues à une séquence donnée, ou détecter des motifs et signaux, notamment représentés par des automates probabilistes ou des modèles de Markov cachés. L'alignement multiple est une tâche fondamentale (pour caractériser une famille ou en retracer l'évolution) qui nécessite toujours des développements algorithmiques. Plus récemment, de nombreux efforts se sont portés sur les réarrangements génomiques et le calcul de distances de réarrangements ; ces travaux ont entraîné des développements fondamentaux sur les permutations, qui forment aujourd'hui un champ très actif de la recherche en algorithmique.

L'algorithmique bioinformatique concerne aussi d'autres structures discrètes. De nombreux problèmes en évolution sont liés aux arbres (phylogénies) et à leur combinatoire, notamment dans le but d'assembler l'Arbre de la Vie à partir des arbres partiels que l'on peut déjà trouver dans les banques en grande quantité. Les arbres et leur algorithmique sont également centraux pour l'étude et la comparaison des structures d'ARN. L'algorithmique des graphes est elle aussi largement mise à contribution et développée, en raison de la quantité sans cesse croissante de données de type graphe, qu'il s'agisse d'interactions protéiques, ou de réseaux métaboliques et de régulation, dont on dispose pour de nombreux organismes. L'objectif est alors de comparer, par exemple pour trouver des régularités statistiquement significatives.

D'autres tâches nécessitent clairement des développements informatiques. En particulier, de nombreux problèmes se rattachent à l'optimisation et/ou au calcul massif impliquant le développement de codes parallèles et distribués. C'est notamment le cas en matière de structure et de dynamique moléculaire. Il est significatif que dans ce domaine des travaux prometteurs aient été inspirés par la robotique et le contrôle des bras articulés, en suivant l'analogie mécanistique entre bras et chaîne protéique.

L'informatique doit également répondre aux défis présentés par l'analyse intégrée des données de la biologie à grande échelle. Celles-ci sont non seulement volumineuse et en croissance rapide, mais également très hétérogènes (combinant, par exemple, des données de séquences, des annotations sur celles-ci, et des données textuelles servant de référence pour ces annotations). Les nouvelles méthodes doivent incorporer des composants de fouilles de données développés dans les domaines de la statistique et de l'intelligence artificielle, ainsi que des méthodes de classification. En outre, des techniques d'analyse de l'information doivent être utilisées, allant de la validation et de l'affinement des données, jusqu'à l'extraction des informations pertinentes et leur utilisation dans un cadre d'aide à la décision. L'intégration étroite de ces protocoles et logiciels dans des ensembles entièrement automatiques sera nécessaire. Cette intégration exigera également des formats de données, des modèles et des ontologies standard, afin de rendre le transfert d'informations aussi transparent que possible.

7 – ÉTAT DES LIEUX

Une étude bibliométrique rapide (1) permet de positionner les recherches se faisant en France dans ces domaines. En matière de publications en bioinformatique (2), la France se place en 4^e position (~ 120 publications),

très loin derrière les USA (~ 1000), mais aussi assez loin derrière l'Allemagne (~ 260) et l'Angleterre (~ 220), deux pays qui ont fortement investi le domaine depuis une bonne dizaine d'année, et qui bénéficient de la présence de laboratoires Européens. La France devance légèrement le Japon, le Canada et la Chine (~ 90). Pour ce qui concerne les publications en biologie des systèmes (3), la France (23) est mal placée, loin derrière USA (~ 200), Japon (~ 85) ou Angleterre (~ 75), mais aussi, par exemple, derrière l'Italie (37) et l'Espagne (28). On peut voir dans ces chiffres la conséquence d'une certaine inertie ; la biologie des systèmes est depuis quelques années largement mise en avant au niveau mondial, mais la France n'a pour l'instant fait que peu d'efforts institutionnels (1 appel ANR plutôt restrictif en 2006 et 2007, rien auparavant). Ce même facteur (avec un décalage d'une dizaine d'années, et accentué par le manque de continuité des programmes et appels d'offres) explique sans doute aussi les résultats seulement honorables en bioinformatique, par comparaison avec l'Allemagne par exemple.

Une autre mesure simple est la présence du thème bioinformatique dans les laboratoires du CNRS, telle qu'on peut la trouver dans l'annuaire des laboratoires sur le site du CNRS. En SdV, 63 laboratoires sur 310 sont fléchés bioinformatique, tandis qu'en MPPU on en trouve 12 sur 335, et en ST2I 11 sur 236. Même s'il ne s'agit pas d'une mesure réelle de l'activité, cela montre que les sciences de la vie ont bien compris l'intérêt de ces approches, mais que les mathématiques et l'informatique n'y ont encore mis que peu de forces. Il est également significatif que le thème biologie des systèmes (ou tout autre équivalent) n'apparaisse tout simplement pas sur le site, dans aucun laboratoire. Enfin, on trouve difficilement plus de 4-5 laboratoires dont le nom évoque directement la bioinformatique ou la modélisation des systèmes biologiques, alors que de tels laboratoires existent en grand nombre à l'étranger. Ceci montre, si besoin, que l'effort vers l'interdisciplinarité que constitue la CID 44 doit absolument se poursuivre et s'intensifier.

Depuis 2004, la CID 44 a assuré le recrutement (ou promotion CR1-DR2) de 33 chercheurs (*voir* le bilan 2007 pour plus d'informations). La pression était très forte, puisque nous avons auditionné environ 600 candidats. Quelques autres recrutements sur les thèmes de la CID 44 ont été faits dans d'autres sections (07, 21, 22, 29), mais avec généralement un caractère interdisciplinaire moins marqué. Notamment, ont été recrutés dans les sections de biologie des chercheurs appliquant des méthodes et programmes bioinformatiques, plutôt que contribuant à les faire progresser. Dans le même temps, de nombreux postes de bioinformatiques ont été affichés dans les Universités, pour répondre au besoin d'enseignements dans ces disciplines devenues indispensables à la biologie d'aujourd'hui. Ce bon niveau général de recrutement n'a malheureusement pas toujours été accompagné de la création de fortes équipes ou laboratoires, si bien que certains enseignant-chercheurs sont parfois isolés en ce qui concerne les aspects interdisciplinaires (ça n'est généralement pas le cas pour les recrutements CNRS et CID 44, où cet aspect est pris en compte lors des concours).

8 – RECOMMANDATIONS

On peut retenir des grands axes présentés ci-dessus quelques mots clefs : génomique comparative et fonctionnelle, biologie structurale, réseaux biologiques, environnement et biodiversité, dont le développement dans les années à venir nécessitera à l'évidence des développements spécifiques en modélisation mathématique (EDP, modèles stochastiques), en algorithmique (du texte mais aussi des arbres et des graphes), en classification, et en bases de données et de connaissances. Pour mener à bien ce programme, et maintenir des recherches de premier plan en France, il faut :

Accentuer les efforts en terme de postes interdisciplinaires, avec l'objectif de combiner au mieux :

1. réponses aux grandes questions de la biologie et développement des approches à grande échelle ;

2. avancées des travaux méthodologiques, car ceux-ci accompagnent (*voir précédent*) les progrès en biologie, et présentent souvent un intérêt propre dans leur discipline d'origine.

Ces postes devraient principalement relever d'une section interdisciplinaire (de type CID 4, pour assurer l'intérêt sur les deux versants), mais aussi des sections disciplinaires, et aller vers SdV, mais aussi vers les autres départements, ST2I et MPPU en particulier.

Ne pas oublier, comme c'est encore parfois le cas, le suivi des chercheurs à la marge de plusieurs disciplines, dont la carrière dépend d'une section pouvant lui préférer un chercheur davantage centré sur le cœur de sa discipline. Ce suivi doit bien sûr s'étendre aux équipes et laboratoires interdisciplinaires.

Développer les équipes ou laboratoires clairement situés à l'interface, en mettant l'accent sur ST2I et MPPU (*voir ci-dessus*) et

en multipliant les laboratoires et équipes bi-appartenant. À ce titre, une politique incitative doit être mise en place (relancée, car des efforts avaient été faits au tournant des années 2000), au travers de programmes CNRS, mais aussi de l'ANR. Cette dernière s'est jusqu'alors montrée peu interdisciplinaire, et il conviendrait de la pousser à lancer un programme d'interface ambitieux, entre biologie, mathématique et informatique.

Développer l'activité de services en bioinformatique, qui est indispensable aux biologistes. Pour être performante, cette activité doit absolument être adossée à la recherche. En retour, la recherche bénéficie de services performants, par exemple lorsqu'il s'agit de récupérer des données ou de mesurer les progrès apportés par telle ou telle méthode. Une bonne part de l'interface entre biologistes et chercheurs en modélisation et bioinformatique, passe par les plateformes de services. Le développement de cette activité implique essentiellement des recrutements d'ITA, qui stabiliseront et amplifieront les services aujourd'hui assurés par des CDD en nombre toujours croissant.

Notes

(1) Web of Science de l'ISI, période 2004-2007, nombre d'articles dont au moins un auteur est dans le pays considéré (les résultats sont similaires en considérant des périodes plus larges ou plus restreintes).

(2) Publications dans la revue *Bioinformatics* (Oxford) qui est la plus ancienne et a le facteur d'impact le plus élevé; des résultats similaires sont obtenus avec d'autres revues comme BMC Bioinformatics.

(3) Publications dans *Biosystems*, *Molecular Biosystems*, *Systems Biology*, *Molecular Systems Biology*.