

OFFRE de THÈSE (diffusion le 17/12/2010)

Sujet de thèse :

Modélisation par contraintes pour le contrôle génomique et physiologique de l'adaptation des algues brunes à la salinité de l'eau

Description :

Contexte : Les techniques d'observations du comportement d'une cellule et d'un organisme sont très variées. Deux approches coexistent. La première est focalisée sur la réponse génétique d'un organisme à un stress, on observe alors l'ensemble des composants de la cellule de manière assez grossière mais massive (transcriptomique, protéomique, ...). Les observations sont ensuite traitées avec des méthodes statistiques ou formelles pour identifier les ensembles de produits génomiques caractéristiques de la réponse au stress. La seconde est focalisée sur la mesure fine de composés biologiques en réponse à un stress environnemental. On regarde alors un petit nombre de produits, en mesurant finement leur concentration et de manière temporelle. Les données sont ensuite intégrées dans des modèles numériques « entrées-sorties » permettant de simuler l'impact de différents facteurs environnementaux sur la réponse de l'organisme.

Jusqu'à maintenant, ces deux approches étaient orthogonales, en partie à cause des différences majeures dans la qualité des données, mais surtout parce qu'aucune observation ne permettait réellement de relier les données de type génomique » aux données relatives à la réponse environnementale. Ce manque est de mieux en mieux comblé avec l'arrivée de nouvelles techniques d'observation du métabolisme intra-cellulaire, qui est à la fois une résultante des événements génomiques (via les enzymes qui sont impliquées dans les voies métaboliques) et une entrée pour la réponse à différents stress (puisque les voies métaboliques utilisent les différents nutriments pour finalement produire des composés d'intérêt). Actuellement diverses plateformes de production de données métaboliques en masse se développent, y compris dans le Grand Ouest, ouvrant de nouvelles perspectives scientifiques. Le travail de thèse s'appuiera notamment sur ces données pour identifier quels sont les éléments génomiques et métaboliques qui permettent à une souche d'algues « brunes », dont le milieu naturel est l'eau salée, de s'adapter à l'eau douce (Cock et al. 2010 ; Dittami et al. 2009). Un objectif final du travail sera de développer des méthodes pour prédire la réponse de ces souches à des différences de salinité, puis d'autres facteurs.

Programme de travail : Le but de cette thèse est de développer une chaîne de méthodes pour identifier de tels éléments en développant une approche d'intégration des données avec des approches de modélisation et de résolution par contraintes. Il s'agira de partir des différences entre les phénotypes (différences observées dans la production de métabolites) et de retracer leurs origines au sein de réseaux métaboliques puis de génomes. La principale difficulté ici sera de maîtriser la taille des jeux de données et les erreurs qui s'introduisent lorsqu'on met en relation les différents types d'observations. Concrètement, les recherches porteront sur la définition d'un formalisme en programmation par contraintes qui permettra d'exploiter toutes les informations disponibles sur le processus d'adaptation. Il s'agira aussi de trouver un compromis entre les méthodes formelles utilisées actuellement pour confronter les données grande-échelle de type « omique », et les méthodes numériques utilisées pour exploiter les données physiologiques. Pour cela, on procédera en plusieurs étapes.

- [Construction d'un réseau intégrant plusieurs niveaux de régulation] par alignement de cartes biologiques à l'aide de des méthodes de programmation par contraintes, en s'appuyant sur les travaux de (Van Hentenryck:2005) qui ont développés des méthodes efficaces de résolution.
- [Réduction du réseau au éléments fonctionnels impliqué dans le phénotype observé] Dans le cas des algues brunes, on se concentrera sur le métabolisme du mannitol. Un grand nombre de méthodes, principalement basées sur des approches de programmation par contraintes linéaires, existent pour étudier finement les réseaux métaboliques - notamment ceux de (Larhlmi & Bockmayr:2008) qui permettent d'obtenir une définition synthétique de l'ensemble des flux métaboliques possibles. Ces approches devront être adaptées aux réseaux fonctionnels produits dans le point précédent et s'appuyer sur les observations grande-échelle disponibles sur le métabolisme.
- [Sélection des ensembles régulateurs pertinents] Intégrer les observations transcriptomiques pour identifier les groupes fonctionnels qui s'expriment de manière globalement cohérente avec la topologie du sous-réseau ainsi construit. Les méthodes qui seront développées reposeront sur des tests de

consistances entre modèles et données à l'aide de programmation logique par contraintes, notamment par de la programmation par ensembles-réponses (Gebser et al, 2010).

- [Quantification du rôle des différents éléments] Enfin, exploiter les informations physiologiques disponibles pour quantifier le rôle de chaque groupe fonctionnel dans la production d'un composé donné. Les différents indices ainsi quantifiés pourront être comparés dans les souches considérées. On utilisera des approches probabilistes basées sur des analyses en moyenne de comportements, en revisitant tout particulièrement la question de l'inférence de modèles avec des formalismes d'optimisation sous contraintes mixtes discrètes et continues et des méthodes de recherche locales adaptées.

Finalité et perspectives : Comme mentionné ci-dessus, différentes méthodes à base de formalismes de contraintes ont déjà été développées pour des questions de bioinformatique. Cependant, ces méthodes ne portent pas sur des représentations des systèmes et des données adéquats pour envisager un procédé d'intégration global. Le travail du doctorant sera d'abord de montrer la faisabilité d'une démarche globale en adaptant et associant toutes ces méthodes. En un second temps, il faudra identifier les caractéristiques d'un formalisme de contraintes adapté pour réaliser l'ensemble de ces tâches de manière groupée. L'objectif final sera de quantifier l'effet de perturbation sur d'autres composés que ceux observés, ce qui pourra être validé expérimentalement via une expertise locale dans des travaux ultérieurs à la thèse.

Lieu : Pour la réalisation des travaux de thèse, l'étudiant sera entouré d'une expertise forte en bioinformatique et systèmes biologique (équipe Symbiose), mais aussi de spécialistes des algues brunes (station biologique de Roscoff) et de spécialistes de programmation par contraintes (équipe Symbiose, équipe Dream, équipes Combi et Tasc/contraintes au LINA-Nantes).

Département : D7 - Gestion des données et de la connaissance

Equipe : SYMBIOSE, <http://www.irisa.fr/symbiose/>

Directeur de thèse : Anne Siegel

Encadrant(s) : Thierry Tonon (*station biologique de Roscoff, UMR "marine plant and biomolecules"*)

Damien Eveillard (*LINA, équipe Combi*)

Contact : anne.siegel@irisa.fr

Début des travaux : Septembre 2011

Bibliographie :

Cock, J. M., L. Sterck, et al. (2010). The Ectocarpus genome and the independent evolution of multicellularity in the brown algae. *Nature* 465: 617-621.

Dittami, S. M., D. Scornet, et al. (2009). Global expression analysis of the brown alga *Ectocarpus siliculosus* (Phaeophyceae) reveals large-scale reprogramming of the transcriptome in response to abiotic stress. *Genome Biology* 10: R66.

Gebser M.; Guziolowski Carito; M. Ivanchev; Schaub Torsten; Siegel Anne; Veber Philippe; Thiele Sven (2010) Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming In Principles of Knowledge Representation and Reasoning, 2010, Toronto, Canada

Larhlmi, Abdelhalim and Bockmayr, Alexander: A new constraint-based description of the steady-state flux cone of metabolic networks. Discrete Applied Mathematics, 157, 2257-2266, 2009

Van Hentenryck, P. and Laurent Michel, Constraint-based Local Search, MIT Press, 2005

Hoos, H. H & Stutzle, T. (2004) Stochastic Local Search : Foundations & Applications (The Morgan Kaufmann Series in Artificial Intelligence). (Morgan Kaufmann),1 edition.