



**HAL**  
open science

## Statistique et génome

Rapport Hcéres

► **To cite this version:**

Rapport d'évaluation d'une entité de recherche. Statistique et génome. 2009, Université Evry-Val-d'Essone - UEVE, Institut national de la recherche agronomique - INRA. hceres-02032856

**HAL Id: hceres-02032856**

**<https://hal-hceres.archives-ouvertes.fr/hceres-02032856>**

Submitted on 20 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



agence d'évaluation de la recherche  
et de l'enseignement supérieur

Section des Unités de recherche

# Rapport d'évaluation

Unité de recherche :

Statistique et Génome – UMR 8071

de l'Université d'Evry-Val d'Essonne



décembre 2008



agence d'évaluation de la recherche  
et de l'enseignement supérieur

Section des Unités de recherche

# Rapport d'évaluation

Unité de recherche

Statistique et Génome

de l'Université d'Evry-Val d'Essonne



Le Président  
de l'AERES

Jean-François Dhainaut

Section des unités  
de recherche

Le Directeur

Pierre Glorieux

décembre 2008



# Rapport d'évaluation

## L'Unité de recherche :

Nom de l'unité : Statistiques et Génome

Label demandé : UMR

N° si renouvellement : UM5

Nom du directeur : M. Bernard PRUM (futur directeur M. Christophe AMBROISE)

## Université ou école principale :

Université d'Evry-Val d'Essonne

## Autres établissements et organismes de rattachement :

INRA/CNRS

## Date de la visite :

24 Novembre 2008



# Membres du comité d'évaluation

## Président :

M. Anestis ANTONIADIS, Université de Grenoble 1

## Experts :

M. Chris HOLMES, Université d'Oxford

M. Alain LECHARNY, Université Paris 10

Expert(s) représentant des comités d'évaluation des personnels (CNU, CoNRS, CSS INSERM, représentant INRA, INRIA, IRD...) :

(au titre du CoNRS) Mme Marie-France SAGOT

(au titre du CNU) M. Charles SUQUET

# Observateurs

Délégué scientifique de l'AERES :

M. Pascal AUSCHER

Représentant de l'université ou école, établissement principal :

Mme Jeanine TORTAJADA, Vice-Président Recherche

Représentant(s) des organismes tutelles de l'unité :

CNRS : M. Jean-Marc GAMBAUDO, DSA CNRS

INRA : M. Bruno GOFFINET, Directeur Département MIA, INRA



# Rapport d'évaluation

## 1 • Présentation succincte de l'unité

L'unité a fusionné en 2006 avec l'équipe CNRS-Université Génome et Informatique. Cette unité atteint, au 1<sup>er</sup> Octobre 2008, un effectif de 18 titulaires composé de 3 chercheurs (dont 2 émérites), 11 enseignants-chercheurs, 3 ingénieurs et une gestionnaire CNRS. Il convient d'ajouter 1 post-doc et 4 thésards. Chaque année l'unité accueille entre 5 et 8 stagiaires. Durant la période concernant l'évaluation une HDR a été soutenue au sein de l'unité. Le nombre de thèses soutenues lors des 4 dernières années est de 6 (et 1 abandon). La plupart de ces thèses ont été encadrées sur une durée moyenne de 3 ans. L'abandon est dû à la fermeture et la délocalisation de l'entreprise qui finançait cette thèse CIFRE. Il y a 3 membres parmi les enseignants-chercheurs qui bénéficient d'une PEDR. Parmi les 14 membres publiants (au sens des critères de l'AÉRES), il convient de signaler 9 enseignants chercheurs, 3 chercheurs (dont les 2 émérites) et les 2 ingénieurs de recherche de l'unité.

## 2 • Déroulement de l'évaluation

Le comité s'est réuni le 24 Novembre 2008 à Evry dans les locaux de l'unité afin, selon sa lettre de mission, d'évaluer le laboratoire Statistiques et Génome dirigé par Monsieur Bernard Prum. Nous avons procédé à l'évaluation du bilan de l'unité de recherche pour les quatre années écoulées ainsi qu'à l'évaluation du projet pour le quadriennal suivant. L'organisation de la visite par le laboratoire fut excellente. La visite commença le lundi matin par une présentation du laboratoire par son directeur ainsi que par un exposé synthétique des réalisations effectuées au sein de l'unité au cours des 4 dernières années. Elle fut suivie par une série de 4 exposés scientifiques, dont un sur les perspectives futures du laboratoire, faisant apparaître la richesse et la variété des thèmes de recherche du laboratoire. Le comité a particulièrement apprécié la clarté dont ont su faire preuve tous les orateurs, et le choix délibéré de la jeunesse dans la sélection des conférenciers.

L'après-midi commença par 3 exposés de collaborateurs biologistes témoignant des collaborations existantes entre généticiens, biologistes et médecins avec les chercheurs de l'unité sur le versant de la statistique. À la suite de ces exposés le comité rencontra les personnels I.T.A et I.A.T.O.S.S. puis s'entretint avec les thésards. Il rencontra ensuite le Conseil de Laboratoire, la discussion portant en particulier sur l'attractivité de l'unité pour le recrutement de nouveaux doctorants.

La fin de la visite fut consacrée à un entretien avec les tutelles, l'Université d'Evry étant représentée par son président. Le président a souligné toute l'importance du laboratoire Statistique et Génome pour l'Université. Il s'est félicité tout particulièrement du développement par le laboratoire de recherches mathématiques en interactions avec la génomique humaine et la génétique médicale et de son intégration dans le cadre de la Génopole. Le représentant du CNRS insista sur le caractère unique de ce laboratoire, qui est un endroit idéal pour le développement de l'interface des mathématiques avec les sciences du vivant qui est un développement essentiel pour le CNRS.

Enfin le comité s'est réuni à huis clos pour un échange de vues avant de se séparer vers 17 heures.



### 3 • Analyse globale de l'unité, de son évolution et de son positionnement local, régional et européen

Les paragraphes qui suivent proposent une synthèse des avis exprimés par les membres du comité d'évaluation. Chacun de ces évaluateurs a fourni un rapport succinct contenant ses impressions sur les activités de l'unité de recherche dans son ensemble ainsi qu'un avis plus détaillé sur le projet de développements futurs. La synthèse qui suit a été effectuée par le Président du comité et «amendée» par tous les évaluateurs.

Les évaluateurs tiennent en premier lieu à souligner la *haute qualité scientifique* de la recherche qui leur a été présentée. Ce laboratoire de mathématiques, spécialisé en statistique, a une longue histoire d'interaction avec la biologie et la génomique. Depuis 2006 l'interdisciplinarité qui le caractérisait déjà a été renforcée avec l'accueil en son sein d'un groupe de bioinformaticiens ayant une formation initiale principalement en biologie mais également en informatique. Bien que deux ans soit une période de temps trop courte encore, il ressort que la structure en une équipe unique présentée par l'unité et couvrant quatre directions de recherche n'est pas artificielle. Elle donne une bonne image de l'intégration entre les «anciens» et les «nouveaux membres» du laboratoire dans la thématique majeure des réseaux biologiques, avec, déjà, des projets intéressants en commun. La production scientifique est d'excellente qualité (60 publications listées sur la période et plusieurs développements de logiciels) et l'équipe semble avoir une vision scientifique claire de son développement à court terme. Soulignons que ce laboratoire occupe une place nationale centrale d'un point de vue thématique qui le conduit à interagir fortement avec plusieurs autres laboratoires de la région parisienne, notamment le laboratoire MIG (Mathématiques et Informatique pour la Génomique), unité propre de l'INRA et le laboratoire de Mathématique et Informatique Appliquées d'AgroParisTech. La collaboration avec l'IBISC à travers un projet ANR est à souligner ainsi que la participation active au groupe SSB qui structure remarquablement les statistiques pour la biologie dans la région parisienne. Les interactions avec les biologistes sont donc remarquables et se retrouvent dans l'organisation et la présentation des enseignements au niveau des masters.

L'opportunité d'un développement en bioinformatique a été saisie avec l'arrivée des membres du laboratoire Génome et Informatique depuis 2006. Cette nouvelle composante du laboratoire devrait permettre un élargissement des collaborations avec les biologistes. Les approches bioinformatiques ont donné lieu à des applications et des publications sur des sujets très compétitifs. C'est le cas pour la collaboration avec le Centre de Génétique Moléculaire sur l'occupation de l'ADN par les nucléosomes, la classification des séquences sans alignement pour la classification des virus HIV/SIV et des topoisomérases IA et les réarrangements de domaine chez les protéines bactériennes. Ces exemples montrent que les méthodes mathématiques développées peuvent participer à la résolution de questions biologiques actuelles concernant les relations entre d'un côté, la structure ou l'évolution des génomes et, de l'autre, le fonctionnement dynamique des génomes.

Une impression qui s'est dégagée est celle d'une visibilité internationale qui n'est pas à la hauteur de la valeur des membres de l'unité. Ce phénomène pourrait s'expliquer par un problème de langue mais c'est un aspect que le laboratoire devrait améliorer très activement afin de ne pas nuire à l'excellence de sa recherche. Le comité a également apprécié l'effort réel qui a été fait pour rendre accessibles les outils développés (MixNet, Mixer) et les logiciels développés et publiés sous forme de packages R sur les serveurs du CRAN, mais encourage les membres de l'équipe à situer clairement les codes développés dans le monde de la biologie en les rendant aussi disponibles sur les serveurs de Bioconductor.

Les projets de recherche proposés pour les prochaines années sont évidemment liés à un élargissement des thématiques sur les réseaux biologiques, la génétique génomique et la génétique évolutive. Ils s'équilibrent bien entre théorie et application, et abordent des problèmes importants et notoirement difficiles en mathématiques aussi bien qu'en biologie ou en santé. L'accent est mis sur les réseaux et les graphes aléatoires d'un côté, et sur l'évolution de l'autre, notamment la co-évolution des génomes avec le métabolisme ou la régulation considérée dans son ensemble et modélisée en réseau d'interactions. Ce dernier point est un problème bioinformatique particulièrement intéressant et permettra au laboratoire de ne pas oublier complètement ses anciennes préoccupations avec l'analyse de séquences. Un ancien thème semble plus en péril du fait de la retraite annoncée du directeur du laboratoire.



Il s'agit du versant statistique génétique et de la collaboration avec les généticiens et médecins. Il nous semblerait toutefois qu'une partie importante du groupe au moins aurait intérêt à explorer ce domaine de recherche qui va être conduit à se développer considérablement les prochaines années avec les nouvelles techniques de séquençage, et où de très nombreux problèmes de statistiques vont se poser. Si certains membres du groupe choisissaient effectivement de le faire dans les années à venir, nous leur recommanderions cependant très vivement de ne pas se contenter de recueillir des données pour y appliquer des méthodes classiques ou novatrices, mais de s'investir à fond dans le détail des diverses techniques employées et des biais introduits par chacune pour un maximum d'efficacité mathématique et bioinformatique.

Une relative distance avec la nature des données, et avec le détail des questions biologiques qui amènent à les traiter, semble encore exister, notamment chez certains plus jeunes et ceux plus portés vers des développements théoriques. Cela est dommage y compris pour les mathématiques. Il est notoire qu'un contact plus intime avec la biologie amène souvent aux problèmes les plus intéressants d'un point de vue même purement formel.

#### 4 • Analyse équipe par équipe et par projet

Comme dit plus haut, le laboratoire se présente sous la forme d'une équipe unique dont l'analyse a été présentée dans le paragraphe précédent.

#### 5 • Analyse de la vie de l'unité

- En termes de management :
- En termes de ressources humaines :
- En termes de communication :

Le comité a également apprécié la bonne ambiance qui règne au sein de l'unité. Aucune difficulté en termes de management et de communication n'est à signaler. Les trois ingénieurs sont parfaitement intégrés et la présence de deux d'entre eux est imputable à la co-tutelle actuelle par l'INRA qui semble très judicieuse pour le laboratoire et pour l'INRA. Ce serait une erreur à notre sens grave de la part de ce dernier de ne pas la maintenir. Si cette dernière n'était pas maintenue cela risquerait d'affecter des points clés du laboratoire en particulier par le départ de ces ingénieurs vers des unités de l'INRA et ceci pour faire évoluer favorablement leur carrière. Cependant, si cela était à envisager, comme de tels mouvements affectent des points clés du laboratoire, il serait impératif que les remplacements éventuels soient assurés par l'Université.

#### 6 • Conclusions

Le comité d'évaluation a été particulièrement impressionné tant par la qualité des recherches effectuées au Laboratoire Statistiques et Génome que par le dynamisme du laboratoire. Sans délaisser les domaines qui ont fait sa renommée, le laboratoire a su développer avec succès de nouvelles thématiques et dispose maintenant d'enseignants-chercheurs de premier plan international tant en mathématiques appliquées qu'en bioinformatique.





Cette évolution positive a été possible grâce à la conjonction de plusieurs facteurs : une politique de recrutement concertée et privilégiant à tout instant la qualité scientifique et l'appui indispensable des autorités de tutelle. Les projets de développement en cours sont également de très grande qualité et le comité les appuie avec la plus grande vigueur, certain qu'il en sera de même pour les autorités de tutelle, Université, CNRS et INRA.

En résumé on pourra souligner :

– **Points forts :**

La qualité des recherches effectuées au sein de l'unité (nombreuses publications dans des journaux de premier plan et plusieurs développements de logiciels). Le dynamisme du laboratoire (développement de nouvelles thématiques et organisations de sessions dans des congrès internationaux importants, des montages de journées thématiques et co-animation du groupe SSB). Des applications exemplaires.

– **Points à améliorer :**

La visibilité internationale qui n'est pas à la hauteur de la valeur des membres de l'unité. C'est un aspect que l'unité devrait améliorer très activement afin de ne pas nuire à l'excellence de sa recherche. Une relative distance avec la nature des données, et avec le détail des questions biologiques qui amènent à les traiter, semble encore exister, notamment chez certains plus jeunes et ceux plus portés vers des développements théoriques.

– **Recommandations :**

Le comité encourage les membres de l'équipe à situer clairement les codes développés dans le monde de la biologie en les rendant aussi disponibles sur les serveurs de Bioconductor. L'unité ou une partie importante de cette dernière aurait intérêt à continuer d'explorer le versant statistique génétique et la collaboration avec les généticiens et médecins car ce domaine de recherche va être conduit à se développer considérablement les prochaines années avec les nouvelles techniques de séquençage, et de très nombreux problèmes de statistiques vont se poser. Si certains membres du groupe choisissaient effectivement de le faire dans les années à venir, nous leur recommanderions cependant très vivement de ne pas se contenter de recueillir des données pour y appliquer des méthodes classiques ou novatrices, mais de s'investir à fond dans le détail des diverses techniques employées et des biais introduits par chacune pour un maximum d'efficacité mathématique et bioinformatique.

| Note de l'unité | Qualité scientifique et production | Rayonnement et attractivité, intégration dans l'environnement | Stratégie, gouvernance et vie du laboratoire | Appréciation du projet |
|-----------------|------------------------------------|---|--|------------------------|
| A               | A                                  | A   | A+   | A                      |

**Bernard PRUM**  
Professeur à l'Université d'Évry  
Directeur de Laboratoire  
"Statistique et Génome"  
UMR CNRS-INRA-UEVE  
prum@genopole.cnrs.fr  
01 60 87 38 06 (sec : 38 00 – fax 38 09) \*



Évry, le 24 février 2009

Mesdames et Messieurs les responsables de l'AERES,  
Mesdames et Messieurs les membres du comité d'experts

L'Université d'Évry a reçu le rapport établi par l'Aeres concernant le Laboratoire «Statistique et Génome», qui lie l'UEVE au Cnrs (UMR 8071) et à l'Inra (UMR 1152). L'Université remercie tout d'abord les experts et les membres de l'Aeres pour le considérable travail fait lors de l'étude du dossier, la visite du laboratoire et l'élaboration du rapport.

Ce rapport a été transmis au Directeur du Laboratoire, Bernard Prum, et au futur Directeur, Christophe Ambroise, et, c'est en accord avec eux que nous réagissons à ce document.

Nous sommes évidemment très satisfaits du ton généralement élogieux de ce rapport, qui encourage tous les membres du laboratoire à poursuivre leur recherche dans les axes qu'ils se sont définis.

Le laboratoire s'efforcera de suivre au mieux les recommandations faites dans le rapport, en particulier :

- en approfondissant ses collaborations avec les biologistes dans l'esprit souligné par le rapport de l'Aeres : développement de méthodes mathématiques pouvant participer à la résolution de questions biologiques ;
- en impulsant l'axe de recherche "statistique génétique" en lien avec l'implantation de nouvelles structures de recherche à Évry, en notant bien que le rapport recommande de concevoir des méthodes originales réunissant innovation mathématique et efficacité informatique pour obtenir des résultats des plus performants ;
- en renforçant sa visibilité internationale par des contacts renforcés, y compris par des accords de collaboration, avec des collègues étrangers.

La vision qu'a le laboratoire de la mission qui lui a été confiée et qui vient d'être rappelée (développement de mathématiques nouvelles au service de la biologie) l'a amené à s'interroger sur la critique que l'on peut percevoir dans ce rapport à propos d'«une relative distance avec la nature des données (qui) semble exister chez certains jeunes plus portés vers les développements théoriques».

L'équilibre abstraction théorique/applications au vivant est très délicat à trouver, que ce soit au niveau individuel ou au niveau d'un laboratoire. Tous les membres de l'unité *Statistique et génome* cherchent à développer des outils mathématiques dans une perspective à court terme de leur application par des biologistes. Ils ne sont pas nécessairement au même point d'équilibre, ce qui semble à la fois inévitable et souhaitable. De bonnes procédures d'analyses, de bons logiciels, ne peuvent être fondés que sur une base théorique solide – base que cherchent à renforcer les "jeunes théoriciens" (dont la connaissance de la génomique est bien sûr très au dessus de celle du mathématicien appliqué standard).

Les tutelles – le CNRS entre autres – ont fixé à ce laboratoire un double rôle : faire profiter la biologie des outils de la mathématique et alimenter la recherche en mathématiques en tirant des problèmes de la modélisation biologique. Nous pensons qu'il doit conforter cette approche, en maintenant une qualité scientifique aussi élevée que possible.

Bernard PRUM  
Christophe AMBROISE



UNIVERSITÉ D'EVRY  
VAL D'ESSONNE

**Université d'Evry-Val-d'Essonne  
Cabinet de la Présidence**

Affaire suivie par :  
Emery Olivier  
Téléphone :  
01.69.47.80.46

Evry, le 2 mars 2009

Le Président de l'Université

à

Monsieur Jean-François DHAINAUT  
Directeur de l'AERES

Objet : Rapport d'Evaluation de l'UMR 8071 – S2100015380

Monsieur le Directeur,

Vous m'avez transmis le rapport d'évaluation du laboratoire « Statistiques et Génome » (UMR 8071 ; CNRS-INRA-UEVE ; 0911975C) dirigé aujourd'hui par le Professeur Bernard PRUM, qui lui a été adressé.

Vous trouverez ci-joint, en réponse à ce rapport, les remarques et commentaires des co-signataires de ce courrier, les Professeurs Bernard PRUM et Christophe AMBROISE futur directeur de cette Unité de Recherche.

Je m'associe à leurs remerciements pour la qualité de cette expertise qui reflète la somme très importante de travail accomplie par les experts et les représentants de l'AERES.

Je souscris à l'ensemble de l'analyse de cette Unité et prends note des recommandations émises. L'Université s'efforcera d'accompagner au mieux ce laboratoire dans ses projets de développer des outils mathématiques novateurs et performants pour des applications biologiques et médicales avec en ligne de mire une meilleure visibilité internationale.

Je vous prie d'agréer, Monsieur le Directeur, l'expression de mes salutations distinguées.

Le Président de l'Université  
d'Evry-Val-d'Essonne

Richard MESSINA

