

## APPEL A PROJETS

**DATE LIMITE : 18 JUIN 2009**

1. **Présentation**
2. **Procédure**
3. **Dossier de candidature**

### PRESENTATION

En complément à la politique des PFR, au croisement des frontières disciplinaires traditionnelles, notre établissement souhaite encourager l'émergence de thématiques pluridisciplinaires. Dans ce but, riche de la diversité de ses disciplines fondamentales, l'Université Montpellier 2 a souhaité se doter de programmes pluridisciplinaires dont les finalités sont :

- contribuer à la structuration de la politique scientifique de l'établissement ;
- soutenir l'émergence de nouveaux champs thématiques en réponse à de grands enjeux de société ;
- accroître ainsi la lisibilité et la visibilité de l'établissement en termes de recherche et formation ;
- travailler au rayonnement et à l'insertion régionale et nationale de la thématique du programme.

Les programmes pluridisciplinaires aujourd'hui identifiés sur la base des compétences présentes au sein des PFR sont :

- Modélisation et simulation des sciences du vivant et de l'environnement ;
- Energie : sources, conversion, gestion et stockage ;
- L'eau: ressources, qualité, usages.

Une information spécifique a d'ores et déjà été fournie sous la forme d'appels à lettre d'intention ou à manifestation d'intérêt comme au travers de journées d'échanges scientifiques (<http://www.blog-um2.univ-montp2.fr>).

De manière générale, le présent appel d'offres vise notamment à :

- participer au développement de projets innovants dont la prise de risque associée est élevée par rapport à des demandes de financements traditionnels ;
- favoriser la constitution et aider au démarrage d'équipes pluridisciplinaires, susceptibles de perdurer au-delà de la proposition initiale.

#### **Communautés concernées**

Les projets proposés devront intégrer au moins une équipe rattachée à un laboratoire de l'Université Montpellier 2.

**Moyens alloués**

Les moyens demandés au titre de cet appel d'offres pourront inclure :

- une allocation doctorale, éventuellement environnée (10K€/an) ;
- une décharge d'enseignement, éventuellement associée au recrutement d'un demi ATER ;
- un soutien financier (fonctionnement, missions) de l'ordre de 10 à 30 k€.

**Evaluation**

La sélection sera effectuée par le Conseil Scientifique qui s'entourera d'experts et de rapporteurs extérieurs. Les critères essentiels de sélection des projets seront :

- la qualité scientifique du porteur ;
- le caractère novateur et la pertinence scientifique ;
- la prise de risque ;
- la complémentarité des partenaires ;
- le non recouvrement avec d'autres appels d'offres potentiels ;
- éventuellement, l'adéquation à la stratégie en matière d'ouverture internationale de notre établissement telle que définie dans la lettre de cadrage du volet recherche pour le prochain contrat quadriennal (<http://www.blog-um2.univ-montp2.fr>).



**Conseil Scientifique**  
**Appel d'offres 2009**  
**Programmes pluridisciplinaires**

**APPEL A PROJETS**  
**DATE LIMITE : 18 JUIN 2009**

**PROCEDURE**

**1. TELECHARGEMENT**

- le texte de présentation,
- la procédure à suivre,
- le dossier de candidature,

sont téléchargeables sur le blog UM2: <http://www.blog-um2.univ-montp2.fr>  
ou sur le site web onglet Recherche / Appels à projets

**2. DEPOT DE CANDIDATURE**

Les dossiers sont à déposer ou à faire parvenir **au plus tard le 18 Juin 2009**

**EN 2 exemplaires papier** à adresser (cachet de la poste faisant foi) ou à apporter au :  
Secrétariat du Conseil Scientifique - Bâtiment 7 – 1<sup>er</sup> étage –  
Université Montpellier 2, Case courrier 404  
Place Eugène Bataillon  
34 095 Montpellier Cedex 5

**ET 1 version électronique** à envoyer à : [recherche@univ-montp2.fr](mailto:recherche@univ-montp2.fr)  
Dans tout courrier électronique nommer votre document suivant la forme : Nom de  
l'appel à projets – Nom du demandeur  
Acronymes à utiliser : Modelisation ou Energie ou Eau – Nom du demandeur

**ET 1 version électronique** à envoyer aux Directeurs de Pôles Formation Recherche dont  
dépendent les équipes impliquées, ainsi qu'au Directeur de l'Ecole Doctorale concernée  
dans le cas d'une demande d'allocation doctorale.

**Tout dossier incomplet ou envoyé au-delà de cette date ne sera pas pris en considération**

**3. COMMUNICATION DES RESULTATS : le 26 Juin 2009 au plus tard**

**4. RESTITUTION DES TRAVAUX A L'ISSUE DU PROJET**

Les travaux réalisés feront nécessairement l'objet d'un rapport et d'une présentation orale en  
Conseil Scientifique par le porteur de la demande.



**Conseil Scientifique**

**Appel d'offres 2009  
Programmes pluridisciplinaires**

**APPEL A PROJETS  
DATE LIMITE : 18 JUIN 2009**

**DOSSIER DE CANDIDATURE**

**1. RENSEIGNEMENTS ADMINISTRATIFS**

**PROGRAMME PLURIDISCIPLINAIRE CONCERNE**

Modélisation et simulation des sciences du vivant et de l'environnement

**INTITULE DU PROJET DE RECHERCHE**

Classification et identification de données de VIH

**MOTS CLES**

Algorithmique, phylogénie, évolution, épidémiologie, VIH

**POLES DE RECHERCHE ET DE FORMATION DE RATTACHEMENT DE L'UM2**

Pôle MIPS (Math, Informatique, Physique, Structure et Systèmes)

**PORTEUR DE LA DEMANDE (ET COORDINATEUR DU PROJET)**

ARIGON Anne-Muriel

Corps : Maître de Conférences

Adresse @ : anne-muriel.arigon@lirmm.fr

Téléphone : 0467418655

Adresse postale : Equipe MAB - LIRMM

UMR CNRS-UM2 5506

161 rue Ada

34392 Montpellier Cedex 5 - France

Unité de recherche

- LIRMM – UMR CNRS/UM2 5506 – Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
- Directeur : ROBERT Michel
- Tutelle principale : Université Montpellier 2 / CNRS

**AUTRES PARTENAIRES**

CHEVENET François

Corps : Ingénieur de Recherche IRD

Adresse @ : francois.chevenet@gmail.com

Téléphone : 0467416180

Adresse postale : GEMI - IRD Montpellier

911 avenue Agropolis, BP 64501

34394 Montpellier cedex 5 - France

Unité de recherche

- GEMI, IRD Montpellier – UMR 2724 CNRS/IRD/UM2/UM1 – Laboratoire de Génétique et Evolution des Maladies Infectieuses
- Directeur : RENAUD François

PEETERS Martine  
Adresse @ : martine.peeters@mpl.ird.fr  
Adresse postale : IRD Montpellier  
911 avenue Agropolis - BP 64501  
34394 Montpellier cedex 5 - France

Corps : Directeur de Recherche IRD  
Téléphone : 0467416161

Unité de recherche

- SIDA ET MALADIES ASSOCIEES - IRD Montpellier – UMR 145 IRD/UM1 – Institut de recherche pour le développement
- Directeur : DELAPORTE Eric

### **COLLABORATEURS**

GASCUEL Olivier  
Adresse @ : gascuel@lirmm.fr  
Adresse postale : Equipe MAB – LIRMM  
UMR CNRS-UM2 5506  
161 rue Ada, 34392 Montpellier Cedex 5 - France

Corps : Directeur de Recherche CNRS  
Téléphone : 0467418547

Unité de recherche

- LIRMM – UMR CNRS-UM2 5506 – Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
- Directeur : ROBERT Michel

LEFORT Vincent  
Adresse @ : vincent.lefort@lirmm.fr  
Adresse postale : Equipe MAB - LIRMM  
UMR CNRS-UM2 5506  
161 rue Ada, 34392 Montpellier Cedex 5 - France

Corps : Ingénieur de Recherche CNRS  
Téléphone : 0467418547

Unité de recherche

- LIRMM – UMR CNRS-UM2 5506 – Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
- Directeur : ROBERT Michel

### **COLLABORATIONS NATIONALES ET INTERNATIONALES**

BRETON Vincent  
Adresse @ : breton@clermont.in2p3.fr  
Adresse postale : Laboratoire de Physique Corpusculaire  
Université Blaise Pascal, I.N.2.P.3./C.N.R.S.  
24 avenue des Landais  
63 177 Aubière Cedex - France

Corps : Chargé de Recherche CNRS  
Téléphone : 0473407219

Unité de recherche

- LPC de Clermont-Ferrand – UMR 145 Université Blaise Pascal , I.N.2.P.3./C.N.R.S – Laboratoire de Physique Corpusculaire de Clermont-Ferrand
- Directeur : BALDIT Alain

DE OLIVEIRA Tulio  
Adresse @ : tulio@sanbi.ac.za  
Adresse postale : South African National Bioinformatics Institute, Private Bag X17,  
Bellville, 7535, The University of the Western Cape, South Africa.

Corps : Senior Lecturer  
Téléphone : +27219592611

Unité de recherche

- South African National Bioinformatics Institute, Afrique du Sud

VANDAMME Anne-mieke  
Adresse @ : annemie.vandamme@uz.kuleuven.ac.be  
Adresse postale : Katholieke Universiteit Leuven  
Laboratory for Clinical and Epidemiological Virology  
AIDS Reference Laboratory  
Rega Institute & University Hospitals  
Minderbroedersstraat 10  
3000 Leuven - Belgium

Corps : Professor  
Téléphone : +3216332160

Unité de recherche

- Rega Institute for Medical Research , Katholieke Universiteit Leuven, Leuven, Belgique

LIBIN Pieter  
Adresse @ : pieter.libin@mybiodata.eu  
Adresse postale : Katholieke Universiteit Leuven  
Laboratory for Clinical and Epidemiological Virology  
AIDS Reference Laboratory  
Rega Institute & University Hospitals  
Minderbroedersstraat 10  
3000 Leuven - Belgium

Corps : Ingénieur  
Téléphone : +32497402888

Unité de recherche

- Rega Institute for Medical Research , Katholieke Universiteit Leuven, Leuven, Belgique

#### NATURE DE LA DEMANDE exprimée en TTC

Soutien financier : 5000 € / an + 30x40,26 € / an = 6 200 € / an

Demande d'une allocation doctorale :

~~OUI~~

NON

Demande d'une allocation doctorale environnée:

~~OUI~~

NON

Décharge d'enseignement (1) : nombre d'heures (2) : 30h ETD montant financier (3) : 1200€/an

Décharge d'enseignement associée au recrutement d'un ½ ATER

~~OUI~~

NON

(1) La décharge d'enseignement devra être incluse dans le budget, sous réserve de l'accord du directeur de la composante d'enseignement

(2) maximum 1/3 de service, à savoir 64h ETD

(3) 1 heure TD = 40,26 €

Ce financement de 5000 euros / an pourra être utilisé pour une mission en Afrique du Sud (1500 euros), en Belgique (1000 euros), pour participer à des congrès, par exemple Bamako ISCB-Africa (1500 euros) et pour l'achat de matériel informatique et de consommable (1000 euros).

#### CV DU DEMANDEUR ET DU COORDONATEUR DU PROJET (si différent)

Le CV succinct comportera en particulier la liste des thèmes scientifiques étudiés en relation ou non avec le projet. Il sera accompagné de la liste des publications du demandeur.

#### Curriculum

Anne-Muriel ARIGON  
15 avenue Dumont D'Urville  
Port-Camargue 30240 Le Grau u Roi  
Tel: 0661262611

28 ans (née le 26/11/1980)  
Française  
Célibataire

- 2008 Maître de Conférences en Informatique - Université Montpellier 2 - Equipe de recherche MAB (Méthodes et Algorithmes pour la Bioinformatique) du LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier)
- 2006-2008 Attaché Temporaire d'Enseignement et de Recherche en Informatique à l'Université Lumière Lyon 2 (371h ETD d'enseignement en informatique)
- 2003-2006 Doctorat en Bioinformatique – Université Claude Bernard Lyon1  
Titre: « Développements d'outils pour l'aide à l'identification dans les grandes banques de familles de gènes »  
Directeurs de thèse: Manolo GOUY (DR CNRS) et Guy PERRIÈRE (DR CNRS)  
Lieu: Laboratoire de Biométrie et Biologie Évolutive - UMR CNRS 5558 - Lyon 1  
Financement : Allocation de recherche MENRT et allocation de moniteur  
Monitorat : 215h ETD d'enseignement en informatique
- 2003 DEA d'Informatique DISIC (Diplôme d'Etude Approfondies en Documents Multimédia, Images et Systèmes d'Information Communicants), INSA de Lyon (préparé en parallèle à la 5<sup>ème</sup> année du cycle ingénieur de l'INSA de Lyon).  
*Mention Assez Bien*
- 2000-2003 Diplôme d'ingénieur INSA de Lyon (Institut National des Sciences Appliquées), spécialité Informatique.
- 1998-2000 Diplôme d'Etudes Universitaires Générales de Mathématiques et Informatique Appliquées aux Sciences (MIAS) option préparation aux Ecoles Nationales Supérieures d'Ingénieurs (ENSI). Université d'Avignon et des Pays de Vaucluse.
- 1998 Baccalauréat Scientifique spécialité mathématiques, Avignon.

*Activités de recherche associées :*

- 2009 Membre du comité de programme de EDA 09 (journées francophones sur les Entrepôts de Données et l'Analyse en ligne) et de MEDES-SW 09 (Management of Emergent Digital EcoSystems – Student Workshop)
- 2008 Membre du comité de relecture de la conférence MajecSTIC 2008 (Manifestation des Jeunes Chercheurs en Sciences et Technologies de l'Information et de la Communication)

*Thèmes scientifiques étudiés :*

- Etude de l'évolution de pathogènes, notamment le VIH (SIDA)
- Classification et identification de données biologiques
- Phylogénie et évolution
- Comparaison et recherche de similarité entre séquences
- Alignements multiples de séquences
- Détection de séquences chimères dédiée aux séquences d'ARN ribosomique 16S
- Entrepôts de données appliquées à des données médicales

*Thèmes méthodologiques associés:*

- Algorithmique du texte et des arbres
- Statistiques et probabilités
- Modèles multidimensionnels

*Liste des publications :*

*Publications dans des revues internationales à comités de lecture*

1. Penel S., Arigon A.M., Dufayard J.F., Sertier A.S., Daubin V., Duret L., Gouy M. and Perrière G. (2009), Databases of homologous gene families for comparative genomics. *BMC Bioinformatics*, sous presse.
2. Arigon A.M., Perriere G., Gouy M. (2009), Classification methods for biological data. Accepté pour publication dans le numéro spécial "Warehousing and Mining Complex Data: Applications to Biology, Medicine, Behavior, Health and Environment" de *International Journal of Biomedical Engineering and Technology*.
3. Arigon A.M., Perriere G., Gouy M. (2008), Automatic identification of large collections of protein-coding or rRNA sequences. Sous presse dans *Biochimie* 2008, doi:10.1016/j.biochi.2007.08.006.
4. Arigon A.M., Miquel M. and Tchounikine A. (2007) Multimedia data warehouses: a multiversion model and a medical application. *Multimedia Tools and Applications*. 2007 October, 35(1):91-1008.
5. Arigon A.M., Tchounikine A. and Miquel M. (2006) Handling multiple points of views in a multimedia data warehouse. *ACM Transactions on Multimedia Computing, Communications and Applications*. 2006 August; 2(3):199-218.
6. Arigon A.M., Perriere G., Gouy M. (2006) HoSeqI: automated homologous sequence identification in gene family databases. *Bioinformatics*. 2006 Jul 15; 22(14):1786-7.


*Conférences internationales à comités de lecture*

7. Arigon A.M., Perrière G. and Gouy M., Automated homologous sequence identification. *SMBE*, Tempe, Arizona, USA, 2006. (Communication orale)
8. Arigon A.M., Tchounikine A. and Miquel M. A Proposal for a Multimedia Data Warehouse, 11th International Workshop on Multimedia Information Systems, MIS 2005 19th-21st September, 2005 Sorrento, Italy, Lecture Notes in Computer Science (LNCS), Springer Verlag Publishers, 2005, Volume 3665, 49-62, ISBN: 3-540-28792-2. (Communication orale)
9. Arigon A.M., Tchounikine A. and Miquel M. A multiversion model for multimedia data warehouse, MDM/KDD 2005, In proc of the Sixth International Workshop on Multimedia Data Mining, 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 21-24, 2005, Chicago, IL, USA. (Communication orale)

*Conférences nationales à comités de lecture*

10. Simon Penel, Anne-Muriel Arigon, Vincent Daubin, Pascal Calvat, Stephane Delmotte, Jean-Francois Dufayard, Manolo Gouy, Guy Perriere, Anne-Sophie Sertier and Laurent Duret. Databases of homologous gene families for comparative genomics. Dans les Actes des Journées Ouvertes Biologie Informatique Mathématiques - JOBIM 2009, Nantes, France, E. Rivals, I. Rusu (editeurs), Juin 2009, p.123-124. (Communication orale)
11. Arigon A.M., Perrière G. and Gouy M., Identification de séquences à partir de banques de famille de séquences. Dans les Actes des Journées Ouvertes Biologie Informatique Mathématiques - JOBIM 2005, Lyon, France, G. Perrière, A. Guénoche, C. Gourgeon (editeurs), Jul 2005, p. 287-289. (Communication orale)
12. Arigon A.M., Miquel M. et Tchounikine A., Intégration de versions fonctionnelles dans les entrepôts de données multimédias, 1ère journées francophones sur les entrepôts de données et l'Analyse en ligne, EDA'2005, Lyon, juin 2005, Revue des Nouvelles Technologies de l'Information (RNTI), Vol. B-1, 91-108, Editions Cepaduès, numéro spécial, 2005, I.S.B.N. : 2.85428.698.7. (Communication orale)

**VISA DU DEMANDEUR**



NOM : ARIGON      Prénom : Anne-Muriel

**VISA DU OU DES DIRECTEUR(S) D'UNITE(S) DE RECHERCHE**

NOM Prénom

**VISA DU DIRECTEUR DE COMPOSANTE D'ENSEIGNEMENT**  
(Dans le cas d'une demande de décharge d'enseignement)

NOM Prénom

**2. RENSEIGNEMENTS SCIENTIFIQUES**

**INTITULE DU PROJET DE RECHERCHE**

**Classification et identification de données de VIH**

**DESCRIPTIF DU PROJET (4 pages, références bibliographiques incluses)**

*La présentation (en français ou en anglais) devra clairement faire apparaître les objectifs du projet et la méthodologie utilisée, son caractère novateur dans le contexte de la recherche, la complémentarité des partenaires impliqués, les avancées attendues.*

L'évolution et la diffusion de l'épidémie du SIDA est un problème d'importance mondiale. L'infection par le VIH est devenue une pandémie qui infecte environ 40 millions de personnes (www.unaids.org), plus particulièrement en Afrique subsaharienne (70% des personnes vivant avec le VIH). En Afrique du Sud, 1 adulte sur 5 est infecté. Ce virus est à la fois simple et complexe. Simple parce que son génome est très réduit, et nous disposons aujourd'hui de centaines de milliers de souches partiellement séquencées. Complexe parce qu'il évolue extrêmement vite, avec un taux très élevé de mutations et recombinaisons, ce qui lui permet d'échapper au système immunitaire et de développer des résistances aux thérapies antirétrovirales. L'analyse génétique du VIH est importante pour le développement de vaccins mais aussi pour les traitements, la surveillance de l'apparition de nouveaux variants du virus et

pour l'aide au diagnostic. Chaque année, la communauté des chercheurs sur le VIH produit une très grande quantité de données. Des modèles et des méthodes automatisés et efficaces sont nécessaires pour analyser ces grands jeux de données rapidement et prendre en considération l'ensemble des informations qu'elles contiennent. Dans ce contexte, je m'intéresse à la problématique de l'identification et la classification de souches de VIH, point central dans l'analyse génétique des données de VIH et le suivi épidémiologique de l'infection. Mon projet est constitué de plusieurs axes de recherche autour de cette problématique.

## **Diversité génétique du VIH**

Le premier axe de mon projet est l'étude de la diversité du VIH. Actuellement, je débute une collaboration avec Martine Peeters qui s'intéresse à cette problématique dans les pays du Sud et aux implications pour la prise en charge des patients (équipe de Eric Delaporte, UMR145, IRD et Université Montpellier 1, Montpellier). En effet, sur le plan virologique, l'une des caractéristiques majeures du virus VIH est son extrême diversité (Peeters et al., 2003) surtout en Afrique. Pour que les avancées scientifiques profitent aux pays du Sud, et notamment à l'Afrique, il faut que les spécificités de l'infection par le VIH dans ces pays soient mieux connues afin de mettre en place des stratégies de prévention et de prise en charge efficaces et adaptées.

L'UMR 145 a déjà accumulé depuis plus de 10 ans de nombreuses données sur la diversité génétique et la réponse aux traitements ARV (antirétroviraux) en Afrique. La masse de données à gérer est trop importante pour des traitements et des analyses manuels. Ainsi des modèles et des méthodes rapides sont nécessaires pour gérer et analyser les nouvelles données. Pour cela nous travaillons en coordination avec Pieter Libin et Anne-Mieke Vandamme (Rega Institute for Medical Research, Leuven, Belgique, <http://www.kuleuven.be/regarei/>) qui ont développé, notamment, un système de gestion et de traitement de données du VIH (RegaDB : <http://www.rega.kuleuven.be/cev/regadb/>). Ce système, en cours d'installation à l'UMR145, permet le stockage de données dans une base de données ainsi que des analyses détaillées dédiées au VIH (résistance au médicament, identification de marqueurs,...). La base de données est adaptée aux données du VIH et est centrée sur le patient. Pour chacun des patients sont stockées des informations personnelles ainsi que des données concernant les thérapies, le suivi (résultats d'analyse et de différents tests), les isolats viraux (séquences),... Ce système propose aussi des analyses et des services comme des graphes récapitulant les données présentes dans la base de données, un outil permettant d'effectuer des requêtes, des algorithmes de résistance génotypique, des rapports de résistance aux médicaments, ...

Cependant ce système, centré sur le patient, ne prend pas en compte les souches et leur diversité au sein d'un même patient. Par exemple, il ne permet pas de vérifier et de valider les nouvelles données lors de la mise à jour de la base de données. Ainsi il n'est actuellement pas possible de savoir, lorsque l'on ajoute un nouveau patient et la séquence correspondante (la souche), si ce patient existe déjà dans la base de données sous un autre identifiant, s'il existe des patients de la même famille, s'il y a plusieurs souches pour un même patient, s'il y a eu contamination, ... L'objectif de nos travaux communs est d'apporter à ce système une vue différentes, non pas centrée uniquement sur le patient mais mettant également en avant la diversité génétique. La première étape consiste à intégrer à ce système un module bioinformatique de validation de données capable d'identifier les similarités entre les nouvelles données et celles déjà existantes. Pour cela, des méthodes algorithmiques de comparaison efficaces et précises sont nécessaires. Ceci permettra d'améliorer le processus de classification de ces données (séquences) et d'adapter cet outil (initialement développé pour les patients suivis dans les pays occidentaux) à la problématique de la diversité génétique et du traitement du VIH dans les pays du Sud en collaboration avec l'UMR145, IRD/UM1.

## **Identification de souches du VIH**

Un autre axe de mon projet est l'identification de souches du VIH-1. En collaboration avec François Chevenet (IRD, Montpellier), Olivier Gascuel (LIRMM, Montpellier) et Tulio de Oliveira (South African National Bioinformatics Institute, Afrique du Sud), je travaille sur l'analyse phylogénétique des souches du VIH-1 (de Oliveira et al., 2005). Les objectifs de ces travaux sont d'affiner les classifications des souches du VIH-1 et de préciser l'origine, le tempo et la dynamique des épidémies. Pendant ma thèse je me suis intéressée à l'identification et à la classification de séquences dans les bases de données de familles de gènes homologues et les bases de données de séquences d'ARN ribosomique 16S (Arigon et al. 2006). Dans ce projet, j'utiliserai ces compétences pour traiter les données du VIH.

Ainsi, je développe une application web permettant à un utilisateur de classer des séquences de VIH par rapport à des séquences de référence de manière très simple. Une fois que l'utilisateur a soumis ses séquences et choisi un ensemble de référence, l'application construit l'alignement puis l'arbre phylogénétique incluant les nouvelles séquences et les séquences de référence. L'arbre phylogénétique obtenu permet d'identifier les séquences analysées. J'envisage également d'automatiser le choix de l'ensemble de référence en se basant sur des annotations fournies par l'utilisateur. Les méthodes utilisées sont basées sur l'application de techniques phylogénétiques qui permettent de retracer l'histoire évolutive, en se basant sur des données de références du VIH-1 représentant l'épidémie à l'échelle mondiale. Ces méthodes doivent être adaptés aux données du VIH, ainsi le modèle d'évolution utilisé pour la reconstruction phylogénétique doit tenir compte du caractère complexe des souches de VIH. Les méthodes de maximum de vraisemblance sont des techniques très précises pour l'analyse phylogénétique. Le logiciel de reconstruction phylogénétique PHYML (Guindon and Gascuel, 2003) sera donc au cœur du projet.

## **Typage de données du VIH et autres pathogènes**

Enfin, je souhaite participer à des projets plus larges dans lesquels j'apporterai mes compétences et appliquerai les méthodes et outils déjà développés.

Tout d'abord, en collaboration avec François Chevenet (IRD, Montpellier), Olivier Gascuel (LIRMM, Montpellier) et Tulio de Oliveira (South African National Bioinformatics Institute, Afrique du Sud), nous travaillons sur une méthode de typage de données du VIH ayant pour objectif de détecter l'apparition de nouvelles souches présentant un caractère particulier sur le plan épidémiologique (par exemple dans un pays peu touché jusque là) ou thérapeutique (par exemple présentant une résistance à certains traitements). Le principe est de labéliser chacun des nœuds et chacune des feuilles de l'arbre avec les caractères à étudier. Une méthode basée sur la parcimonie (algorithme de Fitch, 1971) est utilisée pour reconstruire les états ancestraux. Des critères combinatoires et statistiques permettent alors de sélectionner dans l'arbre labélisé les groupes de feuilles présentant des caractères particuliers. La visualisation est ici importante car elle permet de distinguer facilement dans de grands jeux de données celles qui partagent des caractères spécifiques communs. Il sera également intéressant d'intégrer cette méthode de typage à l'application d'identification de souches du VIH précédemment décrite afin de compléter le processus de traitement et proposer à l'utilisateur des informations concernant l'apparition de nouvelles souches présentant un caractère particulier sur le plan épidémiologique ou thérapeutique. Le VIH ayant un taux très élevé de mutations et recombinaisons, un point essentiel dans ces études sera la modélisation (combinatoire, probabiliste) des effets de la recombinaison sur nos constructions et inférences.

Par ailleurs, les méthodes et modèles bioinformatiques développés pour le VIH peuvent être appliqués à d'autres virus analogues, comme la grippe. Ainsi une collaboration démarre entre Vincent Lefort (LIRMM, Montpellier), Olivier Gascuel (LIRMM, Montpellier), moi-même et Vincent Breton (IN2P3, Clermont-Ferrand) sur l'analyse phylogénétique systématique et en temps réel du virus de la grippe A/H1N1. L'objectif est de suivre et comprendre l'évolution de ce virus dans le temps et géographiquement. Pour cela, nous utiliserons les ressources de la grille EGEE (Enabling Grids for E-sciencE, <http://www.eu-egee.org/>) pour construire la phylogénie de ce virus, en la mettant à jour à chaque nouvelle donnée obtenue. Plusieurs études peuvent alors être envisagées, notamment la comparaison du mode d'évolution du virus (taux et pattern de substitutions, apparition de souches recombinantes ...) suivant les zones géographiques et socio-économiques, ainsi que la détection de nouveaux types épidémiologiques en utilisant l'outil précédemment décrit.

*Références bibliographiques :*

1. Arigon A.M., Perriere G., Gouy M. (2006) HoSeqI: automated homologous sequence identification in gene family databases. *Bioinformatics*. 2006 Jul 15; 22(14):1786-7.
2. Fitch W. M.. Toward defining course of evolution: minimum change for a specified tree topology. *Systematic Zoology*, 20:406–416, 1971
3. Guindon S. and Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003 Oct;52(5):696-704.
4. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AM, van de Vijver DA, Boucher CA, Camacho R, Vandamme AM. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics*. 2005 Oct 1;21(19):3797-800. Epub 2005 Aug 2. (doi:10.1093/bioinformatics/bti607)
5. Peeters M, Toure-Kane C, Nkengasong J. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS* 2003, 17: 2547-2560.

Cachet et date d'arrivée du dossier au conseil scientifique