

1. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@network protein sequence analysis. *Trends Biochem Sci* 2000 ; 25 : 147-50.
2. Combet C, Jambon M, Deleage G, Geourjon C. Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* 2002 ; 18 : 213-4.



## HCVDB : base de données de séquences du virus de l'hépatite C et outils bio-informatiques d'analyse

Christophe Combet, François Penin, Christophe Geourjon, Gilbert Deléage

IBCP, 7, passage du Vercors, 69367 Lyon, France. [g.deleage@ibcp.fr](mailto:g.deleage@ibcp.fr)

> Une base de données de séquences du virus de l'hépatite C (VHC) (HCVDB : <http://hepatitis.ibcp.fr>) a été développée pour offrir à la communauté des biologistes, virologistes et cliniciens un accès facile à l'ensemble des séquences publiques du VHC (nucléiques et protéiques) disponibles à l'EMBL (*European molecular biology laboratory*). La base de données est organisée en 5 modules interconnectés : HCVWEB (interface HTML et programmes), HCVREF (33 génomes de

référence annotés), HCVSRS (système d'interrogation), HCVSA (outils intégrés d'analyse de séquence) et HCVDB (banques de séquences). A partir de l'extraction des séquences en utilisant le mot clé *hepatitis C virus* dans le champ OS de l'EMBL, 4 bases sont automatiquement engendrées : HCVUFN (14 817 entrées nucléiques de l'EMBL), HCVUSN (24 044 entrées nucléiques découpées selon l'organisation du génome), HCVUFP (13 801 entrées protéiques par traduction automatique de HCVUFN) et HCVUSP (22 230 entrées protéiques résultant du clivage de la polyprotéine). HCVDB est mise à jour chaque mois (version 16) et contient actuellement 174 génomes complets. Chaque entrée de la base est automatiquement annotée en comparant la séquence avec celles de 33 génomes de référence annotés par un réseau de spécialistes du VHC. Le génotype (*clade* et sous-type) est déterminé pendant l'annotation en utilisant les séquences nucléotidiques des 3 régions Core, E1 et NS5B du génome. HCVDB est interrogeable grâce au système SRS (*sequence retrieval system*, LION Biosciences) qui permet d'effectuer des requêtes combinées au moyen d'opérateurs logiques sur les différents champs de la base (auteur, description, génotype,...). Les résultats des requêtes

sont directement transférables vers un système d'analyse de séquence inspiré du serveur NPS@ [1] (<http://npsa-pbil.ibcp.fr>) ou de modélisation moléculaire à travers le serveur Geno3D [2] (<http://geno3d-pbil.ibcp.fr>). Parmi les méthodes disponibles (28 à ce jour), les recherches de similitude, les alignements multiples, les recherches de signatures de fonctions biologiques et les prédictions de structures secondaires sont utilisables de manière successive et transparente pour l'utilisateur.

Ainsi, en utilisant HCVDB, l'utilisateur peut très facilement (en moins de 5 minutes) se constituer avec HCVSRS une banque de données personnelle de toutes les séquences protéiques publiques correspondant à l'une des protéines du VHC d'un génotype donné et procéder (par exemple) à leur alignement multiple (affichage en couleur) via HCVSA.

La base HCVDB est opérationnelle depuis mars 2000 et bénéficie du soutien de l'ANRS et du Réseau National Hépatites mis sur pied par G. Inshauspé et C. Bréchet et piloté par J. Dubuisson, J.M. Pawlotsky et G. Inshauspé. <

## ALLO-GÈNES

> 20 000 appels du grand public et des professionnels en 6 ans d'existence

Numéro Azur

08 10 63 19 20

(prix d'une communication locale)

Allo-Gènes, centre national d'information sur les maladies génétiques, est né en 1995 de la réflexion menée par l'association Française contre les Myopathies (AFM) avec le ministère chargé de la Santé et les généticiens cliniciens.

Il répond aux questions des personnes confrontées à une maladie génétique ou à une maladie rare non génétique et est accessible aux personnes concernées personnellement ou familialement comme aux professionnels de santé (médecins et paramédicaux).

**Contact** : Professeur Marie-Louise Briard/directrice d'Allo-Gènes

Tél. : 01 56 53 81 32

