Tetraodon nigroviridis,

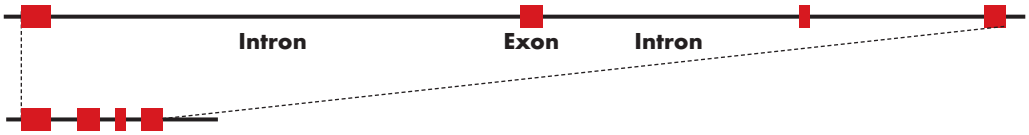
un poisson à génome compact



Dans les génomes de tous les vertébrés, il semble que l'on retrouve à peu près le même ensemble de gènes. Chez les mammifères, ces gènes sont dispersés dans un génome d'environ 3 milliards de nucléotides, si bien que leur identification chez la souris n'est pas moins fastidieuse que chez l'homme. Chez certains poissons de la famille des poissons-ballons, cet ensemble de gènes se retrouve sous une forme beaucoup plus compacte, dans un génome de 400 millions de nucléotides seulement.

Une telle différence est due aux forces qui modèlent les génomes. La plupart des vertébrés ont acquis une forte proportion d'ADN apparemment superflu, que l'on retrouve notamment dans les introns *(voir la fiche* Interpréter les séquences*)*. Cette évolution semble avoir été contrariée, voire inversée, chez les poissons à génome compact qui, de ce fait, possèdent des introns beaucoup moins longs que les mammifères, et des gènes beaucoup moins espacés. En séquençant un fragment d'ADN de taille donnée, on a huit fois plus de chances de trouver un morceau d'instruction (exon) chez l'un de ces poissons que chez un mammifère. Séquencer l'ADN d'un poisson à génome compact est donc la façon la moins coûteuse d'accéder aux gènes des vertébrés.

Gène humain



Gène de *Tetraodon*

En 1997, nous avons entrepris le séquençage du génome d'un de ces poissons. Notre choix s'est porté sur *Tetraodon nigroviridis*, un hôte des cours d'eau d'Asie du Sud-Est, bien connu des aquariophiles. L'objectif principal était d'utiliser les séquences génomiques de *Tetraodon* dans des comparaisons avec la séquence du génome humain, afin de repérer les gènes humains *(voir plus loin)*.

Pour venir à bout des 380 millions de bases de ce génome, nous avons retenun la stratégie du séquençage aléatoire global, qui consiste à faire des lectures d'un très grand nombre de séquences tirées au hasard *(voir la fiche* Stratégies de séquençage*)*. Dans le cas du génome humain, cette approche est rendue très difficile par l'importance des séquences répétées. Ces séquences, présentes à l'identique en divers endroits du génome, induisent des erreurs lorsqu'on tente d'assembler les séquences lues à l'échelle du génome entier. Chez *Tetraodon*, la pauvreté du génome en séquences répétées facilite grandement l'approche du séquençage aléatoire global.

En 2001, le Whitehead Institute (Cambridge, Etats-Unis) s'est joint à l'entreprise du Genoscope. L'ensemble des séquences de *Tetraodon* lues par les deux centres représente désormais 8,3 fois la taille du génome (8,3 X), et a pu être assemblé en une ébauche où presque tous les gènes de *Tetraodon* devraient être représentés par un ou plusieurs exons. Il y a d'ores et déjà matière à comparaisons : l'ébauche de la séquence génomique d'un autre poisson à génome compact, le fugu, a été publiée en août 2002, et le génome d'un troisième poisson, le poisson zèbre, est en cours de séquençage. Ces trois projets ne sont pas redondants : avec plusieurs génomes de poissons à notre disposition au lieu d'un seul, nous pouvons tirer davantage de conclusions sur l'évolution des vertébrés. *(suite au dos)*

Tetraodon nigroviridis,

un poisson à génome compact (suite)



La continuité de l'assemblage de *Tetraodon*, supérieure à celle de l'ébauche plus fragmentée de fugu (bâtie avec moins de 6 X), va permettre de comparer avec précision ce génome de poisson aux génomes de l'homme et de la souris. Cette comparaison éclairera les événements qui ont façonné les génomes dans chaque lignée, tôt dans l'histoire des vertébrés. Par ailleurs, l'étude des régions non-codantes dans le génome de *Tetraodon* offrira peut être des indices sur les raisons et les mécanismes de leur réduction chez ce poisson, et, au delà, sur la dynamique et l'éventuelle importance fonctionnelle de cet ADN parfois qualifié de "poubelle".

Les séquences non-codantes de *Tetraodon* ont un autre attrait : elles permettent de rechercher, dans des intervalles de longueur réduite, des régions de régulation des gènes également présentes chez l'être humain, mais dans des intervalles bien plus grands *(voir la fiche* Exploiter la séquence III). Certaines de ces régions sont en effet assez conservées pour être détectées par la comparaison des séquences génomiques de l'homme et du poisson, en dépit des 450 millions d'années écoulés depuis la divergence des deux lignées. Bien entendu, ce même type de comparaison détecte aussi les régions codantes, mieux conservées, et a déjà servi à améliorer l'annotation du génome humain *(voir la fiche* Comparer les génomes).

Chez fugu, les gènes ont été recherchés au moyen de comparaisons avec les séquences des gènes de l'homme et d'autres organismes. Dans le cas de *Tetraodon*, l'annotation a bénéficié d'une ressource supplémentaire : les séquences d'ADN complémentaires, copies des ARN messagers issus de l'expression des gènes du poisson *(voir la fiche* ADN complémentaires). On peut ainsi accéder à des gènes sans aucun homologue chez les autres vertébrés, et que les comparaisons n'auraient pas détectés.

Un autre aspect intéressant de l'annotation de *Tetraodon* est l'importance relative des familles de gènes, certaines comptant beaucoup plus de membres chez les autres vertébrés séquencés, d'autres s'avérant au contraire plus développées chez le poisson. L'amplification de telles familles par duplication de gènes a une signification évolutive importante, et peut éclairer les adaptations de certaines lignées.

Enfin, le génome annoté de *Tetraodon* servira à interpréter celui du poisson zèbre *(Danio rerio)*. Ce poisson est devenu une espèce modèle en biologie du développement, car il est facile à élever et à observer et se prête bien aux études génétiques. De nombreuses pathologies chez l'homme affectent des étapes du développement que l'on retrouve chez le poisson zèbre et que l'on peut étudier chez cet animal par une inactivation ciblée des gènes. Les génomes de poissons montreront alors toute leur valeur pour la compréhension d'aspects fondamentaux de la biologie humaine, communs à tous les vertébrés.