

Mon génome à poil sur l'Internet

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 juillet 2011. Dernière mise à jour le 18 juillet 2011

<https://www.bortzmeyer.org/23andme.html>

Je viens de faire faire une analyse génomique personnelle en utilisant le service de 23andMe. Je vous rassure tout de suite, je ne mets pas mon génome en donnée ouverte sur ce blog, malgré le titre de l'article.

Au début de l'analyse du génome humain, il fallait des années et des sommes considérables pour analyser le génome d'un seul individu. Les prix de l'analyse baissant à vue d'œil, un particulier peut désormais faire analyser son propre génome. J'ai utilisé le service de 23andMe (pour tester vos connaissances en génétique, demandez-vous, « pourquoi 23? ») et cela ne m'a coûté que 200 dollars (une fois mon analyse faite, les prix ont été divisés par deux...). Pour ce prix là (et 60 dollars de frais postaux), on obtient son génome sous forme d'un fichier de données brutes (plus exactement, on obtient un ensemble de SNP, voir le résumé technique à la fin) et tout un tas d'analyses faites par 23andMe. Parmi lesquelles :

- Des informations de santé : risques (estimés) de développer certaines maladies, lorsque leur cause est partiellement génétique, réponses probables de l'organisme à certains médicaments (ce qu'on nomme la médecine personnalisée),
- Des informations sur ses ancêtres, leur région probable d'origine,
- Et plein d'autres données, comme s'il n'y avait pas déjà assez de choses à lire sur l'Internet...

Je n'ai pas l'intention de distribuer sur l'Internet mon génome complet (certains l'ont fait <<http://manu.sporny.org/2011/public-domain-genome/>> et on trouve même un site rassemblant tous ces génomes publics, SNPedia <<http://www.snpedia.com/>>), les risques n'étant pas encore bien mesurés. Vous allez donc devoir vous contenter de certains extraits.

23andMe propose des outils d'analyse sur le site Web. J'apprends ainsi que mes risques médicaux les plus élevés sont le cancer de la prostate et surtout la dégénérescence maculaire (presque quatre fois le risque moyen). Du point de vue des remèdes, je n'ai pas de résistance génétique au SIDA (mais j'en ai une aux norovirus), et le traitement PEG-IFNalpha/RBV contre l'hépatite C n'aura sans doute que peu d'effet sur moi...

Je trouve, dans la rubrique sur les ancêtres, que les miens viennent sans doute d'Europe du Nord (les Alpes et les Pyrénées étant la frontière génétique de l'Europe du Nord). Mes lignées paternelles et maternelles sont proches, et la comparaison avec les génomes de gens célèbres qui ont été analysés montre des

traits communs avec Warren Buffett... Les beaux graphiques de cette rubrique exigent malheureusement le logiciel non-libre Flash.

Enfin, question physique, j'apprends que mon génome indique des yeux sans doute bleus (c'est pas faux), une tolérance au lactose (j'ai avalé un bol de céréales avec du lait, pour fêter ça), et que je résiste bien au café (qui augmente le risque d'attaque cardiaque pour certaines personnes), ...

À chaque fois, on a droit à une page détaillée sur la maladie, des explications, des pointeurs vers les articles scientifiques, etc. On peut donc passer sa vie à s'informer sur son corps.

Pour les maladies plus graves, et pour lesquelles on ne connaît pas de remède, comme l'Alzheimer, le rapport n'est pas accessible directement, il faut confirmer deux fois qu'on veut bien le lire : 23andMe estime sans doute que le jeu n'en vaut pas la chandelle puisqu'il n'y a pas d'action préventive possible. Autant rester dans l'ignorance (je n'ai pas insisté).

Dans le plus pur style 2.0, on peut partager son génome avec d'autres utilisateurs de 23andMe, ceux marqués comme « amis », ceux marqués comme « famille ». Je vous préviens tout de suite, je ne partage qu'avec les gens que je connais vraiment. Notez bien que, les gènes de parents proches étant... proches, vos frères et parents peuvent en savoir beaucoup sur votre génome, simplement en analysant le leur... Pareil en sens inverse, les risques pour ma santé s'appliquent aussi à ma famille. La lecture de <https://www.23andme.com/help/#privacy> est très intéressante (toute la FAQ aussi, d'ailleurs).

Bon, évidemment, tout ce genre d'analyses est à prendre avec de sérieuses pincettes (« comme les prévisions météo à dix jours », me souffle une biologiste). D'abord, l'« étude » est entièrement automatisée (pour cent dollars, vous n'avez pas un expert de l'analyse génomique qui se penche sur **votre** génome...). C'est donc du prêt-à-porter plutôt que de la confection. Ensuite, d'autres indices, moins "high-tech", peuvent vous renseigner aussi bien. Pour la dégénérescence maculaire, la connaissance de l'historique familial permet d'avoir une idée du risque, certes approximative, mais presque aussi bonne que les analyses génétiques.

Si vous vous demandez comment se passe le prélèvement, 23andMe vous envoie un colis avec l'éprouvette spéciale, dans laquelle il faudra baver pendant dix à vingt minutes en se stimulant l'intérieur de la joue (pour augmenter la production de salive). Il vaut mieux que personne d'autre ne regarde à ce moment, on a vraiment l'air bête. L'éprouvette est spéciale car le voyage international d'échantillons biologiques est très réglementé (présence obligatoire d'un élément absorbant, au cas où l'éprouvette fuit, par exemple). Il faut ensuite remplir pas mal de papiers. Le prix inclut le voyage retour par DHL.

23andMe propose aussi, en cochant une case, de passer ses données à la recherche scientifique. Je ne l'ai pas encore fait.

À noter que d'autres entreprises proposent de la « génomique directe » comme deCODE. Un bon suivi de cet industrie est fourni par le site "Genomes Unzipped" <http://www.genomesunzipped.org/>.

Et question bricolage, faire ses propres analyses, faire des "mashups" et tous ces trucs modernes? 23andMe permet de récupérer l'intégralité des données (option "Download raw data"). Au moment du téléchargement, 23andMe vous met en garde, notamment sur le fait que le fichier, stocké sur une machine non sécurisée, est vulnérable... Moi, j'ai rapidement fait un `gpg --encrypt --user bortzmeyer genome_Stephane_Bortzmeyer_Full_20110707000202.txt` - un simple `chmod 600` ... ne m'aurait pas satisfait. Notez que la sécurité des échanges semble assez convenable, HTTPS partout, demande

systématique de la question secrète lorsqu'on veut récupérer les données brutes, etc. Cela n'indique évidemment rien sur la sécurité chez 23andMe, aussi bien en cas d'attaque qu'en cas de malhonnêteté de la part de l'entreprise.)

Une fois les données récupérées (vingt-quatre mégaoctets, décompressées), voici à quoi ressemblent les données brutes (merci à Victoria Dominguez pour son aide sur cette section) :

```
# rsid chromosome position genotype
rs4477212 1 72017 AA
rs3094315 1 742429 AG
rs3131972 1 742584 AG
rs12124819 1 766409 AA
rs11240777 1 788822 AG
rs6681049 1 789870 CC
...
```

Le format semble être spécifique à 23andMe (un identificateur, le numéro du chromosome, la position dans le chromosome - en nombre de bases depuis le début - et les valeurs des bases, A, T, C ou G). Il ne semble pas possible d'avoir les données dans un format que comprennent la plupart des outils d'analyse, comme le format FASTA. Les données (des SNP <<https://www.23andme.com/gen101/snps/>>) sont constituées des différences avec le génome humain de référence, dit « build 36 <http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&build=36> ». C'est donc un format équivalent à diff. On pourrait penser que, pour retrouver mon génome complet, il suffit d'appliquer ce "patch" au génome de référence. Mais la génomique est bien plus compliquée que cela. Par exemple, dans mes données envoyées par 23andMe, le chromosome 1 va jusqu'à la position 247 185 615 alors que seules 224 999 719 ont été séquencées. Il y a donc pas mal de trous, empêchant une reconstruction facile.

Évidemment, on ne peut pas en faire grand'chose, pour l'instant. Mais l'analyse dans son garage approche <<http://openpcr.org/2011/07/dna-is-now-diy-openpcr-ships-worldwide/>>. Déjà, des sites proposent une analyse des données existantes, en documentant le cas des clients de 23andMe, par exemple celui de James Lick <<http://vps1.jameslick.com/dna/mthap/>>. Un de ces jours, je lirai le livre de Marcus Wohlsen, « *BioPunk : DIY Scientists Hack the Software of Life* » <<http://spittoon.23andme.com/2011/07/01/biopunks-science-and-discovery/>> ».

À noter un article similaire <<http://www.francetvinfo.fr/sante/tests-genetiques/jai-teste-pour-v-368474.html>> écrit deux ans plus tard, sans information supplémentaire.