Quelques remarques sur les analyses ADN des "momies" de Nazca

mardi 25 juillet 2017, par Irna

Suite à mon dernier article (Une Divulgation qui fait pschitt !) où je relevais quelques éléments plus ou moins intéressants sur les analyses fournies par Thierry Jamin de ses « momies aliens », j'ai eu quelques échanges épistolaires avec un généticien qui a bien voulu m'apporter quelques lumières dans un domaine qui m'est très largement étranger, et qui s'est intéressé plus particulièrement à la seule analyse dont on a les résultats bruts, c'est-à-dire celle commandée par Jaime Maussan au laboratoire mexicain *Instituto de Ciencia* (pages 23 à 25 de ce document - Edit : le document n'étant plus disponible sur le site Alien Project, le lecteur pourra en trouver une copie ici).

La première chose qui a frappé ce généticien c'est l'inadéquation entre cette analyse spécifique et le but apparemment recherché :

« Ça paraît réellement ridicule de procéder de la sorte (avec un kit *QIAGEN* destiné à la police scientifique) sur des échantillons qui seraient réellement d' "origine inconnue", qui auraient résisté à la comparaison morpho-anatomique. Dans ce cas, pourquoi des marqueurs spécifiques ? si oui, pourquoi humains, et pas des microsatellites de pingouins ? Pourquoi comparer deux échantillons avant même de les caractériser indépendamment ? Pourquoi gaspiller un matériel biologique rare pour des questions aussi prématurées ? Ce sont des marqueurs spécifiques. Ce qui implique qu'on sait à l'avance qu'il s'agit de matériel d'origine humaine. Or en premier lieu, évidemment, si c'est inconnu, j'aimerais beaucoup qu'un chimiste me confirme qu'il y a bien une chimie carbonée, de l'ADN (ou d'autres supports de codage possibles), protéines ou équivalent, ARN ou équivalent etc… »

Il a ensuite pris la peine d'analyser en profondeur les résultats du labo mexicain :

« L'analyse commanditée par Jaime Maussan est basée sur la caractérisation des échantillons « Maria » et « Cerebro » sur 13 loci microsatellites, grâce au kit commercial *IdPlex Plus*, commercialisé par *Qiagen*, une entreprise majeure du secteur des biotechnologies. Ces analyses ont été réalisées par un laboratoire privé mexicain (*Instituto de Ciencia i Medecina Genomica*, http://institutodeciencia.com/index.htm) spécialisé en partie dans la mise en œuvre de kits de ce type pour la recherche en paternité ou l'investigation policière.

I) Quelques précisions sur les microsatellites

L'intérêt pour les microsatellites a émergé au début des années 90 (e.g. Queller et al. 1993 ; Jarne & Lagoda 1996). Cette technologie a suivi de quelques années l'invention de l'amplification d'ADN par PCR (1983) qui donnait accès à des quantités virtuellement infinies d'ADN à partir d'un échantillon [1].

Le microsatellite est une région particulière du génome, fréquemment rencontrée chez les eucaryotes (groupe comprenant l'Homme), où un motif de 2 à 6 bases est répété en tandem. La séquence génomique, pour rappel, est composée des bases G, C, T, A ; la plupart du temps, les microsatellites d'intérêt sont basés sur le motif CA, GA, GATA, GACA. La répétition donne donc généralement une séquence de type caractéristique GAGAGAGA…, GATAGATAGATA… .

Mais ici, la séquence en elle-même importe peu en réalité. L'information que l'on cherche à obtenir est la taille totale de cette région. Pour un locus (site) microsatellite donné, on peut en effet rencontrer un grand nombre de séquences de tailles différentes. La raison en est le comportement de l'enzyme servant à la réplication ; dans les régions monotones comme celles à forte répétition en tandem, les erreurs de réplication sont fréquentes, et des répétitions du motif sont oubliées ou ajoutées. S'agissant d'ADN non codant (c'est-à-dire en dehors des gènes), ces mutations n'engendrent aucun phénotype délétère. La fréquence de ces régions répétées évolue de façon neutre sans pression de sélection (sauf cas particulier, et indirectement). C'est pourquoi à chaque marqueur, le polymorphisme de taille peut donc être très important, et l'on peut fréquemment y observer une quinzaine de tailles différentes. Ainsi, pour un motif trinucléotide, on peut par exemple trouver des tailles de 150, 153, 156,… jusqu'à 195 bases (150 + 15 x 3). Du fait des contraintes de la PCR et méthodes de génotypage, la gamme des tailles prises en considération est néanmoins limitée : elle est généralement comprise entre 80 et 450 bases.

N'importe quel individu, pour peu qu'il appartienne à une espèce où l'on a décrit des marqueurs satellites, peut donc être caractérisé par la longueur de séquence à ces locus. Mais il y a plus : dans une espèce diploïde comme l'Homme, chaque individu possède toute région de son génome en double, avec un exemplaire d'origine maternelle, l'autre d'origine paternelle. Sauf anomalie, pour chaque site en particulier, l'individu possède deux copies (ou allèles). C'est son « génotype » à ce locus (noté A//A par exemple). S'il s'agit d'un marqueur microsatellite avec polymorphisme de taille, l'individu aura soit deux allèles de tailles différentes (hétérozygote), soit identiques (homozygote).

Caractériser un individu, définir un génotype, ne sont pas des notions qui sont apparues avec l'étude des microsatellites. On peut le faire avec les groupes sanguins par exemple. Mais les microsatellites ont de nombreuses propriétés qui en font un marqueur intéressant en génétique des populations, mais également pour les tests de parenté ou les investigations de la police scientifique :

– (1) ils sont codominants ; quand l'individu est hétérozygote, on peut identifier sans ambiguïté les deux allèles. Ce n'est pas le cas avec les marqueurs enzymatiques, où l'expression du dominant masque la présence du récessif. L'information est donc ici plus précise et riche.

– (2) ils sont suffisamment polymorphes à l'intérieur d'une espèce pour distinguer deux individus, pour peu que l'on cumule l'information obtenue sur un petit nombre de marqueurs indépendants (de l'ordre de 10 ; le kit *QIAGEN* en utilise 13). On n'y arriverait pas avec des marqueurs trop peu polymorphes, ou trop peu nombreux (on n'a pas nécessairement de relations de parenté avec quelqu'un qui a le même groupe sanguin que soi).

– (3) Les régions flanquantes d'un microsatellite, même sans répétitions en tandem, sont également assez variables : il s'agit toujours d'ADN non codant non soumis à la sélection. Mais la réplication n'y est pas aussi aléatoire, on peut trouver des fragments conservés à l'intérieur d'une espèce – ou au moins partagés par la plupart de ses individus. Ceci signifie que, dans une espèce donnée, on

peut donc définir des séquences spécifiques, en amont et en aval du microsatellite, pour en réaliser l'amplification par PCR (i.e. les amorces). C'est à cause de la spécificité que le kit *QIAGEN* est dédié à l'étude d'échantillons « humains » ; sans qu'il soit nécessaire de préciser, par exemple, que cet humain doit préférentiellement être de telle ou telle origine…

– (4) … mais sans garantie, non plus, que cela marche pour génotyper un chimpanzé. Les régions flanquantes sont généralement trop polymorphes pour être utilisées dans différentes espèces, sauf très occasionnellement, quand celles-ci sont proches génétiquement. Dit autrement, les amorces sont suffisamment spécifiques d'une espèce pour réaliser l'amplification en présence de contamination par de l'ADN d'autres espèces (dans des proportions raisonnables, bien entendu). Les échantillons de terrain sont de fait très fréquemment contaminés avec des bactéries ou champignons divers ; néanmoins les kits comme ceux de *Qiagen* restent utilisables en routine sur de l'ADN extrait à faible coût, parfois sur les échantillons eux-mêmes.

– (5) Comme il ne s'agit que d'amplifier un petit fragment particulier (quelques centaines de bases), même l'ADN dégradé et fragmenté peut donner de bons résultats. Cela arrive fréquemment avec des échantillons de terrain ou anciens.

II) Intérêt de l'approche pour la comparaison « Maria »/ « Cerebro »

Source

Pour revenir à l'analyse des échantillons « Maria » et « Cerebro » : la comparaison microsatellite qui a été réalisée peut être justifiée principalement d'un point de vue économique. Le prix catalogue de *Qiagen* est de 2700 CHF/100 échantillons, soit environ 24€ par échantillon. Le même fabricant propose des kits d'extraction d'ADN à moins de 4€ par échantillon. Le prix de revient d'une telle analyse est donc du même ordre, cela même en prenant en compte les coûts de main d'œuvre, d'amortissement, les marges commerciales ; et sans prendre en compte les écarts entre les prix catalogues, et les prix réels pratiqués par des fournisseurs comme *Qiagen* auprès des clients majeurs que sont les laboratoires d'analyse [2].

Il est hors de doute qu'avec les sommes collectées grâce aux Ululeurs (35 000€) Thierry Jamin est à même de mener, en toute indépendance, de nombreuses analyses microsatellite. De nombreux laboratoires à travers le monde proposent ces prestations.

En plus d'être bon marché, ce sont aussi des analyses rapides. La durée de l'ensemble des étapes,

extraction (quelques heures), PCR (1h30), passage au séquenceur (quelques heures) puis lecture du chromatogramme (presque instantanée) se compte en heures, a fortiori dans un laboratoire spécialisé.

Enfin, la méthode donne des résultats y compris avec de l'ADN obtenu et des échantillons manipulés sans précaution.

Néanmoins, il est très difficile de comprendre quelle est la question à laquelle on entendait répondre ici. Jaime Maussan comme Thierry Jamin laissent facilement entendre qu'il s'agit d'échantillons « extraterrestres ». Or, le kit n'est spécifique qu'à des échantillons humains : autrement dit, il y a de fortes chances que ce test échoue et donne des résultats non interprétables (notion d' « allèle nul », e.g. Jarne et Lagoda 1996) s'il s'agissait réellement d'échantillons extraterrestres. En première approche, pour aborder les relations génétiques à plus large échelle, à un coût équivalent, un marqueur à évolution plus lente semblait plus indiqué (e.g. un fragment d'ADN mitochondrial - voir plus bas sur l'analyse canadienne). Ceci aurait pu permettre d'inclure dans la comparaison des données publiées sur des primates et différents populations humaines contemporaines ou anciennes, par exemple celles vivant autour du site. Car en effet, dans les résultats présentés, aucun échantillon témoin ne vient en contrepoint de l'étude de similarité « Maria »/ « Cerebro ».

Certes, encore aurait-il fallu préalablement s'assurer de la présence d'ADN qui, sauf erreur, n'est pas garantie s'agissant de matériel « extraterrestre » (le succès très partiel des amplifications PCR, qui en atteste pourtant, étant donc déjà un résultat en soi) ; de plus, quelques indications sur la quantité et qualité de cet ADN en sortie d'extraction n'auraient également pas été superflues.

III) Interprétation des chromatogrammes dans les analyses microsatellites

L'image qui suit (Fig. 1) présente l'aspect d'un chromatogramme, en sortie du séquenceur automatique (ici, un *Applied Biosystems 3500 Genetic Analyzer*), après analyse des produits PCR généré avec le kit de *Qiagen IdPlex Plus de Qiagen*. Le logiciel utilisé pour identifier les allèles est *Investigator IDproof Software*. Cette image est issue du manuel du kit *Qiagen* [3] et peut être comparée aux chromatogrammes produits par le laboratoire *Instituto de Ciencia i Medecina Genomica* sollicité par Jaime Maussan.

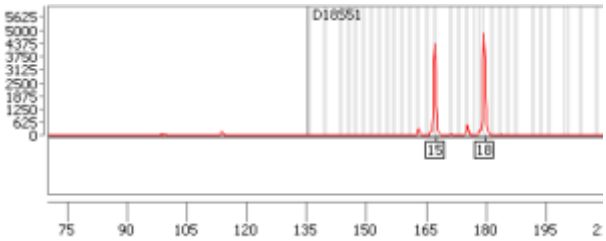


Fig. 1 - Source

L'illustration permet de dégager certaines caractéristiques habituelles d'un chromatogramme dans les analyses microsatellites.

- En abscisse, la taille du fragment amplifié par PCR. On est ici au début de la gamme de taille, de 75 à 195 bases (on monte habituellement jusqu'à 450-500 bases).

- En ordonnée l'intensité du signal fluorescent, donc la quantité de fragments détectés (en unité RFU, Relative Fluorescent Unit).

- Deux pics ressortent avec une intensité de signal de l'ordre de 4000-5000 ; ils sont notés « 15 » et

« 18 », et correspondent aux tailles de 165 et 180 (approximativement). On détecte donc ici des fragments (allèles) de tailles 165 et 180 bases, allèles appelés 15 et 18. Le marqueur (locus) microsatellite considéré est ici appelé D18S51 comme cela est indiqué au dessus. La hauteur idéale des pics est comprise entre 4000 et 5000 RFU selon le séquenceur utilisé.

– Les bandes verticales grises indiquent l'emplacement auquel on aurait pu tout aussi bien détecter un allèle pour ce locus. On parle ici de locus très bien connus, car utilisés dans le cadre d'un kit commercial. Les enjeux, autour de la mise au point de tel kit, ne sont pas négligeables ; paternité, conclusion d'enquête criminelle… Un locus ambigu car trop polymorphe, ou peu informatif, car insuffisamment polymorphe, ne serait pas retenu. C'est pourquoi on s'assure d'avoir une bonne idée de ce polymorphisme et on peut donc ensuite prédire la gamme de taille des allèles qu'on va rencontrer.

– Il apparaît parfois des pics artefactuels, comme ici entre 105 et 120. Leur hauteur est très faible, entre 10 et 100, et ils n'ont aucune signification. Il peut s'agir de fragments d'ADN dégradé, non amplifié, ou du bruit de fond du séquenceur automatique. Quand ils sont en trop grand nombre, cela questionne la qualité de l'ADN analysé.

– Les deux pics « 15 » et « 18 » sont précédés d'un petit pic, quelques bases sur leur gauche. Il s'agit d'un artefact qui résulte du comportement des enzymes de réplication et leur propension à sauter (parfois ajouter) des répétitions en tandem (en anglais *« stutter peaks »*). L'écart avec le pic proprement dit correspond donc à la taille du motif (généralement entre 2 à 4 bases, selon le marqueur).

Ainsi, l'individu caractérisé en Fig 1. est certainement humain, car on a une bonne amplification avec un des marqueurs spécifiques inclus dans le kit *kitIdPlex Plus* ; il possède le génotype 15/18 au locus D18S51 ; il est donc aussi hétérozygote à ce locus.

Il est important de comprendre que toutes ces caractéristiques de taille et de forme de pics s'appliquent à n'importe quel chromatogramme réalisé en analyse microsatellite. Toute déviation par rapport à ces critères est en principe un motif suffisant pour conclure à l'échec du génotypage.

IV) Comparaison « Maria »/ « Cerebro »

Considérons maintenant un extrait du chromatogramme produit par l'*Instituto de Ciencia i Medecina Genomica* sur la base des échantillons de Jaime Maussan (Fig 2), avec un matériel équivalent (*ABI Prism 310*). Le logiciel d'analyse n'est pas précisé, et les chromatogrammes semblent avoir été annotés manuellement. Trois marqueurs sont représentés ici, Amel, HUMTH01 et D3S1358, pour l'échantillon « Cerebro » (haut) et « Maria » (bas). HUMTH01 ne peut désigner ici que TH01, marqueur dont la taille est intermédiaire entre Amel et D3S1358 dans le kit *IdPlex Plus*. Le génotype est directement reporté, semble-t-il manuellement aussi, à coté du nom du locus (par exemple, XY et XX pour Amel).

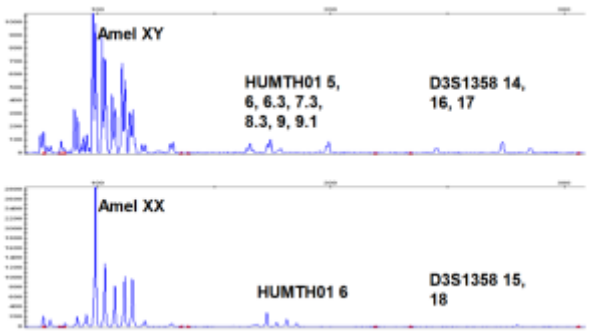


Fig. 2 - Source

On peut faire les remarques suivantes :

– Le nombre de pics pour Amel contredit l'affirmation des auteurs selon laquelle il s'agit d'un génotype diploïde (i.e. XY ou XX). On devrait en avoir au maximum deux, or on en compte 5, quoiqu'assez faibles, de tailles équivalentes et supérieures à 300.

– Aux autres locus, les auteurs suggèrent qu'il y a polyploïdie, ou de multiples duplications du locus (« Cerebro » aurait 7 allèles distincts pour TH01, soit au minimum 4 locus). Ceci n'est pas conforme au comportement attendu pour un kit spécifique d'échantillons humains. L'interprétation mériterait d'être annoncée avec prudence et explicitée.

– Aucun pic significatif n'est visible pour D3S1358 et HUMTH01, alors qu'un génotype est décrit (par exemple 14//16//17 pour « Cerebro » sur D3S1358). Il est vraisemblable qu'il n'y a eu ici aucune amplification : il est difficile de comprendre ici la méthode employée pour détecter les allèles (logiciel éventuel, seuil). D'autant plus que, pour d'autres marqueurs, non représentés ici, un profil en apparence tout à fait similaire est interprété correctement (« non amplifica », N.A sur les chromatogrammes fournis par Thierry Jamin).

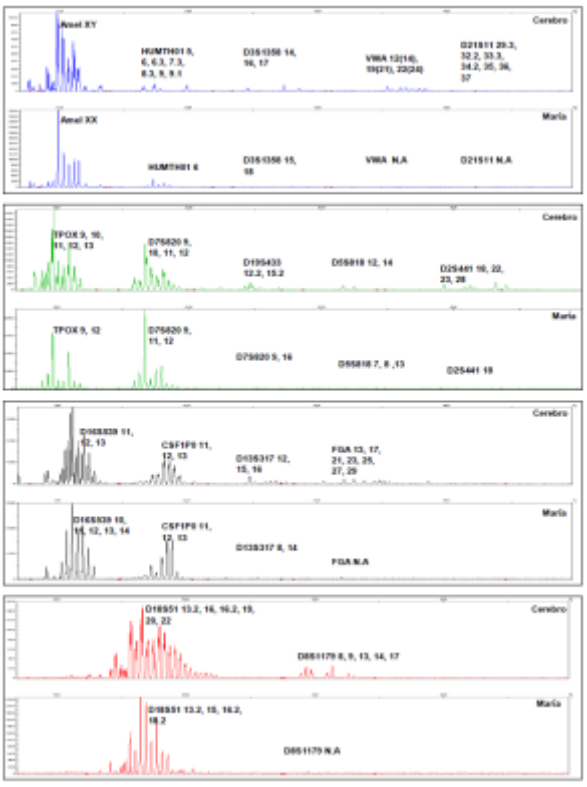
– Quand ils ne sont pas de hauteurs trop faibles, la forme des pics n'est globalement pas conforme aux attendus, excepté peut être pour Amel chez « Maria ». Pour Amel chez « Cerebro », le sommet des pics semble tronqué, avec des « épaules » ou « vallées » (*shoulder peak/valley peak*), la base est large et sa taille exacte à reporter en abscisse est douteuse. Ce signal suggère notamment un mix de fragments de tailles trop proches ou des conditions de température inadéquates.

– L'interprétation qui est faite du génotype en Amel de « Maria » est différente entre l'annotation du chromatogramme (XX) et le compte rendu des résultats (XY).

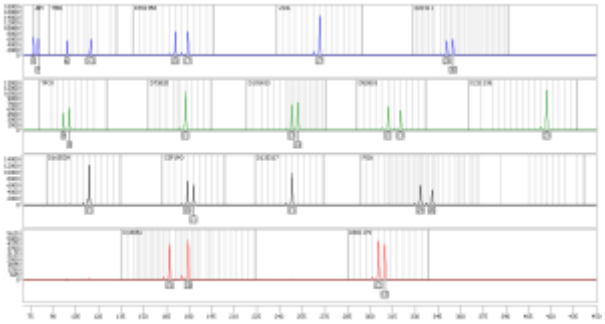
– L'interprétation générale est extrêmement confuse : la surabondance de pics en début de spectre, et la faiblesse du signal sur la majeure partie des spectres évoquent fortement un ADN de mauvaise qualité et contaminé par de multiples sources. Les tailles des allèles détectés ne sont pas toujours conformes aux attendus du kit. Ainsi, l'échantillon « Cerebro » est caractérisé 12(14)//19(21)//22(24) pour le marqueur VWA (non représenté ici). Cette interprétation est basée sur la présence de pics très faibles compris entre 350 bases et 400 bases. Or la taille attendue pour ce marqueur est au maximum de 300 bases approximativement. Il semble que cette interprétation aurait donc dû, au minimum, être proposée avec prudence, ce qui n'est pas le cas.

Au total, peu de certitudes émergent de l'analyse qui a été réalisée, si ce n'est qu'il s'agit sans doute d'ADN humain, ou très fortement apparenté, pour lequel l'amplification a été possible. Beaucoup d'éléments suggèrent que les résultats sont inexploitables. L'hypothèse d'une origine extraterrestre, aux implications immenses, mériterait nécessairement d'autres analyses et une mise en évidence beaucoup plus convaincante. »

Le lecteur intéressé peut approfondir par lui-même la comparaison entre le chromatogramme fourni par Thierry Jamin, page 25 de son document de résultats d'analyses :



et l'exemple de chromatogramme fourni par le manuel du kit *Qiagen* page 52, figure 2 :



La deuxième analyse, celle fournie (gratuitement semble-t-il) par le laboratoire canadien *Paleo-DNA* à Thierry Jamin, a suscité moins d'observations de la part de ce lecteur :

« La méthode choisie (ADNmt) serait (légèrement) plus appropriée dans le cas présent, pour aborder le polymorphisme à large échelle… à condition bien sûr de faire une quantité si grande d'hypothèses (présence d'ADN, d'ADNmt, code génétique identique…) qu'elles reviennent à dire qu'on n'a aucunement affaire à un extraterrestre, mais à un bon vieux eucaryote terrestre. Si on néglige la circularité remarquable du raisonnement, il faut d'abord remarquer que les résultats bruts d'analyse (dosage, chromatogramme) ne sont ici pas produits. Contrairement à l'analyse microsatellite, on ne dispose ici que de l'interprétation qu'en fait le laboratoire, même si, bien sûr, il n'y a pas de raison d'en douter a priori.

Source

Les alignements réalisés par le laboratoire sont corrects – même s'ils sont présentés curieusement, pas de % de coverage ni e-value – et on est bien en présence indiscutablement d'ADN 16S mitochondrial humain. La manipulation et l'extraction d'ADN ont été réalisées correctement comme le prouveraient les contrôles négatifs ; quant aux contrôles positifs d'éventuel ADN humain contaminant, à commencer par celui de Thierry Jamin et de son équipe, il est bien mentionné que rien n'a été fourni pour les réaliser. En leur absence on peut très bien supposer que c'est un ADN contaminant qui est ici analysé. Certaines difficultés rencontrées (échec de l'amplification du 12S notamment) ne sont pas claires ; elles pourraient être liées à une mauvaise qualité de l'ADN mais il n'y a malheureusement aucune information à ce sujet.

Au total c'est une prestation qui semble globalement correctement réalisée, qui conclut sans aucun doute possible qu'on a affaire à du matériel biologique humain. Il est à noter que ce sont des analyses très simples et bon marché ; à savoir une journée de travail environ si le séquençage est réalisé dans la foulée de l'extraction, et pour un coût modique (de l'ordre de 100€ en incluant les coûts de main d'œuvre). On voit de pareils exploits réalisés régulièrement par des stagiaires universitaires dès le niveau L3. »

Du coup, avec les milliers d'euros que Thierry Jamin était censé consacrer aux analyses, on aurait dû pouvoir déjà obtenir pas mal de résultats du même type, par exemple sur les grandes mains ou les « poupées » Alberto, Josefina et Victoria, qu'on s'obstine à nous présenter comme des êtres intacts ayant été vivants malgré les aberrations révélées par les radios…

De même, on peut se poser quelques questions lorsqu'il annonce qu'il faudra « des mois, voire des années », pour avoir des analyses ADN complètes. Le coût du séquençage baisse très rapidement, on en est aujourd'hui aux environs de 1000 $ pour un séquençage complet de génome humain (voir ce graphique), auxquels il faut ajouter les coûts d'extraction : https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/repli-g-mini-kit/#orderinginformation ; quant aux délais pour réaliser un séquençage, on est à trois jours sur certaines machines, voire encore moins sur les plus récentes, donc même en comptant les délais d'acheminement des échantillons et la paperasse, c'est une affaire de quelques semaines, pas de plusieurs mois ou années [4]…

Notes

[1] Jarne, P., & Lagoda, P. J. (1996). « Microsatellites, from molecules to populations and back ». *Trends in ecology & evolution*, 11(10), 424-429.
http://entocourses.tamu.edu/ento601/pdf/Jarne_Lagoda.pdf
Queller, D. C., Strassmann, J. E., & Hughes, C. R. (1993). « Microsatellites and kinship ». *Trends in Ecology & Evolution*, 8(8), 285-288.
http://www.sciencedirect.com/science/article/pii/0169534793902560

[2] https://www.qiagen.com/ch/shop/cart/investigator-idplex-plus-kit/#orderinginformation
https://www.qiagen.com/fr/shop/sample-technologies/dna/genomic-dna/dneasy-blood-and-tissue-kit/#orderinginformation

[3] https://www.qiagen.com/ch/resources/resourcedetail?id=72de9514-100b-4174-aaf4-49b523230986&lang=en page 52, figure 2.

[4] Reste bien sûr la question de l'analyse proprement dite de la séquence obtenue, pour laquelle il faut disposer de vraies compétences que n'ont ni Thierry Jamin ni aucun des « scientifiques » de l'équipe de Jaime Maussan ; sans oublier bien sûr le problème fondamental de toute cette démarche soi-disant scientifique, comme l'écrit mon correspondant : « Ces outils sont dédiés à du matériel qui a d'innombrables propriétés valables sur Terre, mais qui n'ont aucune raison d'être vraies ailleurs. A partir du moment où on séquence, c'est parce qu'on sait d'avance que c'est du matériel terrestre ; si c'est aussi la conclusion à laquelle on arrive, c'est un raisonnement circulaire ».