



## Point de vue

### Le séquençage complet des génomes, un investissement rationnel pour le développement d'outils de diagnostic et d'épidémiologie : l'exemple de la métrite contagieuse équine

Sandrine Petry (sandrine.petry@anses.fr)

Anses, Laboratoire de pathologie équine de Dozulé, Goustranville, France

**Le nombre de génomes complets disponibles augmente de jour en jour avec la réduction continue des coûts et des délais de séquençage. De nombreux laboratoires, aussi bien publics que privés, participent à cet effort qui révolutionne la recherche fondamentale et tous les domaines en lien avec un processus biologique. Il est d'ailleurs à parier que dans un futur proche, le séquençage des génomes sera un outil de routine pour les laboratoires de diagnostic et de référence en microbiologie.**

#### Introduction

La génomique a connu d'importantes révolutions technologiques depuis la publication des premières méthodes de séquençage de l'ADN et du premier génome du bactériophage  $\Phi$ X174 en 1977. Nous parlons aujourd'hui de séquencer le génome d'un individu en quelques heures pour 1 000 dollars alors que le premier génome humain complet a été obtenu en 2003 après 13 ans de travail pour un coût d'environ 2,7 milliards de dollars. Rien que dans le monde bactérien, près de 4 000 génomes complets sont maintenant disponibles et quatre fois plus sont incomplets ou en cours de séquençage et ce, moins de 20 ans après le séquençage du premier génome d'un organisme vivant (*Haemophilus influenzae*, génome de 1,83 Mb). Cet essor vertigineux est motivé par la volonté de connaître, traiter et prédire les pathologies, d'innover dans les domaines des biotechnologies, de l'environnement, de l'agronomie..., mais aussi par un aspect plus fondamental lié à la connaissance de la vie et à la compréhension de son évolution.

#### La métrite contagieuse équine (MCE) en quelques mots

La MCE constitue un problème économique en filière équine (élevage, exportation, vente). Cette infection bactérienne, sexuellement transmissible et contagieuse est apparue fin des années 1970 dans plusieurs régions du globe concernées par des mouvements importants de chevaux. Aujourd'hui, elle est toujours mondialement répartie mais son dépistage obligatoire limite les foyers à quelques cas. Elle se traduit cliniquement par une inflammation de l'endomètre des juments généralement accompagnée d'une infertilité temporaire. En absence de traitement, les animaux (mâles et femelles) peuvent rester porteurs de l'agent pathogène durant plusieurs années. Les échecs de traitement semblent fréquents bien que les souches ne sont apparemment pas résistantes aux antibiotiques employés. D'abord classé dans le genre *Haemophilus*, l'agent causal de la MCE a finalement été reclassé en 1985 dans un nouveau genre bactérien nommé *Taylorella*, aujourd'hui constitué de deux espèces : *T. equigenitalis* dont la présence entraîne la déclaration d'un cas de MCE, et *T. asinigenitalis*,

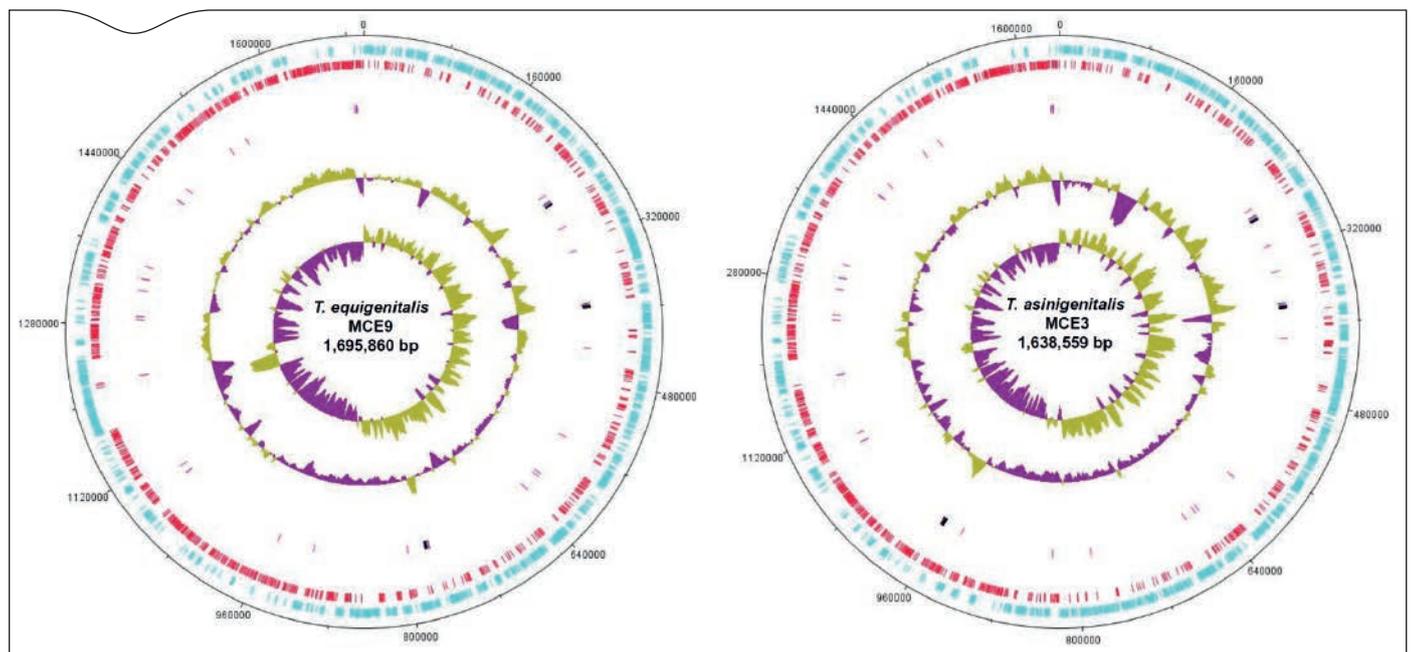


Figure 1. Cartes génétiques des génomes de *T. equigenitalis* MCE9 et *T. asinigenitalis* MCE3.



## Point de vue

considérée comme non pathogène malgré la présence de signes cliniques de métrite après des infections expérimentales intra-utérines chez plusieurs juments. Malgré la croissance lente du genre *Taylorella* sur les milieux de culture actuels par rapport à la flore commensale du tractus génital, le diagnostic officiel de la MCE repose sur l'isolement et l'identification biochimique de l'agent pathogène; de plus, la confirmation de l'espèce par PCR est techniquement nécessaire puisque *T. equigenitalis* et *T. asinigenitalis* sont phénotypiquement très comparables.

Le laboratoire national de référence (LNR) et les activités européennes de référence (dans le cadre du laboratoire de référence de l'Union européenne pour les maladies équine) pour la MCE sont hébergés par l'unité Bactériologie et Parasitologie du laboratoire de pathologie équine de Dozulé de l'Anses.

### Le séquençage complet des génomes au service de la référence, l'exemple de la MCE

Pour un laboratoire de référence dont le rôle est de développer des outils plus fiables de détection/caractérisation des agents pathogènes ainsi que des nouvelles méthodes de typage, le manque d'information génétique peut être un facteur très limitant. L'éloignement génomique du genre *Taylorella* par rapport à des genres bactériens théoriquement proches (exemple de *Bordetella* et *Haemophilus*) et l'absence de données de séquences en dehors des opérons ribosomiques en est un bon exemple. Pour y remédier, nous avons entrepris le séquençage *de novo* des génomes des deux espèces du genre *Taylorella* en collaboration avec l'unité Micalis UMR1319 de l'Inra (Jouy-en-Josas, France); deux technologies de séquençage haut débit ont été combinées : la technologie 454 (Margulies *et al.*, 2005), que nous avons couplé à une lecture « paired-end » (Fullwood *et al.*, 2013) pour faciliter l'assemblage des deux génomes, et la technologie Solexa/illumina (Bentley *et al.*, 2008). Les deux génomes ont ensuite été annotés et comparés (Hébert *et al.*, 2011; Hébert *et al.*, 2012; **Figure 1**).

La disponibilité des séquences génétiques facilite grandement le développement d'outils de diagnostic moléculaire. Pour améliorer la fiabilité du dépistage de la MCE, nous collaborons actuellement à la mise au point d'une PCR multiplexe en temps réel qui s'affranchira des ADN ribosomiques 16S pour gagner en spécificité. La connaissance de la séquence des génomes profite aussi au développement d'outils de diagnostic bactériologique ou sérologique. Par exemple, la reconstruction *in silico* des voies métaboliques du genre *Taylorella* nous permet d'appréhender ses besoins nutritionnels et d'envisager le développement d'un milieu de culture plus performant pour augmenter la sensibilité du diagnostic officiel de la MCE. La mise au point de ce milieu de culture a débuté en janvier 2013 en collaboration avec la société AES chemunex (bioMérieux industry, France).

Plusieurs outils d'épidémiologie moléculaire sont disponibles et pertinents en fonction du pathogène et de leur utilisation (Sabat *et al.*, 2013). Pour le genre *Taylorella*, notre choix s'est porté sur la MLST (*MultiLocus Sequence Typing*), qui, en plus d'être robuste, simple d'exécution et portable, est une méthode de référence pour la surveillance épidémiologique globale d'une infection et l'étude de l'évolution des populations bactériennes. À partir des génomes annotés de *T. equigenitalis* et *T. asinigenitalis*, nous avons donc sélectionné plusieurs gènes communs aux deux espèces du genre *Taylorella* dont les produits assurent les fonctions indispensables à la survie des

cellules. Sept d'entre eux ont été validés chez 113 *T. equigenitalis* et 50 *T. asinigenitalis* provenant de six pays sur une période de 35 ans. À terme, cet outil d'épidémiologie moléculaire sera transféré aux LNRs de l'Union européenne pour la MCE afin de réaliser un état des lieux des souches du genre *Taylorella* circulantes en Europe ainsi que des études rétrospectives en fonction des collections bactériennes disponibles. Les données de MLST (épidémiologie des souches et séquences d'ADN des sept marqueurs MLST) seront mutualisées et en libre accès sur une base spécifique du genre *Taylorella* (<http://pubmlst.org/taylorella/>) hébergée par le Département de zoologie de l'Université d'Oxford (Jolley and Maiden, 2010).

En parallèle du développement d'outils de diagnostic et d'épidémiologie moléculaire, les laboratoires de référence mènent des recherches approfondies pour accroître les connaissances fondamentales sur la biologie et le pouvoir infectieux des agents pathogènes, en interaction avec leur hôte et l'environnement. La connaissance de la séquence complète des génomes facilite grandement la programmation de cette recherche au long court. La comparaison des génomes annotés de *T. equigenitalis* et de *T. asinigenitalis* nous a, par exemple, permis de lister les potentiels facteurs de colonisation et de virulence communs au genre *Taylorella* et spécifiques de chacune des deux espèces. En perspective, et pour avoir une vision plus fine de la diversité génomique du genre *Taylorella*, nous nous apprêtons à séquencer les génomes d'une dizaine de souches en collaboration avec la plateforme génomique/transcriptomique de l'Anses (unité Génétique Virale et Biosécurité - laboratoire de Ploufragan-Plouzané).

### Utilisation du séquençage complet des génomes en routine

L'utilisation du séquençage complet des génomes en routine a récemment fait l'objet de plusieurs articles et revues avec, entre autres, une étude pilote réalisée chez *Staphylococcus aureus* et *Clostridium difficile* (Eyre *et al.*, 2012). Tous s'accordent à dire que la réduction continue des coûts et des délais de séquençage représente un changement radical dans les capacités des laboratoires de diagnostic et de référence en microbiologie. Les applications les plus évidentes sont l'identification de microorganismes (notamment s'ils sont non cultivables, à culture difficile ou hautement pathogènes) et le génotypage de souches avec une résolution optimale en « temps réel ». L'épidémie de *Escherichia coli* entérohémorragique O104:H4 en Allemagne en 2011 (Mellmann *et al.*, 2011) et la récente identification de l'agent étiologique de la maladie de Theiler (Chandriani *et al.*, 2013) pour laquelle les techniques classiques d'identification avaient échouées depuis près d'un siècle, illustrent parfaitement ces applications. Les données de séquençage des génomes et les nouvelles technologies de séquençage peuvent aussi être utilisées pour analyser spécifiquement certains gènes cibles (par exemple des gènes de résistance aux antibiotiques) et ainsi réduire les coûts de préparation, d'analyse et de stockage des données.

Il est cependant à préciser que, même si les contraintes techniques et les coûts liés à la génération des données de séquences ne sont plus réellement des facteurs limitants, comme illustré par l'apparition de matériels de séquençage « presse-bouton », il semble difficile d'envisager la généralisation de l'utilisation du séquençage complet des génomes comme méthode de diagnostic et d'épidémiologie à court terme. En effet, plusieurs avancées technologiques



## Point de vue

restent à réaliser pour faciliter l'analyse des données à l'aide de logiciels d'interprétation automatique utilisables par des microbiologistes non bio-informaticiens et pour mutualiser ces données à l'aide d'interfaces communes suffisamment conviviales et performantes pour permettre l'incrémentation de données et leur comparaison avec celles existantes. Ce dernier point implique que les formats et logiciels issus des différentes plateformes de séquençage soient homogénéisés et les infrastructures informatiques adaptées pour le stockage et le transport de telles quantités de données générées.

En conclusion, bien qu'il soit peu probable que le séquençage complet des génomes se substitue totalement aux méthodes de diagnostic et d'épidémiologie actuelles, il s'agit d'une méthode d'avenir pour les laboratoires de référence ouvrant des possibilités bien supérieures à celles actuelles.

### Remerciements

Je tiens à remercier Marie-France Breuil, Fabien Duquesne et Laurent Hébert pour leur implication quotidienne à la référence et à la recherche sur la MCE, le service administratif et financier du laboratoire de pathologie équine de Dozulé ainsi que son Directeur, Claire Laugier. Je remercie également L. Hébert pour son implication active et ses conseils dans la relecture de cet article.

Liste des financeurs de la recherche sur la MCE présentée dans cet article : Anses, Conseil régional de Basse-Normandie, FEDER, Institut français du cheval et de l'équitation, Union européenne.

### Références bibliographiques

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgman JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218): 53-9.

Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, Van Blaricum AJ, Tennant BC, Kistler AL. 2013. Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15): 1407-1415.

Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TE, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open*. 2012, 2(3): e001124.

Fullwood MJ, Lee J, Lin L, Li G, Huss M, Ng P, Sung WK, Shenolikar S. 2011. Next-generation sequencing of apoptotic DNA breakpoints reveals association with actively transcribed genes and gene translocations. *PLoS One*, 6(11): e26054.

Hébert L, Moumen B, Duquesne F, Breuil MF, Laugier C, Batto JM, Renault P, Petry S. 2011. Genome sequence of *Taylorella equigenitalis* MCE9, the causative agent of contagious equine metritis. *Journal of Bacteriology*, 193(7): 1785.

Hébert L, Moumen B, Pons N, Duquesne F, Breuil MF, Goux D, Batto JM, Laugier C, Renault P, Petry S. 2012. Genomic characterization of the *Taylorella* genus. *PLoS One*, 7(1): e29953.

Jolley KA., Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC bioinformatics*, 11, 595.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, 437(7057): 376-80.

Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One*, 6(7): e22751

Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijk Jm, Laurent F, Grundmann H, Friedrich AW, on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM). 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance*, 18(4): 20380.