

Sujet de stage M2 : Annotation et structure génétique des Foc-TR4 et lignées proches, pour l'exploration du pangéome

Encadrement : Sébastien RAVEL (sebastien.ravel@cirad.fr) et Emmanuel WICKER (emmanuel.wicker@cirad.fr), UMR PHIM, Montpellier

Les techniques de séquençage de nouvelle génération ont permis de mieux cerner les mécanismes moléculaires liés à l'adaptation des pathogènes de plante à de nouveaux hôtes ou de nouveaux environnements. Les relations entre variation de structure génomique et évolution des champignons pathogènes sont ainsi de plus en plus documentés (Eschenbrenner *et al.*, 2020, Plissonneau *et al.*, 2018).

Fusarium oxysporum f.sp. *ubense* (Foc) est un champignon ascomycète haploïde, responsable d'une maladie majeure du bananier (Fusariose, aussi appelée « Fusarium wilt of Banana » ou maladie de Panama). Cette espèce est assez diverse, se structurant en 3 clades (A, B, C), 8 clusters génétiques (SC01 à 8), et 39 groupes de compatibilité végétative (VCG) (Mostert *et al.*, 2022). L'émergence de la lignée « Tropical Race 4 » (FOC-TR4, Clade A/SC02) suscite de fortes inquiétudes pour la pérennité de cette culture (Ordonez *et al.*, 2015), d'autant plus que son expansion mondiale a brusquement accéléré depuis sa sortie d'Asie en 2011. Nous faisons l'hypothèse que cette accélération brusque n'est pas seulement liée à la globalisation des échanges, mais aussi à des changements génomiques dans cette lignée. Des « chromosomes accessoires » ont d'ailleurs été décrits dans le génome de *F.oxysporum* (Yang *et al.*, 2020). A ce jour, plusieurs génomes de FOC-TR4 et non-TR4 sont accessibles sur NCBI, mais les génomes de qualité chromosome-level manquent, à part Fol (Ma *et al.*, 2010, Ayhan *et al.*, 2018) et Fo 47 (Wang *et al.*, 2020).

Notre équipe a récemment généré plusieurs séquences génomiques en technologie long-read (Oxford Nanopore, ONT) : deux génomes de TR4 (issus respectivement de la phase asiatique de la maladie, et de la phase mondiale) et deux génomes non-TR4 du même clade A (SC01 et 02), assemblés grâce au pipeline CulebrONT (Orjuela *et al.*, 2022). Cette première étude a d'abord optimisé la qualité des assemblages génomiques, puis comparé différents outils d'annotation structurale et fonctionnelle. Nous avons ainsi mis en évidence plusieurs variations de structure génomique, en identifiant 12 core-chromosomes et un nombre variable de chromosomes accessoires, caractérisés par une plus faible densité en gènes et un enrichissement en effecteurs de pathogénicité et séquences répétées (Bache, 2022). Ces premiers résultats suggèrent que les génomes de Foc suivent le modèle « two-speed genome » (Dong *et al.*, 2015, Moller & Stukenbrock, 2017).

L'objectif de ce stage est de préciser la définition du core-génome et pangéome structuraux et du « pan-gene atlas » (tels que définis par (AgBioData, 2020) et cité par (Durant, 2022)) chez la Foc-TR4 et les lignées proches dans le clade A, par l'analyse d'un plus grand nombre de génomes séquencés en technologie ONT.

Matériel biologique, ressources génomiques

Outre les quatre génomes assemblés et annotés en 2022 qui serviront de références, le laboratoire dispose de données génomiques short-read (120 isolats) et de SNPs spécifiques de la TR4.

Seize nouvelles séquences génomiques ONT vont être générées, dont une dizaine de TR4, et des lignées proches dans le clade A. Certaines TR4 et VCG0121 ont été échantillonnées dans le même pays (Taiwan) et dans la phase très précoce de l'émergence (fin des années 1980).

Les workflows d'assemblage et analyses génomiques sont disponibles : CulebrONT (Orjuela et al., 2022), PodiumASM, EffiCAZ (Bache, 2022).

Programme de stage

1. Assemblages, annotation structurale et fonctionnelle

Les assemblages, annotations structurales et fonctionnelles seront menés selon l'approche décrite dans (Bache, 2022). Les workflows SNAKEMAKE développés à ce moment-là seront à optimiser.

Il s'agira notamment de :

- Faire un inventaire des effecteurs et des « carbohydrate-active enzymes » (CAZymes) grâce au workflow EffiCAZ
- Cet inventaire sera éventuellement d'autres outils (funannotate, autres selon (Achari *et al.*, 2021))
- Etudier la distribution de ces effecteurs sur le génome : localisation chromosomique, proximité de zones répétées (MIMPs)
- Faire un inventaire des zones répétées et éléments transposables
- Etudier leur distribution sur les chromosomes, par famille de transposons
- Etudier leur association avec les gènes de virulence
- Identifier le core-génome et pangénome fonctionnel (ou « pan-gene atlas ») en établissant la distribution des orthogroupes par famille fonctionnelle (Gene Ontology, KEGG)

2. Comparaison des structures génomiques

L'objectif de cette partie est d'établir et comparer les structures génomiques des VCG0120, VCG0121, et VCG01213/16. Sur chaque assemblage, il s'agira de déterminer le nombre de chromosomes présents à l'aide des différents outils mis en œuvre dans PodiumASM (repérage des variants structuraux, identification des zones télomériques par Tapestry, ...).

Les chromosomes accessoires (« lineage-specific »), éventuelles ruptures de synténie, seront identifiées par différents outils (SynVISIO, CIRCOS,..), et interprétées en tenant compte des relations phylogénétiques entre lignées.

Le nombre de génomes étudié ici permettra également de visualiser les échanges génétiques récents entre TR4 et lignées proches, par une approche de chromosome painting probabiliste (Rogerio *et al.*, 2022).

Rémunération :

Gratification mensuelle de 600 €.

Références

- Achari, S. R., Kaur, J. K., Mann, R. C., Sawbridge, T., Summerell, B. A. and Edwards, J. (2021) Investigating the effector suite profile of Australian *Fusarium oxysporum* isolates from agricultural and natural ecosystems. *Plant Pathology*, **70**, 2. doi: <https://doi.org/10.1111/ppa.13303>.
- AgBioData, T. C. (2020) Pan-genome viewers (May 2020). [online video] 2020 [cited 2022 June 16]. Available from: <https://www.youtube.com/watch?v=ATPzVlrTW0s>.
- Ayhan, D. H., Lopez-Diaz, C., Di Pietro, A. and Ma, L. J. (2018) Improved Assembly of Reference Genome *Fusarium oxysporum* f. sp. *lycopersici* Strain Fo4287. *Microbiol Resour Announc*, **7**, 10. doi: 10.1128/MRA.00910-18.
- Bache, S. (2022) Analysis of the genomic structures of several strains of the fungal plant pathogen *Fusarium oxysporum* f.sp. *cubense*. Master Thesis Thesis, Université de Strasbourg. Montpellier, Strasbourg, 30 pages.
- Dong, S. M., Raffaele, S. and Kamoun, S. (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.*, **35**. doi: 10.1016/j.gde.2015.09.001.
- Durant, E. (2022) Design of novel visual representations and tools applied to plant pangenome visualization. PhD thesis Thesis, Université de Montpellier. 252 pages.
- Eschenbrenner, C. J., Feurtey, A. and Stukenbrock, E. H. (2020) Population Genomics of Fungal Plant Pathogens and the Analyses of Rapidly Evolving Genome Compartments. *Methods Mol Biol*, **2090**. doi: 10.1007/978-1-0716-0199-0_14.
- Ma, L. J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W. B., Woloshuk, C., Xie, X., Xu, J. R., Antoniw, J., Baker, S. E., Bluhm, B. H., Breakspear, A., Brown, D. W., Butchko, R. A., Chapman, S., Coulson, R., Coutinho, P. M., Danchin, E. G., Diener, A., Gale, L. R., Gardiner, D. M., Goff, S., Hammond-Kosack, K. E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C. D., Koehrsen, M., Kumar, L., Lee, Y. H., Li, L., Manners, J. M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S. Y., Proctor, R. H., Regev, A., Ruiz-Roldan, M. C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D. C., Turgeon, B. G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C. A., Kistler, H. C. and Rep, M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, **464**, 7287. doi: 10.1038/nature08850.
- Moller, M. and Stukenbrock, E. H. (2017) Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol*, **15**, 12. doi: 10.1038/nrmicro.2017.76.
- Mostert, D., Wicker, E., de Jager, M. M., Al Kaabi, S. M., O'Neill, W. T., Perry, S., Li, C., Yi, G., Pegg, K. G., Mostert, L. and Viljoen, A. (2022) A Polyphasic Approach Reveals Novel Genotypes and Updates the Genetic Structure of the Banana *Fusarium* Wilt Pathogen. *Microorganisms*, **10**, 2, 269. doi: 10.3390/microorganisms10020269.
- Ordonez, N., Seidl, M. F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B. P., Ploetz, R. C. and Kema, G. H. (2015) Worse Comes to Worst: Bananas and Panama Disease--When Plant and Pathogen Clones Meet. *PLoS Pathog*, **11**, 11. doi: 10.1371/journal.ppat.1005197.
- Orjuela, J., Comte, A., Ravel, S., Charriat, F., Vi, T., Sabot, F. and Cunnac, S. (2022) CulebrONT: a streamlined long reads multi-assembler pipeline for prokaryotic and eukaryotic genomes. *Peer Community Journal*, **2**, e46. doi: 10.24072/pcjournal.153.
- Plissonneau, C., Hartmann, F. E. and Croll, D. (2018) Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.*, **16**, 1. doi: 10.1186/s12915-017-0457-4.
- Rogério, F., Van Oosterhout, C., Ciampi-Guillard, M., Correr, F. H., Hosaka, G. K., Cros-Arteil, S., Margarido, G. R. A., Massola, N. S. and Gladieux, P. (2022) Means, motive and opportunity for biological invasions: Genetic introgression in a fungal pathogen. *Mol. Ecol.* doi: 10.1111/mec.16366.
- Wang, B., Yu, H., Jia, Y., Dong, Q., Steinberg, C., Alabouvette, C., Edel-Hermann, V., Kistler, H. C., Ye, K., Ma, L. J. and Guo, L. (2020) Chromosome-Scale Genome Assembly of *Fusarium oxysporum* Strain Fo47, a Fungal Endophyte and Biocontrol Agent. *Mol Plant Microbe Interact*, **33**, 9. doi: 10.1094/MPMI-05-20-0116-A.
- Yang, H., Yu, H. and Ma, L.-J. (2020) Accessory Chromosomes in *Fusarium oxysporum*. *Phytopathology*®, **110**, 9. doi: 10.1094/phyto-03-20-0069-ia.