



INTERREG V A FRANCE-SUISSE 2014-2020

(2014 - 2020)

Axe prioritaire-Priorité d'investissement-Objectif spécifique 2-1-2

AXE 2 : Protéger et valoriser le patrimoine naturel et culturel

6.c En conservant, protégeant, favorisant et développant le patrimoine naturel et culturel

OS 4 = Préserver et restaurer les écosystèmes fragilisés de l'espace transfrontalier

SYNAQUA

(Ref : 2369 / 2016-36)

SYNérgie transfrontalière pour la bio-surveillance et la préservation des écosystèmes AQUAtiques

LIVRABLE 39634

WP2.1 : Bases de données de référence

Responsable : INRA – UMR CARRETEL



Bases de données de référence

Une base de données de référence a été développée pour chacun des 2 bio-indicateurs, diatomées et oligochètes, grâce à l'intégration des expertises des chercheurs suisses et français. En effet, ces bases s'appuient sur une expertise déjà ancienne dont le projet SYNAQUA a permis de consolider les acquis. Elles reposent sur des spécimens collectés pour les oligochètes ou isolés et cultivés pour les diatomées. Ces spécimens sont caractérisés de deux façons : 1) par une identification taxonomique (nom d'espèce et de genre) sur la base de critères morphologiques et 2) par un barcode ADN spécifique de ce spécimen. C'est cette double identification qui permet d'assigner une taxonomie à une séquence environnementale.

1- Base de référence « diatomées »

Connaissances et références préalables :

Les diatomées (Bacillariophyta) sont des micro-algues ubiquistes qui produisent un exosquelette siliceux et contribuent de manière importante à la productivité des milieux aquatiques. Elles présentent une grande diversité, ce qui en fait d'excellents indicateurs écologiques des écosystèmes. Habituellement, les diatomées sont identifiées à l'aide des caractéristiques morphologiques de leur exosquelette. Une bibliothèque de référence experte est nécessaire pour associer de façon fiable un identifiant taxonomique à une séquence ADN (barcode). Le barcode *rbcL*, un marqueur chloroplastique, est adapté à l'identification des diatomées au niveau des espèces (Kermarrec et al. 2013a, 2013b, 2014) et a permis une large application de l'approche de metabarcoding à la bioindication des cours d'eau (ex : Vasselon et al 2017). Pour cela, l'INRA a développé R-Syst::diatom, une bibliothèque en accès libre dédiée aux diatomées, maintenue depuis 2012 qui a été créée à l'aide du réseau de systématique de l'INRA, R-Syst (Rimet et al 2016, 2018). Les données proviennent de deux sources (1) la base de données de nucléotides NCBI (Centre national d'information sur la biotechnologie) et (2) les données de séquençage non publiées de collections de cultures en France, au Royaume-Uni et en Russie.

- KERMARREC L., BOUCHEZ A., RIMET F., HUMBERT J.F. (2013) First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing complex (Bacillariophyta). *Protist* 164:686-705
- KERMARREC L., FRANC A., RIMET F., CHAUMEIL P., HUMBERT J.F., BOUCHEZ A. (2013) Next Generation Sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources* 13(4):607-619
- KERMARREC L., FRANC A., CHAUMEIL P., RIMET F., FRIGERIO J.M., HUMBERT J.F., BOUCHEZ A. (2014) A Next-Generation Sequencing approach for river biomonitoring using benthic diatoms. *Freshwater Science* 33(1):349-363
- RIMET F., CHAUMEIL P., KECK F., KERMARREC L., VASSELON V., KAHLERT M., FRANC A., BOUCHEZ A. (2016) R-Syst::diatom: An open-access and curated barcode database for diatoms and freshwater monitoring. *DATABASE The Journal of Biological Databases and Curation* 1-21 DOI 10.1093/database/baw016
- RIMET F., ABARCA N., BOUCHEZ A., KUSBER W.H., JAHN R., KAHLERT M., KECK F., KELLY M., MANN D.G., PIUZ A., TROBAJO R., TAPOLCZAI K., VASSELON V., ZIMMERMANN J. (2018) The potential of high throughput sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. *Fottea* 18(1):37-54
- VASSELON V., RIMET F., TAPOLCZAI K., BOUCHEZ A. (2017) Assessing ecological status with diatom DNA metabarcoding: scaling-up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators* 82:1-12

Base de référence actualisée lors du projet SYNAQUA :

Depuis 2017, plusieurs experts européens ont collaboré pour gérer la base de référence RSyst::diatom devenue récemment « *Diat.barcode* ». Elle rassemble désormais des références issues de « The Culture Collection of Thonon » (TCC – INRA), « The Barcoding Project of Diatoms » (UK) et de données publiques (NCBI). La version V7 de *Diat.barcode* (23/02/2018) a fait l'objet d'une curation par un panel international d'experts en taxonomie des diatomées. La base de données *diat.barcode* est accessible en libre accès, ainsi qu'un sous-ensemble de la base prêt à l'emploi pour les analyses de métabarcoding.

La description de la base de référence *Diat.barcode* est disponible à :

https://www6.inra.fr/carrtel-collection_eng/Barcoding-database

Elle est disponible en libre accès et téléchargeable à :

<https://data.inra.fr/dataset.xhtml?persistentId=doi%3A10.15454%2FTOMBYZ>

Les photos des souches sont accessibles en fichiers zip disponibles à :

https://www6.inra.fr/carrtel-collection_eng/Photos

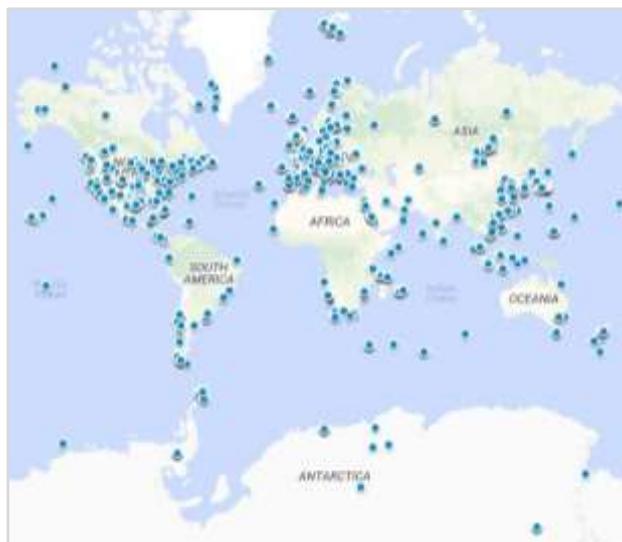
L'appui du projet SYNAQUA au développement de la base *diat.barcode* est cité à :

https://www6.inra.fr/carrtel-collection_eng/Barcoding-database/Fundings

Base de référence complétée à l'issue du projet SYNAQUA :

Dans le cadre de SYNAQUA, des séquençages complémentaires ont été réalisés (prestataire INRA Transfert), afin d'augmenter encore la complétion de cette base de référence en y intégrant des séquences issues d'échantillons environnementaux variés en utilisant la procédure de curation que nous avons développée (Rimet et al 2018). Ces résultats obtenus dans le cadre de SYNAQUA n'ont pu être intégrés qu'à l'issue du projet.

La version 9 de *Diat.barcode* contient désormais plus de 8 000 séquences ADN de référence représentant 1 400 espèces de diatomées et 300 genres, avec une large distribution des échantillons (carte ci-dessous). Cette nouvelle version, plus complète, permet d'améliorer encore l'efficacité de la technologie développée dans SYNAQUA.



La version 9 de *Diat.barcode* est désormais téléchargeable à :

<https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/TOMBYZ>

2- Base de référence « oligochètes »

Connaissances et références préalables :

Les oligochètes aquatiques comprennent un nombre important d'espèces allant de très sensibles à très résistantes aux pollutions chimiques (type toxique et eutrophisation). Ils sont utilisés comme indicateurs de la qualité biologique des sédiments fins/sableux des cours d'eau et des lacs ainsi que des sédiments grossiers et du milieu hyporhéique des cours d'eau. Les oligochètes sont également des descripteurs des échanges hydrologiques entre les eaux de surface et les eaux souterraines et donc du fonctionnement des cours d'eau. L'identification des oligochètes à l'espèce est difficile et possible que pour une partie des spécimens d'un échantillon. Une base de données des séquences du marqueur cytochrome c oxydase (COI) des oligochètes aquatiques sur la base de l'analyse de spécimens récoltés en Suisse a été initiée en 2013 et un seuil de variation génétique de 10% a été proposé pour distinguer les espèces (Vivien et al., 2015). Cette base de données a été utilisée pour tester la capacité de l'approche du séquençage à haut débit à identifier les espèces présentes dans un échantillon et déterminer leurs abondances (Vivien et al., 2016).

- VIVIEN, R., LAFONT, M. (2013) Diversité des oligochètes aquatiques dans la région genevoise (Suisse). *Revue suisse de Zoologie* 120, 161-173
- VIVIEN, R. (2013) Note sur la diversité des oligochètes aquatiques dans la région genevoise (Suisse). *Revue suisse de Zoologie* 120, 415-420
- VIVIEN R., LAFONT M. (2015) Faunistic note on the aquatic oligochaetes of the Geneva area and of Switzerland. *Revue Suisse de Zoologie* 122 (2) : 207-212
- VIVIEN, R., FERRARI, B.J.D., PAWLOWSKI, J. (2016) DNA barcoding of formalin fixed oligochaetes for biomonitoring. *BMC Research Notes* 9:342
- VIVIEN R., WYLER S., LAFONT M., PAWLOWSKI J. (2015) Molecular Barcoding of Aquatic Oligochaetes: Implications for Biomonitoring. *PLoS ONE* 10(4):e0125485.
<https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0125485&type=printable>
- VIVIEN R, LEJZEROWICZ F, PAWLOWSKI J (2016) Next-Generation Sequencing of Aquatic Oligochaetes: Comparison of Experimental Communities. *PLoS ONE* 11(2): e0148644
<https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0148644&type=printable>

Développement de bases de données lors du projet SYNAQUA :

La base de données COI a été complétée en séquençant des spécimens collectés dans diverses régions de Suisse. Un article sur cette mise à jour de la base de données a été publié (Vivien et al., 2017). De plus, une base de données du marqueur 16S (ADN ribosomal) a été créée en utilisant les extraits d'ADN de la collection COI. Par ailleurs, une méthode permettant de préserver à la fois la qualité de l'ADN et la densité et structure des communautés des oligochètes a été développée (Vivien et al., 2018). Cette méthode de préservation permettra de développer efficacement les bases de données.

Les bases de données COI et 16S sont encore en train d'être complétées. En effet, dans le cadre du projet SYNAQUA, deux indices différents basés sur le séquençage à haut débit sont développés. L'un est basé sur le séquençage d'ADN (COI, 16S) extrait du sédiment total alors que l'autre est basé sur le séquençage (COI) de spécimens préalablement triés et marqués génétiquement. Ce dernier indice permet d'obtenir une séquence COI pour chaque spécimen trié et donc de compléter en parallèle les bases de données (la partie antérieure des spécimens séquencés est conservée pour identification morphologique). Les bases de données COI et 16S finales seront disponibles avant juin 2019 et seront par la suite publiées.

- VIVIEN R., HOLZMANN M., WERNER I., PAWLOWSKI J., LAFONT M., FERRARI B.J.D (2017) Cytochrome c oxidase barcodes for aquatic oligochaete identification: development of a Swiss reference database. *PeerJ* 5:e4122; DOI 10.7717/peerj.4122

VIVIEN R, WERNER I, FERRARI BJD (2018) Simultaneous preservation of the DNA quality, the community composition and the density of freshwater oligochaetes for the development of genetically based biological indices. *PeerJ* 6:e6050; DOI 10.7717/peerj.6050

L'article Vivien et al. (2017) est en libre accès au lien suivant : <https://peerj.com/articles/4122.pdf>. Toutes les séquences présentées dans les illustrations de l'article sont disponibles en libre accès au lien suivant : <https://peerj.com/articles/4122/#supp-5>. Toutes les séquences (COI et ITS2) obtenues dans le cadre de ce travail sont disponibles en libre accès au lien suivant : <https://peerj.com/articles/4122/#supp-6>