

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 39635

WP2.2.1 : Présentation des indices de diatomées et d'oligochètes :
à l'intention des professionnels de bio-indication sur le territoire
transfrontalier

Responsables : UNIGE



Projet SYNAQUA – développement de l'indice moléculaire des diatomées et des indices moléculaires d'oligochètes

Développement de l'indice moléculaire des diatomées

Introduction

La législation actuelle suisse recommande d'utiliser les diatomées épilithiques pour évaluer l'état écologique des rivières et des cours d'eau. Les diatomées ont été sélectionnées car elles sont très sensibles aux conditions environnementales et réagissent rapidement aux changements physicochimiques et biologiques. Dans différents pays, divers indices biotiques ont été développés pour évaluer l'impact environnemental à l'aide des diatomées. La plupart de ces indices sont basés sur la fréquence relative des espèces, pondérées par leur valeur autoécologique.

En Suisse, l'indice suisse des diatomées (DI-CH) a été proposé afin de caractériser l'état biologique des cours d'eau en utilisant la fréquence et la distribution de plus de 400 espèces de diatomées et variétés morphologiques (Hürlimann & Niederhauser 2007). Le DI-CH est basé sur des paramètres chimiques indiquant une pollution anthropique et classe la qualité de l'eau en 5 classes écologiques différentes sur une échelle de 1 à 8, correspondant aux critères établis par la législation suisse. Traditionnellement, l'indice suisse des diatomées (DI-CH), comme les autres indices de diatomées, est calculé sur la base d'une analyse microscopique de la communauté de diatomées. Les diatomées sont isolées à partir d'échantillons de biofilms collectés et traités conformément aux descriptions de Hürlimann & Niederhauser 2007.

Dans les pays européens, les indices les plus couramment utilisés sont l'Indice Trophique des Diatomées (TDI, Kelly et Whitton, 1995), l'Indice de Polluosensibilité Spécifique (IPS, CEMAGREF, 1982) et l'Indice Biologique des Diatomées (IBD, Coste et al., 2009), tous acceptés par le règlement européen de la Directive-Cadre sur l'Eau (DCE).

Indices génétiques

Depuis 2010, plusieurs études ont été menées en Europe pour tester l'application du métabarcoding de l'ADN du biofilm afin d'évaluer la composition de la communauté des diatomées (ex : Kermarrec et al. 2013, 2014, Zimmermann et al. 2014, Visco et al. 2015, Rimet et al. 2016). Ces études ont entre autre testé différents marqueurs génétiques, complété la base de données de référence et analysé la corrélation entre les données morphologiques et moléculaires.

En Suisse, depuis 2013, l'indice moléculaire des diatomées (MDI-CH) est en cours d'élaboration. La première étude comparant les données moléculaires et morphologiques des communautés des diatomées dans le canton de Genève a donné des résultats très prometteurs (Visco et al. 2015). La suite de ces travaux a

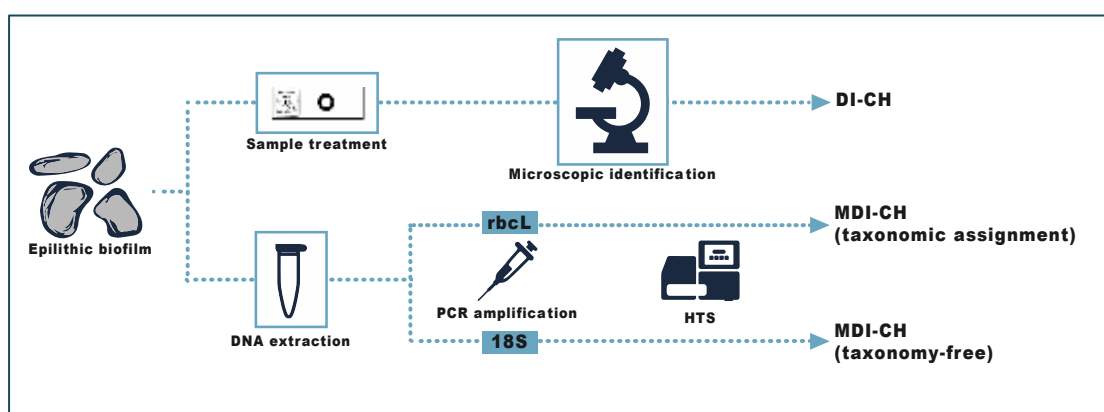
montré une forte amélioration de la corrélation entre les indices morphologiques et moléculaires en utilisant une approche dite « sans taxonomie » (Apothéloz-Perret-Gentil et al. 2017). Les études en cours étendent la portée géographique des sites d'échantillonnage afin de couvrir le territoire. Pour ce faire, des échantillons ont été récoltés lors des campagnes nationales NAWA ainsi que par le biais du programme européen INTERREG SYNAQUA.

Projet SYNAQUA

Le projet SYNAQUA a permis, comme mentionné ci-dessus, d'agrémenter la base d'échantillons d'ADN environnementaux disponibles pour le développement et la mise en place de l'indice moléculaire diatomée. De plus, grâce à la collaboration entre les parties suisse et française, les méthodes principales utilisées par les différents groupes de recherche pour la détection des communautés de diatomées ainsi que pour le calcul de l'indice ont pu être comparées (Apothéloz-Perret-Gentil et al, 2020).

Des échantillons prélevés sur lac et rivières des régions transfrontalières entre la France et la Suisse ont tout d'abord été séquencés sur les plateformes de séquençage de l'UNIGE et de l'INRA pendant le projet SYNAQUA. Ce premier lot d'échantillons a permis le développement des indices moléculaires diatomées. Pour compléter ce travail, un séquençage d'échantillons complémentaires d'origines géographique et environnementale variées a été réalisé dans le cadre de SYNAQUA en 2019 et 2020. Les bibliothèques ont été réalisées à l'INRA de Thonon (extraction ADN, amplification PCR, purification et pool des amplicons) et le séquençage sur la plateforme INRA Transfert (technologie Illumina MiSeq). Ces nouvelles données permettront 1) d'asseoir la méthodologie afin d'aller vers des protocoles standardisés et 2) d'élargir la base géographique et environnementale des échantillons afin d'améliorer la capacité prédictive des indices. Ces travaux sont en cours de valorisation.

Méthodologie



Résumé des étapes des différentes méthodes de calcul de l'indice des diatomées

Échantillonnage



Echantillonnage des diatomées

L'objectif du MDI-CH étant de fournir une évaluation de l'état écologique complémentaire et comparable à celle de l'indice morphologique, le matériel utilisé pour obtenir l'indice moléculaire est le même que pour la méthodologie traditionnelle. Les diatomées devraient être isolées du biofilm qui recouvre les pierres situées dans le courant de la rivière. Il est important de prendre des cailloux dans le courant (radier) et non dans de l'eau stagnante pour éviter l'accumulation d'ADN de cellules de

diatomées mortes qui pourrait fausser les prévisions. Vous trouverez plus d'informations sur le choix des substrats et la procédure de collecte dans le module Diatomées (Hürlimann & Niederhauser 2007).

Laboratoire moléculaire

Le biofilm des pierres doit être gratté et transféré dans un tube pré-rempli avec un tampon pour la préservation de l'ADN. Grâce à ce tampon, les échantillons peuvent être conservés jusqu'à plusieurs mois à -20°C . Une fois envoyés au laboratoire moléculaire, l'ADN des diatomées sera amplifié à l'aide d'amorces spécifiques et séquencé sur une plate-forme de séquençage à haut débit. Deux codes à barres génétiques sont actuellement utilisés pour le métabarcoding des diatomées: le gène chloroplastique *rbcl* et la région hypervariable V4 du gène nucléaire ribosomal 18S.

Analyse de données et calcul d'indice

Les séquences obtenues sont analysées à l'aide d'outils dédiés qui filtre les séquences de bonne qualité et les regroupe en fonction de leur similarité. Finalement, l'indice moléculaire peut être calculé selon la formule.

Deux approches sont actuellement disponibles pour calculer MDI-CH. La première approche est basée sur l'assignation taxonomique des séquences de diatomées, permettant de créer une liste d'espèces similaire à celle obtenue par analyse morphologique. Cette approche est particulièrement utile lorsque vous utilisez le gène marqueur *rbcl*, pour lequel une base de données de référence *Diat.barcode* comprenant un grand nombre d'espèces de diatomées est disponible (Rimet et al. 2016, Rimet et al. 2019).

Une approche alternative consiste à calculer le MDI-CH directement à partir des séquences de diatomée en contournant leur assignation taxonomique (Apothéloz-Perret-Gentil et al. 2017, Cordier et al. 2017). Cette approche a été développée pour le marqueur 18S V4. En revanche, la résolution taxonomique ainsi que la base de données de référence sont moins développées que pour le premier marqueur. Comparées à *rbcl*, les données 18S V4 sont plus simples à

obtenir, mais leur analyse nécessite un jeu de données d'apprentissage afin de connaître certaines compositions des communautés vivant dans la zone géographique.

Interprétation des données et limitations

Les deux méthodes utilisées pour le calcul des indices moléculaires ont leurs avantages et leurs inconvénients. L'indice basé sur *rbcL* peut être utilisé partout, à condition que la base de données de référence soit disponible pour assigner les séquences aux espèces. Certaines améliorations de cet indice ont été proposées ; en plaçant phylogénétiquement les séquences inconnues avec les morpho-espèces connues et en utilisant les valeurs écologiques des espèces proches (Keck et al. 2018). Une autre approche a été d'améliorer l'interprétation des données quantitatives, Vasselon et al. (2018) ont proposé un facteur de correction pour l'indice français IPS basé sur la biomasse de chaque espèce.

L'approche sans taxonomie est théoriquement plus sensible car elle inclut toutes les espèces de diatomées, pas seulement celles pouvant être attribuées sur le plan taxonomique. Cependant, la méthode nécessite un ensemble de données acquises à grande échelle et présentant un large gradient de qualité d'eau dans une vaste zone géographique. La valeur prédictive de l'indice est faible si la communauté de diatomées ne figure pas dans le jeu de données d'apprentissage.

Dans les deux cas, le pourcentage des séquences utilisées pour calculer l'indice peut être généré et utilisé pour ajouter un degré de confiance à la prédiction.

Mise en œuvre l'indice moléculaire des diatomées

Depuis 2018, le métabarcoding de l'ADN environnemental est utilisé pour le calcul de l'indice trophique de diatomée au Royaume-Uni (Kelly et al. 2018). Le TDI moléculaire est basé sur l'analyse des séquences de *rbcL*. La corrélation entre les qualités calculées avec l'indice morphologique et moléculaire est très forte. L'introduction de l'index moléculaire des diatomées est également l'un des objectifs du groupe de travail sur les diatomées associé à l'action COST Action DNAqua-net de l'UE (<https://dnaqua.net>). En Suisse, 400 échantillons ont été analysés à l'aide du marqueur 18S V4. La plupart des échantillons ont été collectés dans la région lémanique, en collaboration avec le Service d'écologie de l'eau du canton de Genève (SECOE).

.....

Développement des indices moléculaires d'oligochètes

Introduction



Echantillonnage
oligochètes

Les sédiments représentent une composante essentielle des écosystèmes fluviaux et lacustres. Ils ont de plus la propriété de stocker certains types de contaminants. Par exemple, les polluants hydrophobes tels que les métaux, PCB et PAH tendent à s'y accumuler. Ces substances peuvent atteindre des concentrations suffisantes pour induire des effets néfastes sur les organismes benthiques et ainsi perturber le bon fonctionnement de l'écosystème. Ces effets peuvent se produire non seulement à l'échelle locale, mais également à l'échelle de l'hydrosystème via le transport des sédiments et le relargage des contaminants. Par exemple, les sédiments contaminés peuvent être remobilisés lors d'épisodes de hautes eaux (Wolfram *et al.* 2012). Donc l'absence de prise en compte du

compartiment sédimentaire dans le cadre des programmes de surveillance peut conduire à des conclusions trop optimistes sur la qualité des cours d'eau (Lafont *et al.* 2010).

Il est donc crucial que la qualité de ce compartiment soit évaluée afin d'identifier s'il peut ou non contribuer à la dégradation d'un écosystème aquatique. Les effets des polluants présents dans les sédiments sur les organismes peuvent être évalués en appliquant des tests écotoxicologiques et/ou des indices biologiques. Divers indices biologiques basés sur l'étude de la structure des communautés d'oligochètes ont été développés pour l'évaluation de la qualité biologique des sédiments de cours d'eau et lac. Les oligochètes représentent de bons bioindicateurs de la qualité des sédiments. Ils sont inféodés à ce compartiment, se déplacent peu, et leur mode d'alimentation est principalement constitué par l'ingestion des sédiments fins. De plus, le groupe comprend un grand nombre d'espèces, allant de sensibles à très résistantes aux pollutions (Rodriguez & Reynoldson, 2011) et les oligochètes sont en règle générale abondants dans les sédiments (Vivien *et al.* 2014).

Parmi les méthodes oligochètes développées, l'indice oligochètes de bioindication des sédiments (IOBS) permet d'évaluer la qualité biologique des sédiments fins/sableux des cours d'eau (AFNOR 2016) et l'indice oligochètes de bioindication lacustre (indice IOBL, pourcentage de taxons sensibles) de décrire à la fois l'état de fonctionnement et la qualité biologique des sédiments (AFNOR 2016). L'indice IOBS est appliqué depuis dix ans en Suisse dans le cadre de programmes d'évaluation de la qualité des cours d'eau et de manière ponctuelle (Vivien *et al.*, 2014 ; 2015). De plus, les communautés d'oligochètes sont régulièrement étudiées en Suisse depuis plusieurs décennies pour évaluer la qualité des sédiments de lacs.

L'application des indices oligochètes requiert un haut niveau d'expertise. De plus, il est possible d'identifier à l'espèce qu'une partie des spécimens d'un échantillon. La plupart des espèces de la sous-famille Tubificinae et des familles Lumbriculidae et Enchytraeidae ne peuvent être identifiées à l'état immature et le pourcentage dans les échantillons de spécimens non identifiables à l'état immature est souvent élevé. Dans ce contexte, le développement d'un indice basé sur l'identification des oligochètes à l'aide de code-barres génétiques permettrait de résoudre les problèmes liés à l'identification des espèces et une plus large utilisation des oligochètes comme bioindicateurs.

Indices génétiques

Un projet de développement d'indices oligochètes génétiques a débuté en Suisse en 2013 et comprend trois axes de recherche :

1. Création de bases de données de référence des séquences de code-barres génétiques (cytochrome c oxydase I (COI) et 16S ADNr) des oligochètes aquatiques basée sur l'analyse de spécimens collectés en Suisse ; détermination de seuils de divergence génétique pour distinguer les espèces.
2. Développement d'approches de (e)DNA metabarcoding, i.e. basées sur l'analyse directe de l'ADN extrait du sédiment total ou de mélanges de spécimens, en utilisant les marqueurs COI et 16S
3. Développement d'indices basés sur le séquençage à haut débit de spécimens marqués génétiquement (high-throughput DNA barcoding).

Projet SYNAQUA

Le projet SYNAQUA a permis de développer significativement chacun des trois axes décrits ci-dessus et de valider l'approche basée sur le séquençage à haut débit de spécimens marqués génétiquement.

Notre base de données des séquences COI initiée en 2013 (Vivien et al., 2015 ; 2017), a été complétée (Vivien et al., 2019) et une base de données des séquences 16S a été établie en analysant les extraits d'ADN de notre collection. Le seuil de divergence génétique de COI de 10 % déterminé pour distinguer les espèces (excepté les espèces du genre *Nais* (8%) (Vivien al., 2017) a de plus été validé. Concernant 16S, un seuil de divergence génétique de 5% a été proposé pour les espèces de toutes les familles/sous-familles à l'exception de celles des *Naidinae* (1%).

Des travaux effectués avant le projet SYNAQUA sur des échantillons de cours d'eau ont montré que, malgré des différences significatives obtenues entre les approches morphologique et moléculaire concernant la présence/absence de taxons et les abondances des taxons, le développement d'approches d'(e)DNA metabarcoding étaient possibles moyennant des adaptations du calcul de l'indice et des délimitations des classes de qualité (Vivien et al., 2016; 2019). Dans le cadre du projet SYNAQUA, des approches de eDNA metabarcoding (analyse du sédiment total) ont été testées sur 10 sites de cours d'eau et 7 sites de lac en utilisant deux marqueurs différents (COI et 16S). Concernant les échantillons de

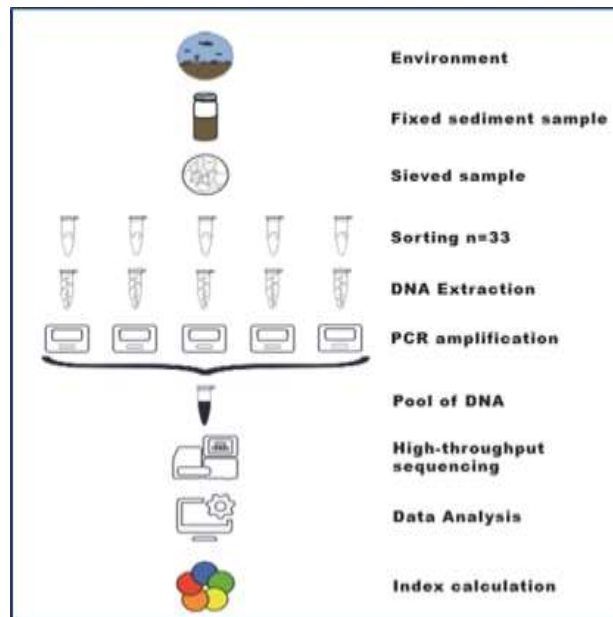
cours d'eau, les diagnostics de qualité biologique établis avec les approches morphologique et moléculaire ont pu être globalement concordants grâce à une adaptation du calcul de l'indice. Par contre, une partie des échantillons lacustres présentaient d'importantes discordances entre les deux types d'approches. Un manque d'homogénéisation des échantillons lacustres et/ou une spécificité insuffisante des primers utilisés ont pu entre autres expliquer ces résultats. L'acquisition de données supplémentaires et des modifications/adaptations des méthodes (choix des primer, etc.) seront nécessaires pour valider des indices basés sur ce type d'approche.

La méthode basée sur le séquençage à haut débit de spécimens marqués génétiquement présente l'avantage, par rapport au (e)DNA metabarcoding, d'estimer de manière fiable et précise les abondances des espèces d'un site, étant donné que chaque spécimen est marqué avant le séquençage. Nous avons tout d'abord mis au point une méthode de traitement des échantillons permettant de préserver simultanément la qualité de l'ADN, la composition des communautés et les abondances des oligochètes aquatiques (utilisation du formol tamponné à pH neutre, Vivien et al., 2018). Cette étude préliminaire était nécessaire pour pouvoir développer des méthodes basées sur le séquençage de spécimens.

La méthode basée sur le séquençage à haut débit de spécimens marqués génétiquement a été appliquée sur 13 sites de cours d'eau et 7 sites de lac. 33 ou 66 spécimens ont été séquencés par site (33 sur 11 sites et 66 sur 9 sites) (Vivien et al., 2019). Au niveau de tous les sites étudiés (cours d'eau et lac), les diagnostics de qualité biologique établis avec les approches morphologiques et moléculaires étaient concordants. De plus, les résultats ont montré que le séquençage de 33 spécimens par site était suffisant pour une évaluation correcte de la qualité biologique des sédiments. Toutefois, nous recommandons de séquencer 45-50 spécimens par site, pour optimiser la détection des espèces et l'évaluation des abondances des espèces. Pour que cette approche puisse remplacer l'approche morphologique, il est nécessaire de la rendre applicable en routine en facilitant et raccourcissant certaines étapes de l'analyse. Parallèlement, nous rédigerons un guide méthodologique que nous proposerons à l'Organisation internationale de normalisation pour l'élaboration d'une norme ISO.

Etapes de la méthode (validée) basée sur le séquençage à haut débit de spécimens marqués génétiquement

- ▶ Les sédiments ont été prélevés à l'aide d'un filet de type Surber (0,2mm de vide de maille) (cours d'eau) ou d'une benne (lac)
- ▶ Le matériel biologique a été fixé avec du formol à pH neutre (tamponné) 20% (concentration finale de formaldéhyde de 4%), puis, de retour au laboratoire, a été conservé à 4°C pendant quelques jours jusqu'au tamisage (Vivien et al., 2018).
- ▶ Les sédiments ont été tamisés sur une colonne de tamis de 5mm et 0,5mm de vide de maille. Le matériel retenu dans le tamis de 0,5mm a été transféré dans une cuve de sous-échantillonnage. Le contenu de cases choisies au hasard a été examiné sous une loupe binoculaire et les oligochètes ont été extraits à l'aide d'une pince fine. 33 ou 66 spécimens ont été extraits par site.



Résumé des étapes de la méthode basée sur le séquençage à haut débit de spécimens marqués génétiquement

- ▶ Un fragment de chacun des 33 ou 66 spécimens a été placé dans un tube (1 tube par spécimen, soit 33 ou 66 tubes par site).
- ▶ L'ADN de chaque spécimen a été extrait (méthode guanidine) et un fragment du gène COI (313 pb) a été amplifié par PCR à l'aide de primers marqués génétiquement (chaque spécimen correspondant à un marquage spécifique). Ce marquage permet, une fois le séquençage à haut débit effectué, de retrouver la séquence correspondant à chaque spécimen (spécimens numérotés de 1 à 33 ou de 1 à 66).
- ▶ Le produit d'amplification de chaque spécimen d'un site a été transféré dans un seul tube (un tube par site). Puis, pour chaque site, une librairie a été préparée avec les produits d'amplification des spécimens en vue du séquençage à haut débit. Chaque librairie a été marquée génétiquement pour pouvoir, une fois le séquençage à haut débit effectué, trier les séquences obtenues par site.
- ▶ Une fois le séquençage à haut débit effectué (Illumina), les séquences brutes ont été triées par librairie (ou site) et par numéro de spécimen.
- ▶ Les séquences (correspondant aux spécimens analysés) ont été assignées (à l'espèce si possible, ou au genre ou à la sous-famille) au moyen de la base de données COI suisse, de Genbank, en construisant des arbres phylogénétiques et par l'analyse morphologique (le fragment antérieur d'un certain nombre de spécimens par site a été conservé pour une éventuelle identification morphologique ultérieure).
- ▶ Une liste d'espèces/taxons a ainsi été obtenue pour chaque site, avec un nombre d'individus pour chaque espèce/taxon.
- ▶ Les indices génétiques ont ensuite été calculés de la même manière que les indices basés sur l'analyse morphologique (calculs identiques).