

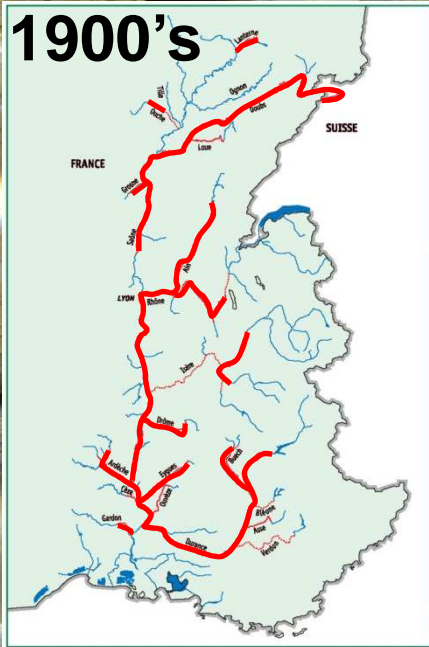
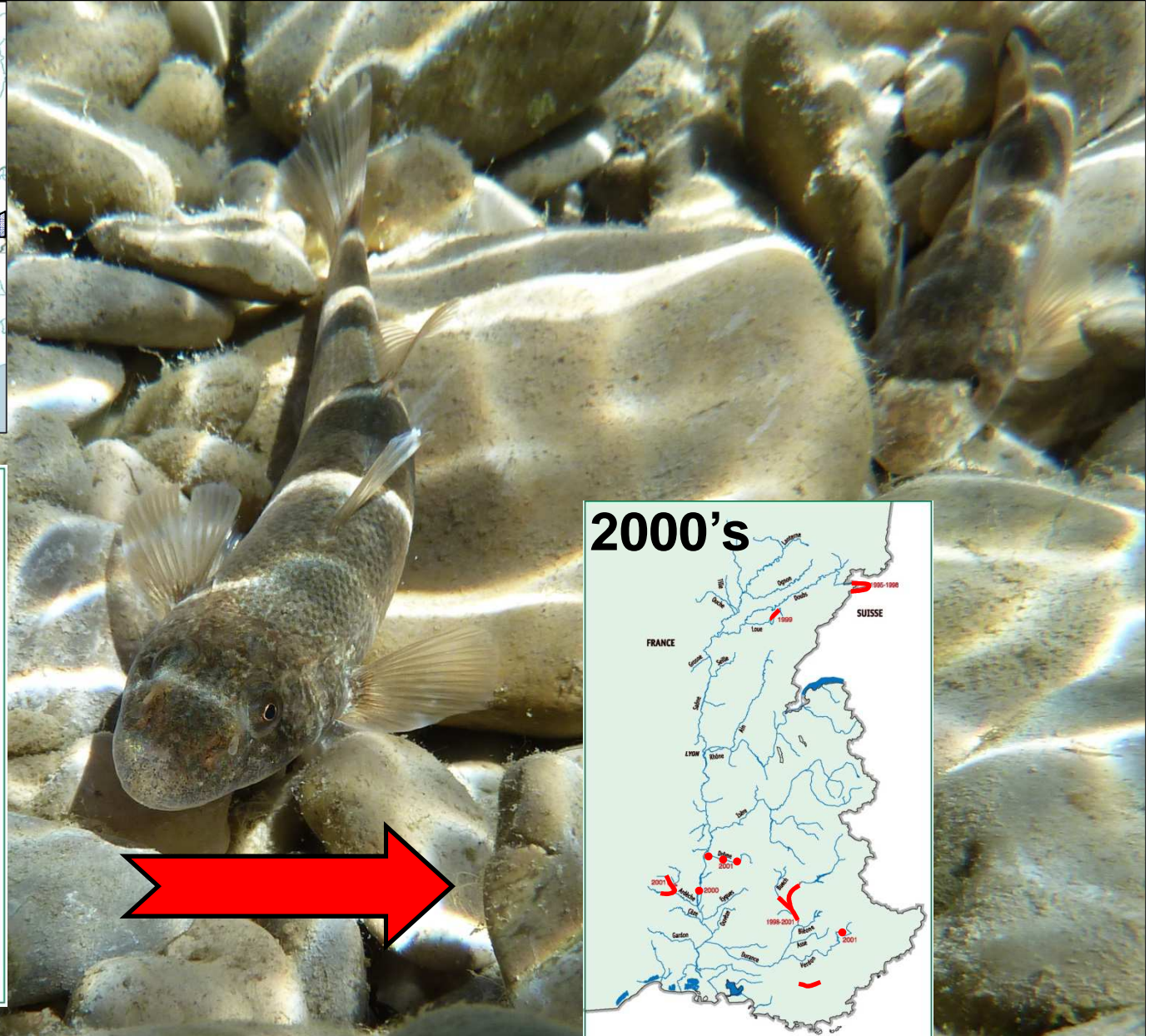
Atelier métagénomique / métabarcoding – Campus St-Charles
25 novembre 2016

Etude de régime alimentaire dans les milieux aquatiques: Barcoding alimentaire de l'apron (*Zingel asper*)

Vincent DUBUT, Emese MEGLE CZ, Emmanuel CORSE



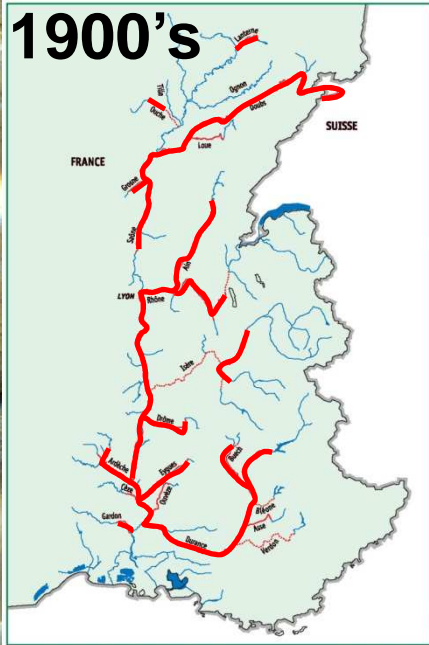
Le modèle de l'étude: L'apron du Rhône (*Zingel asper* L.)



Le modèle de l'étude: L'apron du Rhône (*Zingel asper* L.)



- 1- Life I (1998-2001)
- 2- Life II (2004-2009)
- 3- Plan National d'Action (PNA) 2012-2016



Objectifs

- Plasticité trophique ?
- Sélection des proies ?
- Utilisation de l'habitat ?

Elargir connaissances
biologiques de
l'espèce: plasticité
écologique

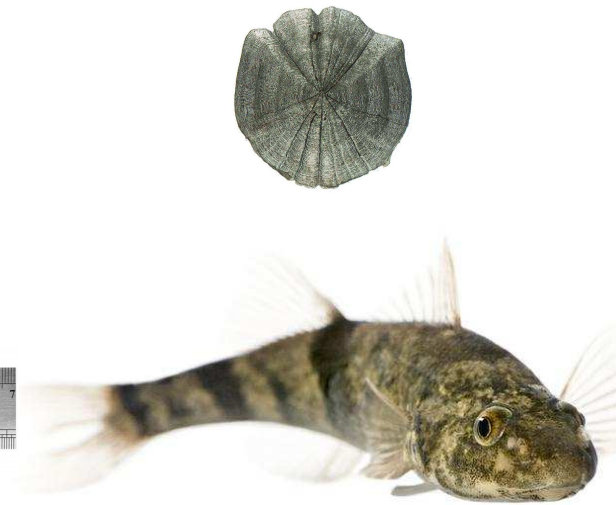
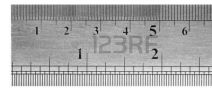
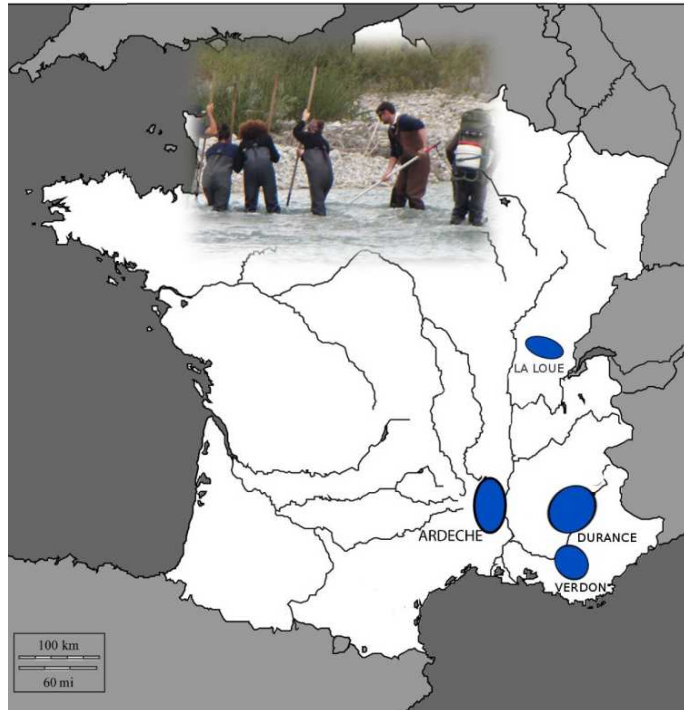


Isoler les facteurs
limitant le
développement des
populations



Aide à la gestion

Méthodologie - Terrain



Méthodologie - Terrain



Echantillonnage qualitatif
(base de donnée ADN)

+

Echantillonnage quantitatif
(90 points / site / campagnes)

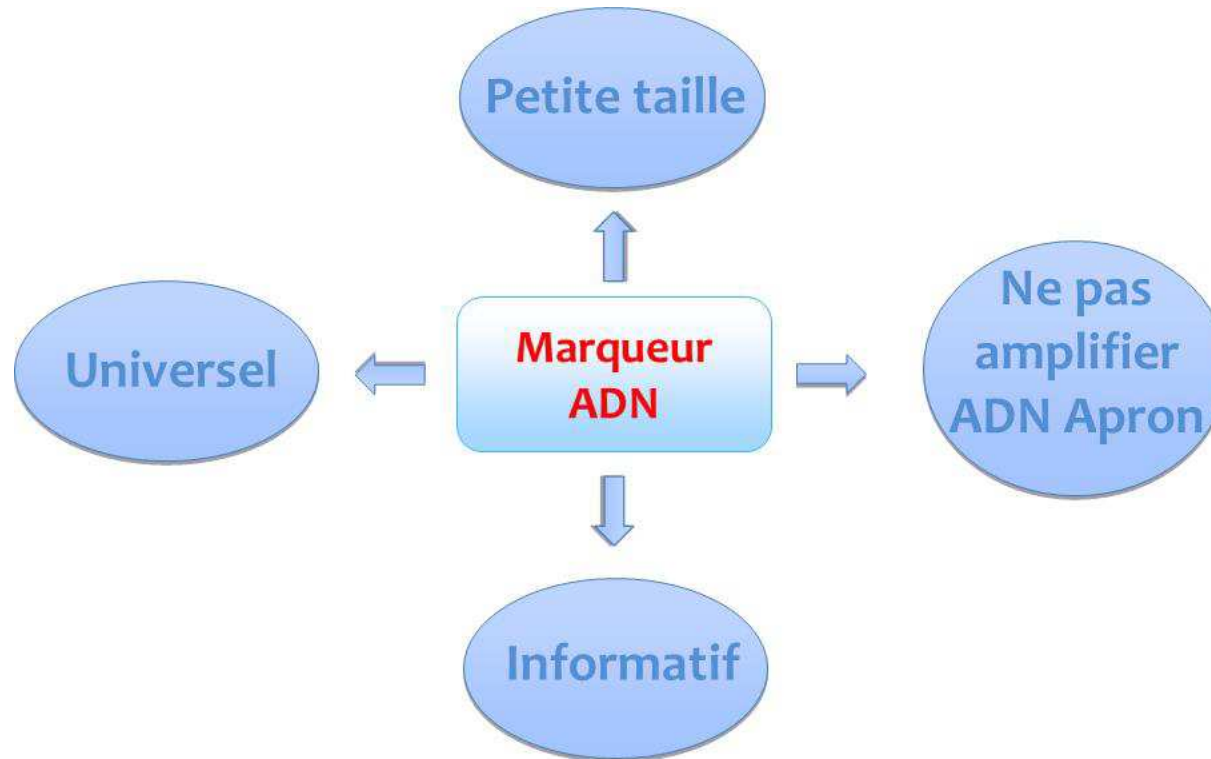
Les défis de metabarcoding alimentaire

- ADN dégradé
- Couverture taxonomique
- Identification taxonomique des séquences
- Elimination des artéfacts :
 - Erreur de séquençage
 - Erreur de PCR + Mis-tagging
 - Contaminations

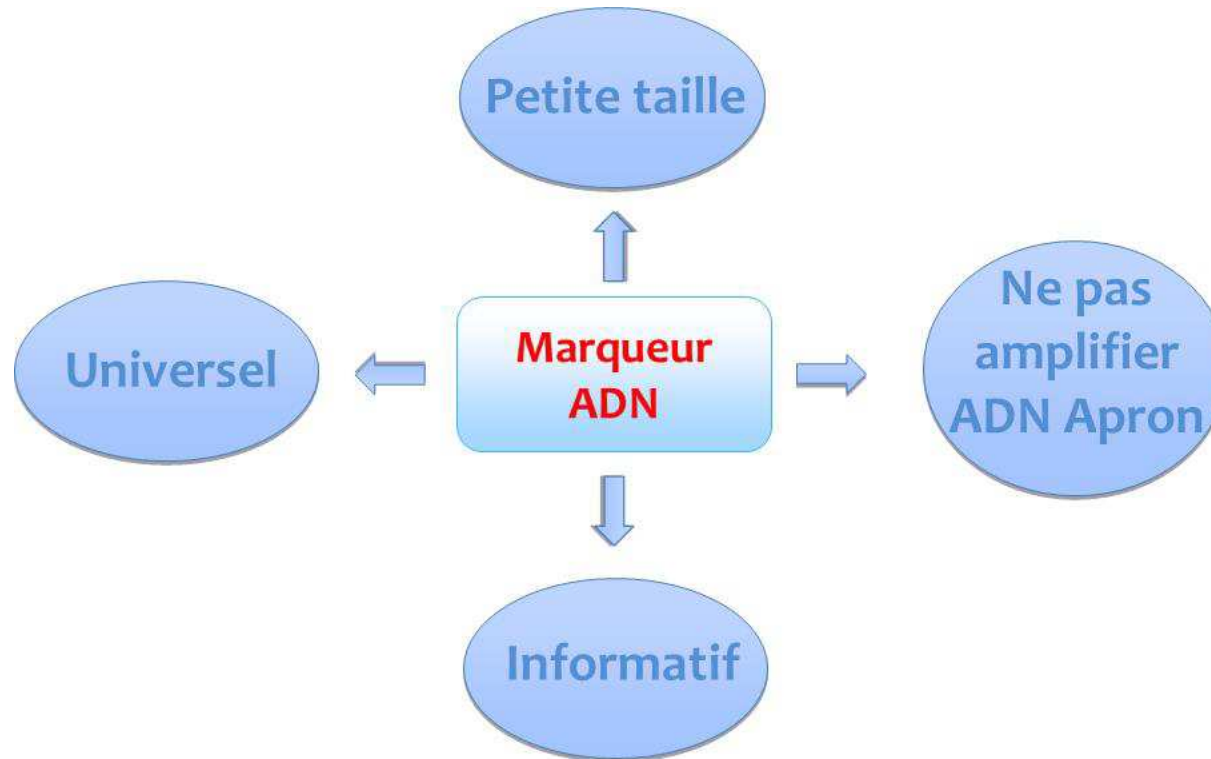
Les défis de metabarcoding alimentaire

- ADN dégradé (Choix des amorces)
- Couverture taxonomique (Choix des amorces)
- Identification taxonomique des séquences (Bases de données + Méthodes d'assignation)
- Elimination des artéfacts : (Bonnes pratiques + Témoins + Bioinformatique)
 - Erreur de séquençage
 - Erreur de PCR + Mis-tagging
 - Contaminations

Enrichissement par PCR : choix des amorces



Enrichissement par PCR : choix des amorces



De 2013 à 2015, 2 marqueurs ADN (MF/ZR et ZF/ZR)

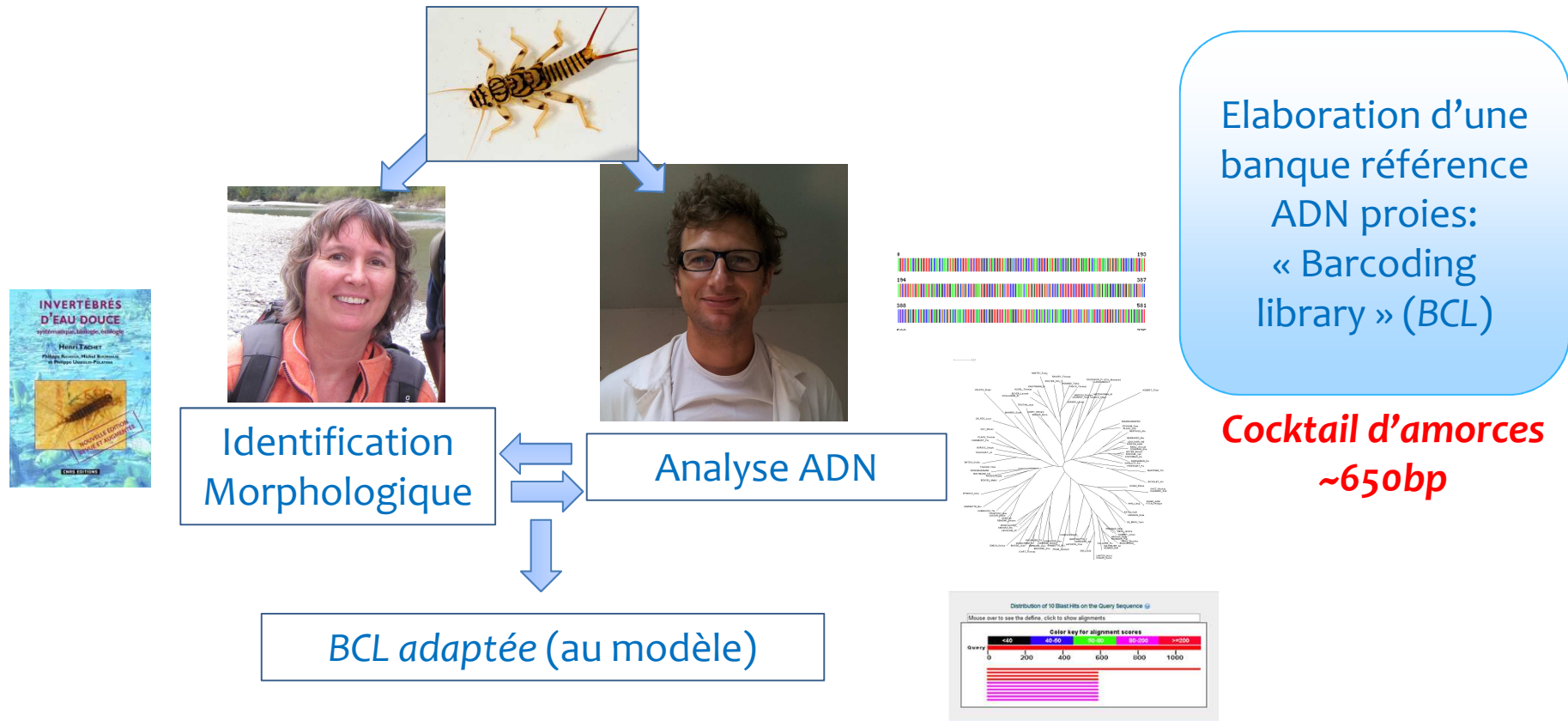
- ➡ Amplification de 96% des organismes d'intérêt
MAIS certaines proies d'intérêt mal détectées (Perlidés, Gammares)
- ➡ Développement d'un nouveau marqueur en 2015 : LF/CR

Banque ADN des proies

Séquences préexistantes (GenBank + BOLD)

+

Echantillonnage qualitatif



Design du run MiSeq

35 excréments Aprons

5 excréments Gobie

2 T+ avec ADN connus

1 Tneg PCR

2 Tneg extraction

2 Tneg pailleasse

1 Ttag

3 réplicas

3 couples d'amorces: ZFZR, MinifZR, Rbcl

1 Run MiSeq

Protocole pré-run MiSeq

- 1- Extraction plateforme ADN dégradé (2*24)
- 2- Dosage au Qubit
- 3- Dilution (max 20ng/μl) et mise en plaque des ADN
- 4- PCR
- 5- Dépôts d'un aliquot sur gel d'agarose.
- 6- Poolage des produits de PCR par réplica en fonction de l'intensité (3 niveaux)
- 7- Migration sur gel, découpage de la bande d'intérêt et purification
- 8- Préparation de la banque 9 index (1 par replica)
- 9- Bioanalyser
- 10- Cirad dosage qPCR
- 11- Run Miseq

Pipeline bioinformatique

- De Barba et al (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. (Mol.Ecol.Res. 14:306-323)
 - Utilisation de répliques
 - Elimination des bruits de faible intensité (LFN: Low Frequency Noise)
 - Elimination des erreurs de séquençage
 - Comparaison des répliques
- ObiTools (Boyer et al. 2015)

Analyse des séquences

Nomenclature

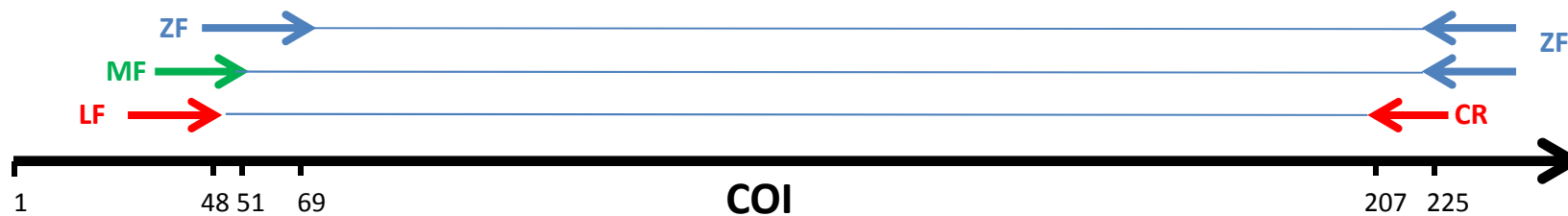
- Échantillons
 - 40 échantillons de faeces d'un individu
 - 35 Aprons
 - 5 Gobies
 - 2 témoins positifs (Tpos)
 - 6 témoins négatifs (Tneg)
- 3 répliques pour chaque échantillon (3 x 48)
- 3 marqueurs (3x3x48 => 432 PCRs)

	Tpos1	Tpos2
<i>Chironomus riparius</i>	+	+
<i>Dinocras cephalotes</i>	+	+
<i>Planorbarius corneus</i>		+
<i>Eisenia andrei</i>	+	
<i>Gammarus pulex</i>		+
<i>Hydropsyche instabilis</i>		+
<i>Hydropsyche modesta</i>	+	+
<i>Oligoneuriella rhenana</i>	+	
<i>Phoxinus cf phoxinus</i>	+	
<i>Serratella ignita</i>	+	+
<i>Velia saulii</i>		+
<i>Zingel asper</i>	+	+

Analyse des séquences

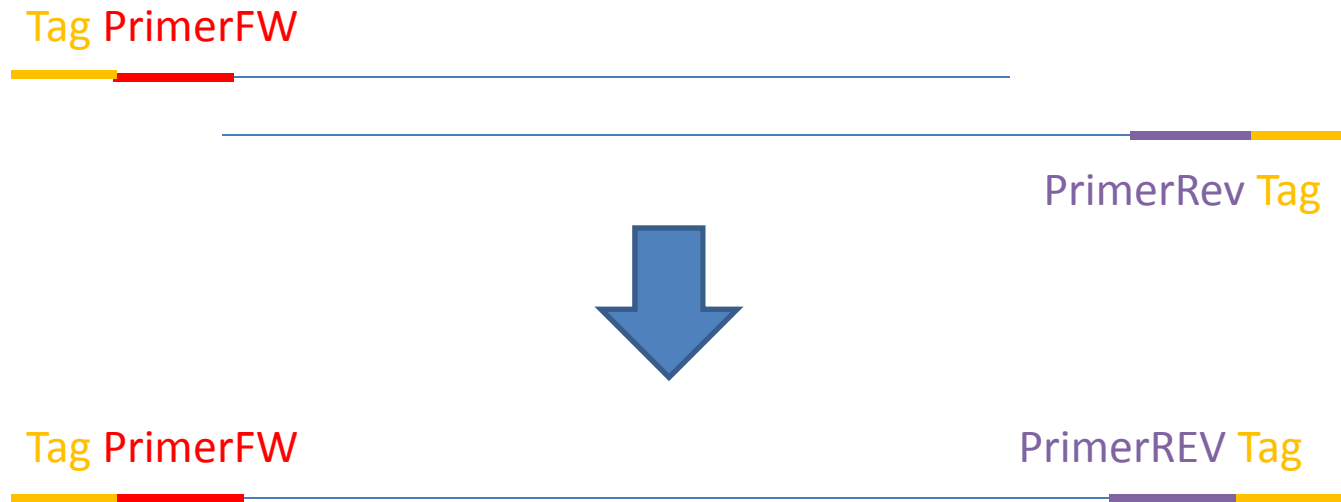
Nomenclature

- Marqueurs
 - LF/CR
 - MF/ZR
 - ZF/ZR



Assemblage des paires de reads

- Merge reads (PEAR, Zhang et al., 2014)



Assignment des reads aux réplicas

- Assign reads to replicate

Tag5 PrimerFW

PrimerREV Tag11

#Tag	1	2	3	4	5	6	7	8	11	12	16	17
1	14Mon01	14Mon10	14Ben07	14Cro06	14Cro15	Tneg PCR	14Mon01	14Mon10	14Ben07	14Cro06	14Cro15	Tneg PCR
2	14Mon02	14Deo01	14Ben08	14Cro07	14Cro16	Tpos1	14Mon02	14Deo01	14Ben08	14Cro07	14Cro16	Tpos1
3	14Mon03	14Deo02	14Ben09	14Cro09	14Cro17	Text1	14Mon03	14Deo02	14Ben09	14Cro09	14Cro17	Text1
4	14Mon04	14Deo03	14Ben06	14Cro10	P1	Text2	14Mon04	14Deo03	14Ben06	14Cro10	P1	Text2
5	14Mon05	14Deo04	14Cro01	14Cro11	P2	Tpos2	14Mon05	14Deo04	14Cro01	14Cro11	P2	Tpos2
6	14Mon06	14Ben01	14Cro02	14Cro12	P3	Tpai1	14Mon06	14Ben01	14Cro02	14Cro12	P3	Tpai1
7	14Mon07	14Ben02	14Cro03	14Cro13	P4	Tpai2	14Mon07	14Ben02	14Cro03	14Cro13	P4	Tpai2
13	14Mon09	14Ben05	14Cro04	14Cro14	P5	Ttag	14Mon09	14Ben05	14Cro04	14Cro14	P5	Ttag
	ZFZR						MiniFZR					

Variants

- Assignation des reads aux réplicas
- Elimination des amorces/tags
- Regrouper les reads identiques en **variants**
 - Comptage du nombre de reads de chaque variant dans chaque réplica
 - 28 000 variants = 3 Millions reads

Filtrage des variants

Vérification des marqueurs

- BLAST variants contre base de données de COI
 - Min coverage: 80%
 - E-value: 1e-10
 - COI db
 - 1 séquence per genre de proie potentielle '**macro-invertébré**'
 - All COI sequenced by EGE (macro-invertebrates)
 - All COI sequences in nt NCBI for **Rotifera**
 - All COI sequences in nt NCBI for **Diatoms** (Bacillariophyta)
 - 1 sequence per class of **Chlorophyta**
- **28 000 variants => 27 000**

Elimination du bruit (Low Frequency Noise ;LFN)

- $N_{\text{réplica-variant}}$
 - nombre de reads du variant dans le réplica
- $N_{\text{réplica}}$
 - nombre total des reads dans le réplica
- N_{variant}
 - nombre total des reads du variant

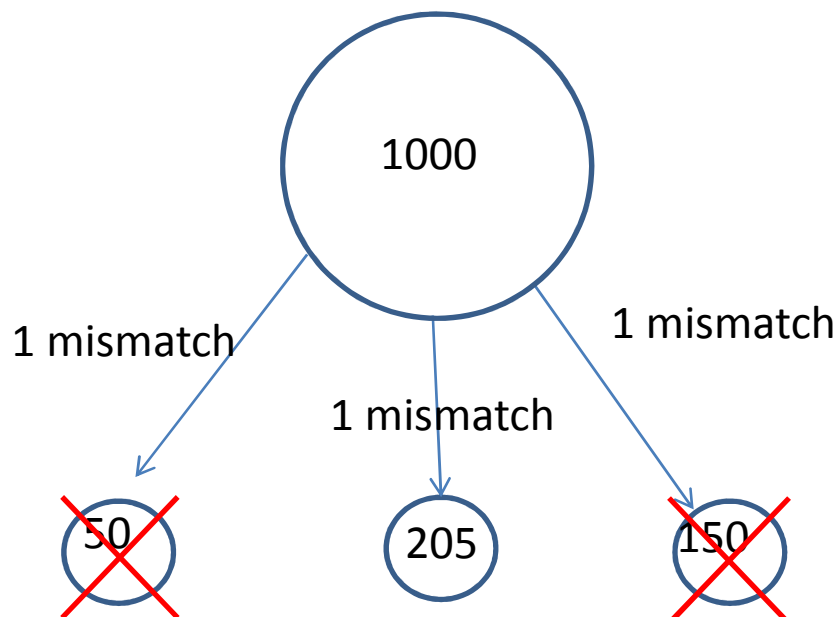
Élimination du bruit (Low Frequency Noise ; LFN)

- LFN_{Tpos}
 - $N_{\text{réplica-variant}} / N_{\text{réplica}} < 0,001$
 - Éliminer les variants peu représentés dans le réplica (contamination, erreurs de séquençage)
- LFN_{Tneg}
 - $N_{\text{réplica-variant}} < 20$
 - Éliminer les variants peu représentés même si peu de reads dans réplica (contamination)
- $LFN_{TagJump}$
 - $N_{\text{réplica-variant}} / N_{\text{variant}} < 0,0045$
 - Éliminer les variants peu représentés dans le replica par rapport aux nombres totaux des reads des variants (Mistagging)
- **3M reads => 2.3 M**
- **27 000 variants => 214**

Cohérence entre répliques (1)

- Eliminer les variants qui ne sont pas présents dans au moins 2 répliques de l'échantillon
- **2.3M reads => 2.3 M**
- **214 variants => 135**

Obiclean



- Diminuer les erreurs de séquençage/PCR
- $r=0,2$ (basé sur Tpos)
- **2,3 M reads => 2,3 M**
- **135 variants => 92**

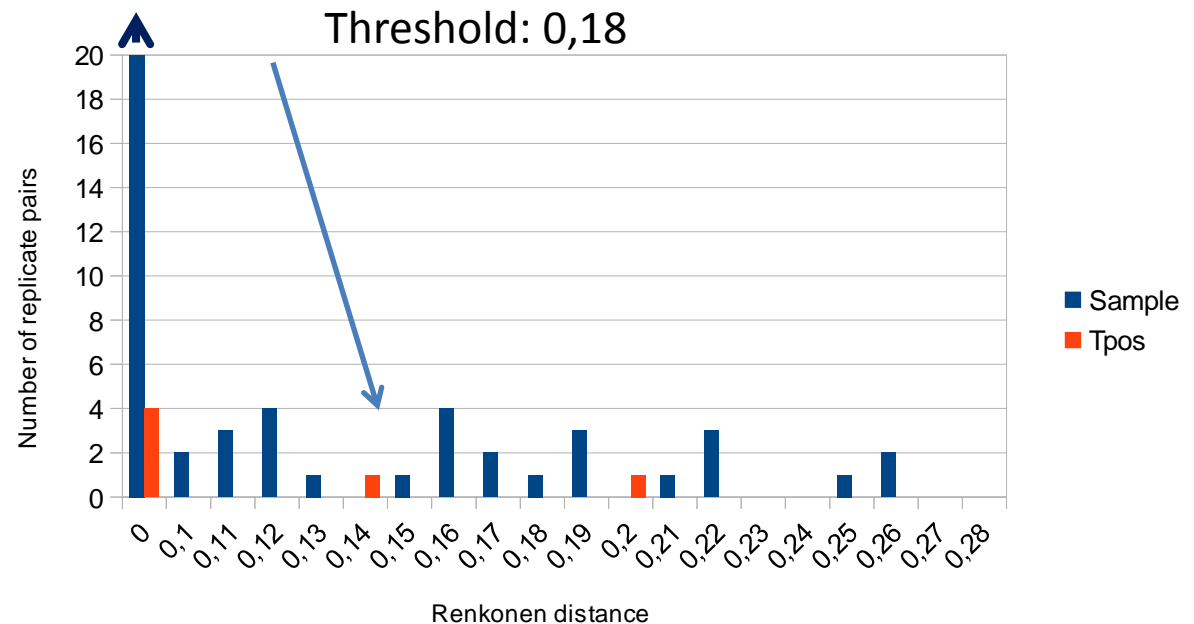
Renkonen index

- Indice de similarité entre 2 répliques (1,2) de même échantillon

$$p_{i1} = \frac{n_{variant_i replicate_1}}{n_{replicate_1}}$$

$$- p_{renkonen_{1,2}} = \text{summ}(\min(p_{i1}, p_{i2}))$$

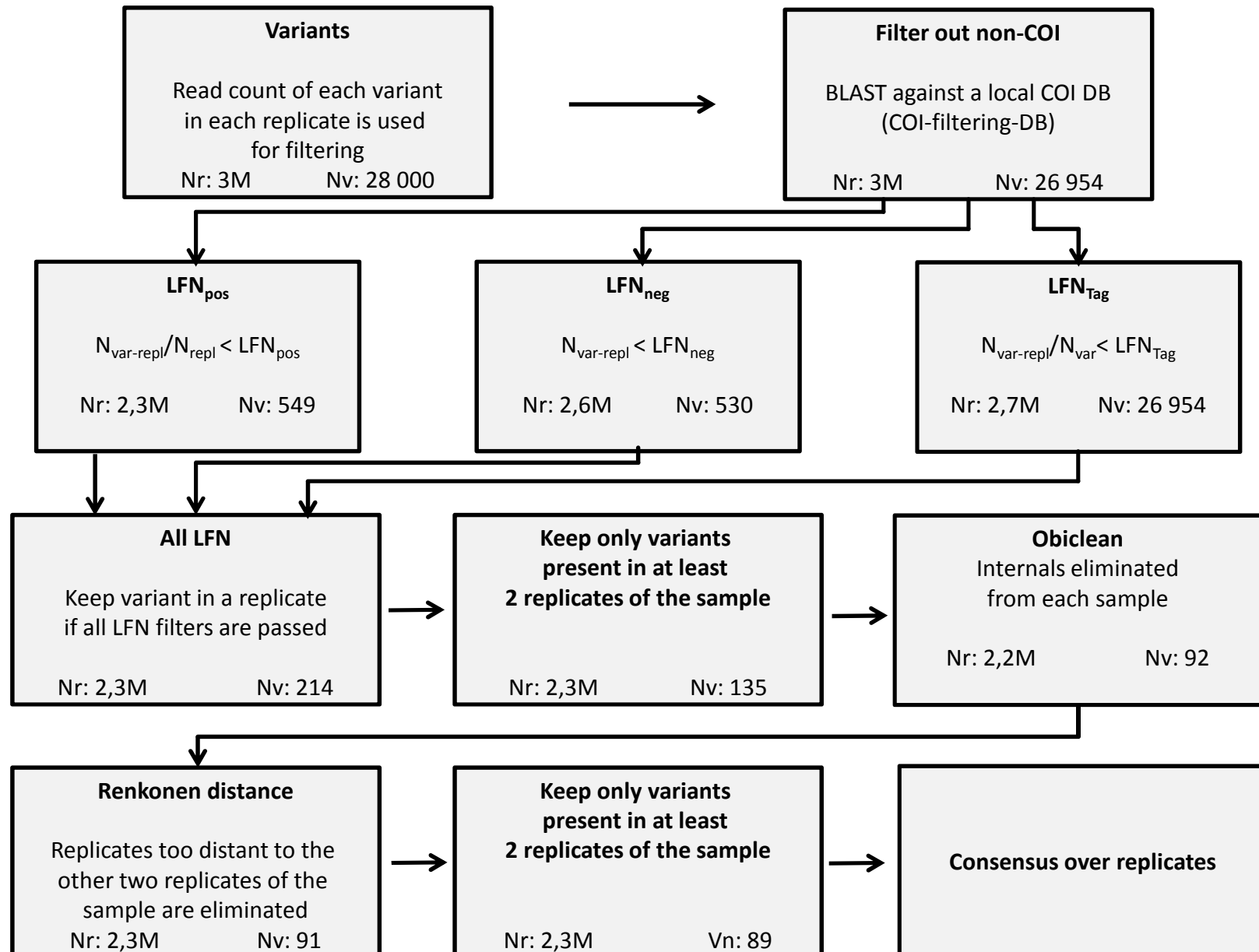
- Calcul de $p_{renkonen}$ chaque pair de répliques



Répétabilité, Consensus

- Eliminer les répliques distants des autres répliques de l'échantillon (distance de Renkonen)
- Eliminer les échantillons avec un seul réplique
- Eliminer les variants présents dans un seul réplique de l'échantillon
- **92 variants => 89**
- **40 échantillons apron + gobie => 38**
- Consensus pour chaque réplique

Filtrage des variants par marqueur

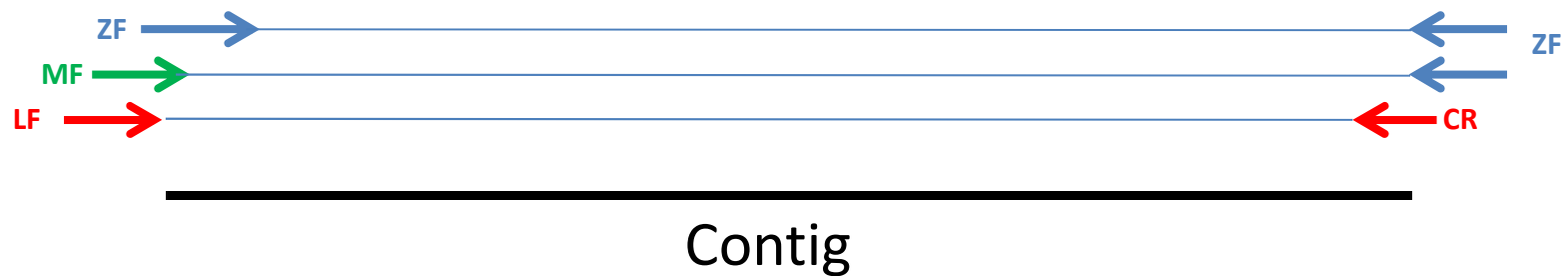


Étapes supplémentaires

- Élimination des chimères
 - Tpos.
 - Uchime, Uparse
- Détection des pseudogènes
 - Par longueur des séquences
 - Présence de codon STOP

Contigs

- 3 marqueurs par échantillons => séquence contig

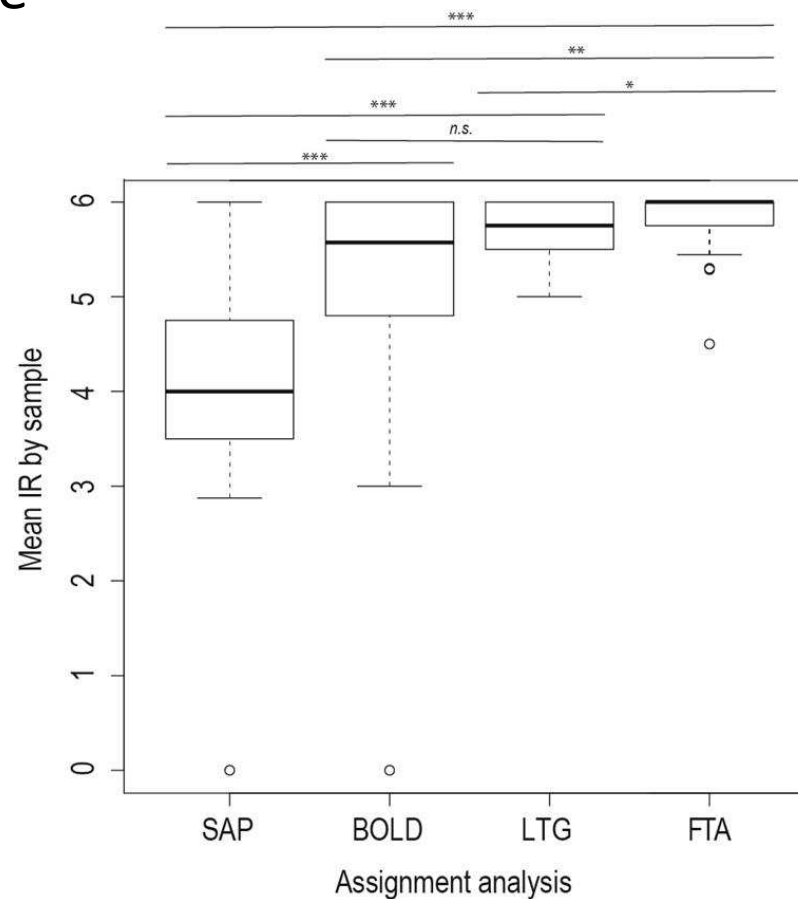


Assignation taxonomique automatique

- BLAST contre BDD locale
 - NCBI nt
 - BOLD publique
 - Séquences privées
- Retenir hits, si
 - % similarité > 100 à 70%
 - Coverture > 90%
 - L'organisme est déterminé au moins au niveau famille
- LTG: Lowest Taxonomique Groupe qui contient au moins 90% des hits
- Retenir LTG avec le meilleur % de similarité (si %similarité < 97%, LTG doit être basé sur au moins 3 taxa)

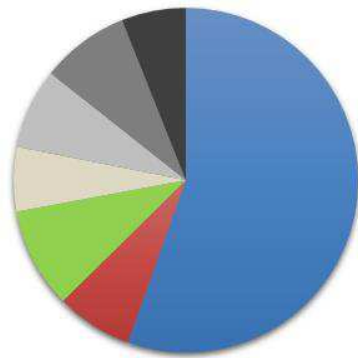
Assignation ajustée

- SAP + LTG + Assignation manuelle par BOLD
- Si incongruence => analyse phylogénétique + information biogéographique

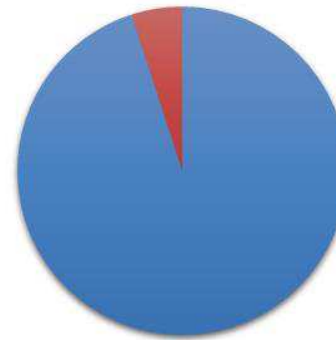


Résolution taxonomique

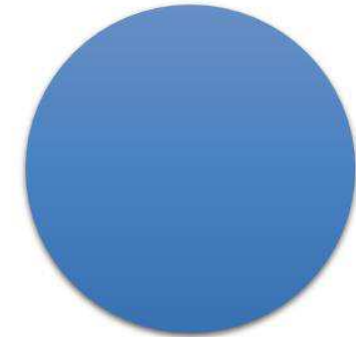
Tout organisme



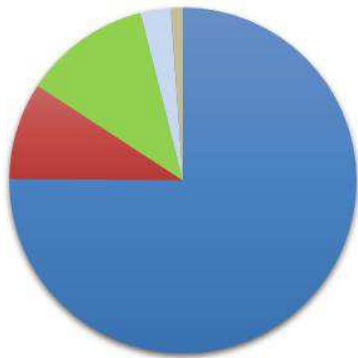
Trichoptères



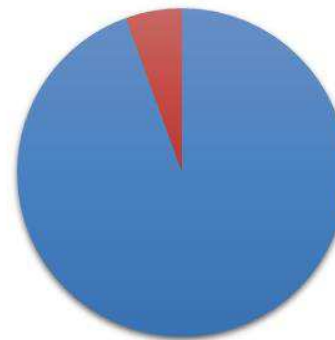
Plécoptères



Macroinvertébrés / Vertébrés



Ephémères

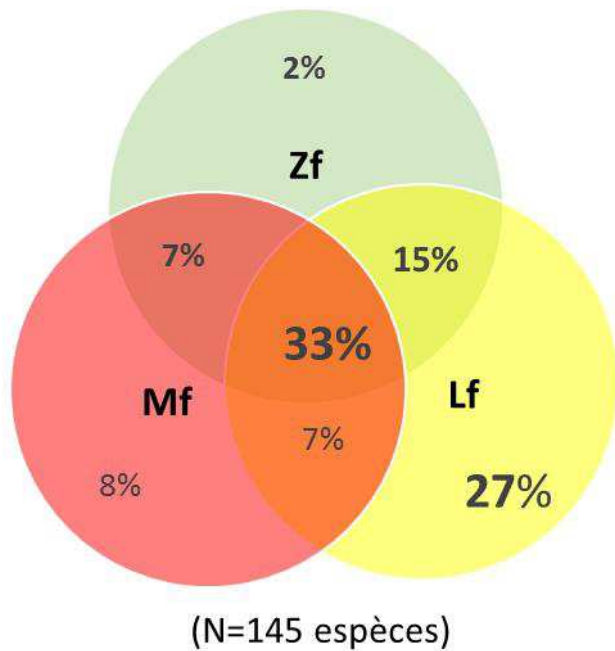


Diptères

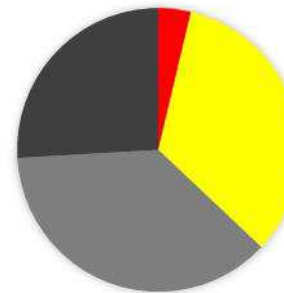


Complémentarité des marqueurs

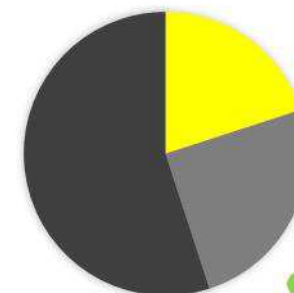
Rendement des marqueurs
sur les macro-invertébrés / vertébrés



DIPTERES



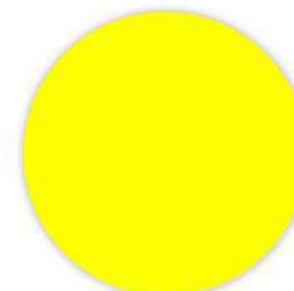
EPHEMERES



TRICHOPTERES



PLECOPTERES



- Zf
- Mf
- Lf
- 2 marqueurs
- 3 marqueurs

Nombre minimum de proies (NMI)

Filtration stringente



Séquence précise de chaque proie

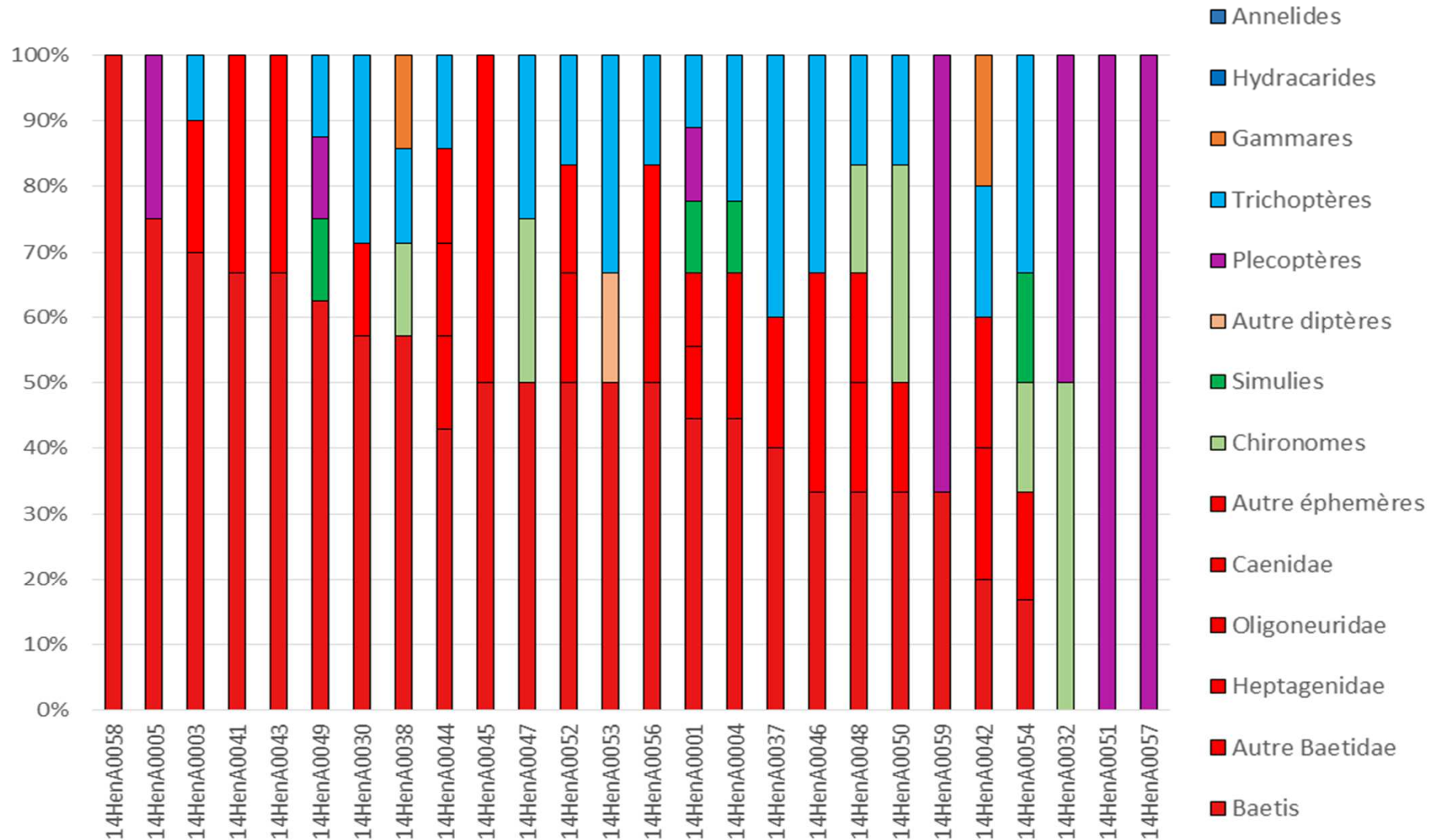


Nombre des variants d'une proie d'un même échantillons indique le nombre minimal de proies consommées par le prédateur

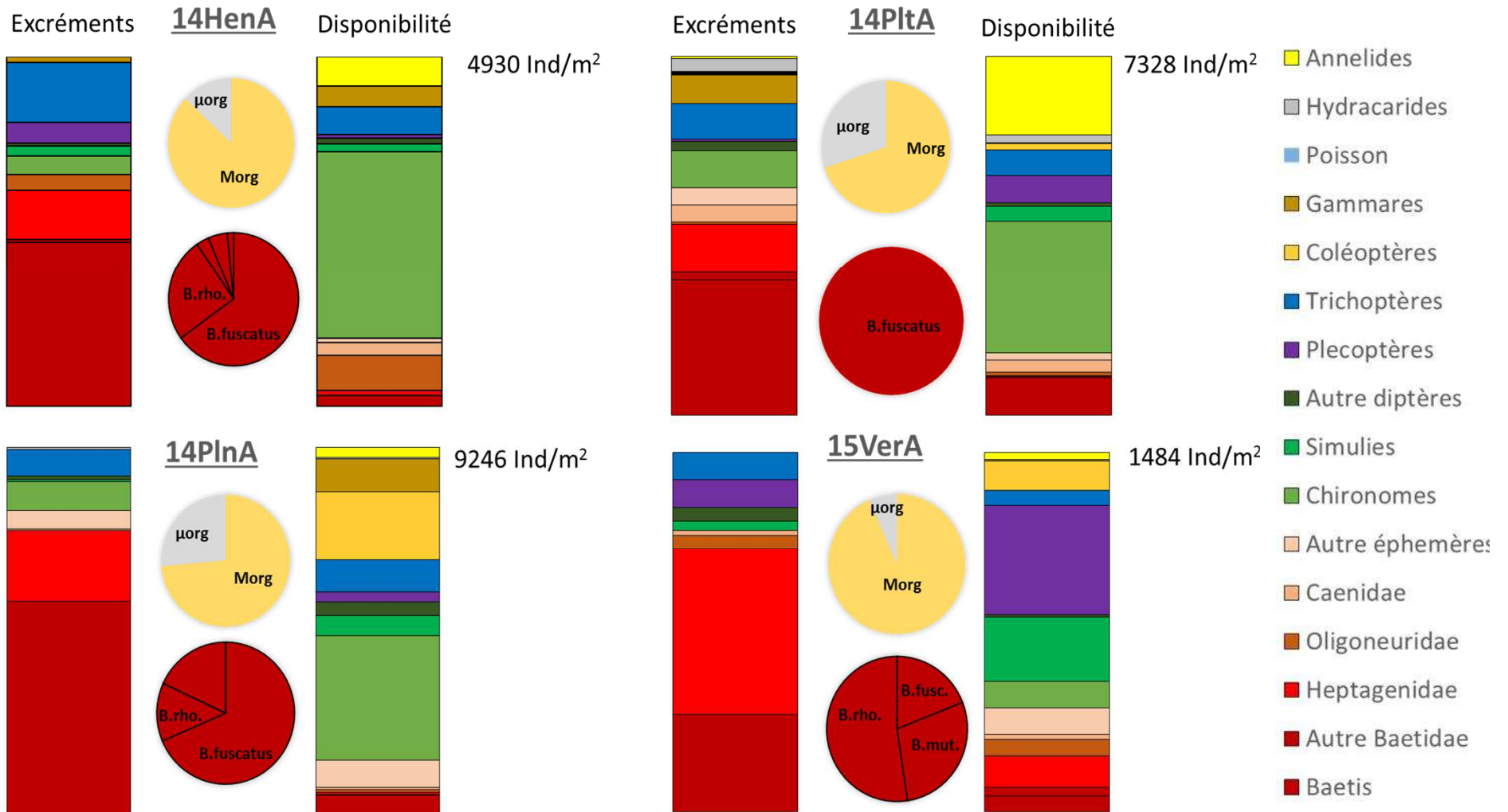


Approche semi-quantitative

Résultats: Occurrence des proies par individu

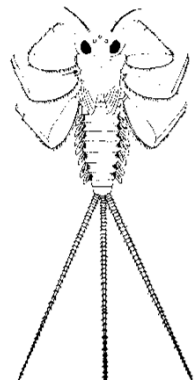
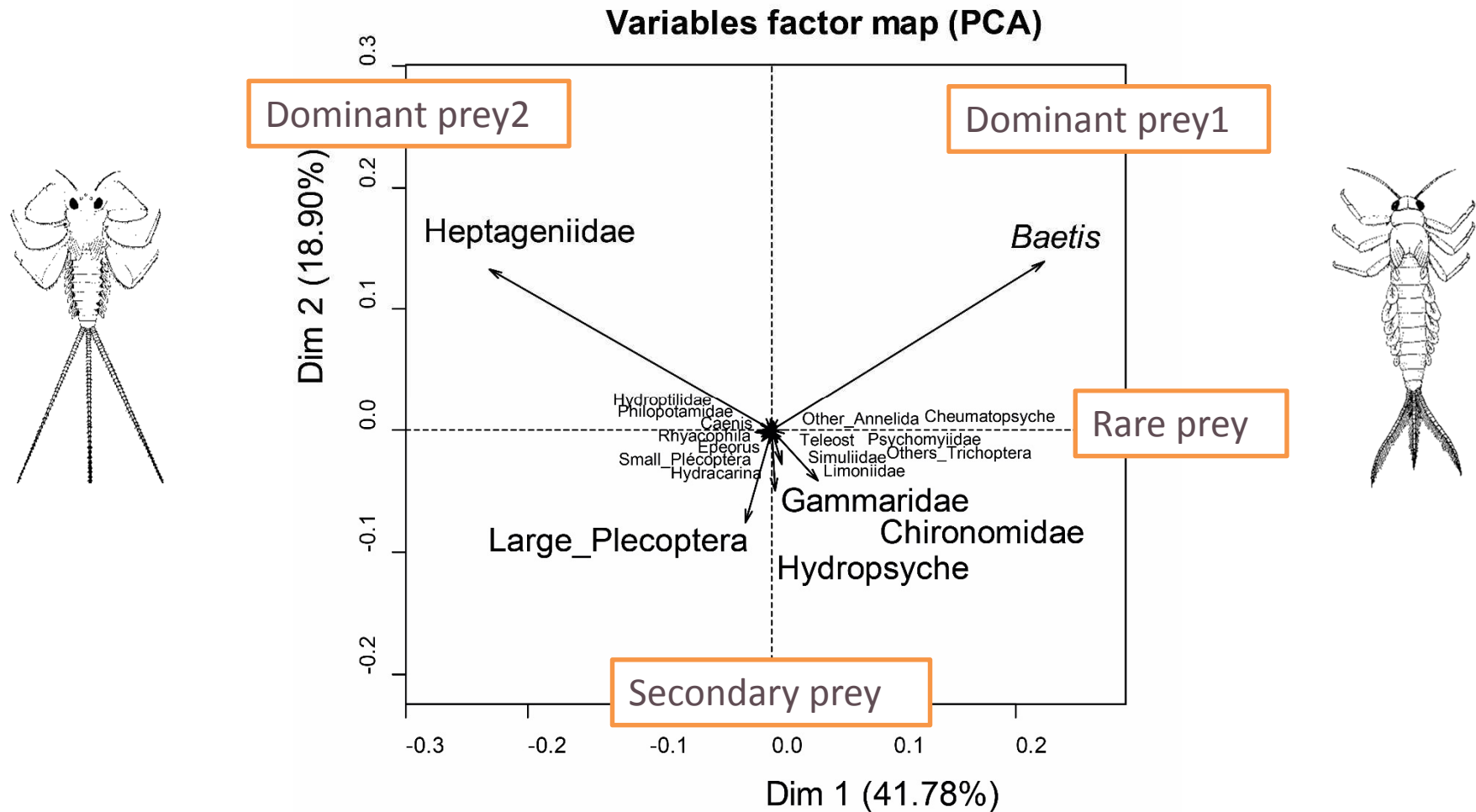


Variabilités spatiales



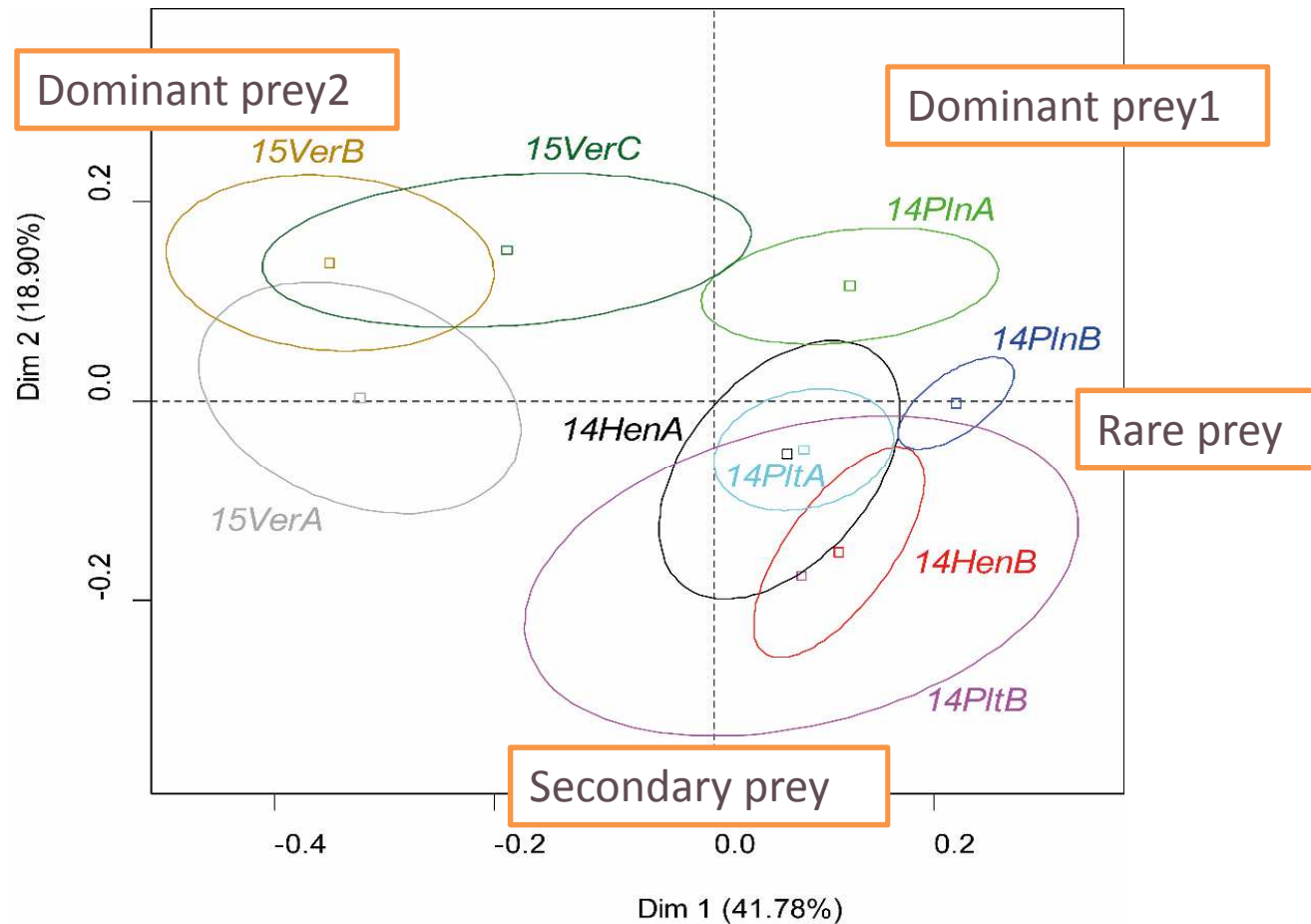
Plasticité trophique

(De Crespin de Billy *et al.*, *Journal of Fish Biology* 2000)



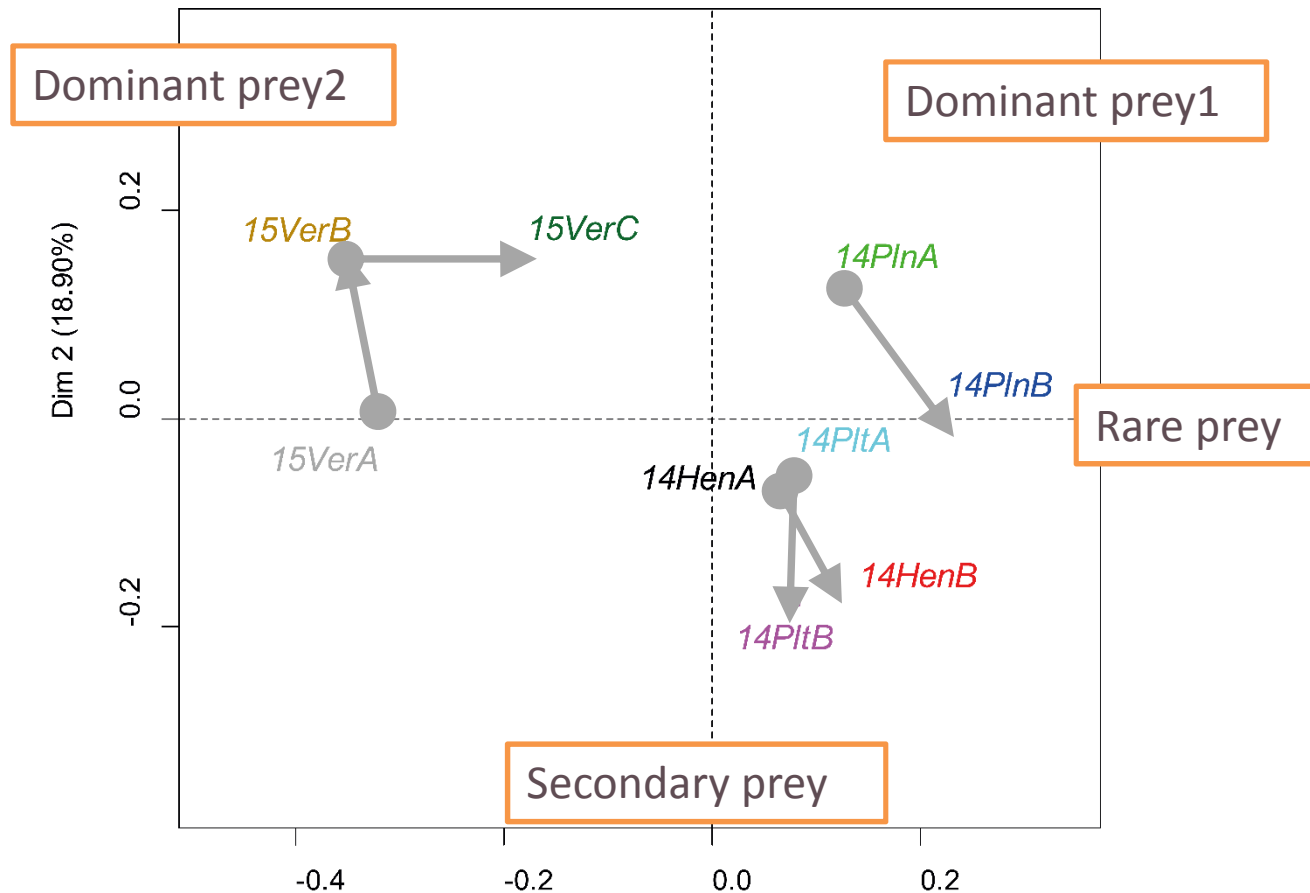
- 2 proies dominantes : *Baetis* + Heptageniidae
- Proies secondaires : *Hydropsyche*, Plecoptera, Chironomidae, Gammaridae

Plasticité trophique



- 2 patrons spatiaux: 1) Proie dominante 1 (*Baetis*) + proies secondaires ou rares
- 2) Proie dominante 2 (Heptageniidae) dans le Verdon

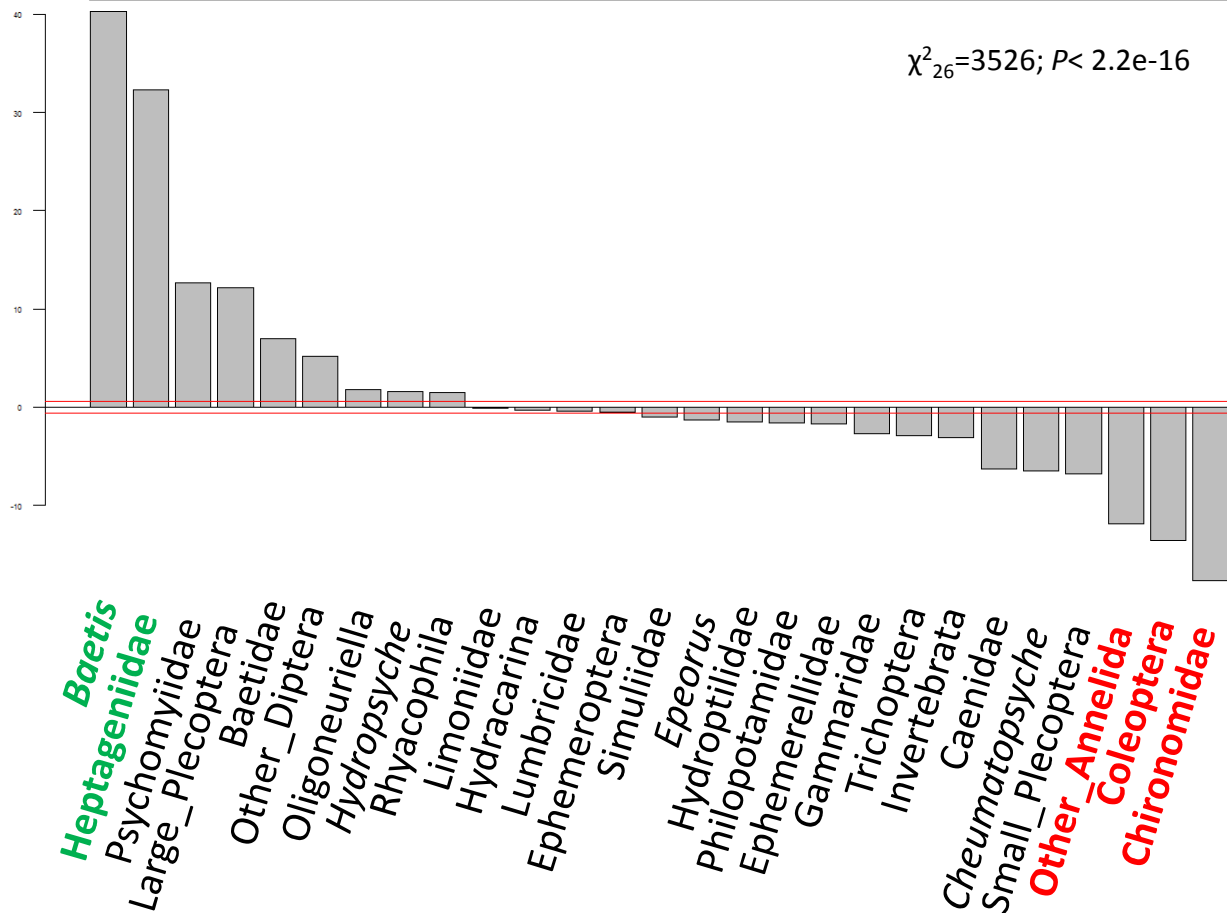
Plasticité trophique



- **2 patrons spatiaux:** 1) Proie dominante 1 (*Baetis*) + proies secondaires ou rares
2) Proie dominante 2 (*Heptageniidae*) dans le Verdon
- **Variations temporelle :** augmentation des proies secondaires en automne

Sélection des proies

χ^2 test; H_0 : prey are ingested according their abundance



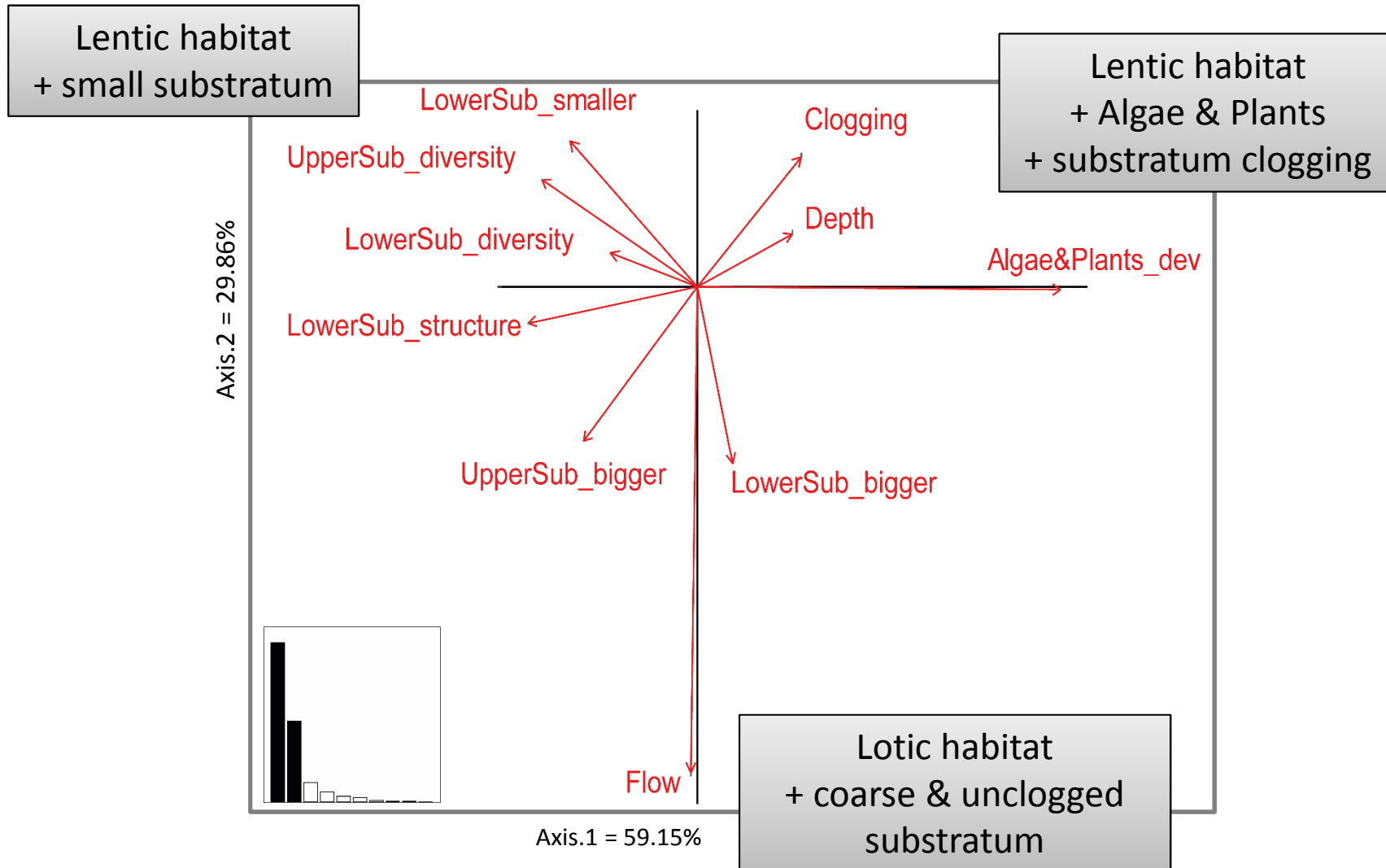
Selection+

Threshold under H_0

Selection -

- Les proies ne sont pas ingérées en fonction de leur abondance
- Les proies dominantes (*Baetis* et Heptageniidae) sont sélectionnées positivement

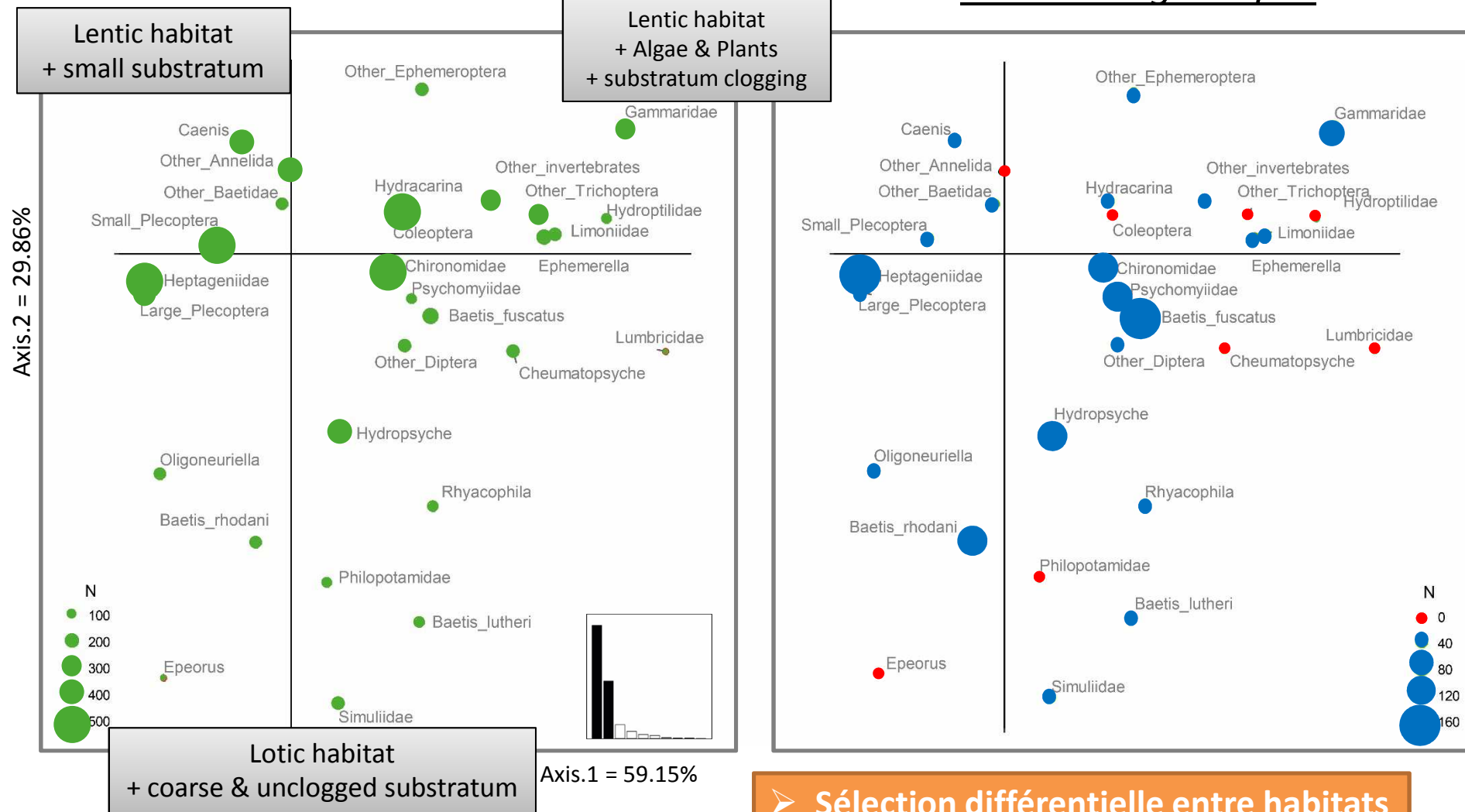
Utilisation de l'habitat



Utilisation de l'habitat

Invertebrates community

Diet of *Zingel asper*



- Sélection différentielle entre habitats
- Sélection différentielle intra-habitat

Conclusions

Que mange l'apron

- Proies dominantes = Ephemeroptera (*Baetis* + Heptageniidae)
- Proies secondaires (ex: Hydropsyche, Plecoptera, Chironomidae, Gammaridae)
- Sélection positive (ex. *Baetis*) et négative (ex. Chironomidae)

Où va-t-il les chercher:

- Habitats lotiques, substrat grossier, non colmaté

Remerciements



Mathilde Bertrand, Jean-Pierre Balmain, Pierre Gibert, Pierre Favrioux, Maxime Logez