

La bioinformatique une aide précieuse pour la paléoprotéomique

Fabrice BRAY

Miniaturisation pour la Synthèse, l'Analyse & la Protéomique UAR 3290

Journée JOLIBIO, Jeudi 9 Novembre 2023

Université de Lille, FST



Introduction

Datation

Âge de l'échantillon

Mesure des rapports isotopiques

Alimentation, climat

Analyse d'ADN

Reconstruction
phylogénétique

Analyse des protéines

Identification de l'espèce,

Conservation,

Modifications,

Identification du sexe,

Phylogénie

Morphologie

Identification du taxon



Mâchoire d'ours des cavernes

Ursus speleus



Introduction

Mesure des rapports isotopiques
Prélèvement et préparation
spécifique

Datation

Méthode plus destructive que la
protéomique

ADN

Limitation avec la
dégradation et les
contaminants

Analyse des protéines
Besoin d'améliorer de la
préparation à l'analyse

Morphologie

Petits fragments
Outils



Problématiques

- Identification d'espèces éteintes et des outils via la morphologie:

Outils en os fabriqués par les Hommes préhistoriques, brûlés et petits fragments



Outils en os de la cave de Scladina



Os brûlés



Petits fragments

- Reconstruction phylogénétique

- Analyse d'un grand nombre d'échantillons par protéomique:

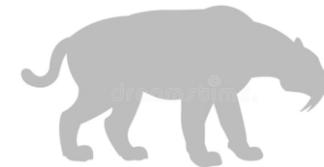
Plusieurs milliers de fragments sur les sites

- Analyse d'échantillons précieux dans les musées:

Prélèvement soumis à autorisation

- Identification du sexe des animaux:

L'Homme de Neandertal mangeait plus d'Aurochs femelle ou mâle?

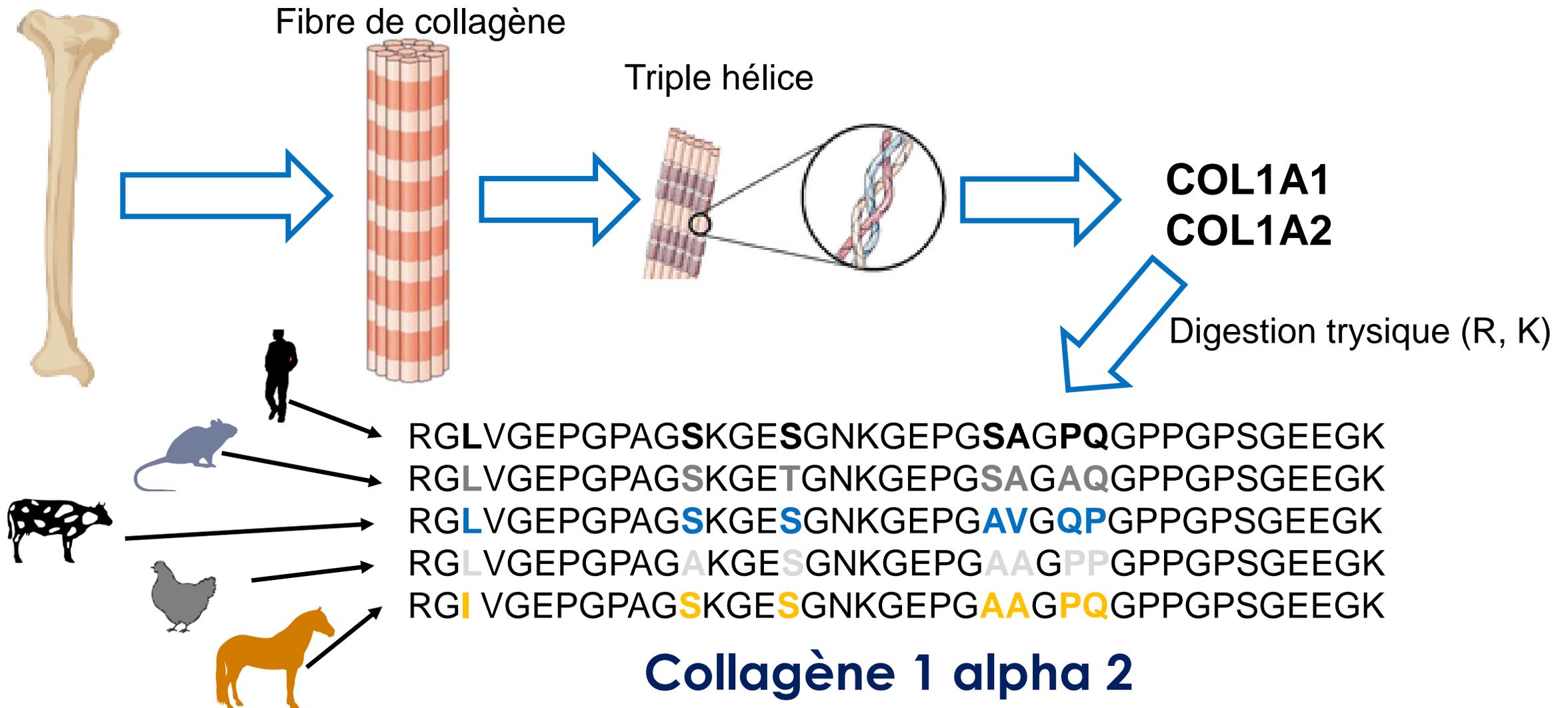


Saber-toothed cat
(*Homotherium latidens*)



Cave bear
(*Ursus spelaeus*)

Identification des taxons (famille, genre, espèces) à l'aide de marqueurs

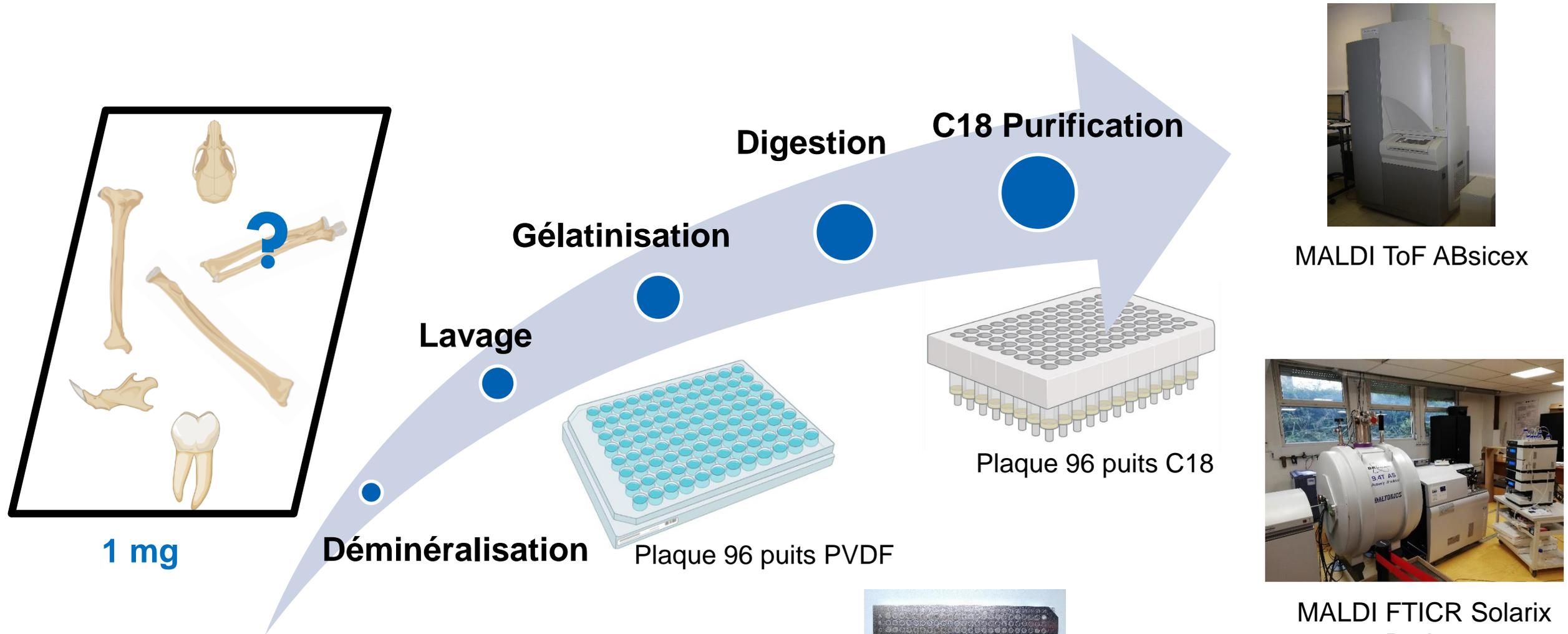


Buckley, M., Collins, M., Thomas-Oates, J., & Wilson, J. C. (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, 23(23), 3843-3854.

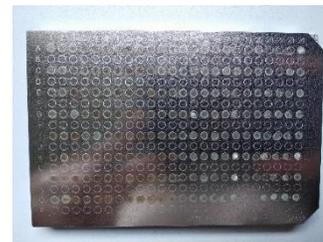
Base de donnée des peptides ZooMS

Species	Sample age	P1	A	A'	B	C	P2	D	E	F	F'	G	G'
<i>Bos primigenius</i>	Modern	1105.6	1192.7	1208.7	1427.7	1580.8	1648.8	2131.1	2792.3	2853.4	2869.4	3017.5	3033.5
<i>Ovis aries</i>	Modern	1105.6	1180.6	1196.6	1427.7	1580.8	1648.8	2131.1	2792.3	2883.4	2899.4	3017.5	3033.5
<i>Capra hircus</i>	Modern	1105.6	1180.6	1196.6	1427.7	1580.8	1648.8	2131.1	2792.3	2883.4	2899.4	3077.5	3093.5
<i>Capra ibex</i>	Modern	1105.6	1180.6	1196.6	1427.7	1580.8	1648.8	2131.1	2792.3	2883.4	2899.4	3077.5	3093.5
<i>Cervus elaphus</i>	Modern	1105.6	1180.6	1196.6	1427.7	1550.8	1648.8	2131.1	2792.3	2883.4	2899.4	3017.5	3033.5

Étapes du protocole

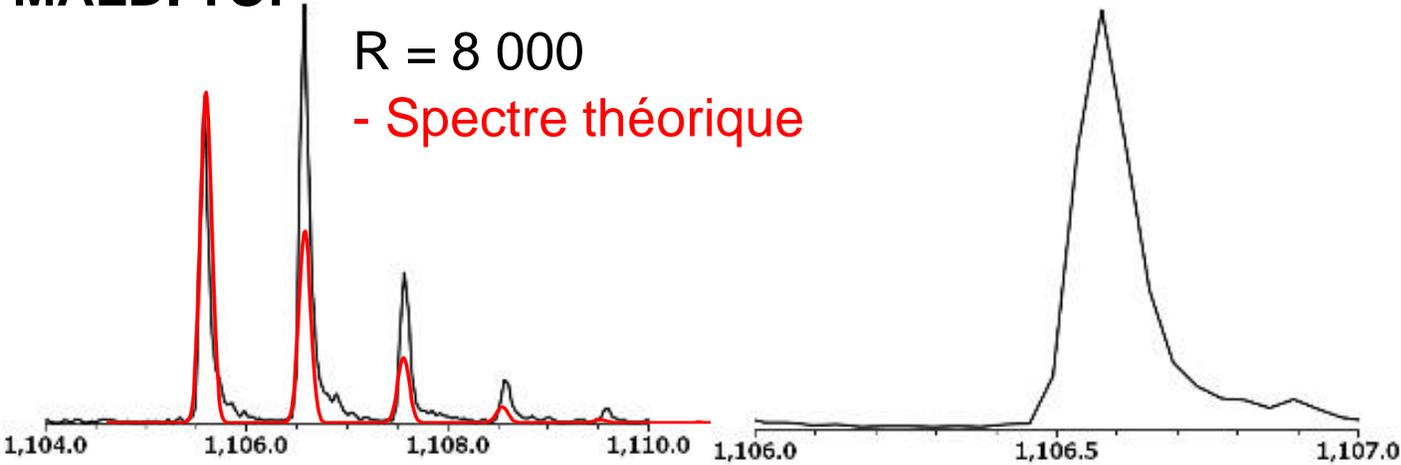


7 Bray, F., Fabrizi, I., Flament, S., Locht, J. L., Antoine, P., Auguste, P., & Rolando, C. (2023). Robust High-Throughput Proteomics Identification and Deamidation Quantitation of Extinct Species up to Pleistocene with Ultrahigh-Resolution MALDI-FTICR Mass Spectrometry. *Analytical Chemistry*, 95(19), 7422-7432.



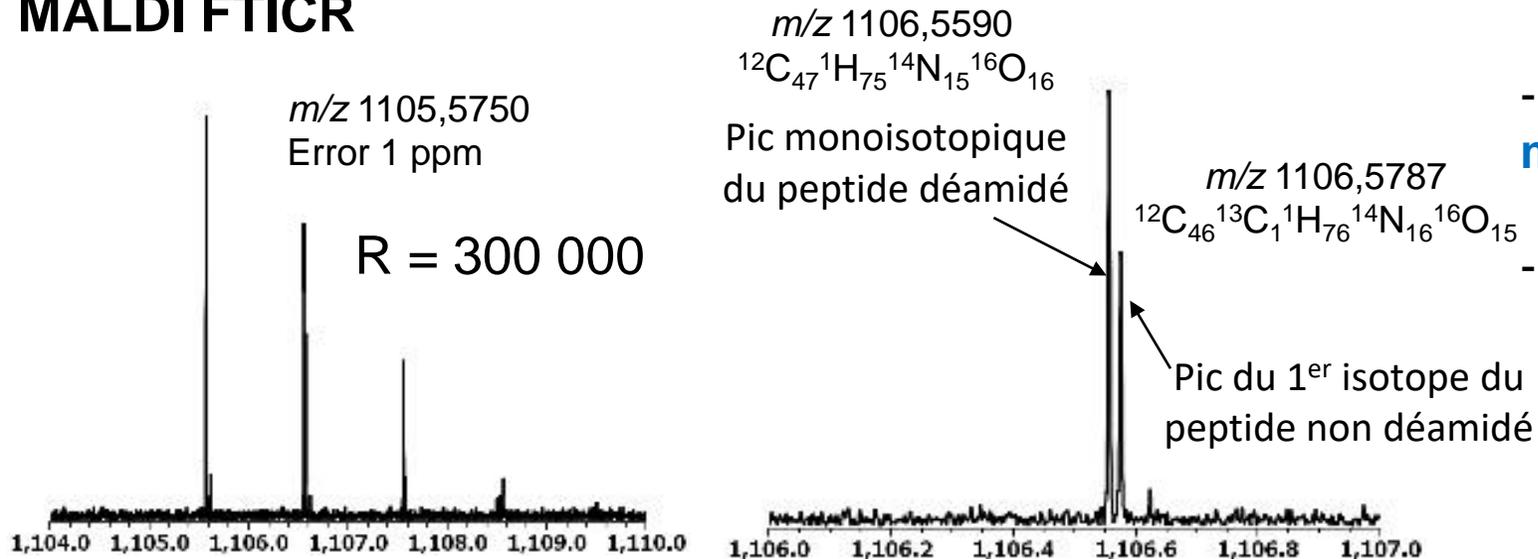
Comparaison MALDI-TOF (ZooMS) vs MALDI FTICR

MALDI TOF



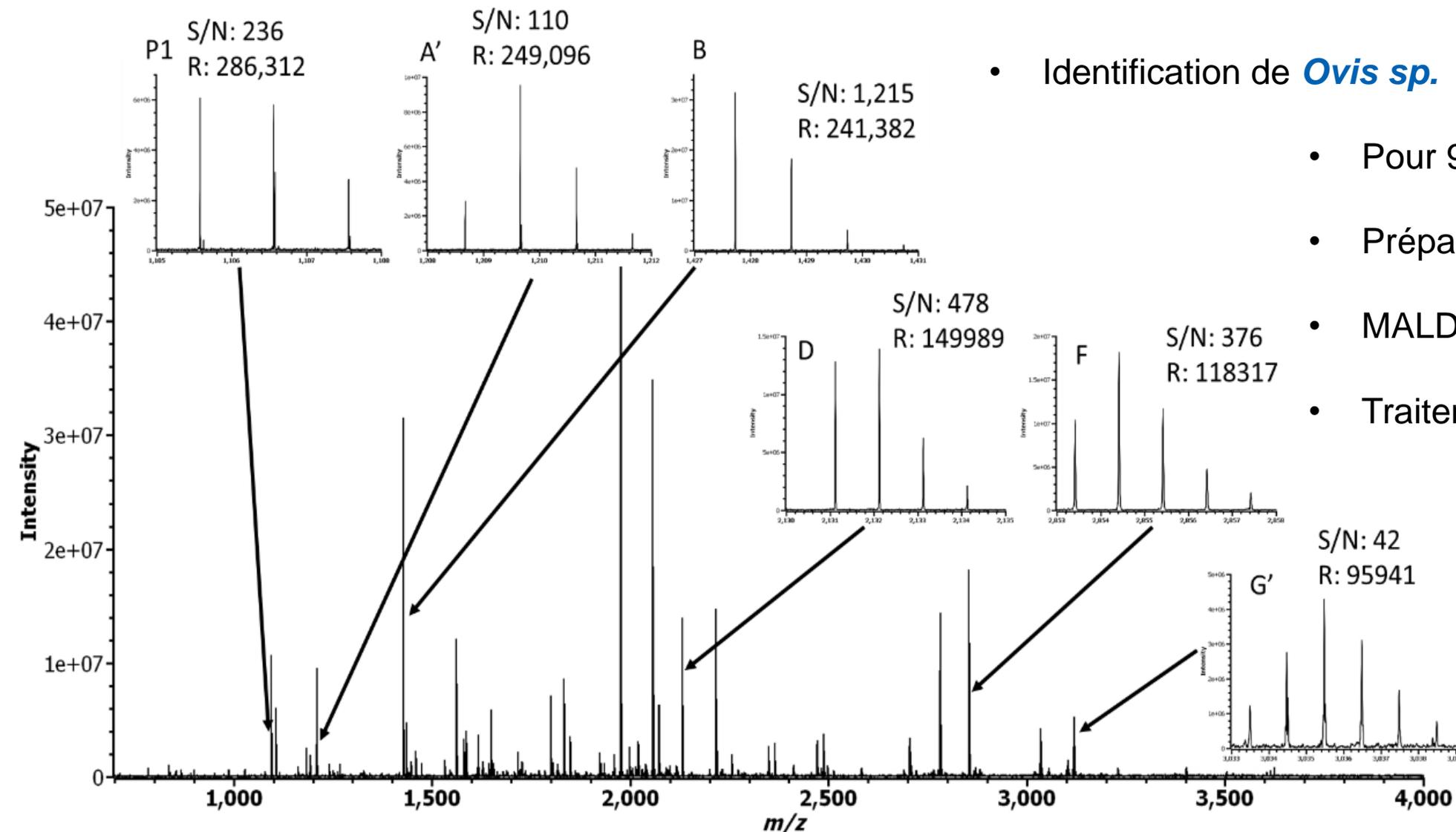
- *Bos primigenius*, digestion de l'os (Ca 101)
- Étude de la séquence GVQGP^{ox}PGPAGPR : peptide dans le collagène des mammifères
- **Peptide déamidés** GVQ^{dem}GP^{ox}PGPAGPR présent dans la digestion de l'os
- **Une acquisition de 2M** est suffisante pour mesurer précisément la déamidation

MALDI FTICR



- FTICR permet d'observer une **différence de masse de 0,019 Da** (déamidation)
- **Erreur de moins d'1ppm**

Spectre MALDI FTICR de mouton



- Identification de *Ovis sp.*

- Pour 96 échantillons
- Préparation = **2 days**
- MALDI FTICR = **2 h**
- Traitement des données = **6 h**

Identification à l'aide de programme

- **En 2018**, Muxin Gu & Michael Buckley utilisent du **machine learning** pour identifier la taxonomie des vertébrés.
- **En 2020**, Simon Hickinbotham et al. ont développés **une méthode de classification** des spectres MALDI-ToF sur des poissons.
- **En 2023**, Andrew Baker et al. ont utilisés un **algorithme d'IA** pour classifier des spectres MALDI ToF de poissons
- **Programmes non utilisés**
- **Nécessité de créer un programme simple, rapide et modifiable**

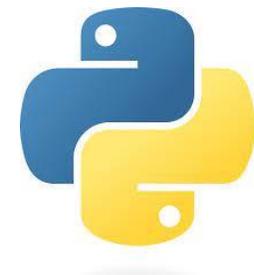
Gu, M., & Buckley, M. (2018). Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. *BMC bioinformatics*, 19(1), 1-9.

Richter, K. K., McGrath, K., Masson-MacLean, E., Hickinbotham, S., Tedder, A., Britton, K., ... & Speller, C. F. (2020). What's the catch? Archaeological application of rapid collagen-based species identification for Pacific Salmon. *Journal of Archaeological Science*, 116, 105116.

10 Baker, A., Harvey, V. L., & Buckley, M. (2023). Machine Learning for collagen peptide biomarker determination in the taxonomic identification of archaeological fish remains. *Journal of Archaeological Science: Reports*, 49, 104001.



Identification des spectres MALDI



NCBI

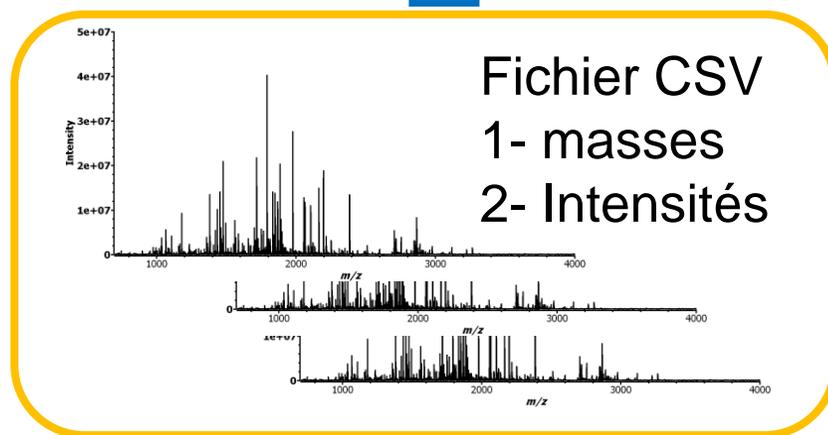
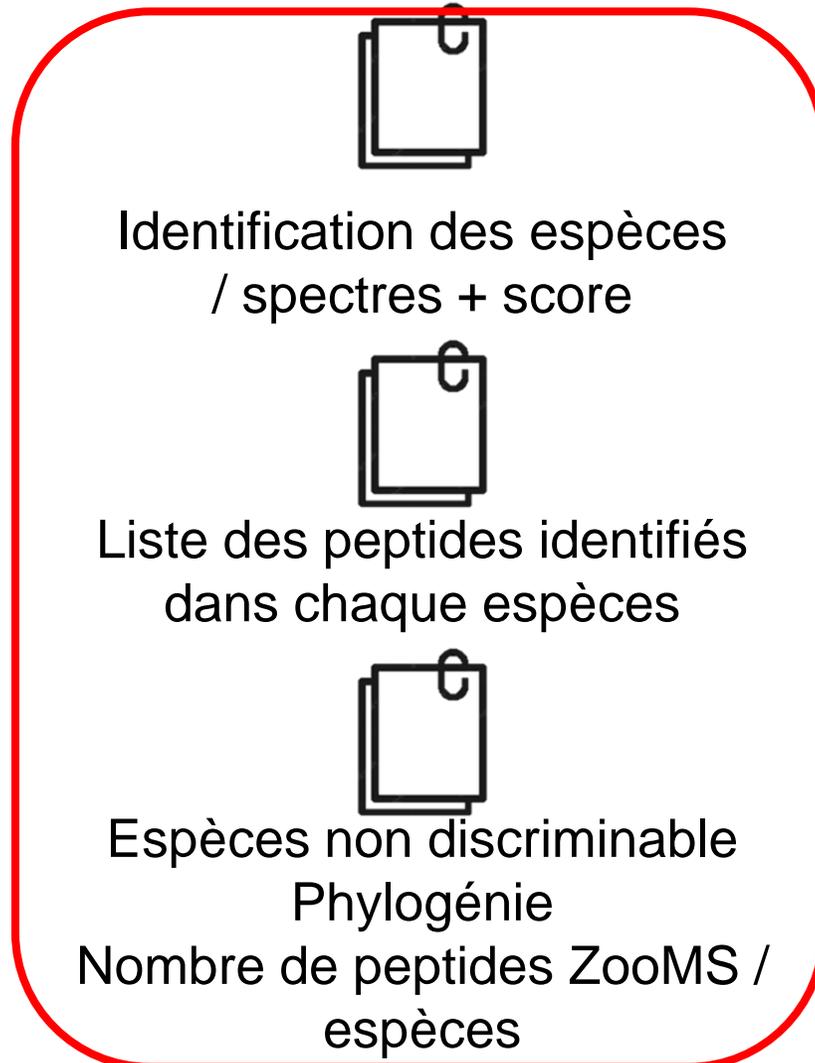
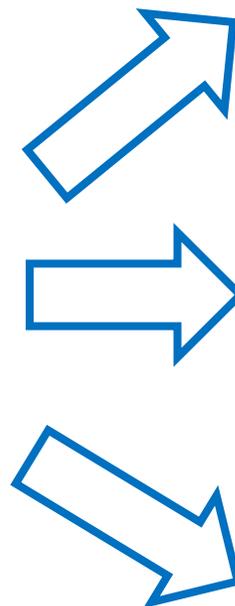
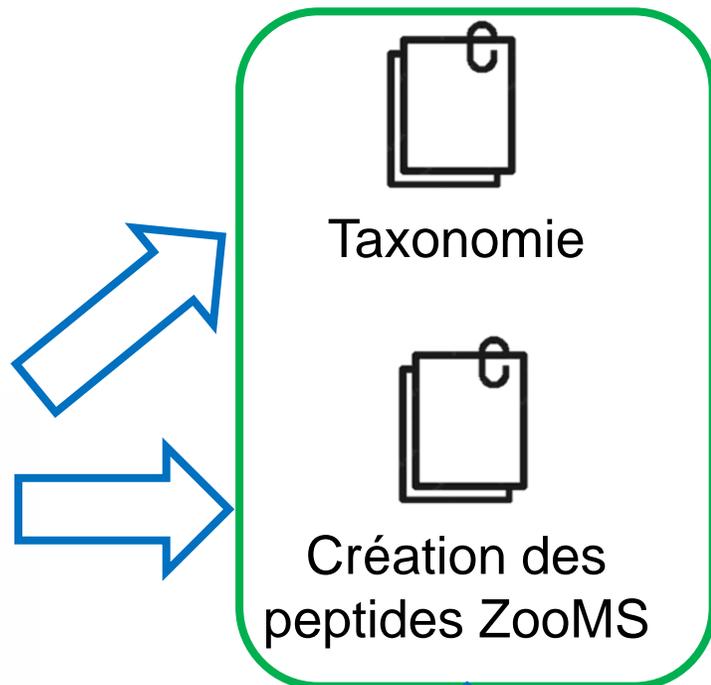
Uniprot



COL1A1
COL1A2



Dr H el ene TOUZET



Fichier taxonomie

Taxon Id	Common name	Scientific name	Parent	Rank
9254		Prototheria	40674	no rank
9255	egg-laying mammal	Monotremata	9254	order
9256		Ornithorhynchidae	9255	family
9257		Ornithorhynchus	9256	genus
9258	Duckbill platypus	Ornithorhynchus ana	9257	species
9259	echidnas	Tachyglossidae	9255	family
9260	short-nosed echidn	Tachyglossus	9259	genus
9261	Short-beaked echid	Tachyglossus aculeat	9260	species
9263	marsupials	Metatheria	32525	no rank
9265	opossums	Didelphidae	38605	family
9266		Didelphis	126287	genus
9267	North American opo	Didelphis virginiana	9266	species
9268	Southern opossum	Didelphis marsupiali	9266	species
9269		Marmosa	126287	genus
9271		Philander	126287	genus
9272	Gray four-eyed opo	Philander opossum	9271	species
9273		Thylacnidae	38608	family
9274		Thylacinus	9273	genus
9275	Tasmanian wolf	Thylacinus cynoceph	9274	species
9277		Dasyuridae	38608	family
9278	quolls	Dasyurus	9277	genus
9279	Eastern quoll	Dasyurus viverrinus	9278	species

Fichier table peptides

Rank	Taxid	Taxon name	Sequence	PTM	Code	Masses	Protein	SeqId	Begin	End	Comment
species	32536	Acinonyx jubatus	GPSGEPGTA(4P		G	2983.51196	COLIA2	XP_01492202	847		879 K - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GEQGPPAGPP 3P		E	2820.35473	COLIA2	XP_01492202	544		573 K - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GPPGESGAA(1P		C	1566.75068	COLIA2	XP_01492202	592		609 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	TGHPGVSVP,1P		Aâ€™	1207.61781	COLIA2	XP_01492202	1068		1080 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GPNGEAGSA 2P		P2	1609.75649	COLIA2	XP_01492202	382		399 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GPSGEPGTA(5P		Gâ€™	2999.50687	COLIA2	XP_01492202	847		879 K - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GLPGVSGSV(3P		D	2163.10404	COLIA2	XP_01492202	883		906 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GLPGEFGLP(2P		B	1453.74341	COLIA2	XP_01492202	574		588 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	TGHPGVSVP,0P1D		A	1192.60691	COLIA2	XP_01492202	1068		1080 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	TGHPGVSVP,0P		A	1191.6229	COLIA2	XP_01492202	1068		1080 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	TGHPGVSVP,1P1D		Aâ€™	1208.60183	COLIA2	XP_01492202	1068		1080 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GVQPPGPA 1P		P1	1105.57488	COLIA1	XP_02688985	682		693 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GLTGP(2P		F	2853.41258	COLIA1	XP_02688985	760		792 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GVQPPGPA 1D1P		P1	1106.5589	COLIA1	XP_02688985	682		693 R - peptide - G, #mismatch 0
species	32536	Acinonyx jubatus	GLTGP(3P		Fâ€™	2869.40749	COLIA1	XP_02688985	760		792 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GVQPPGPA 1D1P		P1	1106.5589	COLIA1	XP_03449645	681		692 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GLTGP(3P		Fâ€™	2869.40749	COLIA1	XP_03449645	759		791 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GVQPPGPA 1P		P1	1105.57488	COLIA1	XP_03449645	681		692 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GLTGP(2P		F	2853.41258	COLIA1	XP_03449645	759		791 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GPSGEPGTA(4P		G	2957.49631	COLIA2	XP_00291827	847		879 K - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GEQGPPAGPP 3P		E	2820.35473	COLIA2	XP_00291827	544		573 K - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GPPGESGAA(1P		C	1566.75068	COLIA2	XP_00291827	592		609 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	TGHPGTVGP,0P1D		A	1206.62256	COLIA2	XP_00291827	1068		1080 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	TGHPGTVGP,1P1D		Aâ€™	1222.61748	COLIA2	XP_00291827	1068		1080 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GPNGEAGSA 2P		P2	1609.75649	COLIA2	XP_00291827	382		399 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GPSGEPGTA(5P		Gâ€™	2973.49122	COLIA2	XP_00291827	847		879 K - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GLPGVSGSV(3P		D	2147.10912	COLIA2	XP_00291827	883		906 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	GLPGEFGLP(2P		B	1453.74341	COLIA2	XP_00291827	574		588 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	TGHPGTVGP,0P		A	1205.63855	COLIA2	XP_00291827	1068		1080 R - peptide - G, #mismatch 0
species	9646	Ailuropoda melanoleu	TGHPGTVGP,1P		Aâ€™	1221.63346	COLIA2	XP_00291827	1068		1080 R - peptide - G, #mismatch 0

Fichier résultat

Spectra:	G:\Bioinfo\2023\Code_pour_Fabric-update\DatabaseFT-SNAP-cal																		
Markers:	table_mammals_with_deamidation.tsv																		
Resolution:	0.008																		
	A - OP	A - OPID	Aâ€™ - 1P	Aâ€™ - 1PID	B - 2P	C - 1P	D - 3P	E - 3P	F - 2P	Fâ€™ - 3P	G - 4P	Gâ€™ - 5P	P1 - 1D1P	P1 - 1P	P2 - 2P	Score	Assignment	Rank	Species
Bison-priscu	1192.67656	1193.66318	1208.67404	1209.65796	1427.72756	1580.76596	2131.11613		2853.40864	2869.40499		3033.49041	1106.55825	1105.5743	1648.82783	13	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
Bison-priscu	1192.68023	1193.66344	1208.67441	1209.65808	1427.72767	1580.76656	2131.11642		2852.40942	2869.4058	3017.49459	3033.49098	1106.55839	1105.57438	1648.82806	14	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
Bison-priscu-CL1-84p-dei	1192.67656	1193.66327	1208.67385	1209.65767	1427.72719	1580.76598	2131.11474		2853.40744	2869.40322		3033.49098	1106.55805	1105.57417	1648.82733	11	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
Bison-priscu	1192.67663	1193.66284	1208.67415	1209.65776	1427.72722	1580.76578	2131.11485		2853.40791	2869.40258		3033.48345	1106.55807	1105.57414	1648.82745	13	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
Bos.d.csv	1192.67974		1208.6751		1427.72918	1580.76775	2131.11935		2853.41366	2869.40962		3033.49092	1105.57495	1105.57495	1648.83031	10	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
BosO.d.csv	1192.67949		1208.6744		1427.72797	1580.76692	2131.11663	2792.32945	2853.41001	2869.40652		3033.4872	1105.57468	1105.57468	1648.82902	11	27592	[Bovini subfamily]	72004 [Bos mutus] 9915 [Bos indicus] 43346 [Bison b
Porc1-1.d.csv	1180.64352	1181.6274	1196.6381	1197.6219	1453.74416	1550.75744	2131.11906	2820.36257	2883.42687		3017.49798	3033.49336	1106.55875	1105.57469	1647.80936	14	35497	[Suina suborder]	9825 [Sus scrofa domestica] 9823 [Sus scrofa] 41426
Porc1-2.d.csv	1180.64328	1181.62726	1196.63637	1197.62168	1453.74322	1550.757	2131.11752	2820.36089	2883.4256	2899.42496	3017.49917	3033.493	1106.55859	1105.5747	1647.80775	15	35497	[Suina suborder]	9825 [Sus scrofa domestica] 9823 [Sus scrofa] 41426
Porc2-1.d.csv	1180.64309		1196.6376		1453.74424	1550.7578	2131.11898	2820.36203	2883.4267		3033.49303		1105.57492			9	35497	[Suina suborder]	9825 [Sus scrofa domestica] 9823 [Sus scrofa] 41426
Porc2-2.d.csv	1180.64326		1196.63577		1453.7442	1550.7575	2131.11866	2820.36152	2883.42685		3017.49989	3033.49261	1105.57487			10	35497	[Suina suborder]	9825 [Sus scrofa domestica] 9823 [Sus scrofa] 41426
Renne 796-1	1150.63273		1166.62763		1427.72864	1580.76649	2131.11779		2883.42573	2899.42419		3093.49847	1105.57487	1648.82952		10	9869	[Rangifera]	9870 [Rangifer tarandus]
Renne 796-2	1150.63092		1166.62784		1427.7288	1580.7674	2131.11819		2883.4258	2899.42459	3077.50024	3093.49915	1105.57498	1648.82918		11	9869	[Rangifera]	9870 [Rangifer tarandus]
act1-1 Mystic	1189.64641		1205.64161		1441.70939	1577.76892			2883.42747	2899.4254	3007.48076	3023.47334	1080.54559	1079.56262	1652.82589	9	9727	[Cetacea infraorder]	9767 [Balaenoptera acutorostrata]
act1-2 Mystic	1189.64651		1205.6414		1441.70925	1577.76856	2135.11243		2883.42717	2899.4254	3007.47354	3023.47275	1080.54559	1079.56262	1652.82561	9	9727	[Cetacea infraorder]	9731 [Globicephala melas]
act10-1Vulpe	1210.65374		1226.64916		1437.71365	1566.75342	2131.12045		2853.42003	2869.41284	2999.51145	3007.47354	1106.56171	1105.57491		10	9627	[Vulpes]	9627 [Vulpes vulpes]
act10-2Vulpe	1210.65225		1226.64957		1437.71471	1566.75452							1106.55824	1105.57447		7	9627	[Vulpes]	9627 [Vulpes vulpes]
act11-1Vulpe	1210.65481		1226.64987		1437.71504	1566.75567		2820.36174					1105.57573			5	9627	[Vulpes]	9627 [Vulpes vulpes]
act11-2Vulpe	1210.65524		1226.64987		1437.7152	1566.75449							1105.57583			4	9627	[Vulpes]	9627 [Vulpes vulpes]
act13-1Oryctolagus cuniculus.d.csv		1221.6341			1453.74383		2129.13701	2836.3473	2883.42114	2899.42057	2957.5038		1106.5611	1105.57529	1592.7639	10	9979	[Leporidae]	9986 [Oryctolagus cuniculus]
act13-2Oryctolagus cuniculus.d.csv		1221.63396			1453.74367		2129.13697	2836.34803	2883.42211	2899.42046	2957.50305			1105.57509	1592.76631	9	9979	[Leporidae]	9986 [Oryctolagus cuniculus]
act19-1Oryct	1205.63925		1221.634		1453.74437		2129.13787	2836.34969	2883.42332	2899.42192			1106.56227	1105.57542	1592.76855	10	9979	[Leporidae]	9986 [Oryctolagus cuniculus]
act19-2Oryct	1205.63801		1221.63412		1453.74421		2129.13784	2836.3494	2883.42329	2899.4219			1106.5604	1105.57536	1592.76631	10	9979	[Leporidae]	9986 [Oryctolagus cuniculus]
act22-1Marmota marmota.d.csv					1427.72856		2133.09581		2883.42159	2899.42049	2983.51578	2999.5073		1105.57508	1637.7892	8	10158	[Octocoridae]	10160 [Octodon degus]
act22-2Marmota marmota.d.csv					1427.72824		2133.09625		2883.42185	2899.42099	2983.51626	2999.5072	1106.56022	1105.57527	1637.7884	9			

Informations du programme

- Temps d'exécution:

30 sec pour 96 fichiers

- **Différenciation** des espèces :

Sus scrofa, Sus scrofa domesticus, Phacochoerus africanus

Bos mutus, Bison bison bison, Bos taurus, Bos indicus

- Information sur la **masses des peptides ZoomS**

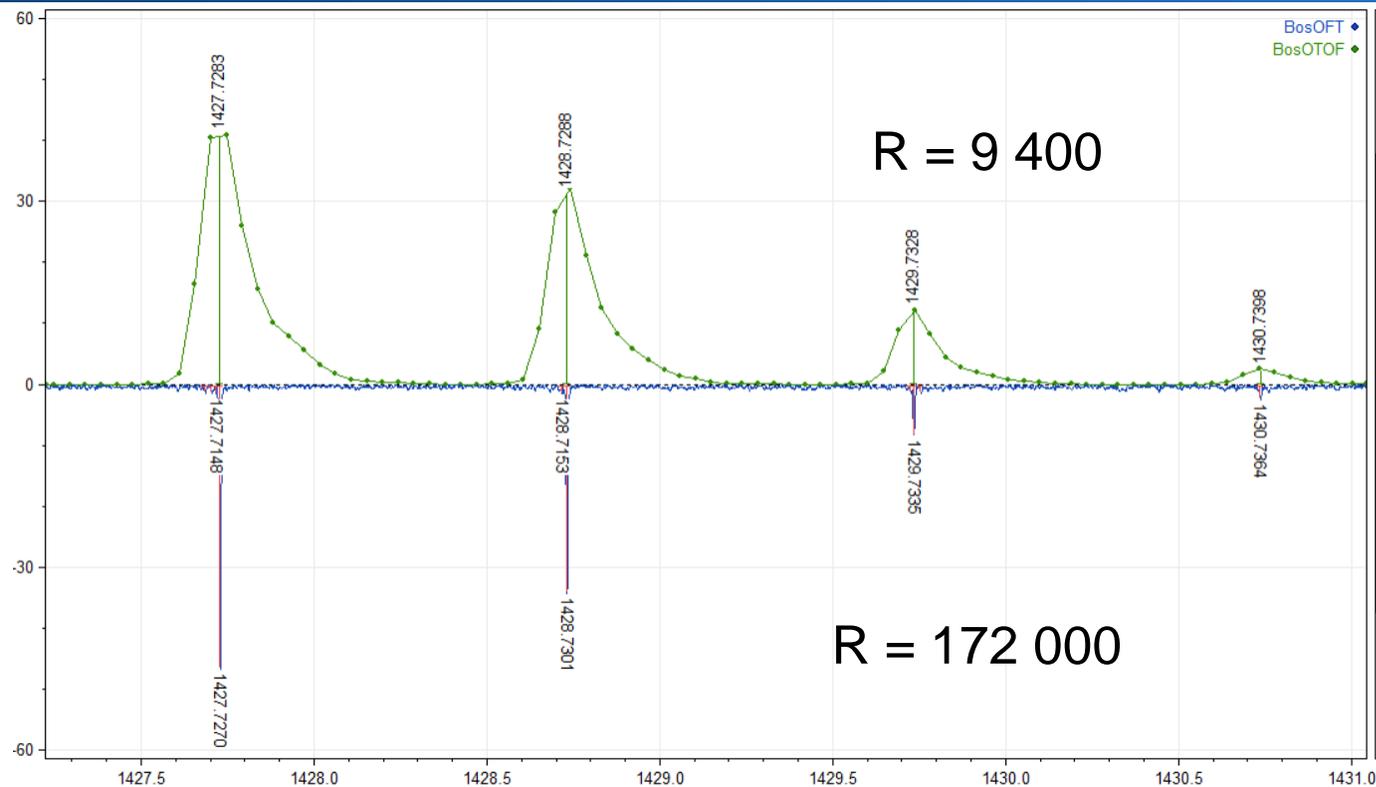
Sequence	Formula	A	A+D	A'	A'+D	Species
TGHPGAVGPAGIR	C51H85N18O15	1189.64363	1190.62765	1205.63855	1206.62256	<i>Orcinus orca</i>
SGHPGTVGPAGIR	C51H85N18O16	1205.63855	1206.62256	1221.63346	1222.61748	<i>Lemur catta</i>
TGHPGTVGPAGVR	C51H85N18O16	1205.63855	1206.62256	1221.63346	1222.61748	<i>Ailuropoda melanoleuca</i>

A = nom de la nomenclature du peptide

A' = peptide A avec une hydroxyproline = + 15,99 Da

D = déamidation, delta de masse +0,99 Da

Comparaison FT ICR et ToF



- Précision TOF 0,1 Da
- Précision FT ICR 0,005 Da

- **100% d'identification** sur espèces **modernes** pour le **FTICR** et **TOF** (n = 25)
- **100%** d'identification sur spectres **FTICR** **archéologiques** (n = 100)
- **Gain de temps** de 2 jours

Conclusions & perspectives

- Paléoprotéomique est en **pleine expansion**
- En France, très **peu de personnes spécialisées** dans le domaine
- Nécessité **d'améliorer** les méthodes et les traitements des données
- Programme permettant d'identifier **les spectres automatiquement**

- **Détection des pics isotopiques** à intégrer
- Mettre au point une **comparaison des profils de masse avec de l'IA**
- Application pour **toutes les espèces animales**

Remerciements

