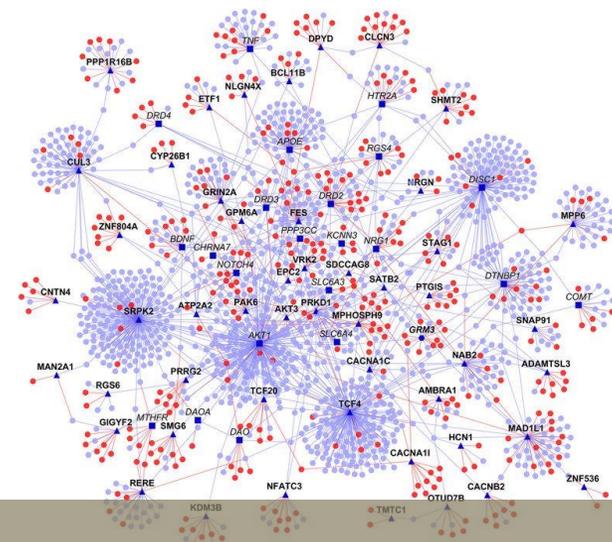
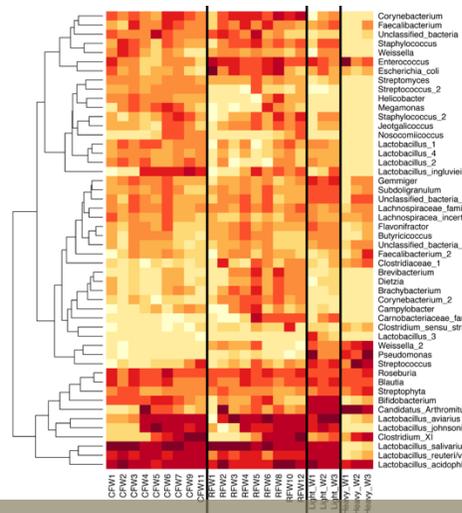
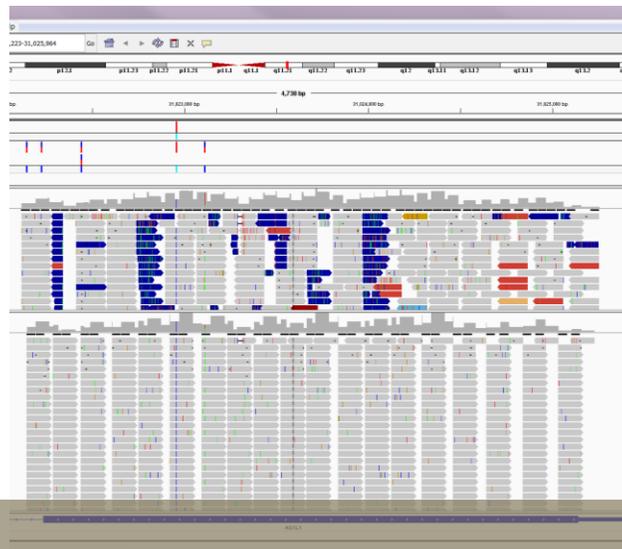


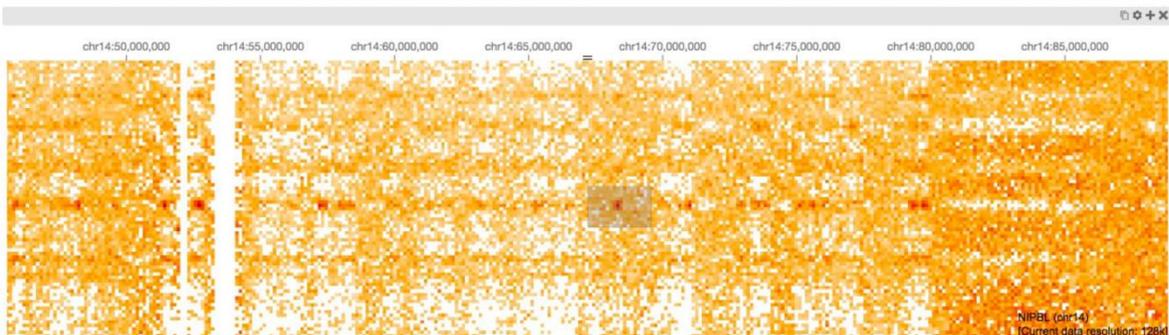
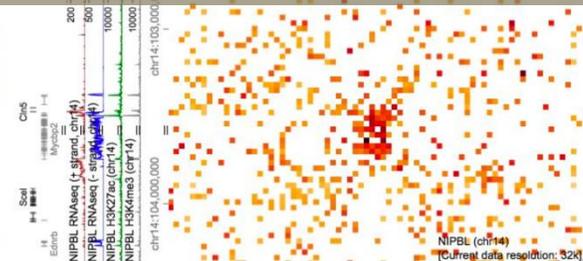
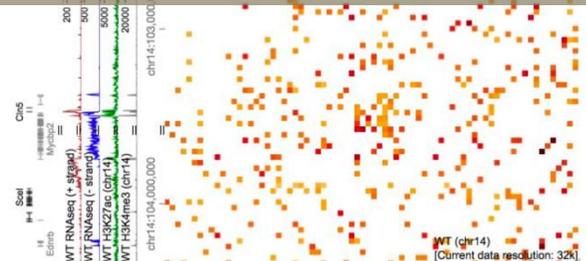
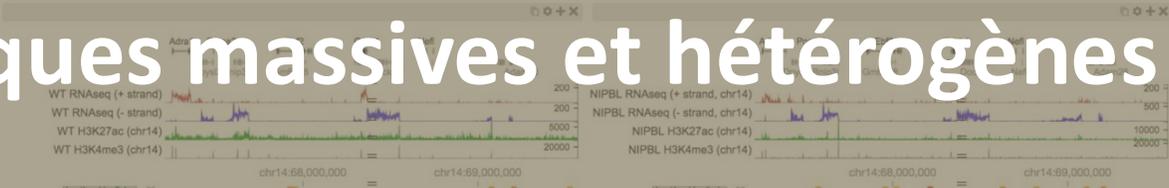
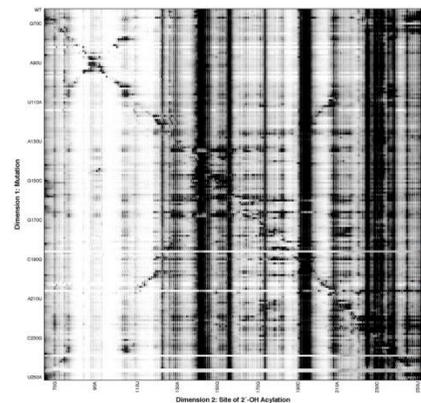
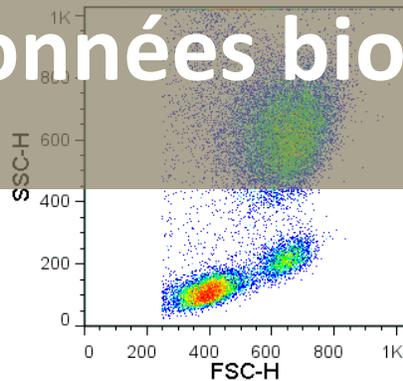
DÉCODER LA BIOLOGIE AVEC LE JEU VIDÉO

Roman Sarrazin-Gendron
Département d'informatique
UQAM

<https://bias.uqam.ca/>



Données biologiques massives et hétérogènes

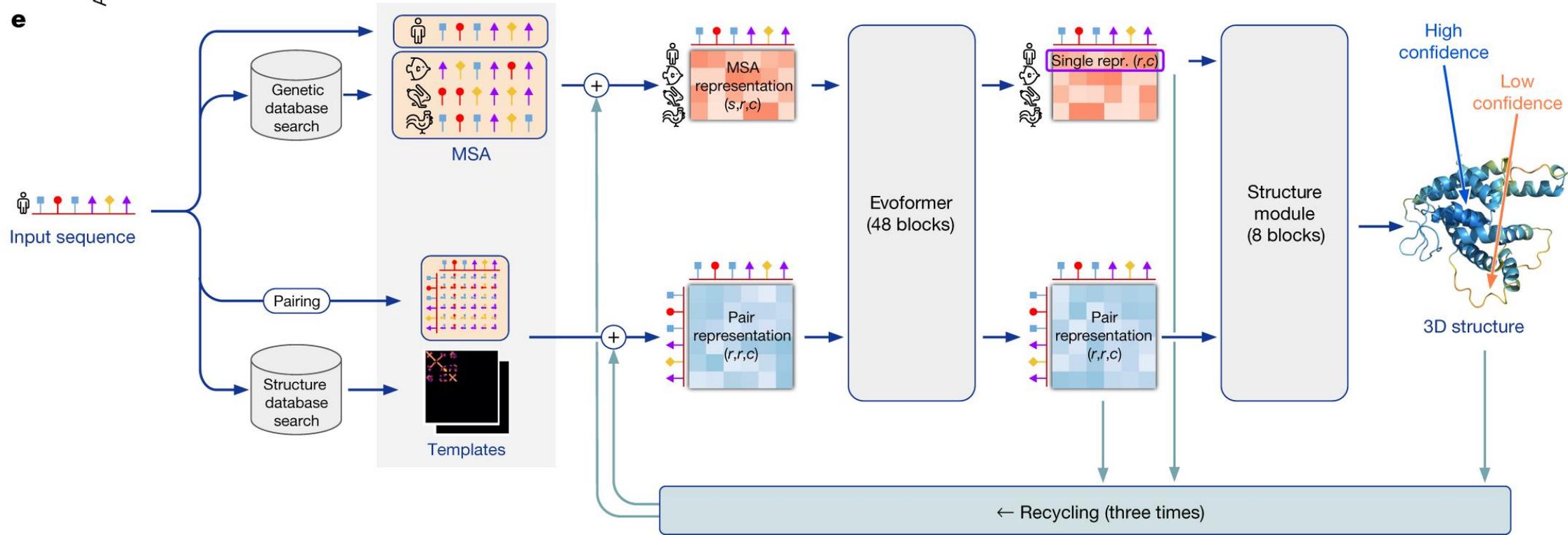
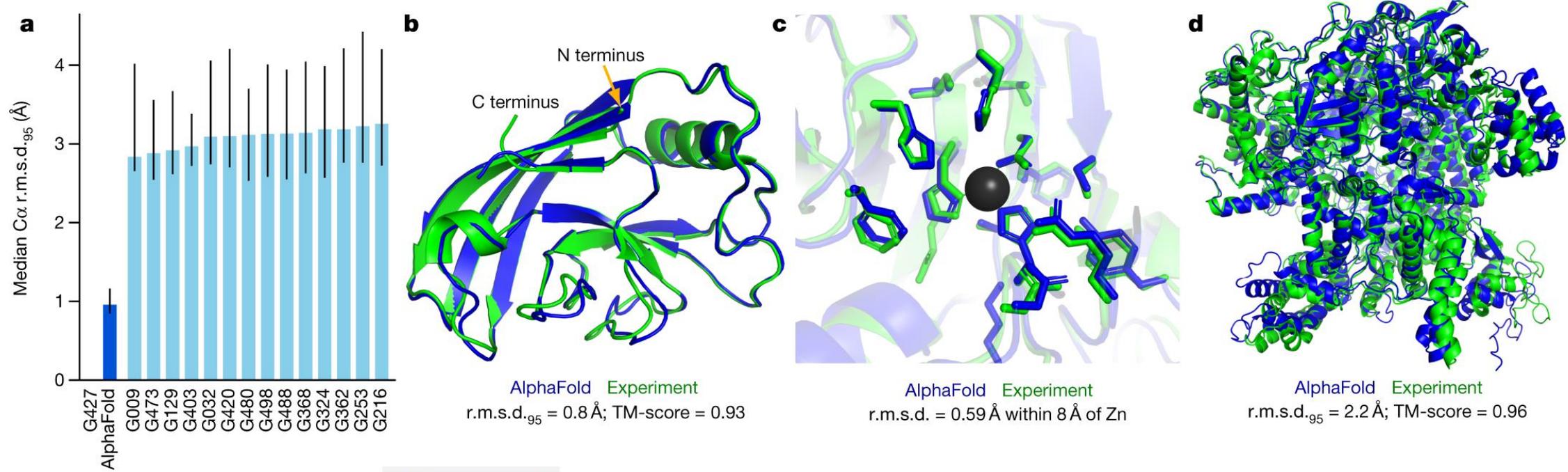


L'INTELLIGENCE ARTIFICIELLE À LA RESCousse

- **2016:** DeepMind lance le projet AlphaFold pour le repliement des protéines
- **2020:** AlphaFold 2 gagne la compétition CASP13.
- **2024:** AlphaFold 3 inclut d'autres molécules que les protéines, mais seulement 100 ARN.

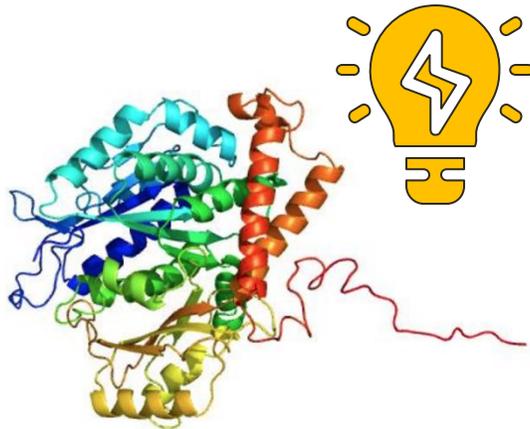


AlphaFold



CONDITIONS FAVORABLES POUR L'IA

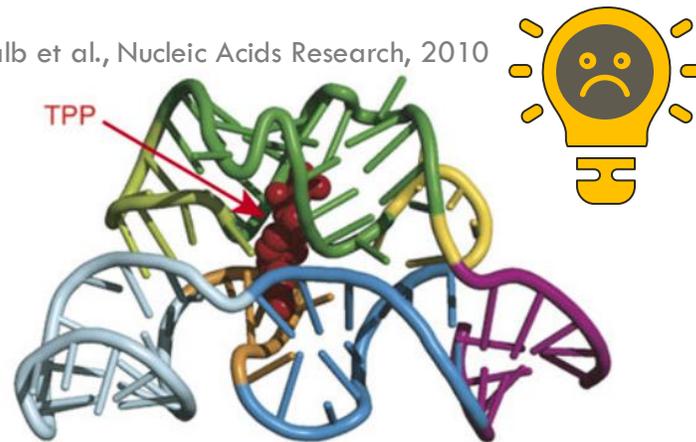
- Données nombreuses et équilibrées
- Annotations de qualité
- Possibilité de formuler une fonction d'optimisation



repliement de protéines

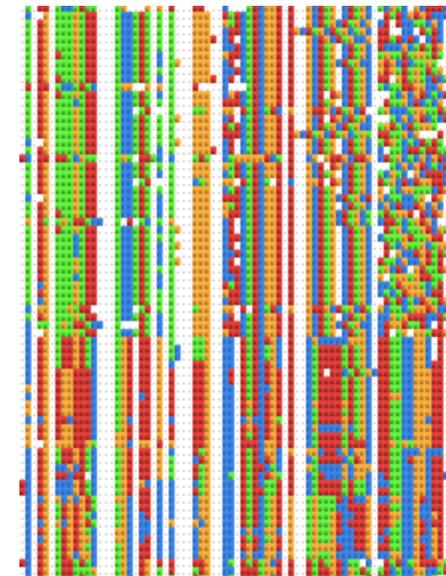
N= 60 000

Falb et al., Nucleic Acids Research, 2010



repliement de l'ARN

N= 1 800



alignement de séquences

N= y en a pas

COMPARER DES SÉQUENCES BIOLOGIQUES

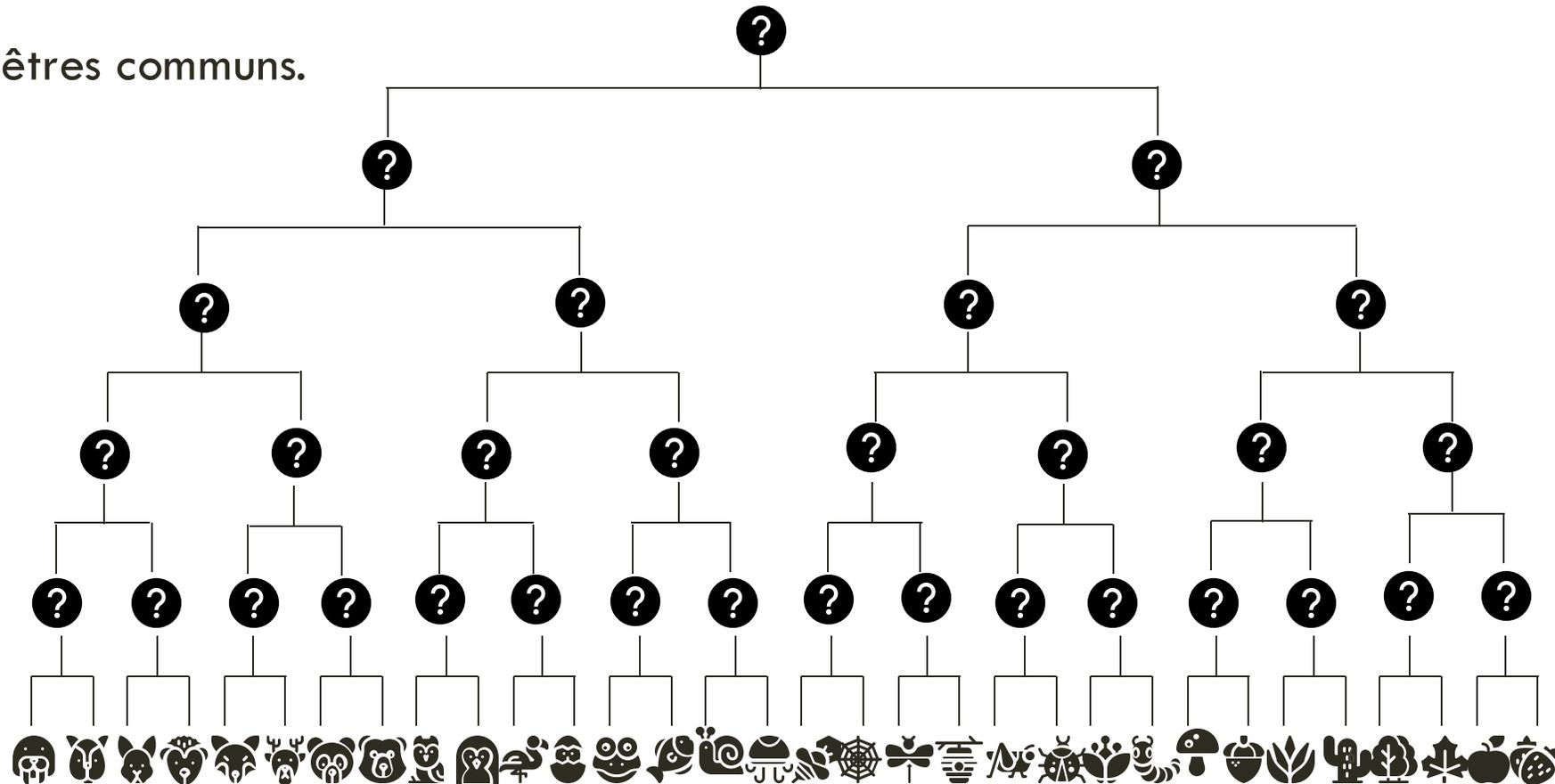


AATTACAAGTTCCTTGGC



AATCAACTTGAC

- Les génomes des organismes vivants sont le fruit de milliards d'années d'évolution.
- On ne connaît aucun des ancêtres communs.
- Comment les comparer ?



Arbre phylogénétique

ALIGNEMENT DE SÉQUENCES

- En biologie, l'alignement de séquences est la mise en correspondance de régions des séquences partagées avec un ancêtre commun (inconnu) selon un modèle statistique.
- Insérer une faille (gap) ou mettre en correspondance des bases différentes correspond à formuler l'hypothèse qu'une mutation s'est produite depuis l'ancêtre commun.
- On peut estimer la proximité de l'ancêtre commun via la similarité des séquences alignées

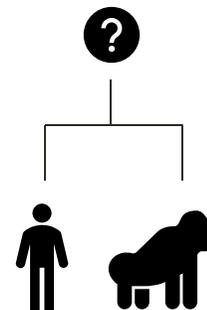


A A T T A C A A G T T C T T G G C

A A T C A - - - - - C T T G A C

- - C A

A A T C A C T T G A C

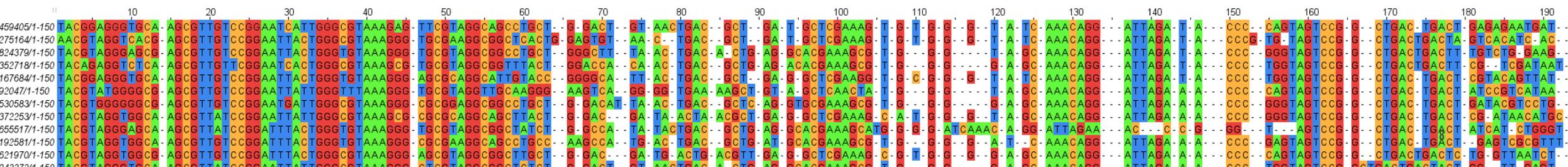


A A T ? A ? ? ? ? C T T G ? C

Estimation de la séquence ancestrale

ALIGNEMENT DE SÉQUENCES MULTIPLES

- Pas d'alignement *optimal* universel, mais on peut définir un *bon alignement*
- Les approches algorithmiques se généralisent difficilement en l'absence d'optimum bien défini.
- Même si on avait un alignement optimal, quand même exponentiel en nombre de séquences
- Beaucoup de biologistes corrigent encore leurs alignements à la main
- Applications:
 - Identifier des mutations responsables de maladies
 - Identifier la proximité évolutive entre espèces
 - ...



GÉNOMIQUE COMPARATIVE



ACCGACATGCATCATGACGCAGGACA



TCCTAATGCATTATTAACCGACA



CTACATGGATTATGACGCGCCT

?



ACCGAC**ATGCATCATGAC**GCAGGACA



TCCTA-**ATGCATTATTA**CC--GACA



--CTAC**ATGTATTATGGC**GC--GCCT



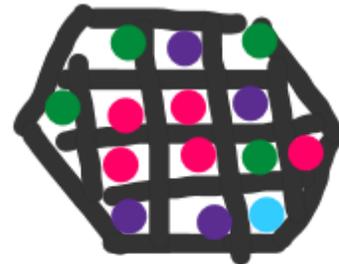
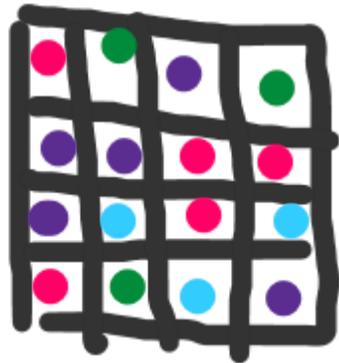
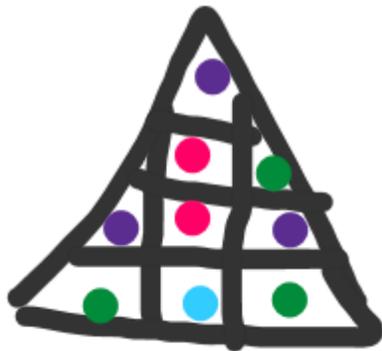
**QUE FAIRE QUAND LES CONDITIONS
FAVORABLES NE SONT PAS RÉUNIES ?**

LES PROBLÈMES DIFFICILES

Exemples de problèmes informatiques «difficiles»:

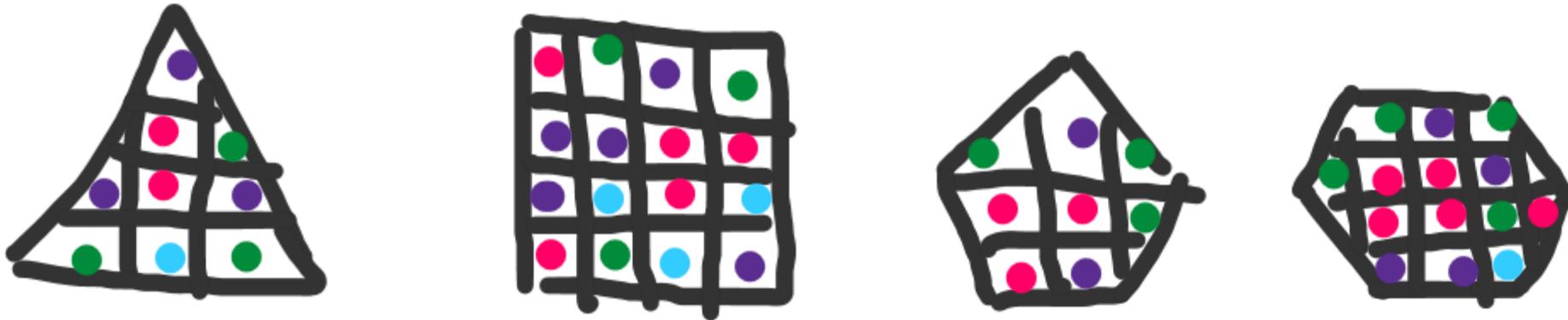
- Les règles changent constamment
- On dispose de très peu de données
- Chaque exemple est différent

Exemple: trouvez le plus grand cluster de points roses



LES PROBLÈMES DIFFICILES?

- Trouver le plus grand cluster de points rose est visuellement TRÈS facile
- Mais pas trivial à coder.



- Beaucoup de tâches sont plus faciles à résoudre visuellement qu'algorithmiquement
- L'intuition visuelle ne requiert aucune information sur le problème

CALCUL HUMAIN

Utilisation de la cognition humaine pour résoudre des problèmes informatiques, le plus souvent à travers un **interface humain-ordinateur**



CROWDSOURCING

(MYRIADISATION)

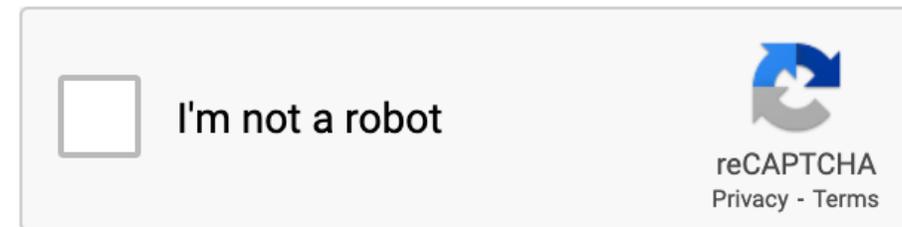
Une approche de type **diviser-pour-régner** pour résoudre un problème informatique

Le problème est divisé en **micro-tâches**, chacune résolues par un ou plusieurs humains, et les solutions sont ensuite **agrégées** pour solutionner le problème.



COMMENT RECRUTER DES CALCULATEURS HUMAINS ?

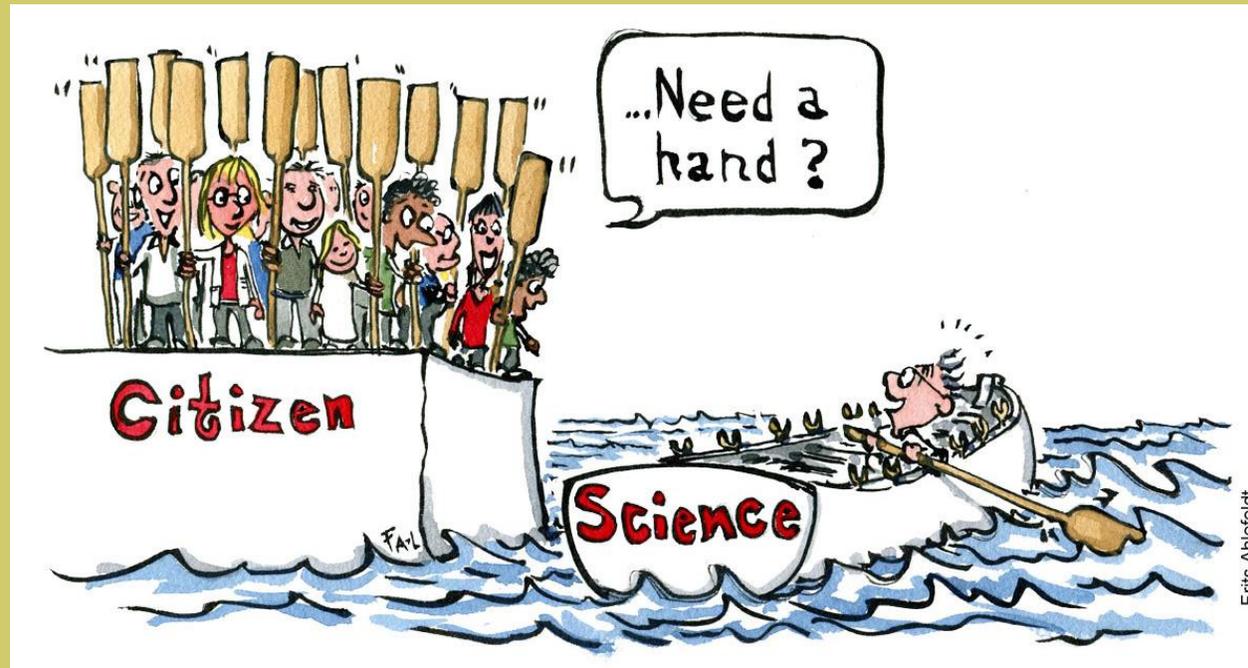
- Les payer
- Les obliger
- Demander gentiment:
- Ludifier la tâche :



science citoyenne



jeux sérieux



JEUX SÉRIEUX ET SCIENCE CITOYENNE

ESP GAME: LE PREMIER JEU SÉRIEUX



(von Ahn and Dabbish, 2004)

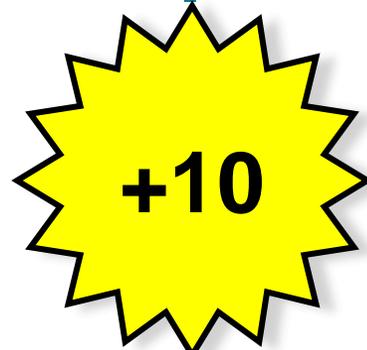
ESP GAME : LE PREMIER JEU SÉRIEUX



Animals

White

Sheeps



Grass

Sheeps

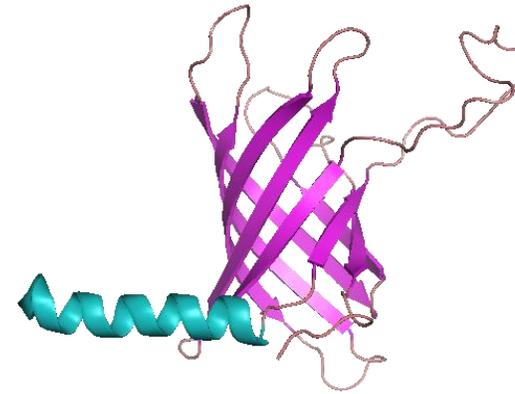
ESP GAME : LE PREMIER JEU SÉRIEUX



New label: **Mouton**

(2008) FOLDIT

ATSTVTGGYAQSDAQGQMNMKG
FNLKYRYEEDNSPLGVIGSFTYTEKSR
TASSGDYNKNQYYGITAGPAYRIND
WASIYGVVGVGYGKFQTTEYPTYK
NDTSDYGFSYGAGLQFNPMENVAL
DFSYEQSRIRSVDVGTWIAGVGYRF



Déterminer la structure des protéines avec un jeu

Rank: 17 Score: 9092

48: Pro Peptide

Group Competition	
#	Group Name
1	The Lone Folder
2	Street Smarts
3	Illinois
4	Berkeley

Player Competition		
#	Player Name	Score
16	psen	9098
17	kathleen	9092
18	weslat82	9081
19	darktorres	9081
20	ccarrico	9032
21	mbojkegren	9048
22	sliderson	9038

Shake sidechains to improve the protein.
Hotkey: S

Shake Sidechains Wiggle Backbone Clear Locks and Bands Reset Puzzle Mouse Help

Actions History View File Pull Tool

<http://fold.it>

REVENONS À L'ALIGNEMENT DE SÉQUENCES: PHYLO

< **CATEGORIES**

Digestive and respiratory system diseases

Gaps - 0

Gap Extends - 0

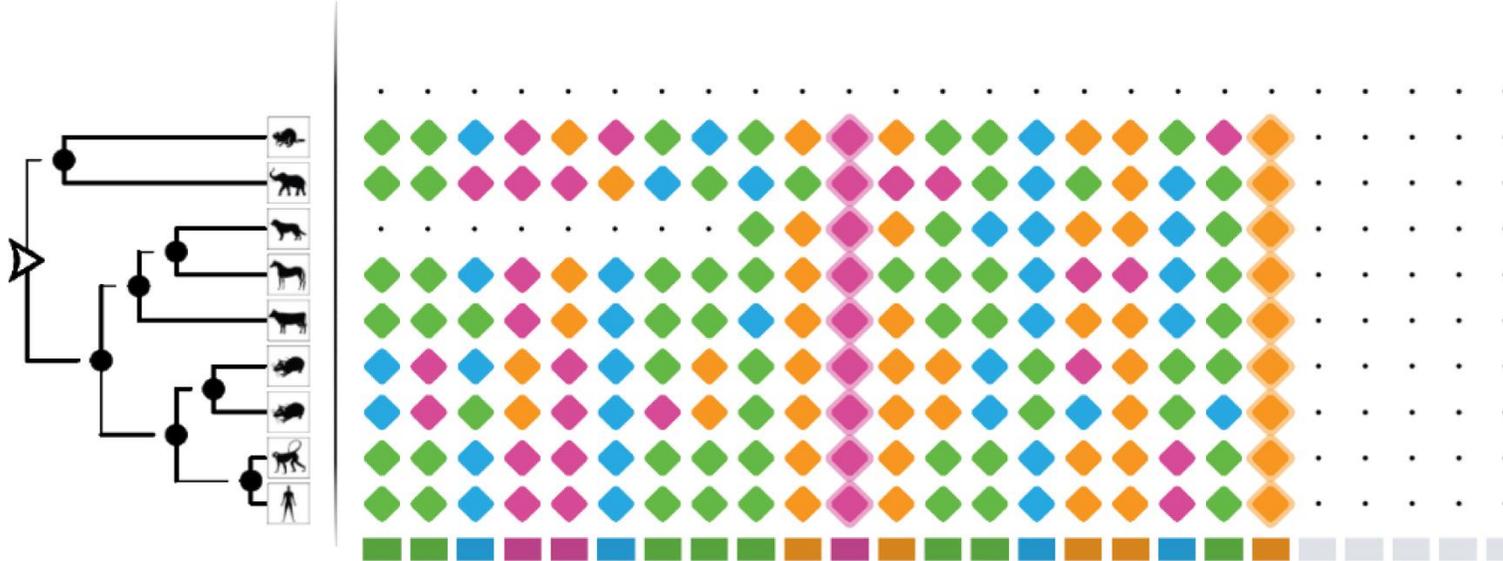
Matches - 261

Mismatches - 51

Goal 202

Score
202

232 Top Score

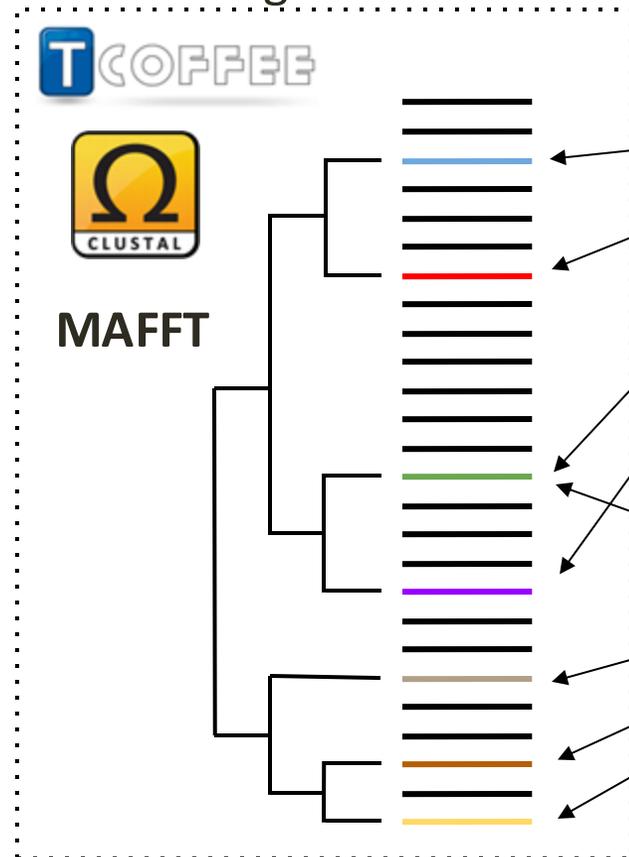


NEXT STAGE

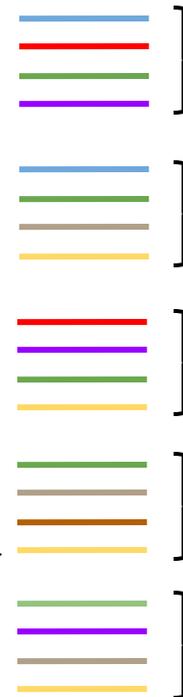


AGRÉGATION & ÉVALUATION

Stratégie 1: utilisation des solutions pour guider l'alignement

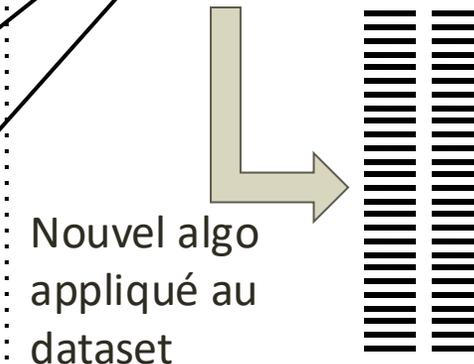
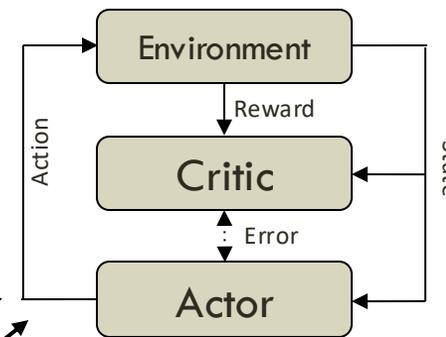


Assemblage des solutions en alignements complets

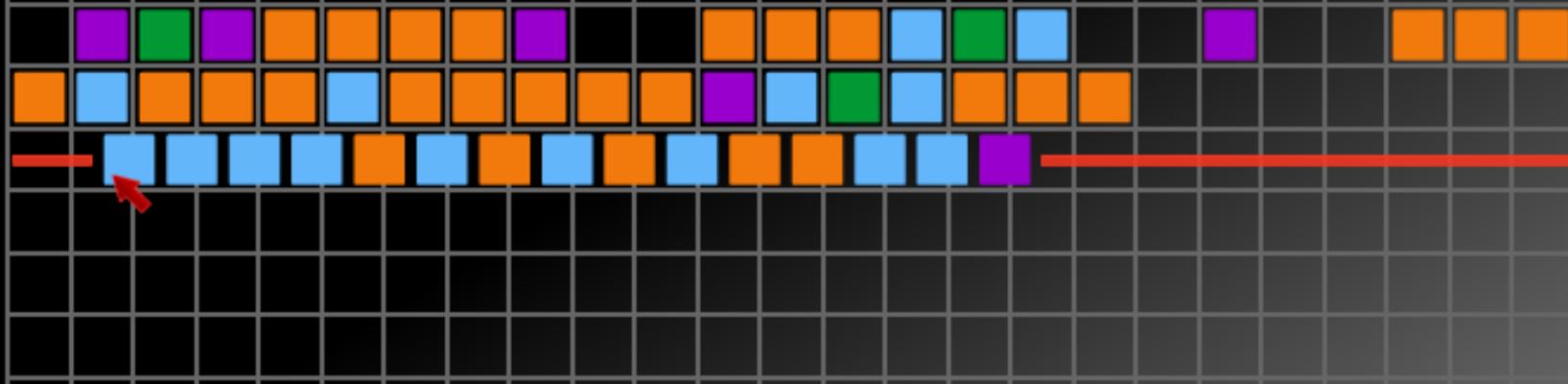
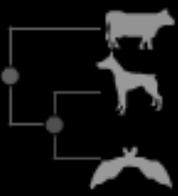


Stratégie 2: nouvel algorithme!

Guidage de l'apprentissage par les solutions



Classement via score ou statistiques bayésiennes

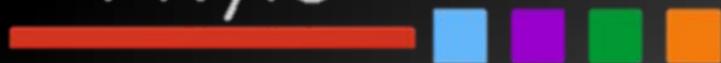


>1 million 

>400,000 

40-95% 

Phylo



Score Now

Best: -1

Score: -4

Stats



LES CLEFS DE PHYLO

Pourquoi? Problème fondamental de la biologie moléculaire

Quoi? Problème bien défini prouvé difficile pour les ordinateurs

Comment? Un jeu de puzzle qui maximise l'accessibilité



<http://thegoddamn90s.com>

LES LIMITES DE PHYLO

- Les puzzles peuvent prendre plus de 30 minutes à compléter
- Une minorité de joueurs complètent plus de 2 puzzles
- Les solutions sont dominées par les experts



<http://thegoddamn90s.com>

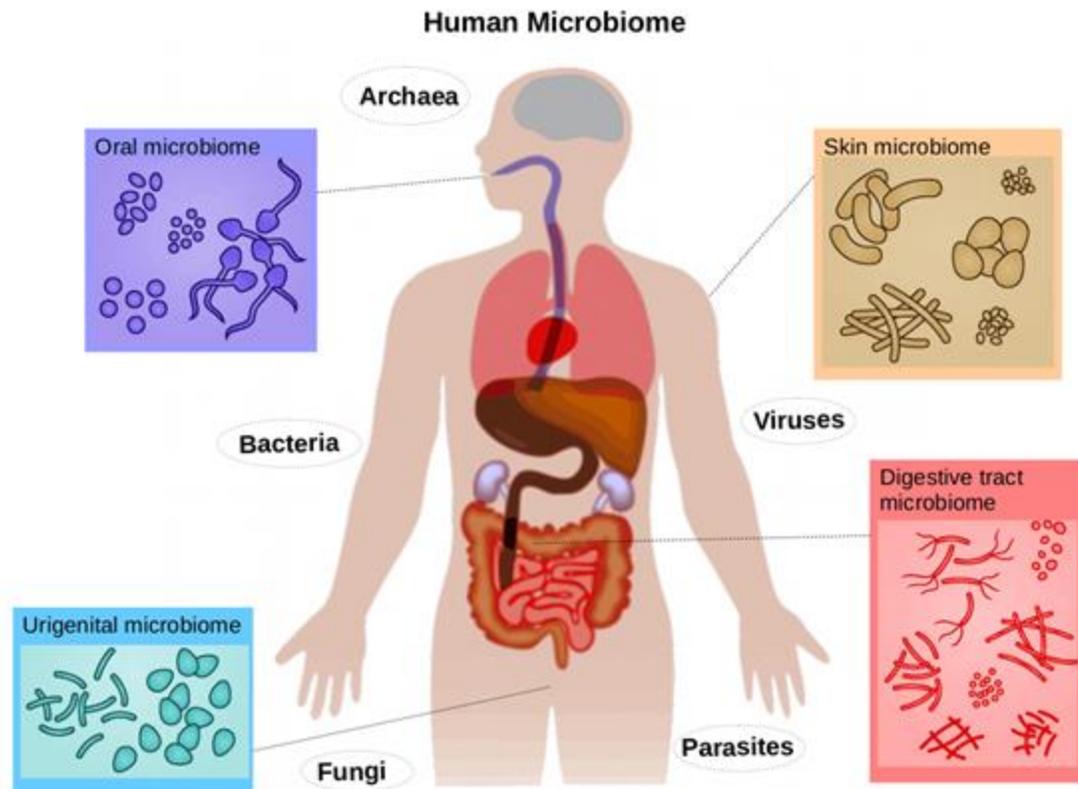


CHANGER D'ÉCHELLE |

Un nouveau jeu de données du microbiome humain

- 23,000 échantillons de selles
- 1 million de séquences ADN à aligner

LES ALIENS PARMIS NOUS: LE MICROBIOME HUMAIN





Comment décupler le nombre de participants ?



THE
MICROSETTA
INITIATIVE™

Borderlands 3: > 18 millions de copies vendues



Borderlands Science



TARGET SCORE

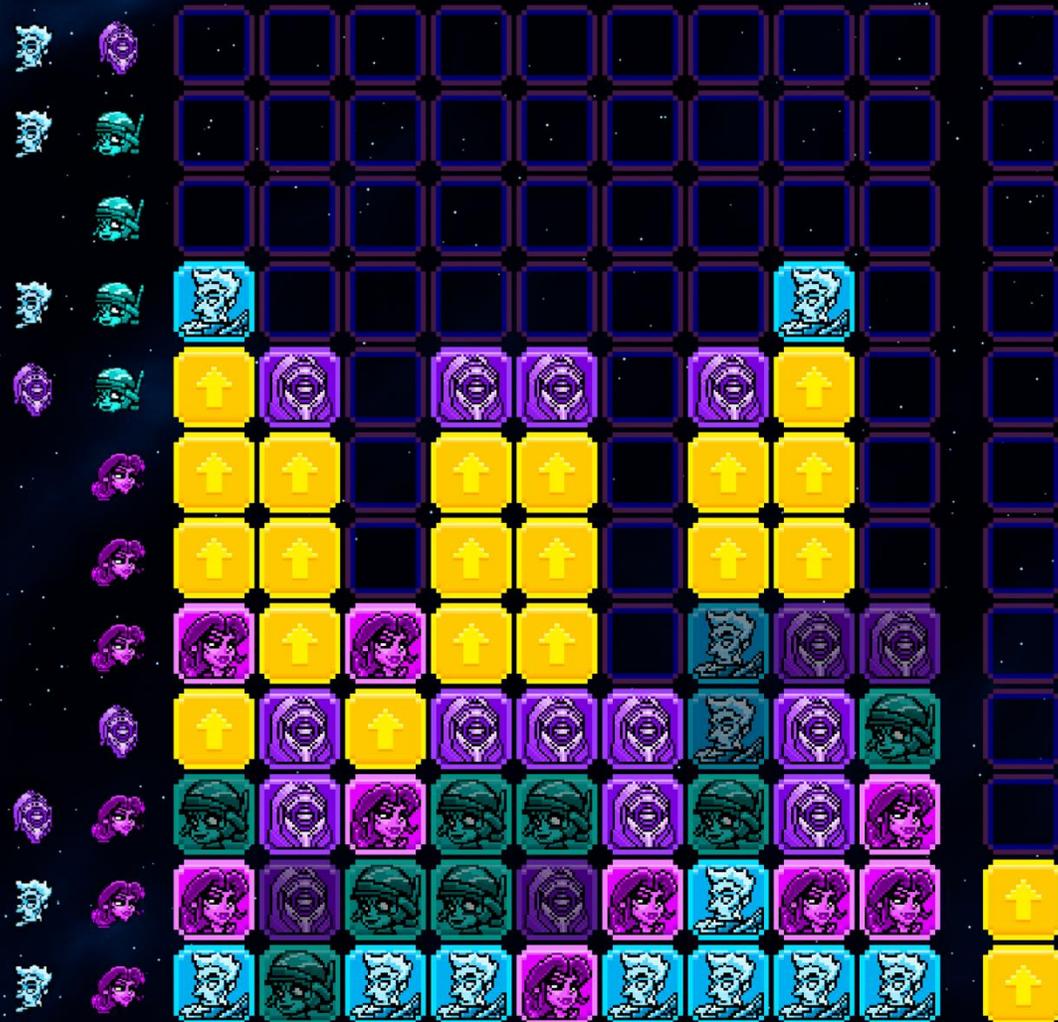
31

YOUR SCORE

31

HIGH SCORE

31



2/21

BONUS

0

SS

SU

BOOSTER SHOP

89485 

Brain Nanobots

+25% XP from combat (2h)



500 

Lucky Jabber Foot

+100% Cash acquired (2h)



500 

Caffeine Caplets

+10% Reload Speed, +15% Run Speed (2h)



500 

Jabber-Cola

+10% Damage dealt (1h)



1000 

Elemental Powder

+25% Elemental chance, +10% Elemental damage (1h)



1000 

Butt Stallion Milk

Upgrade loot quality! (1h)



59:01 
1500 



Statistiques

En quatre ans:

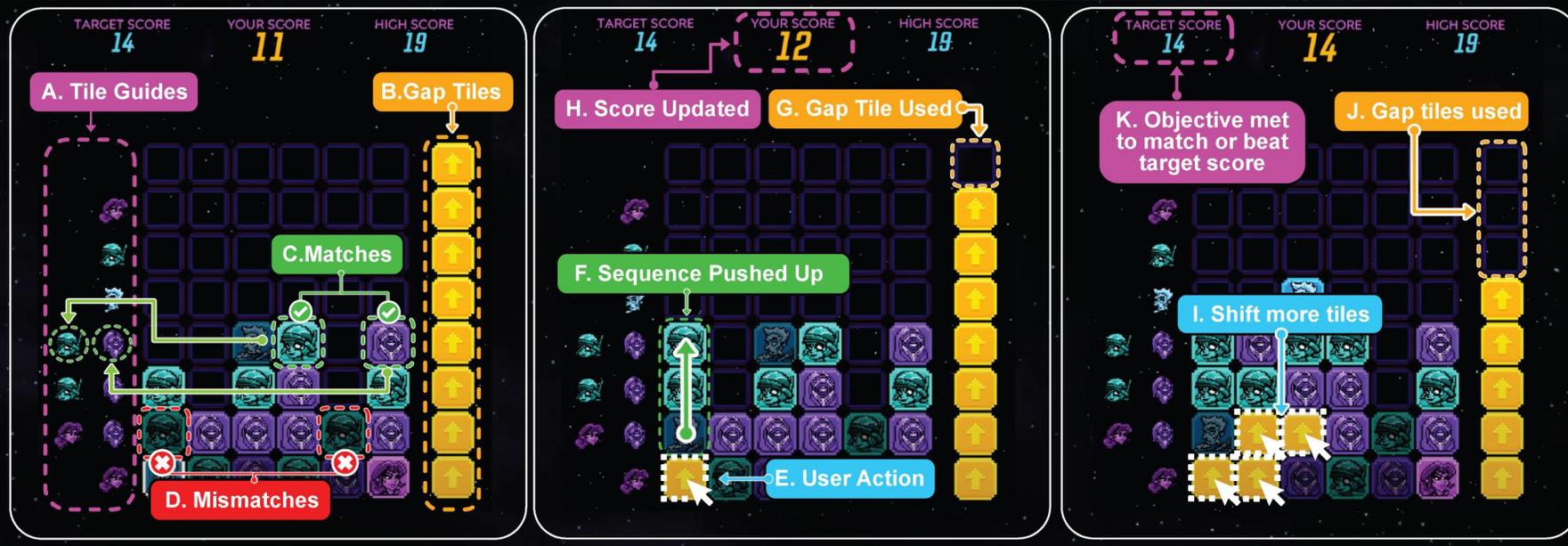
- >5,000,000 participants
- >150,000,000 puzzles résolus

**BORDERLANDS
SCIENCE**

BORDERLANDS SCIENCE: LE JEU

a.

Borderlands Science gameplay



OBJECTIFS DE **BORDERLANDS SCIENCE**

Mieux aligner les séquences du microbiome humain

- Diversité microbienne humaine ➔ infections, cancers
- Phylogénie ➔ diversité ➔ maladies
- **Améliorer les alignements** ➔ améliorer la phylogénie
- 1M de séquences du microbiome humain à aligner.

Mieux comprendre les solutions humaines

- Caractériser les différences entre les joueurs débutants et experts
- Entraîner des IA à reproduire les stratégies des joueurs



APPLICATIONS

[Gastroenterol Hepatol \(N.Y.\)](#), 2012 Mar; 8(3): 191–194.

Fecal Transplantation for the Treatment of *Clostridium difficile* Infection

[Lawrence J. Brandt](#), MD, MACG, AGAF, FASGE*

▶ [Author information](#) ▶ [Copyright and License information](#) ▶ [PMC Disclaimer](#)

PMCID: PMC3365524

PMID: [22675283](#)

Article | Published: 11 March 2020

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

[Gregory D. Poore](#), [Evguenia Kopylova](#), [Qiyun Zhu](#), [Carolina Carpenter](#), [Serena Fraraccio](#), [Stephen Wandro](#), [Tomasz Kosciolk](#), [Stefan Janssen](#), [Jessica Metcalf](#), [Se Jin Song](#), [Jad Kanbar](#), [Sandrine Miller-Montgomery](#), [Robert Heaton](#), [Rana Mckay](#), [Sandip Pravin Patel](#), [Austin D. Swafford](#) & [Rob Knight](#) 

[Nature](#) **579**, 567–574 (2020) | [Cite this article](#)

PERSPECTIVE | BIOLOGICAL SCIENCES | 



Microbiome as a tool and a target in the effort to address antimicrobial resistance

[David A. Relman](#)   and [Marc Lipsitch](#) [Authors Info & Affiliations](#)

Edited by Rino Rappuoli, GlaxoSmithKline, Siena, Italy, and approved October 11, 2018 (received for review April 19, 2018)

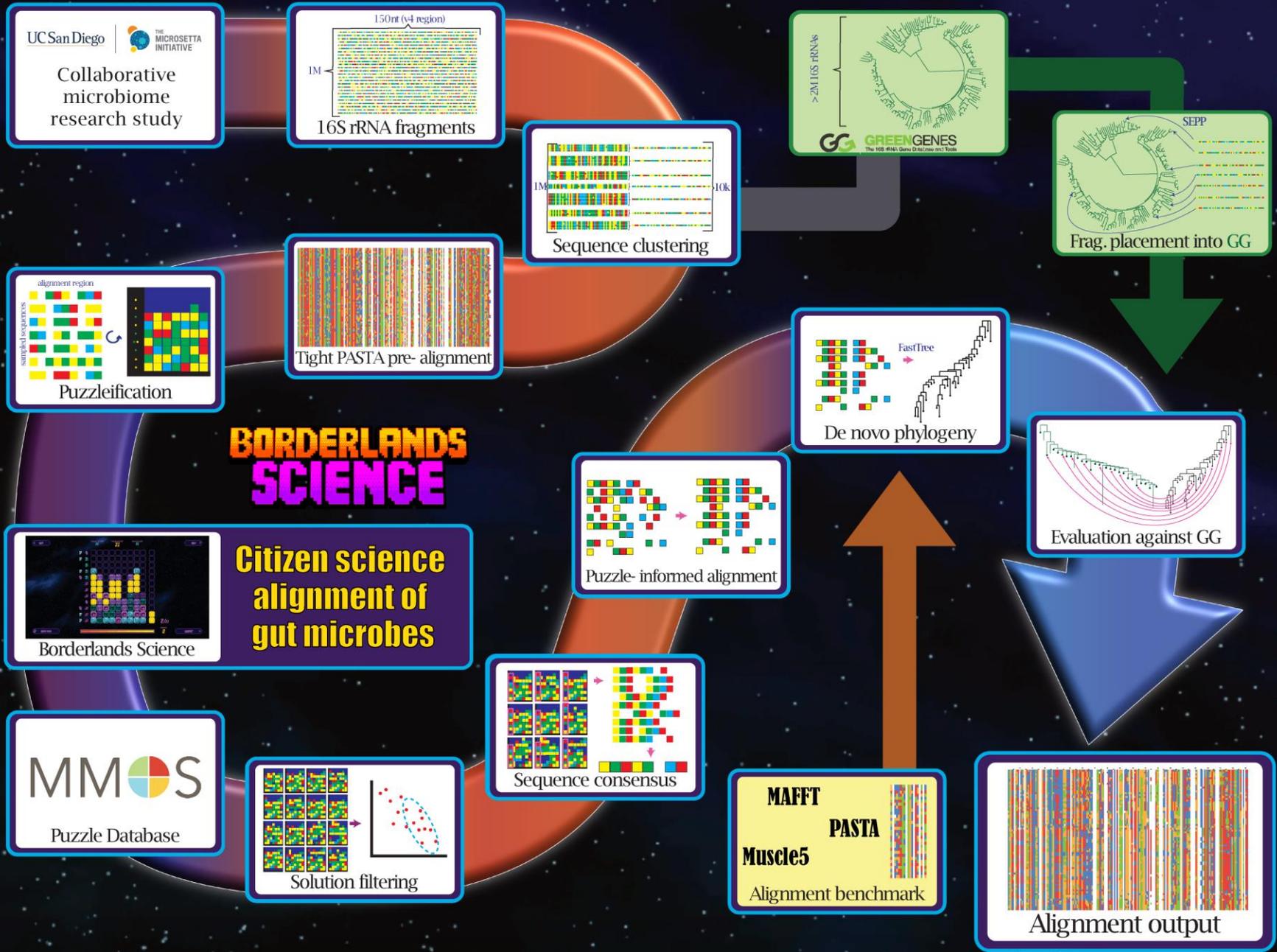
December 17, 2018 | 115 (51) 12902–12910 | <https://doi.org/10.1073/pnas.1717163115>



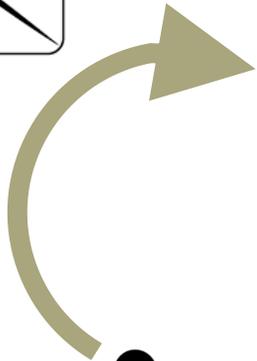
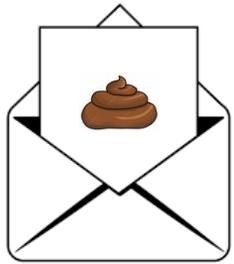
COMMENT/POURQUOI ÇA MARCHE ?

b.

Design and analysis pipeline



LE PROJET AMERICAN GUT



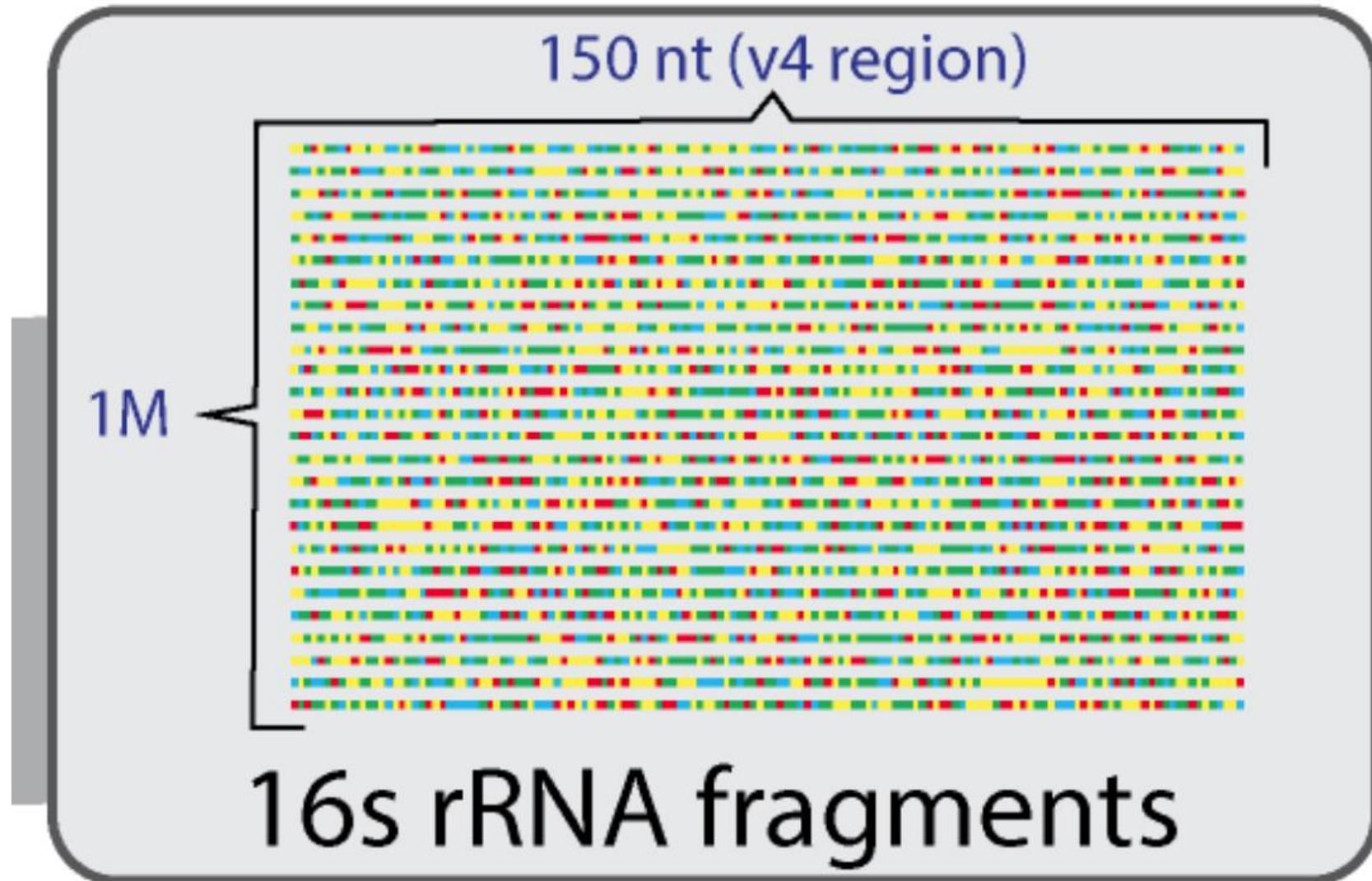
23 000 x



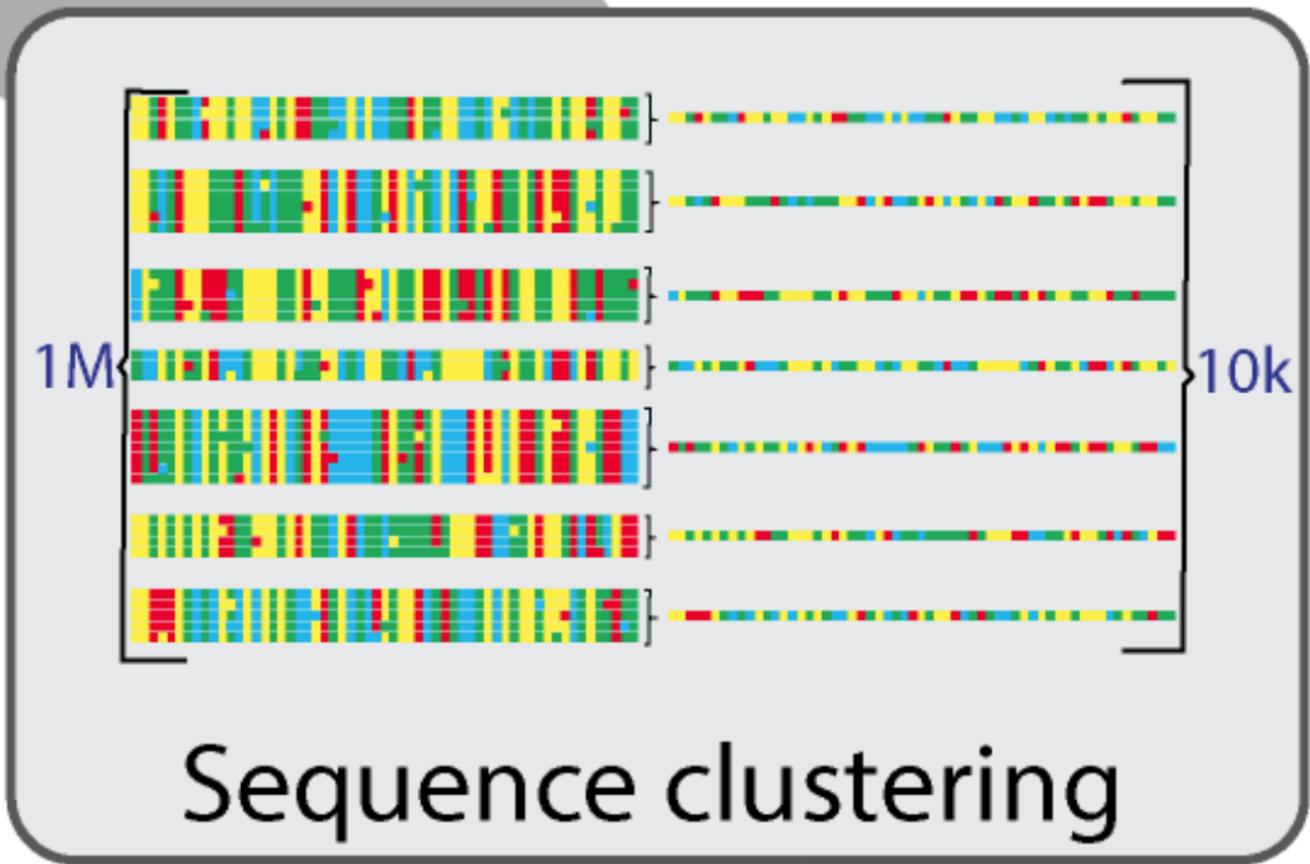
LES DONNÉES DERRIÈRE BORDERLANDS SCIENCE

1 million de séquences

Notre mission: les aligner!



CLUSTERING



Regrouper les séquences similaires pour maximiser l'efficacité du jeu

PRÉ-ALIGNEMENT

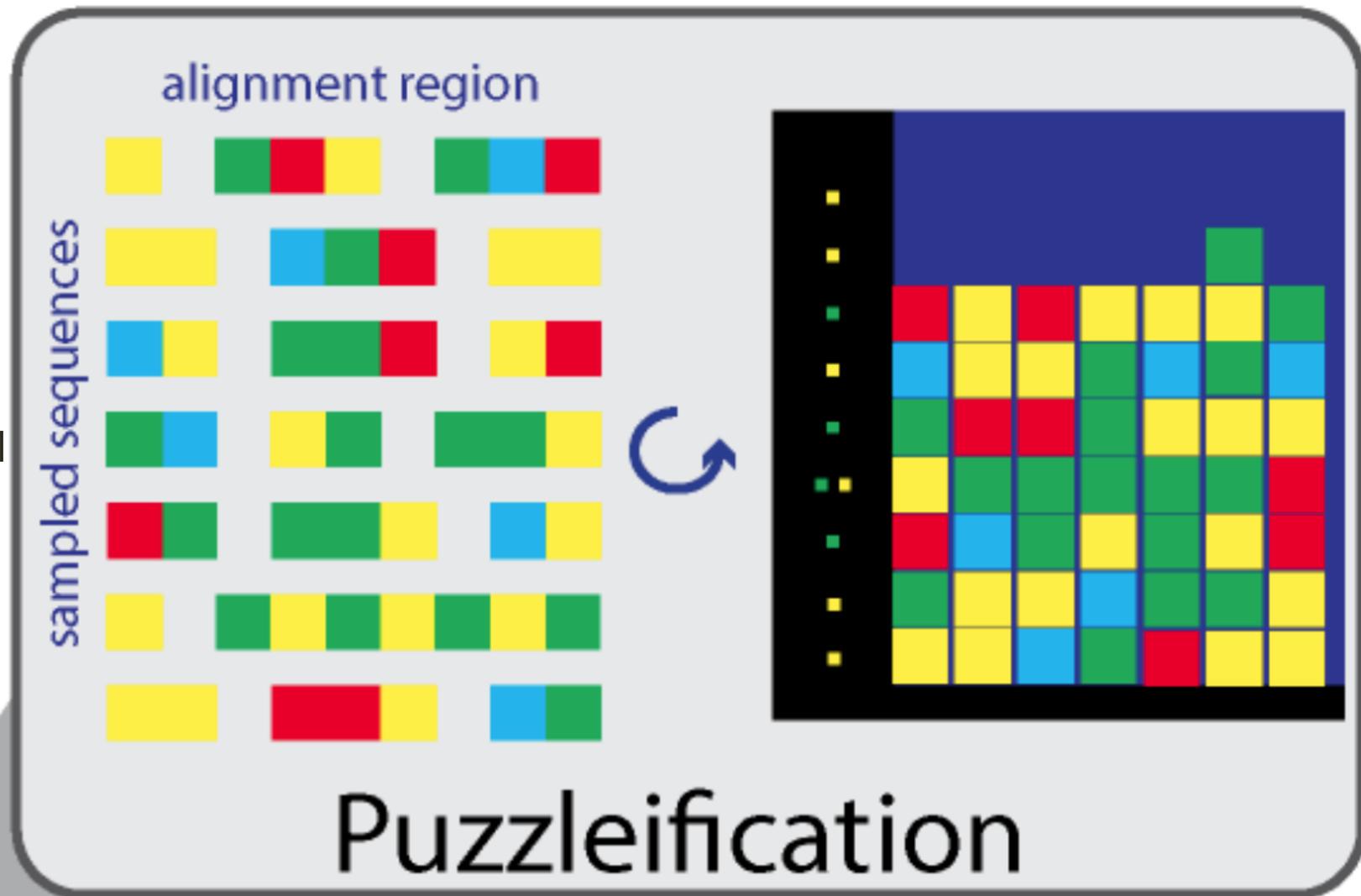
- Exécuté par un logiciel conçu pour les grands alignements
- Bon départ, mais très imparfait



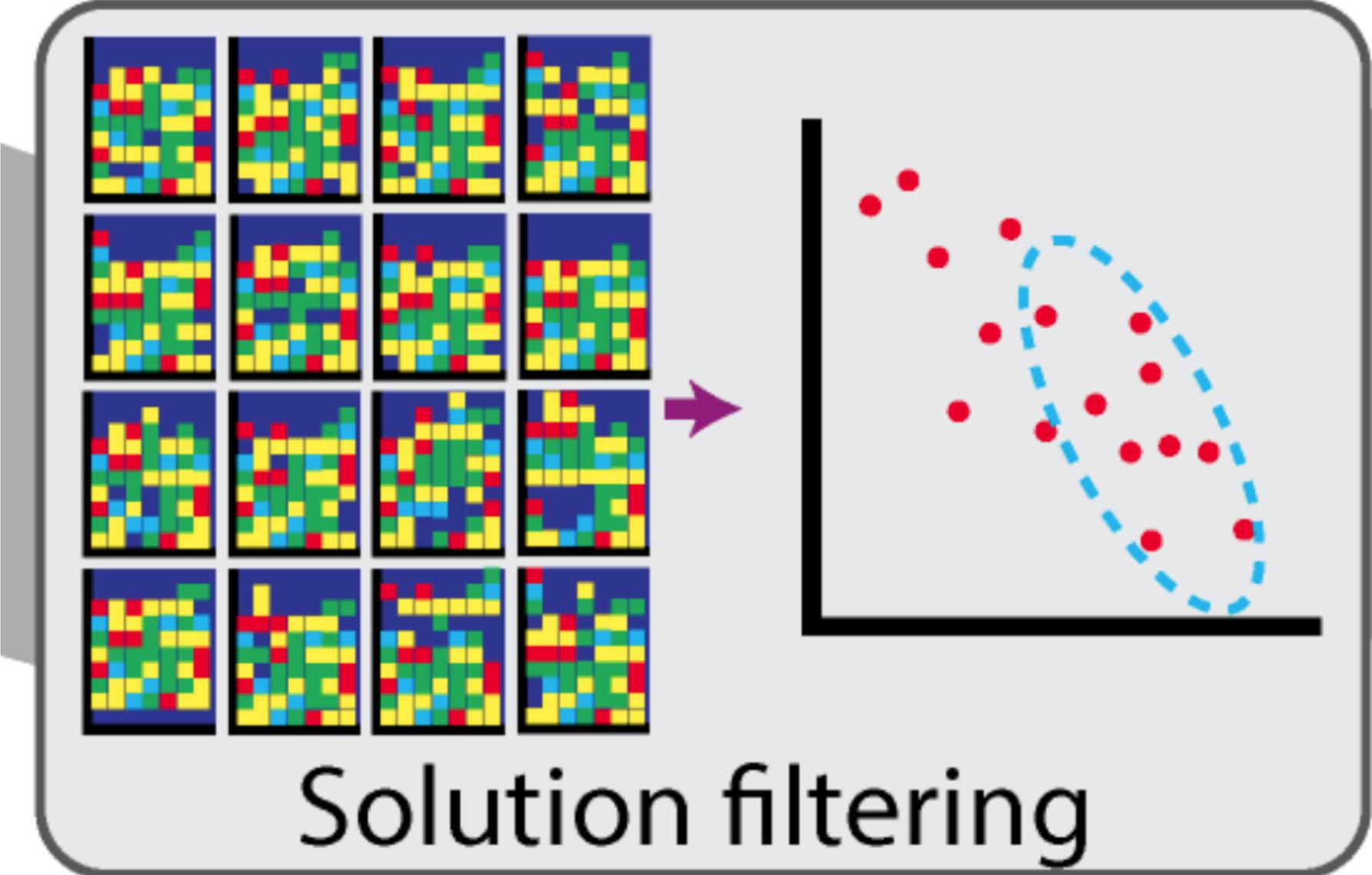
Tight PASTA pre-alignment

PUZZLE-IFICATION

Des échantillons de l'alignement initial sont transformés en puzzles



FILTRAGE DES SOLUTIONS



ÉVALUATION D'OPTIONS DE POULET



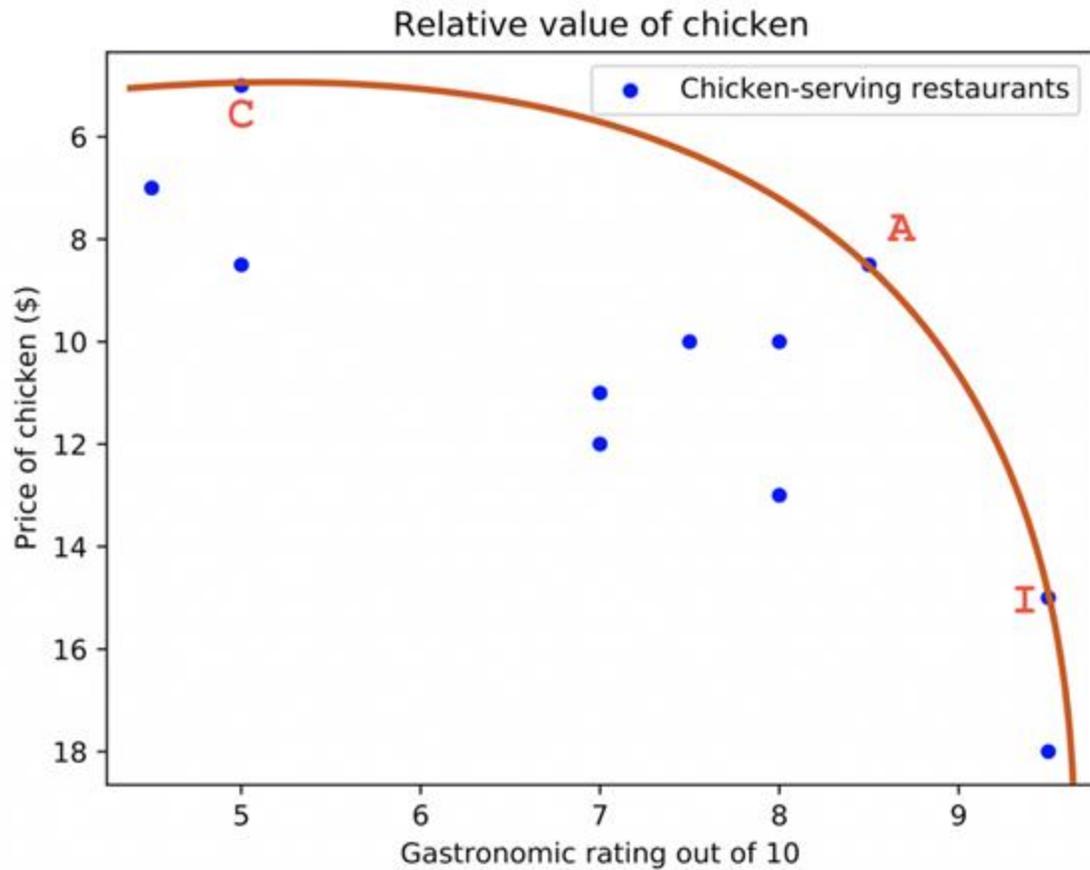
Restaurant	Prix (\$)	Évaluation gustative (/10)
A	8.50	8.5
B	12.00	7
C	5.00	5
D	18.00	9.5
E	11.00	7
F	7.00	4.5
G	10.00	7.5
H	13.00	8
I	14.00	9.5
J	8.50	5
K	10.00	8

- Moyenne des évaluations du prix et de la qualité du poulet dans 11 rôtisseries de Montréal par une équipe de 172 inspecteurs.

- Quel restaurant est le *meilleur*?

- Fonction **bi-objective!**

PARETO-OPTIMALITÉ DU POULET



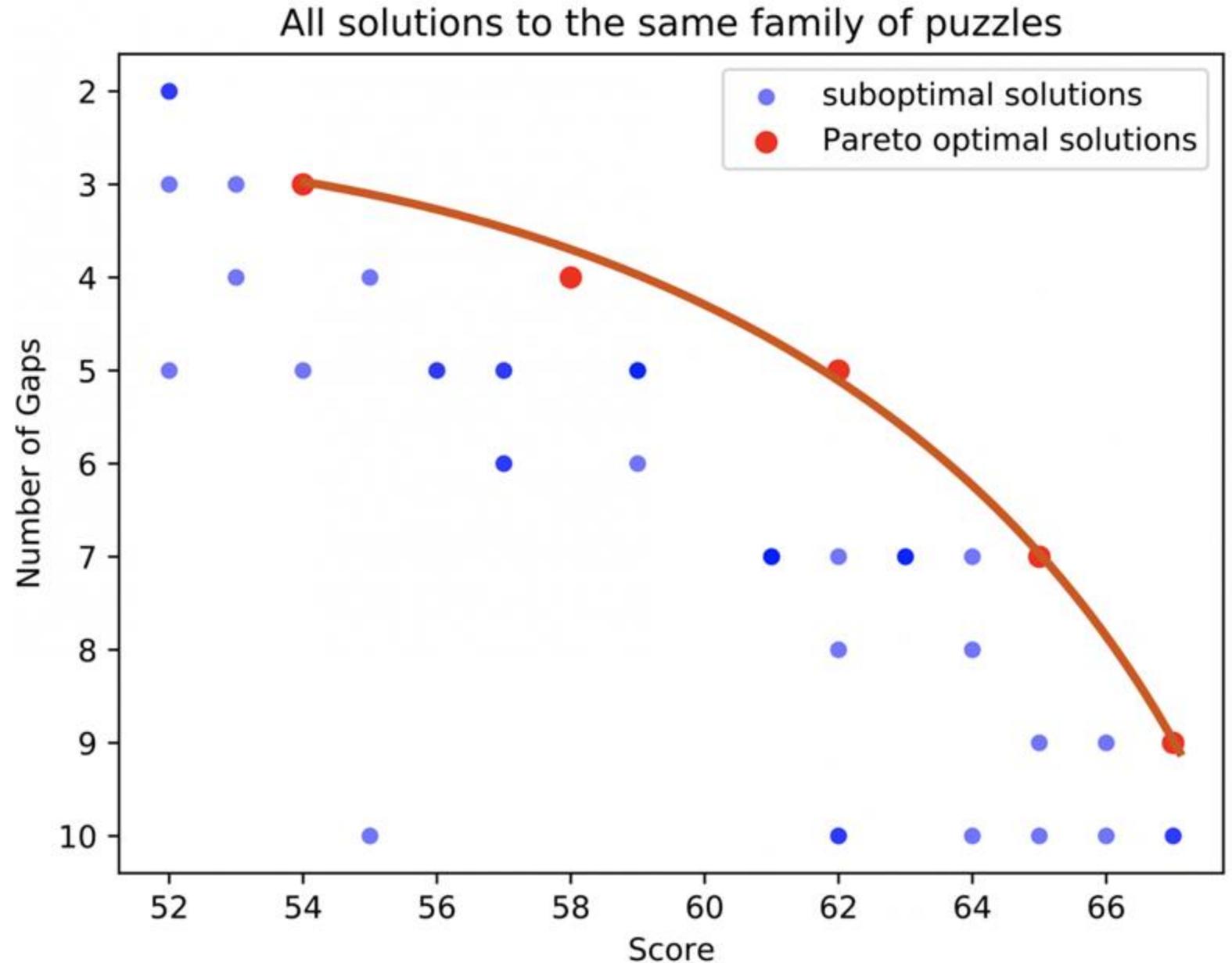
Restaurant	Prix (\$)	Évaluation (/10)
A	8.50	8.5
B	12.00	7
C	5.00	5
D	18.00	9.5
E	11.00	7
F	7.00	4.5
G	10.00	7.5
H	13.00	8
I	14.00	9.5
J	8.50	5
K	10.00	8

Pareto optimal: strictement inférieur à aucune autre option

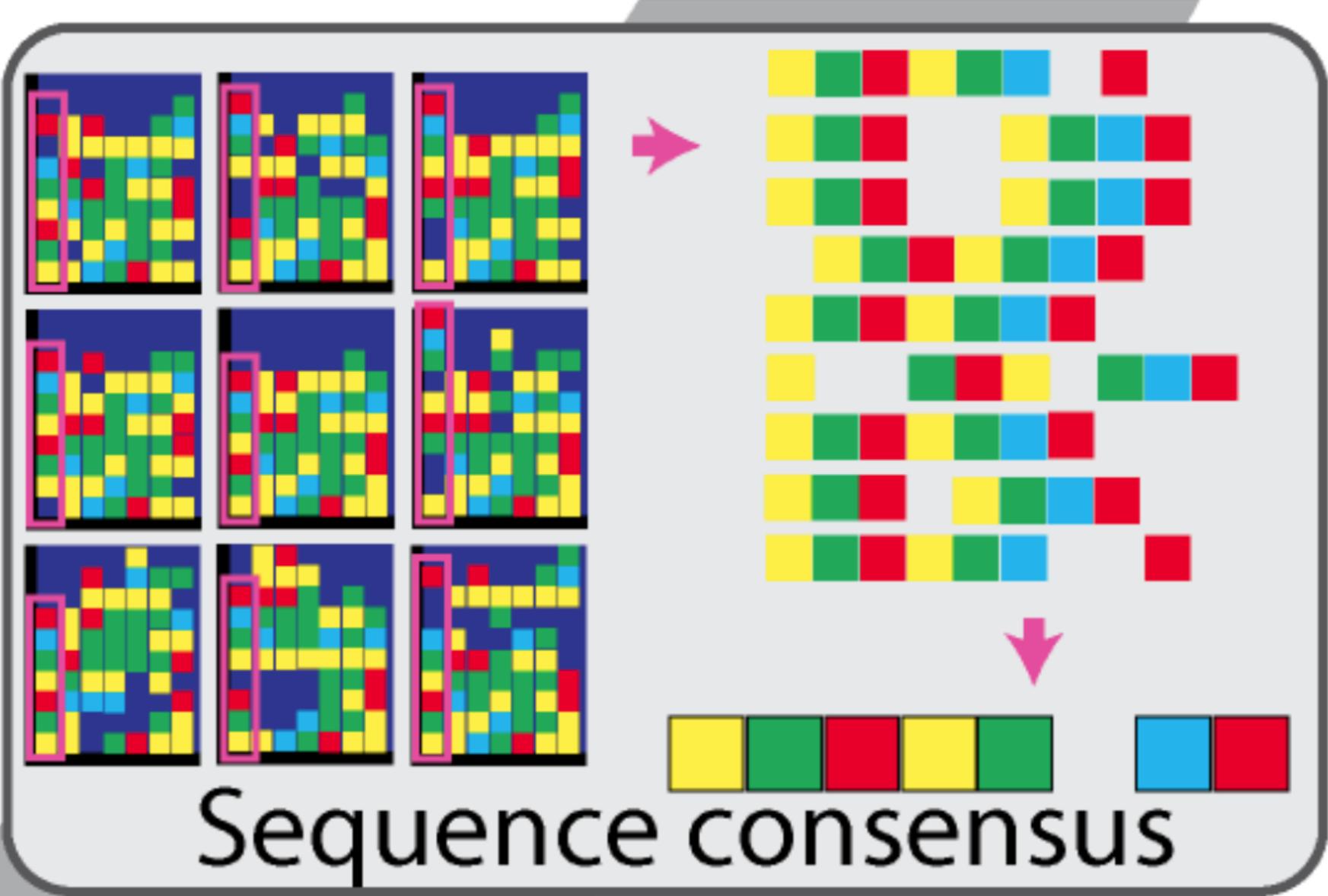
PARETO- OPTIMALITÉ DES SOLUTIONS

On cherche des solutions:

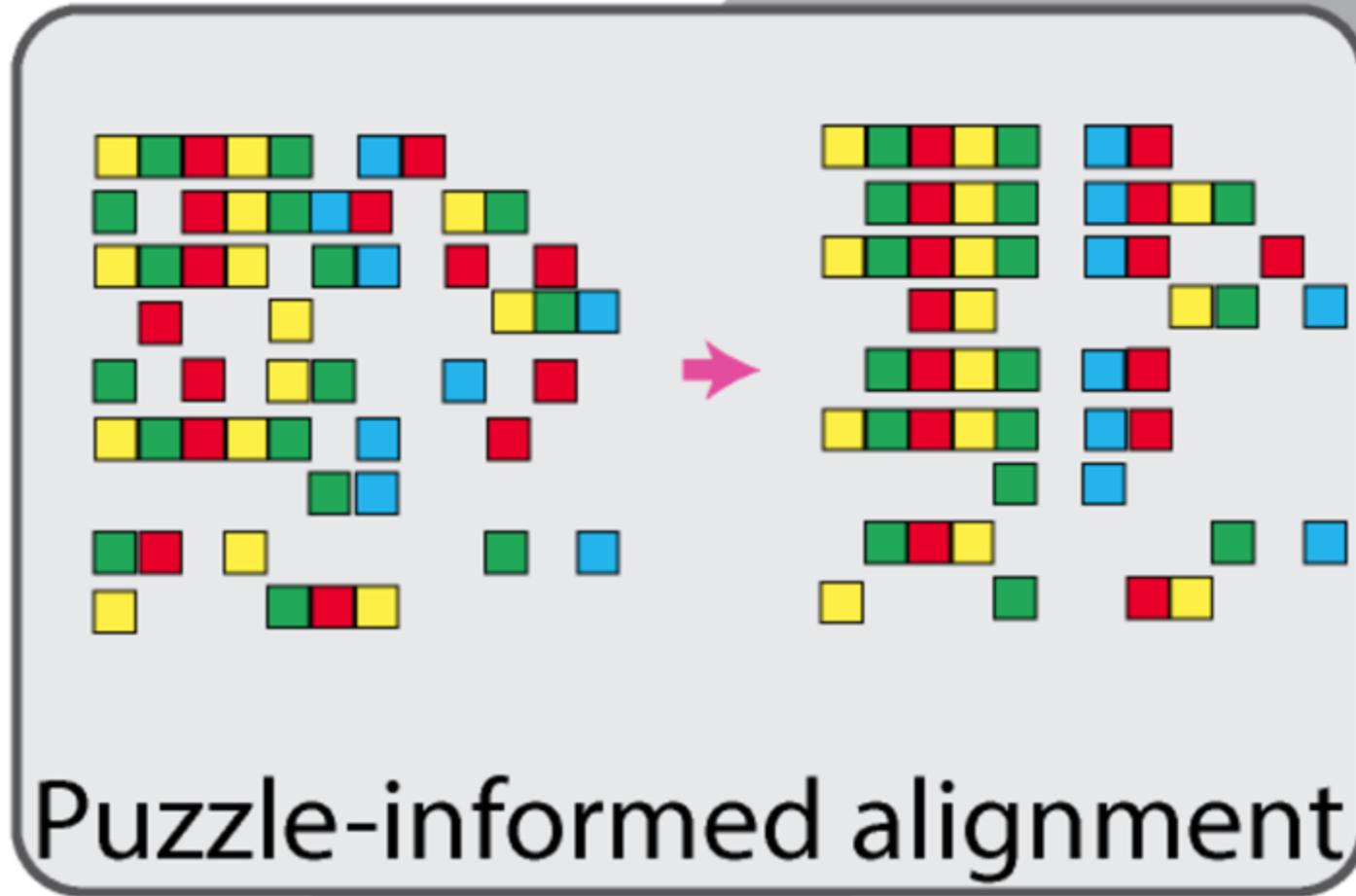
- Près du front de Pareto
- Compatibles avec le consensus des solutions des joueurs



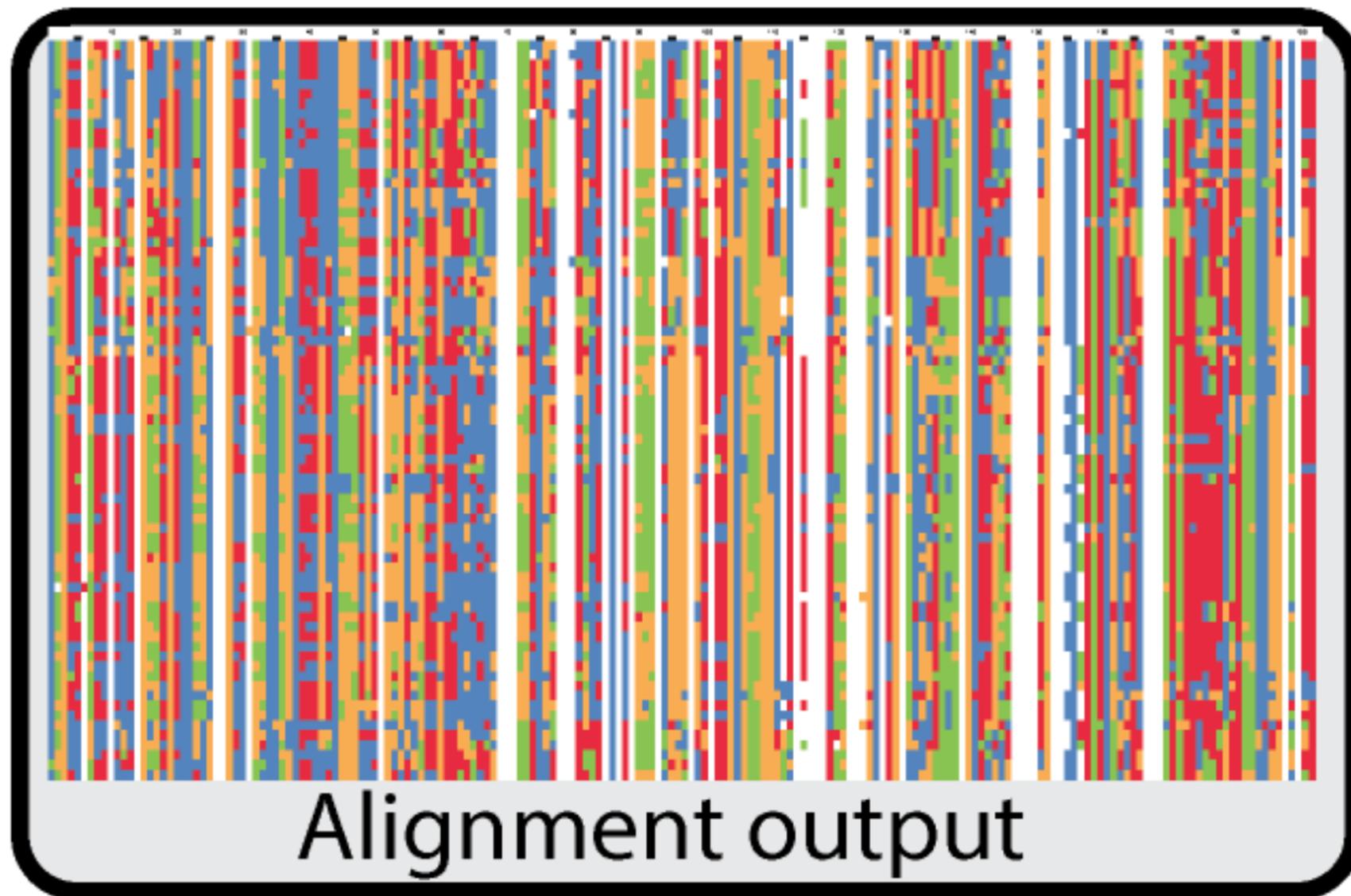
RECONSTRUCTION DES SÉQUENCES À PARTIR DES PUZZLES



ALIGNEMENT DES SÉQUENCES CONSENSUS



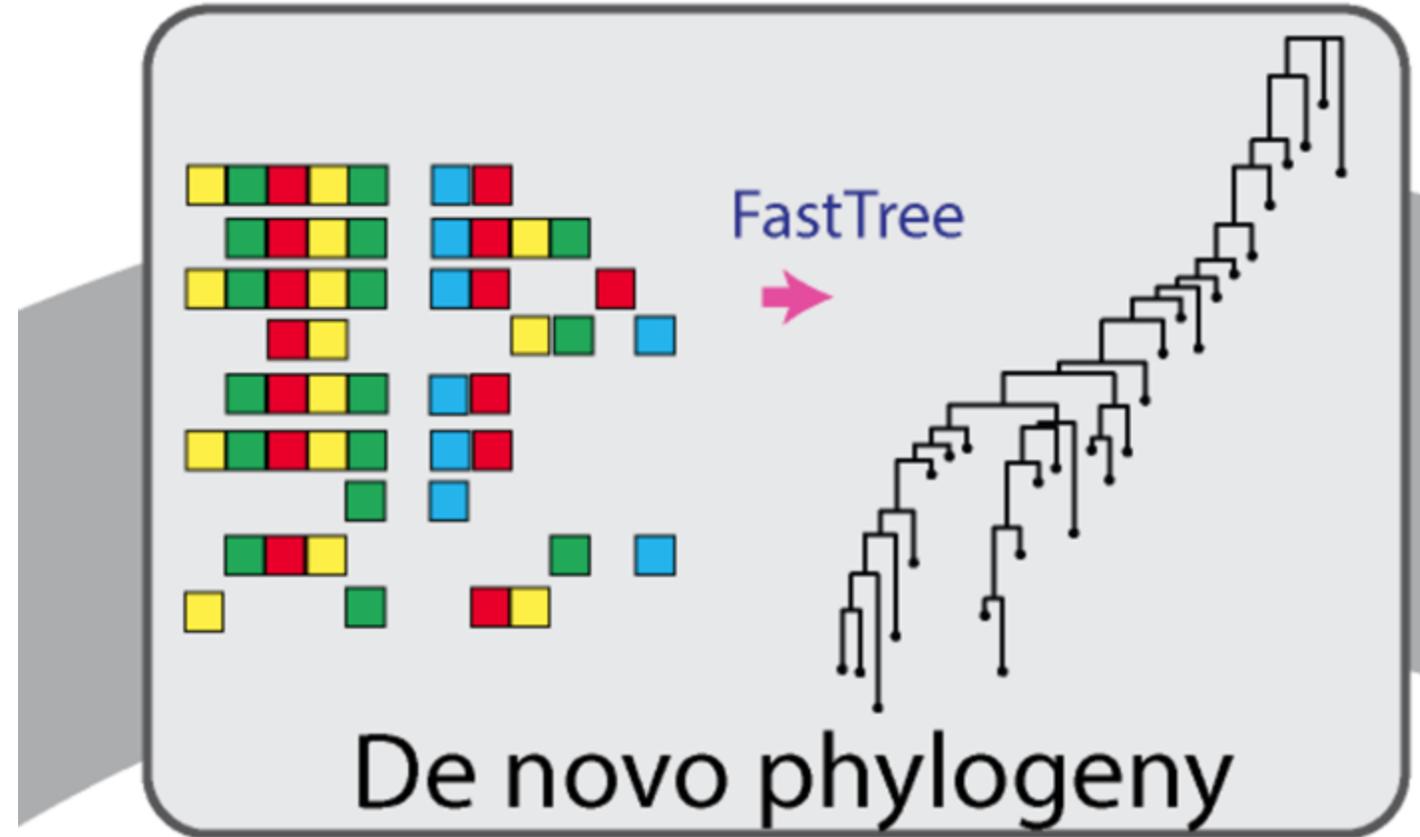
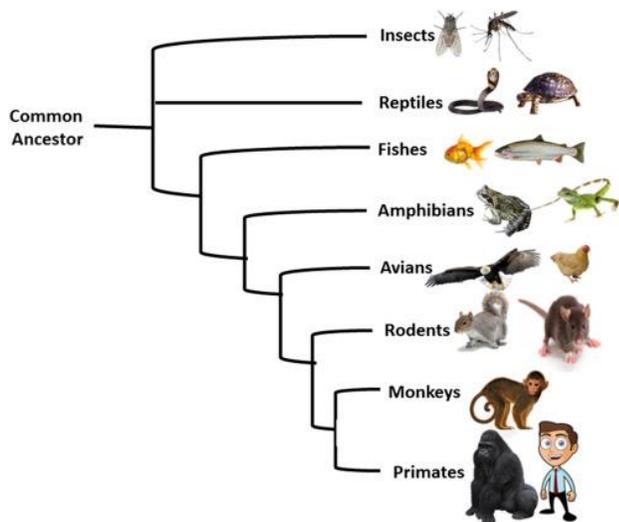
ALIGNEMENT FINAL



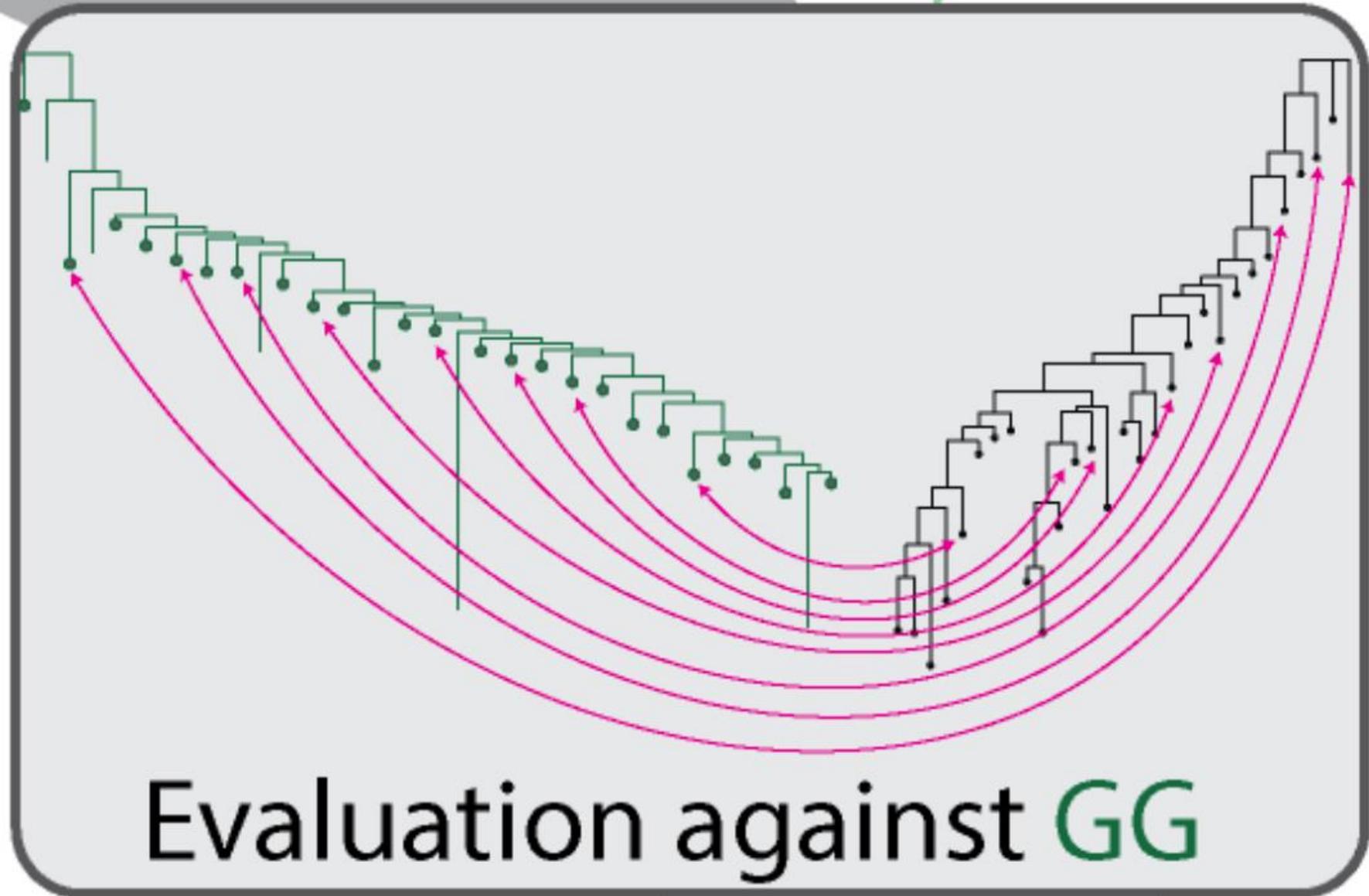
Final ?

UNE PHYLOGÉNIE ESTIMÉE À PARTIR DE BORDERLANDS SCIENCE

- Phylogénie estimée à partir des distances entre les séquences de l'alignement
- Comparée à la phylogénie de référence du sous-ensemble des microbes connus



COMPARAISON AVEC LA RÉFÉRENCE



POURQUOI AMÉLIORER LA PHYLOGÉNIE ?



Échantillon de selles

Comportement/santé de l'humain

- Brossage de dents
- Consommation d'alcool
- Types de plantes
- Antibiotiques

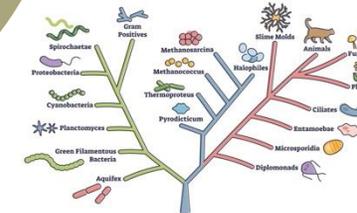


ADNs microbiens



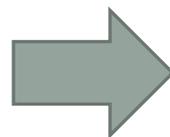
Borderlands Science

Prédiction



Phylogénie

Échantillon de selles



Borderlands Science

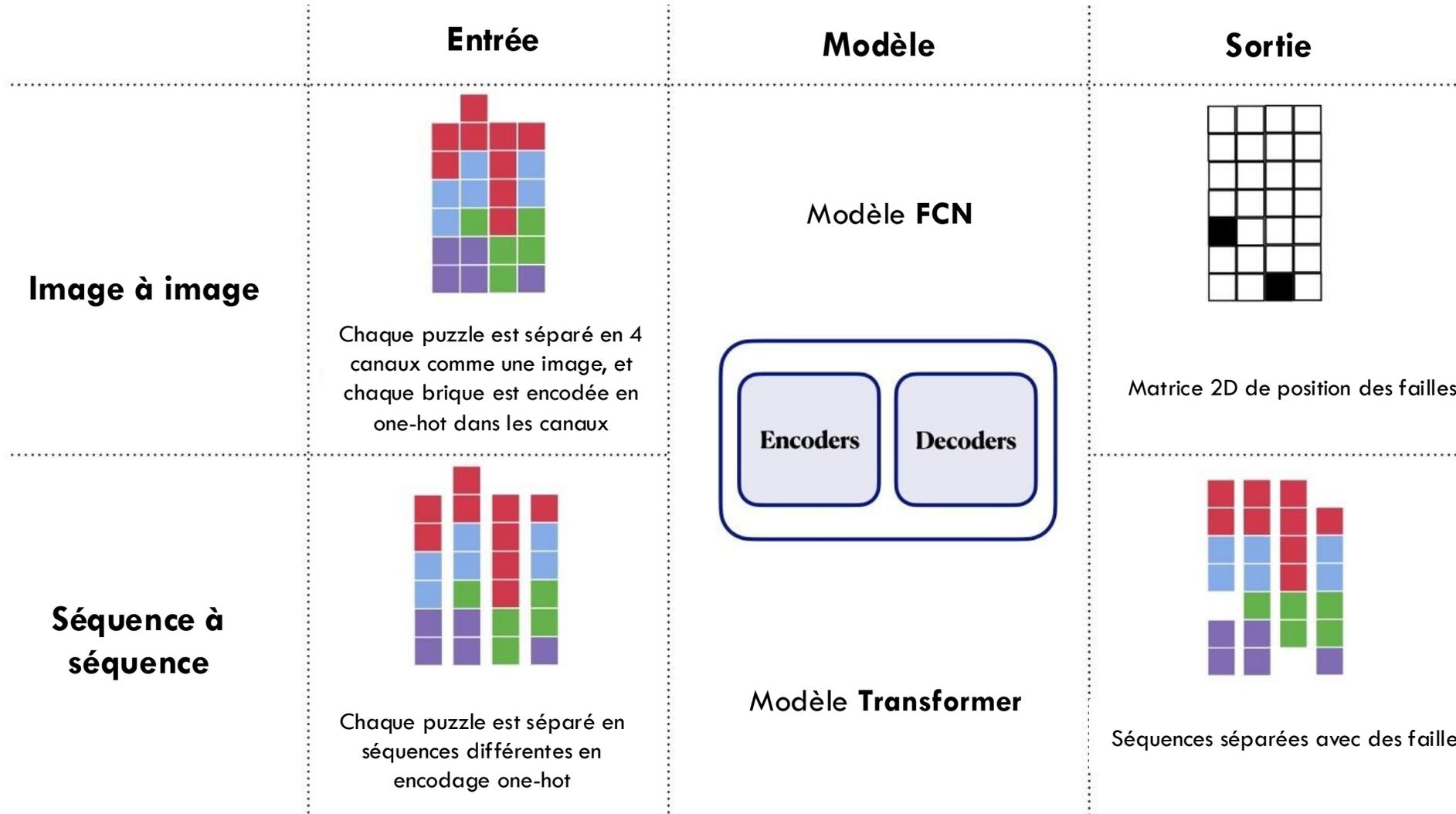


Information sur l'auteur des selles

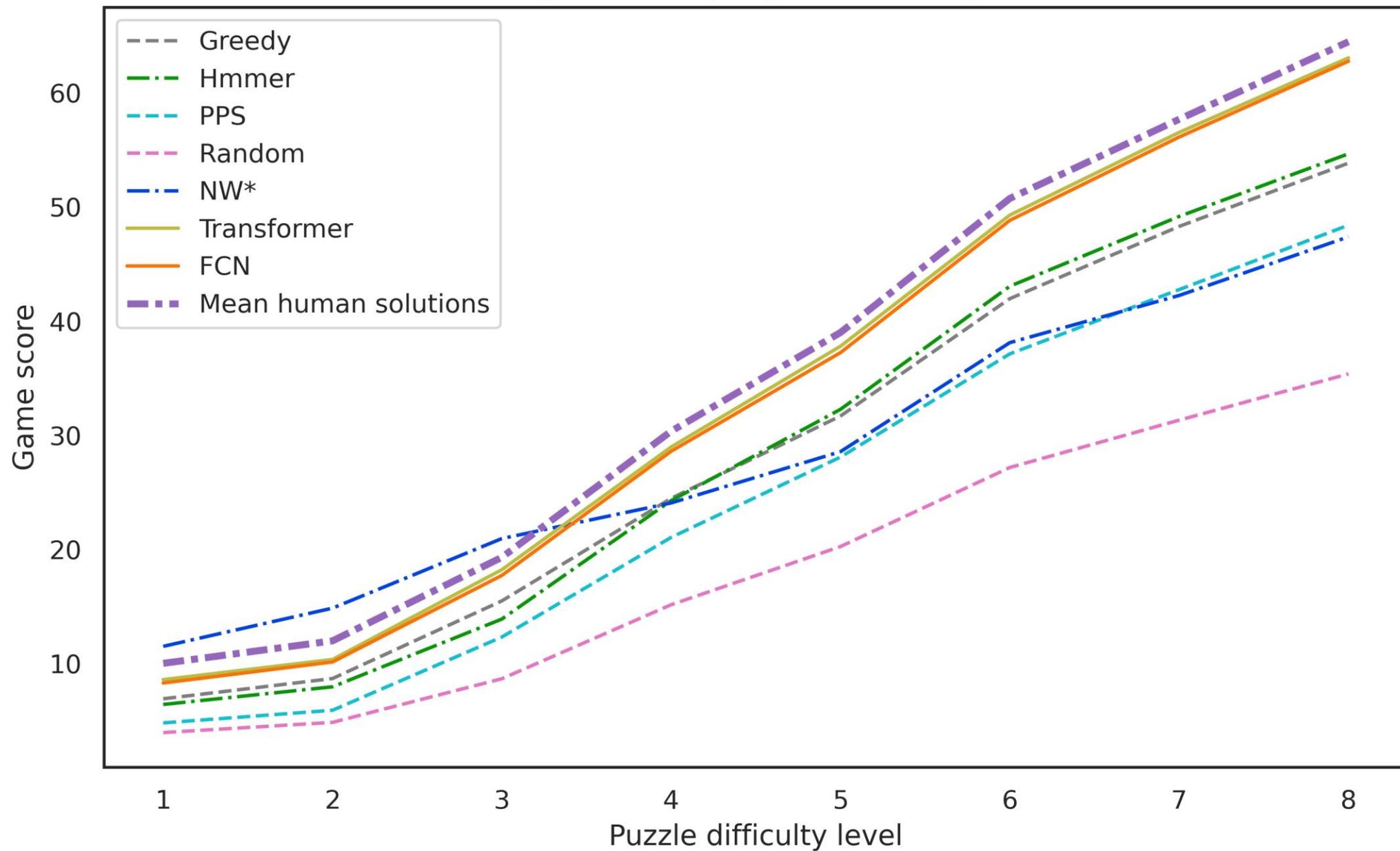


UTILISATION DES DONNÉES GÉNÉRÉES

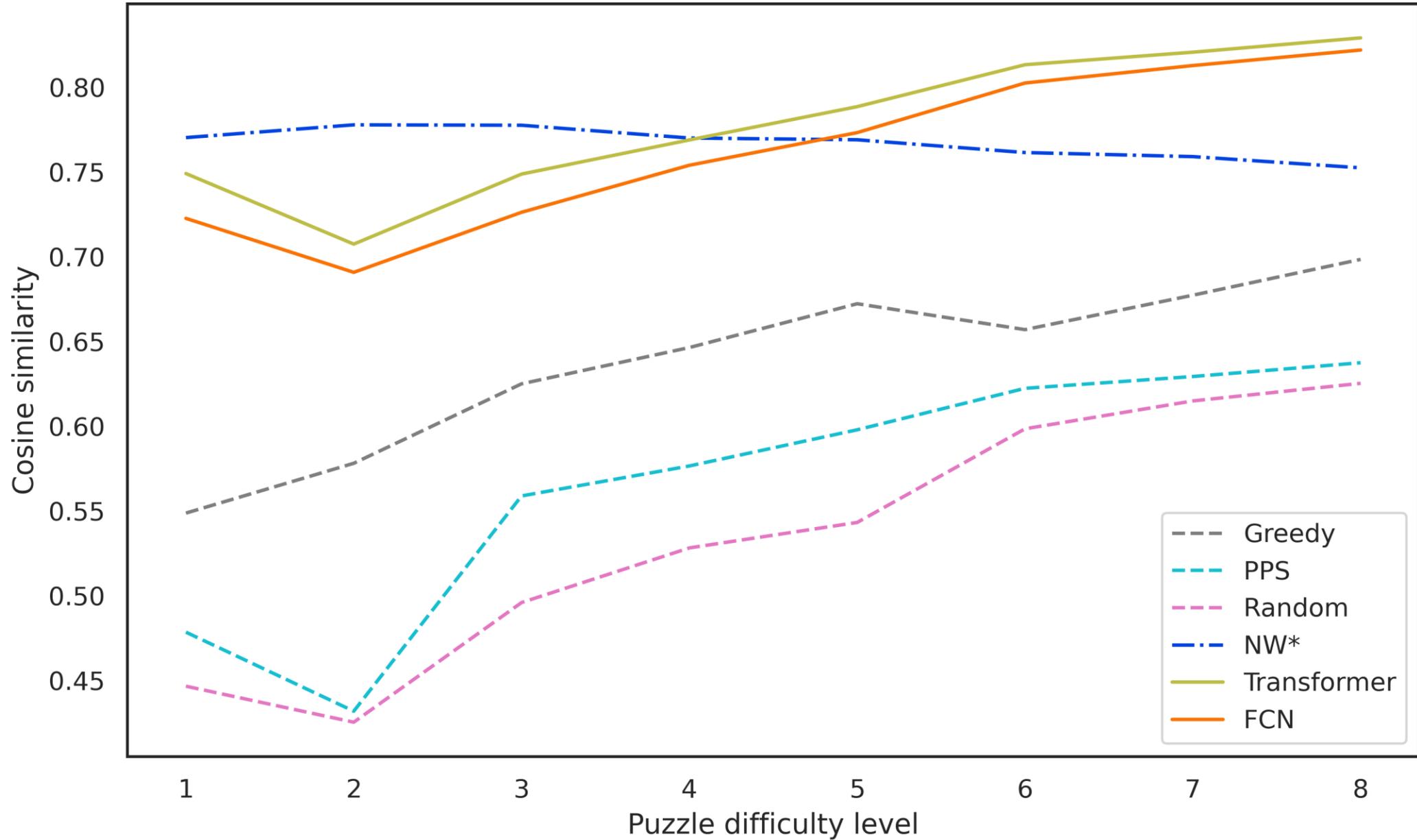
APPRENDRE DES STRATÉGIES DES JOUEURS



Mean game score per puzzle difficulty level



Mean cosine similarity to human solutions per puzzle difficulty level



CONCLUSIONS — APPRENDRE DES STRATÉGIES DES JOUEURS

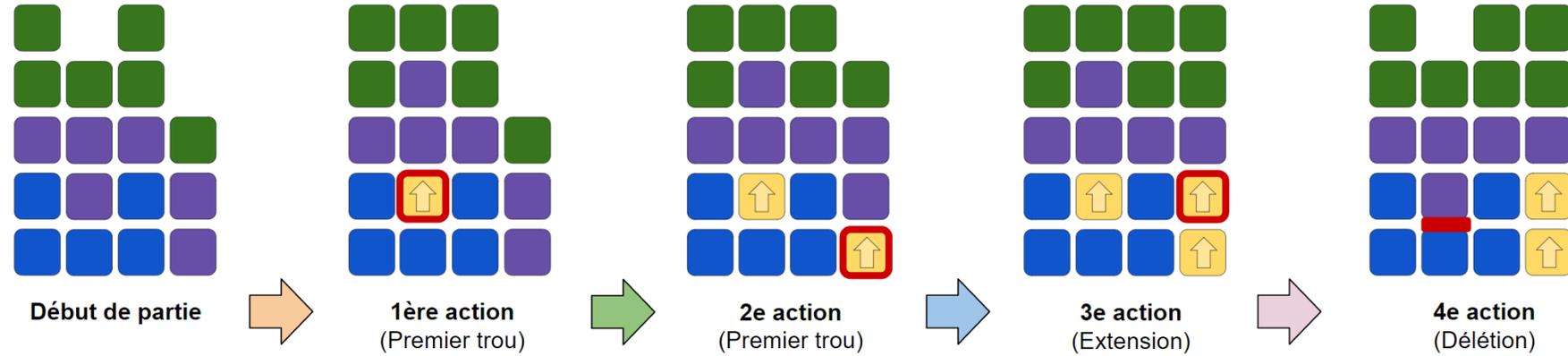
On peut entraîner des agents pour reproduire les solutions humaines

Ces agents sont plus similaires aux humains qu'aux algorithmes traditionnels

Les agents produisent de bonnes solutions aux puzzles (hauts scores)

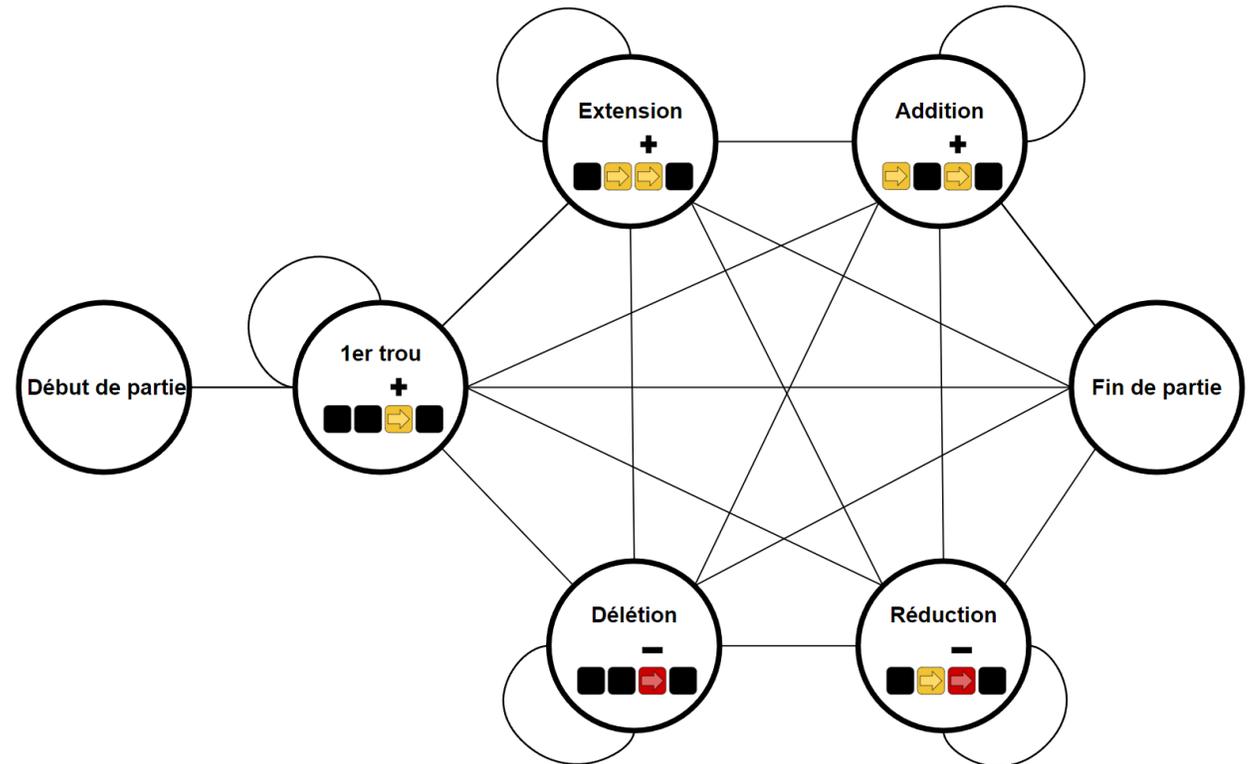
En théorie, on peut donc appliquer ces agents à de nouvelles données

DÉCRYPTER LA CONTRIBUTION HUMAINE



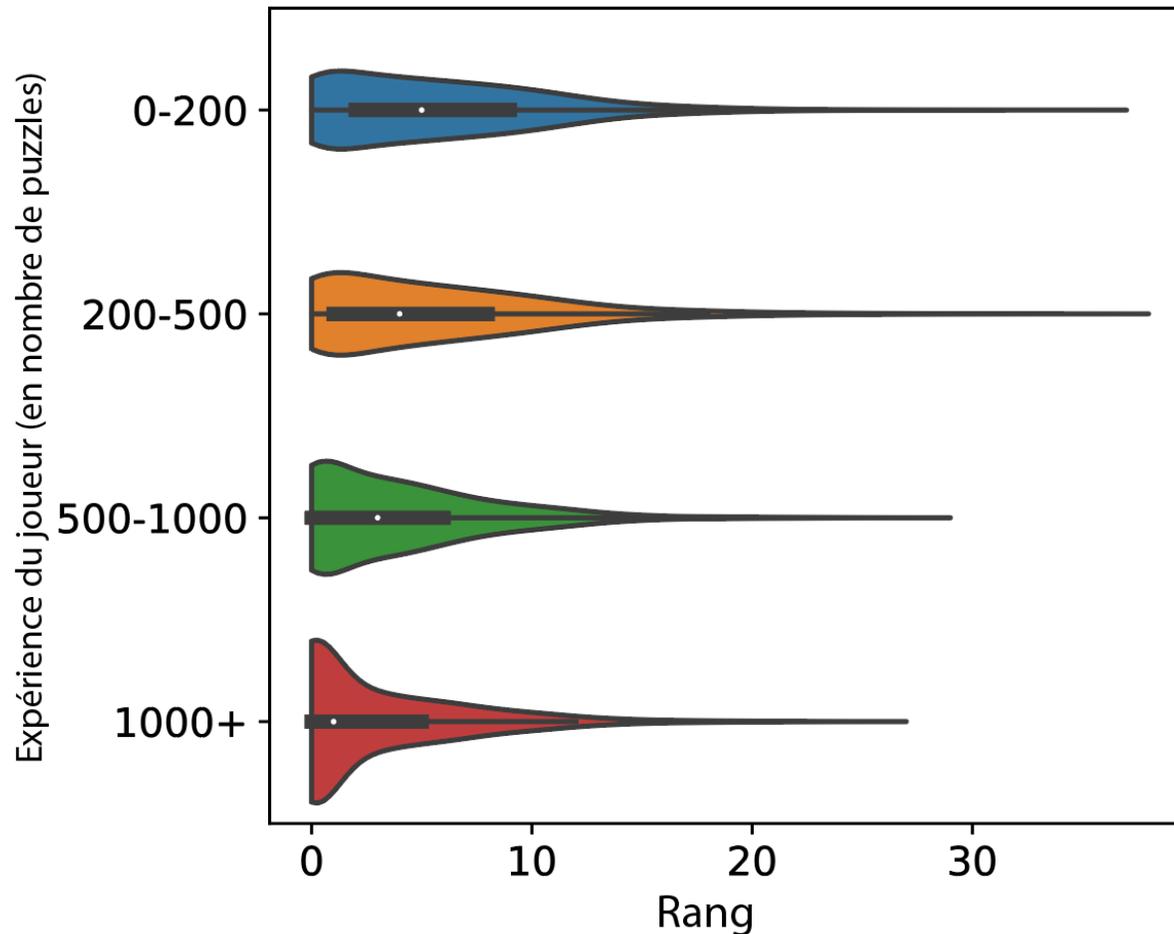
Modéliser la séquence de décisions d'un joueur via modèles Markoviens

Mais est-on sûrs qu'il y a quelque chose à apprendre ?



DÉVELOPPEMENT D'UNE EXPERTISE

Rang des solutions en fonction de l'expérience des joueurs



Plus un joueur joue, plus ses solutions sont hautes dans le classement... donc ils apprennent !

CONCLUSIONS — COMPRENDRE LES STRATÉGIES DES JOUEURS

Les joueurs s'améliorent en jouant

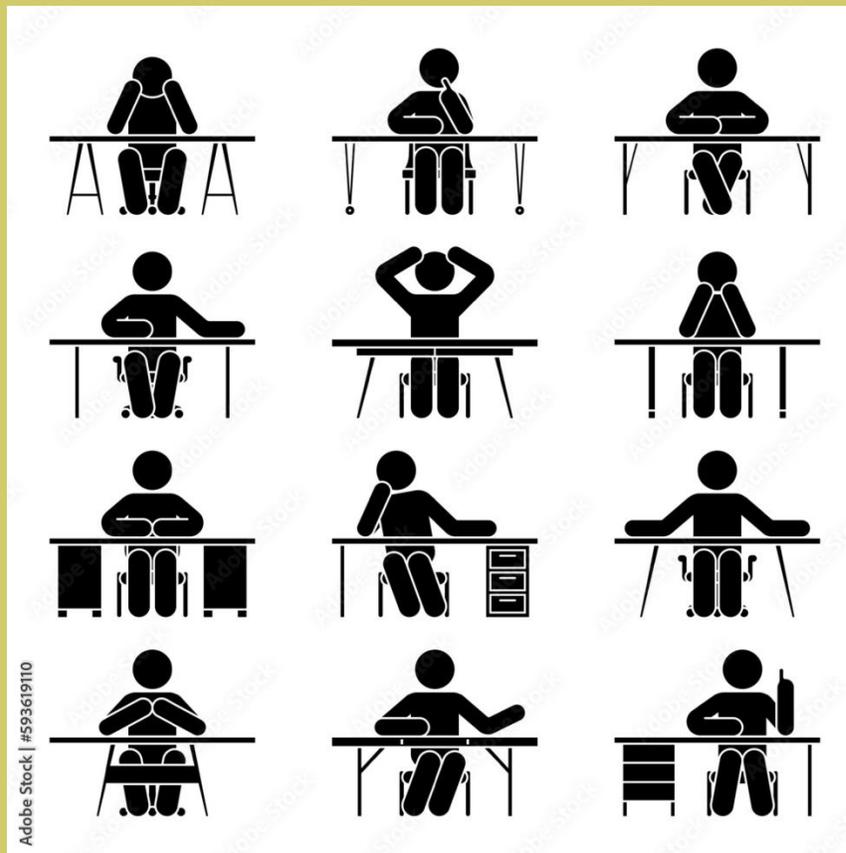
Les experts jouent différemment des débutants

Les experts trouvent des meilleurs coups et jouent plus longtemps

Les non-experts contribuent tout de même significativement à la résolution du problème

POURQUOI LA BIOLOGIE ?

- ❑ Pas besoin de faire des jeux en rapport avec la biologie
- ❑ En pratique, la bio contient plein de problèmes visuels et plein de données partielles et déséquilibrées
- ❑ Il est essentiel d'avoir une expertise dans les problèmes ciblés indépendamment des jeux pour pouvoir en valider les résultats avec les méthodes d'état de l'art du champ de recherche.
- ❑ Mon background est en bio, donc je fais de la bio, mais en théorie on peut travailler sur n'importe quoi à condition d'avoir des collaborateurs avec de l'expertise sur le sujet.



PROJETS EN COURS

- Comment concevoir et développer un jeu vidéo amusant ? **jeu vidéo**

- Comment répartir le travail entre joueurs ?
- Comment agréger des millions de solutions ?

calcul humain

- Comment bien choisir une tâche à transformer en jeu ?
- Comment démontrer la qualité du résultat ? **bioinformatique**

- Comment reproduire les solutions des humains ?
- Comment capturer l'hétérogénéité des stratégies ?
- Comment appliquer les modèles entraînés ?

apprentissage automatique

- Comment distinguer un débutant d'un expert ? **informatique cognitive**
- Qu'est-ce que les joueurs font mieux que les algos ?

JE CHERCHE

- Des étudiants en informatique à **tous les cycles** qui sont intéressés par:
 - La biologie et/ou
 - Le développement de jeux vidéo et/ou
 - Le potentiel de contribution humaine à la résolution de problèmes et/ou
 - L'analyse du comportement humain dans le jeu vidéo
- Pas besoin de connaître la biologie!
- La date limite pour le BRPC est **demain**

DNADLE — UN NOUVEAU JEU SÉRIEUX

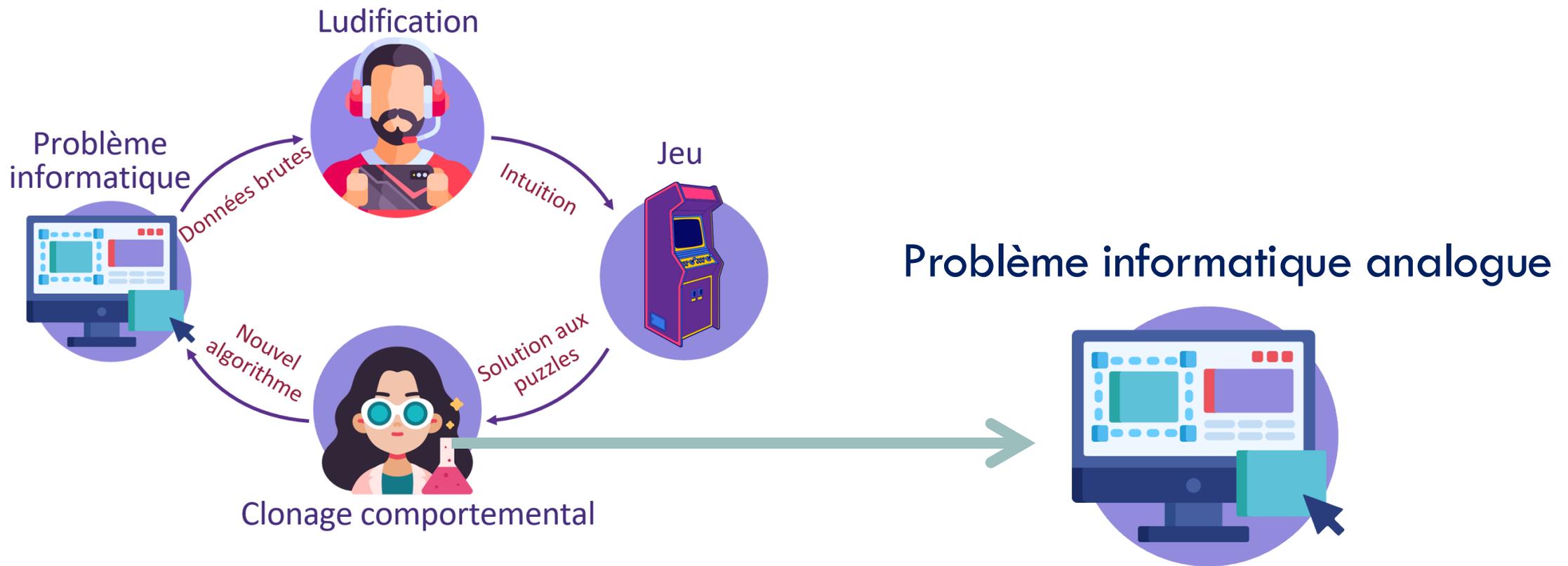


DNADle:

Mini-jeu d'identification de motifs d'ADN/ARN développé à l'UQAM.

But: évaluer la performance des humains dans l'identification de motifs

NOUVELLES APPLICATIONS D'AGENTS ENTRAÎNÉS SUR BORDERLANDS SCIENCE



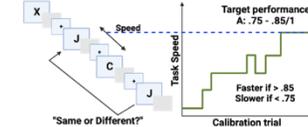
L'HUMAIN ET LE JEU: PERSPECTIVE INFORMATIQUE

- L'utilisateur choisit un jeu de son choix
- On applique de la chaleur sur sa peau
- On évalue son niveau de douleur en jouant, en ne jouant pas, et avec un jeu poche comme contrôles.
- Méthodes statistiques pour déterminer si le jeu vidéo génère de l'analgésie, et comment.
- Nouveau projet similaire avec des exercices d'échecs, dans une machine IRM.
- Mon objectif: déterminer si les joueurs arrêtent d'essayer si le problème est trop dur.

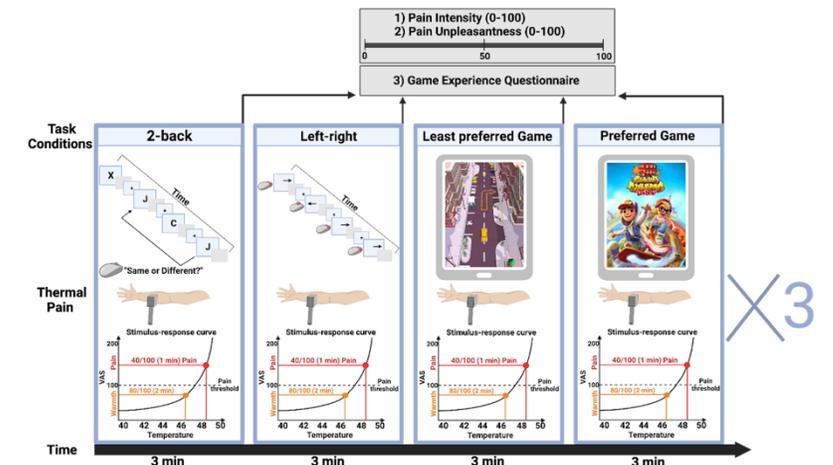
A) Prior to the experiment



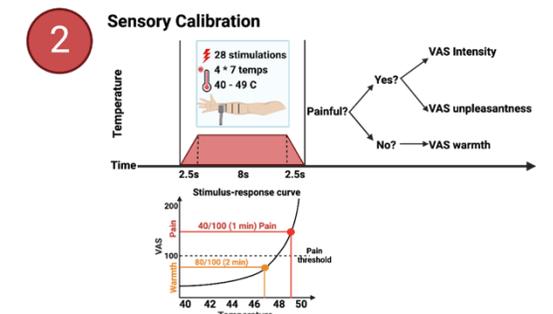
3 Cognitive Calibration



5 Main Task



B) The experiment session

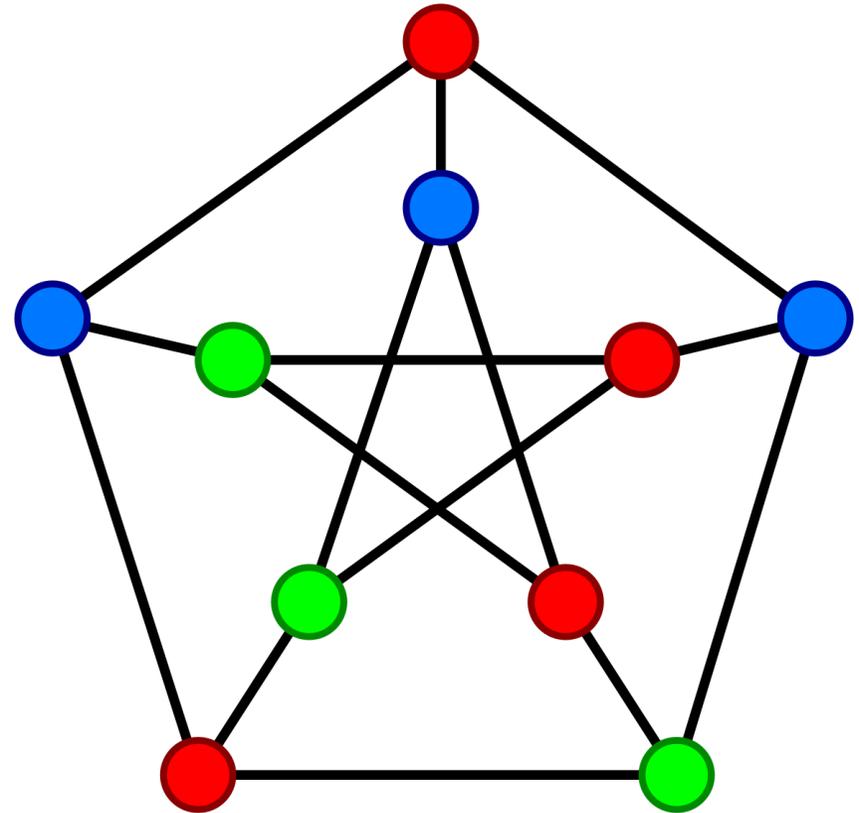


4 Questionnaires:

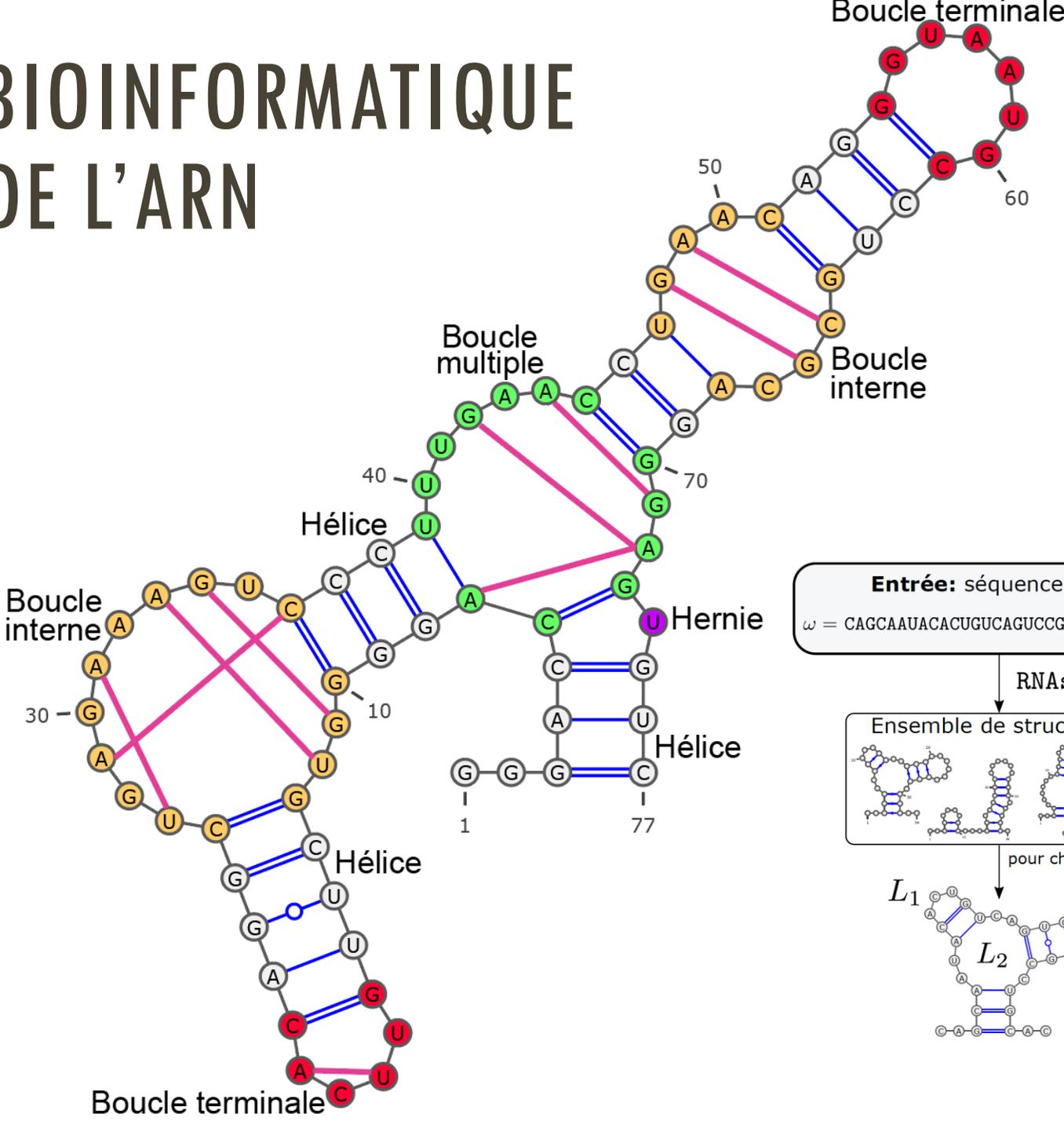
- Pain Catastrophizing questionnaire (PCS)
- Big five inventory (BFI)
- Five Facet Mindfulness Questionnaire (FFBQ)

DESIGN DE NOUVEAUX JEUX SÉRIEUX

- Jeu web de coloriage de graphe
- Objectif: évaluer la performance des humains
- Enjeux:
 - Possibilité de bouger les noeuds pour mieux voir
 - Coloriage intuitif minimisant le nombre de clics
 - Ludification

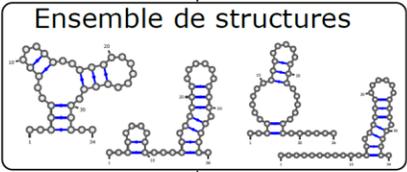


BIOINFORMATIQUE DE L'ARN

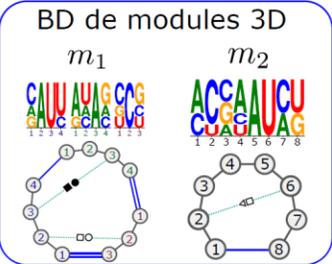
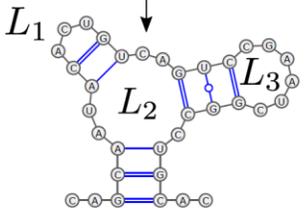


Entrée: séquence ARN
 $\omega = \text{CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC}$

RNAsubopt

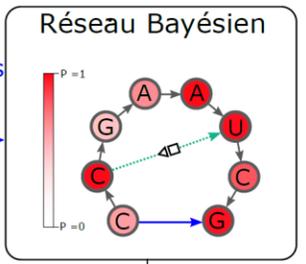


pour chaque structure de l'Ensemble



$m_1(4; 4; 3)$ and $m_2(8)$

fréquences des bases



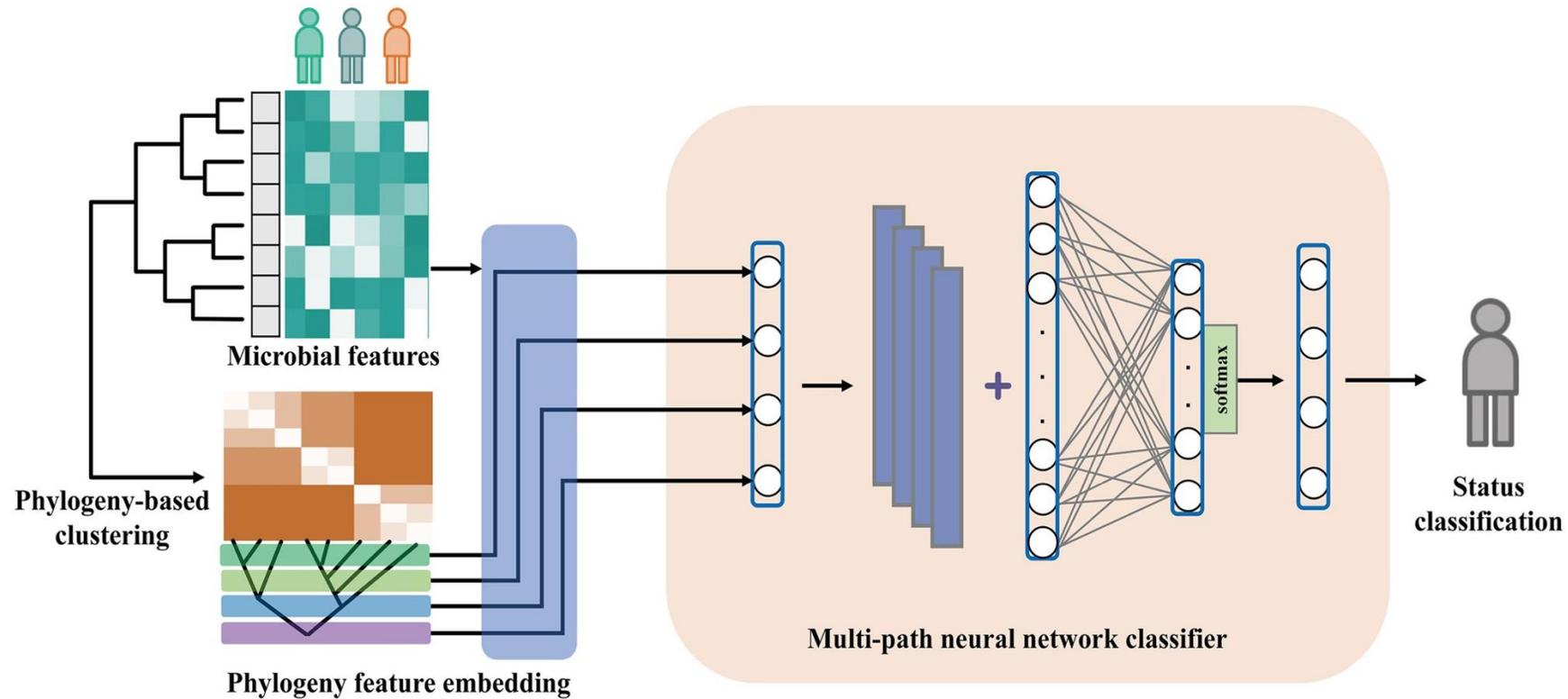
$\mathbb{P}(m_2 | \omega)$

Sortie: score probabilistique
 $MS = \log[\mathbb{P}(m_2 | \omega)\mathbb{P}(L_3 | \omega)^\lambda]$

correspondance de taille

$L_3(8)$

BIOINFORMATIQUE DU MICROBIOME HUMAIN



CONCLUSION

- Certains problèmes sont comparativement plus faciles pour les humains que pour les algorithmes d'état de l'art.
- On peut cibler ces problèmes et combiner la contribution humaine avec les algorithmes pour obtenir un résultat qui dépasse la somme des parties.
- Ce faisant, on génère des données intéressantes à analyser pour mieux comprendre le comportement humain et propager la contribution humaine à de nouvelles données.
- Nécessite de développer diverses expertises et collaborations dans un environnement pluridisciplinaire.
- Peu importe votre background, vous pouvez contribuer.

REMERCIEMENTS

Université McGill

Jérôme Waldispühl
Mathieu Blanchette
Jay White Bear
Alex Butyaev
Parham Ghasemloo Gheidari
Renata Mutalova
Eddie Cai
Chris Drogaris
Tous les auteurs de Borderlands Science

Gearbox Software Québec

Amélie Brouillette
Gabriel Richard
Ludger Saintélieu
Sébastien Caisse
Randy Pitchford
Toute l'équipe de Borderlands Science

Autres collaborateurs

Vladimir Reinharz
Yann Ponty
Hua-Ting Yao

Massively Multiplayer Online Science

Attila Szantner
Kornel Erhart

Université de Californie – San Diego

Rob Knight
Daniel McDonald
Gibs Rahman

