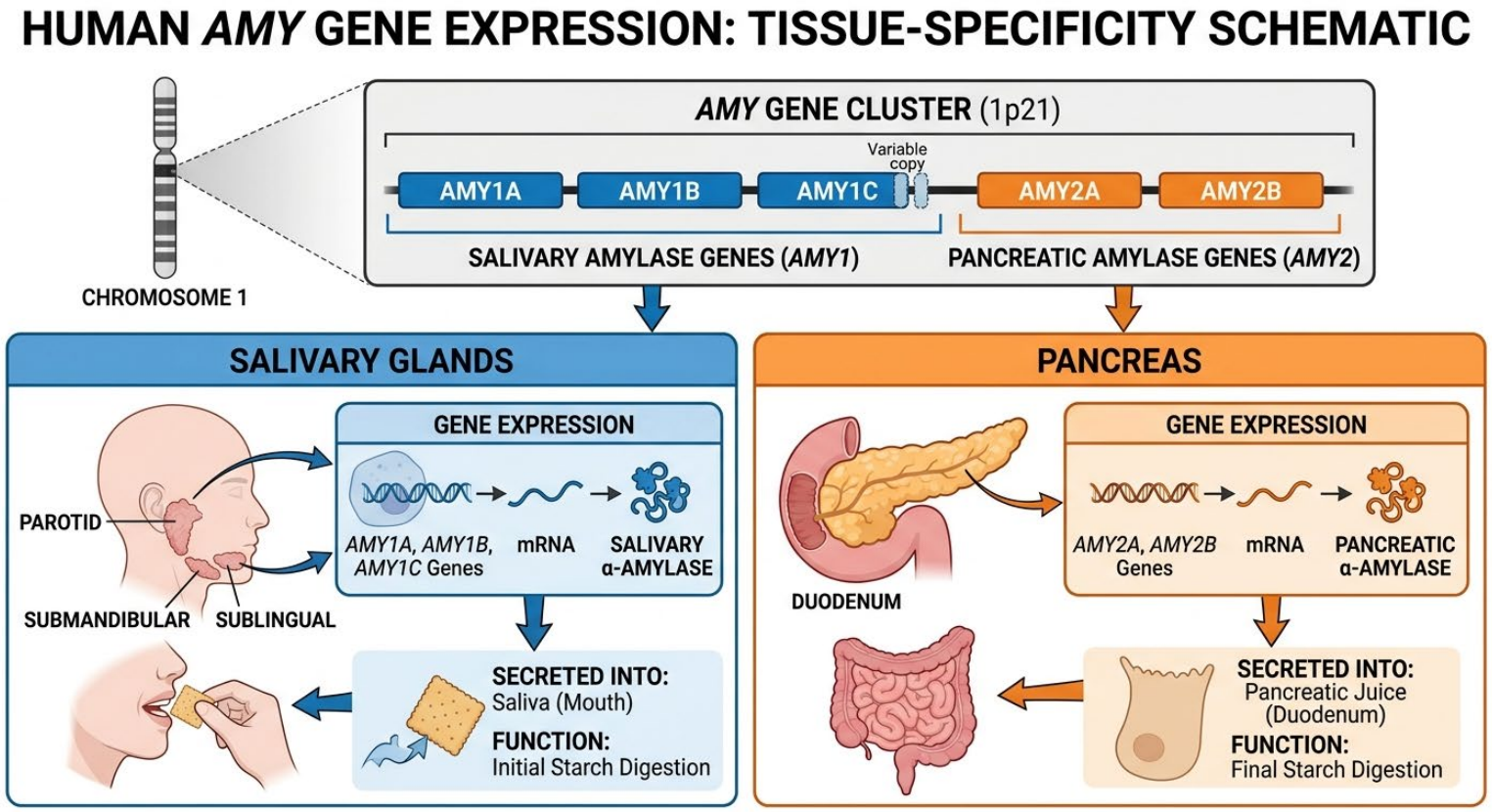
Convergent evolution through independent rearrangements in the primate amylase locus

Quanyu Chen

26.3.20

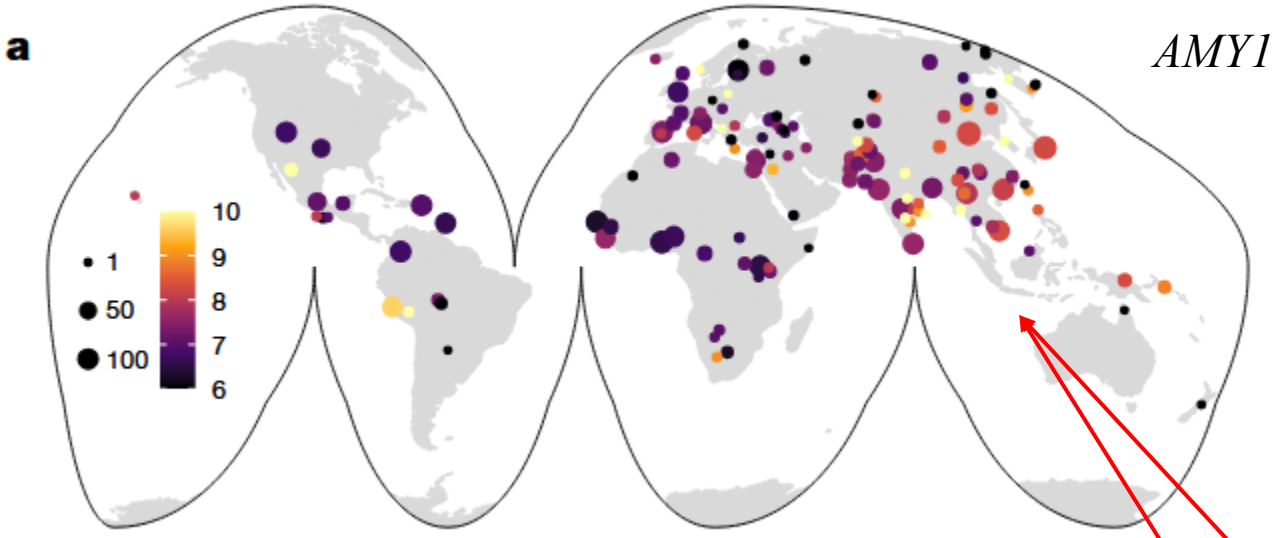
What is the amylase?

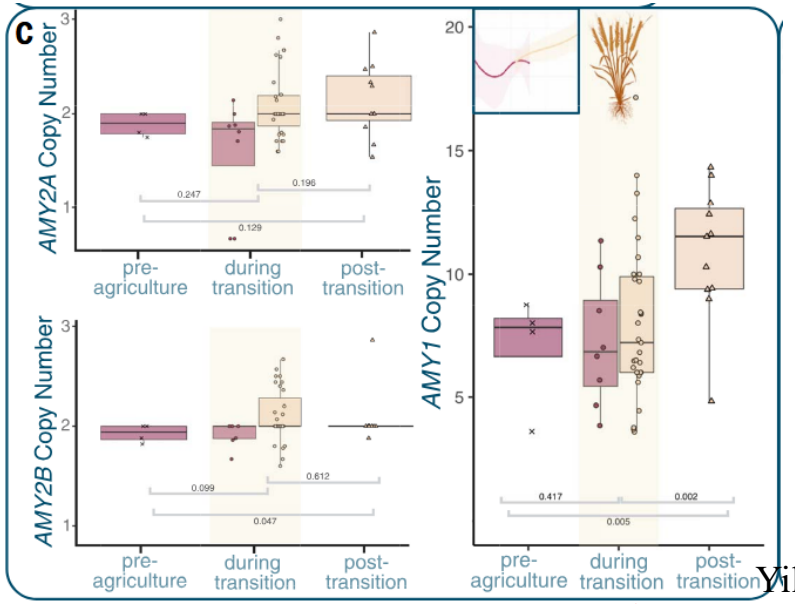


Gemini nano banana v2

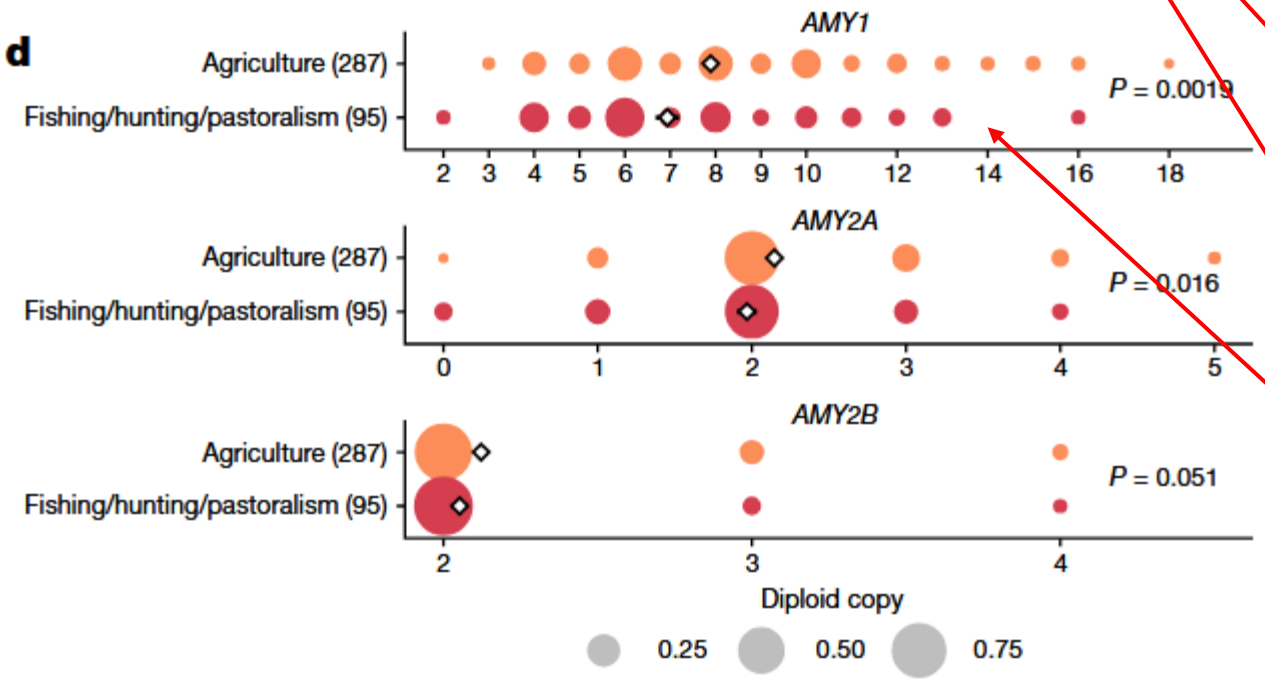
- Amylases are secreted enzymes that catalyze the first step in digestion of dietary starch and glycogen
- There are three types of *AMYs* in human genome: *AMY1s* express exclusively in **salivary glands (唾液腺)**, while *AMY2A* and *AMY2B* in **pancreas (胰腺)**

Why *AMY* has been widely noted?





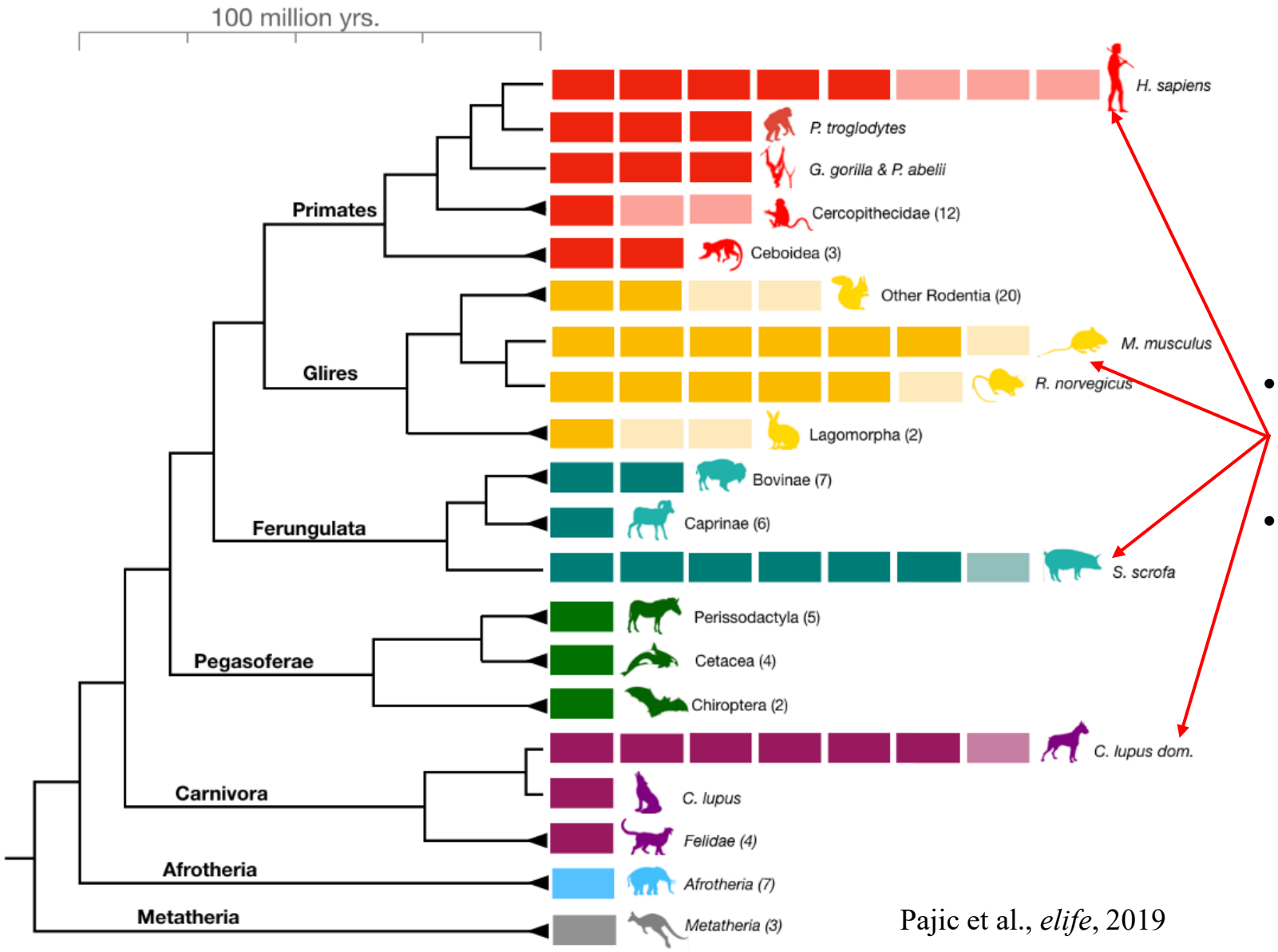
Yilmaz et al., *Science*, 2024



Bolognini et al., *Nature*, 2024

- Extensive **copy number variation (CNV)** has been documented at the amylase locus in humans
- The CNV had significant **population-stratification**, e.g., for *AMY1s,* the East Eurasia significantly harbors higher CN than the West Eurasia --> potential positive selection
- The CN of all three amylase genes was higher in populations with **agricultural subsistence** than in those from fishing, hunting and pastoral groups --> potential positive selection
- The *AMY* CN has increased in the past 4000 years among European **farmers** --> positive selection

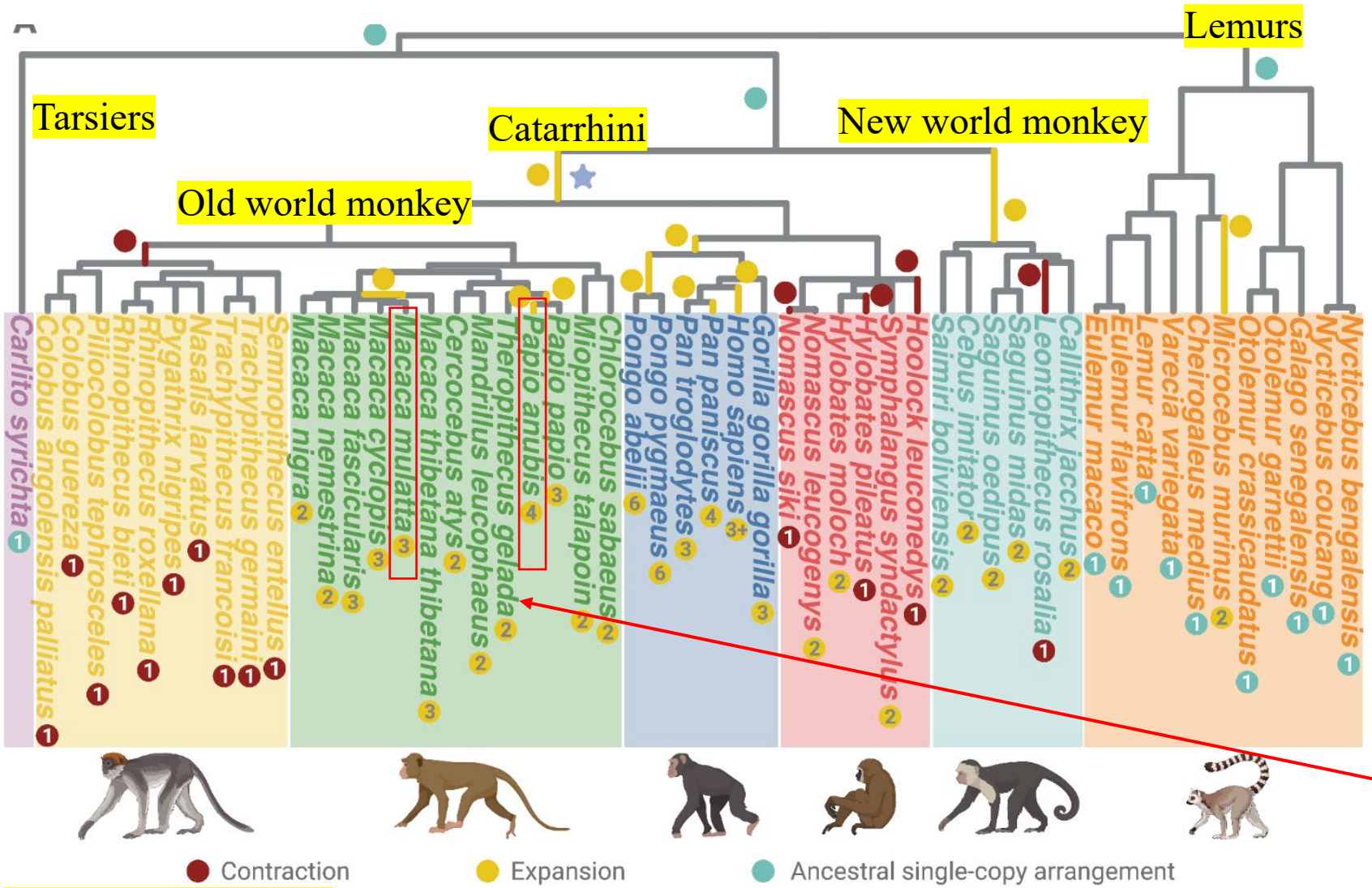
Remaining questions

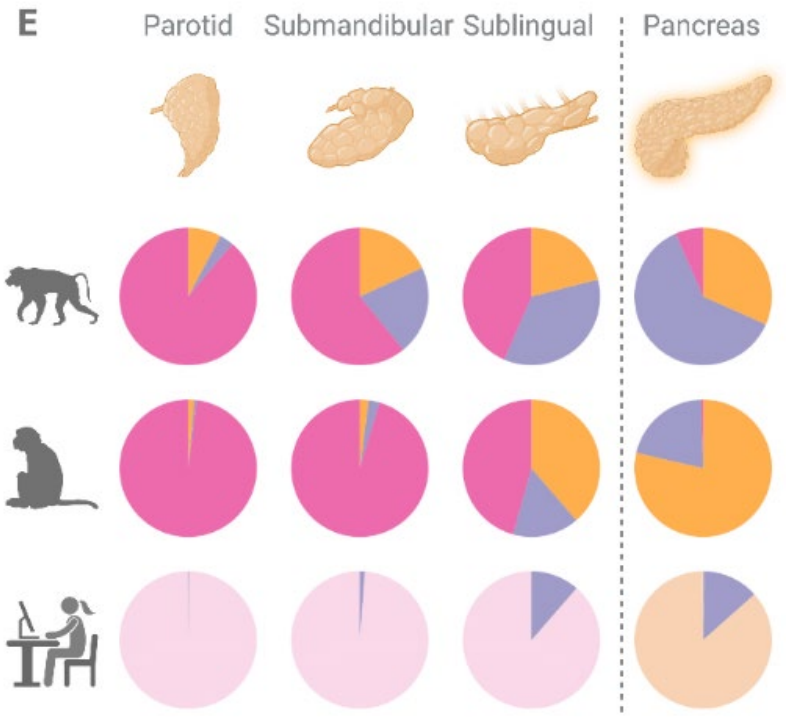


- The CN of *AMYs* in other primates or mammals (or vertebrates) --> **convergent evolution**
- The **regulatory shifts** from pancreas to salivary glands

Pajic et al., *elife*, 2019

Non-human primates





- 53 primates, 69 genomes
- Transcriptome from salivary glands and pancreas from rhesus macaques (M.m.) and olive baboons (P.a.)

Why am I interested?

- *AMYs* is a good example illustrating **gene-culture co-evolution** that links genomic landscapes with the trajectory of human civilization
- **Convergent evolution** will be documented in my study
- Omer Gokcumen is good at telling CNV and diets story

Omer Gokcumen

FOLLOW

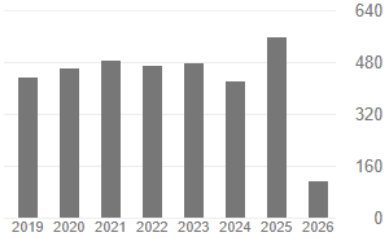
Professor - Biological Sciences, University at Buffalo
Verified email at buffalo.edu - Homepage

Evolutionary genomics human evolution Structural Variation Anthropology ekoevo

TITLE	CITED BY	YEAR
Origins and functional impact of copy number variation in the human genome DF Conrad, D Pinto, R Redon, L Feuk, O Gokcumen, Y Zhang, J Aerts, ... 综合性期刊TOP SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 48.5 SWJTU A++ Nature 464 (7289), 704-712	2418	2010
Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls 综合性期刊TOP SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 48.5 SWJTU A++ Nature 464 (7289), 713-720	979	2010
Landscape of somatic retrotransposition in human cancers E Lee, R Iskow, L Yang, O Gokcumen, P Haseley, LJ Luquette III, JG Lohr, ... 综合性期刊TOP SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 45.8 SWJTU A++ Science 337 (6097), 967-971	888	2012
A highly annotated whole-genome sequence of a Korean individual JI Kim, YS Ju, H Park, S Kim, S Lee, JH Yi, J Mudge, NA Miller, D Hong, ... 综合性期刊TOP SCI升级版 综合性期刊1区 SCI基础版 综合性期刊1区 IF 48.5 SWJTU A++ nature 460 (7258), 1011-1015	417	2009

Cited by VIEW ALL

	All	Since 2021
Citations	8074	2526
h-index	35	25
i10-index	69	57



Public access VIEW ALL

0 articles	72 articles
not available	available

Based on funding mandates

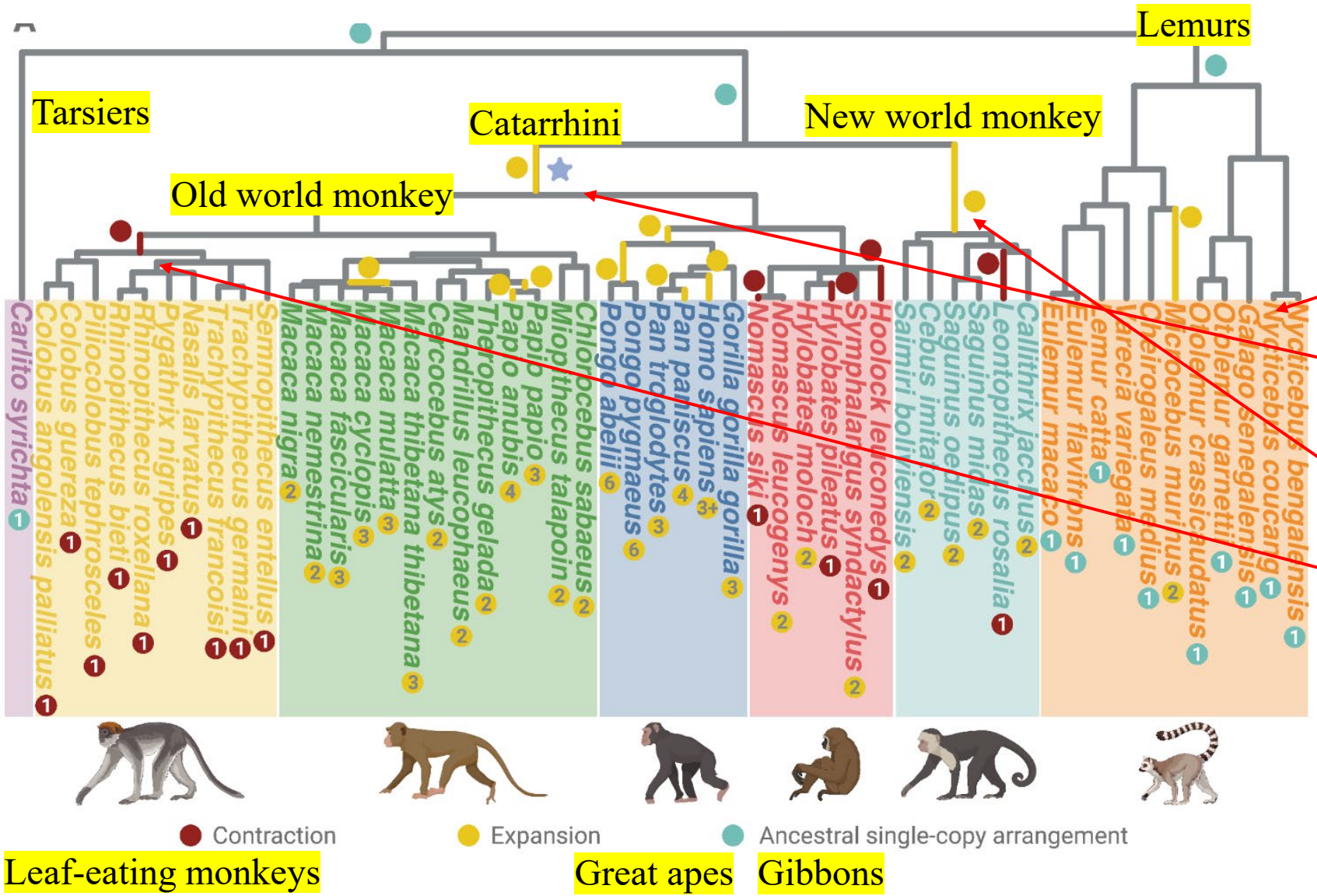
RESEARCH ARTICLE SUMMARY

HUMAN GENETICS

Reconstruction of the human amylase locus reveals ancient duplications seeding modern-day variation

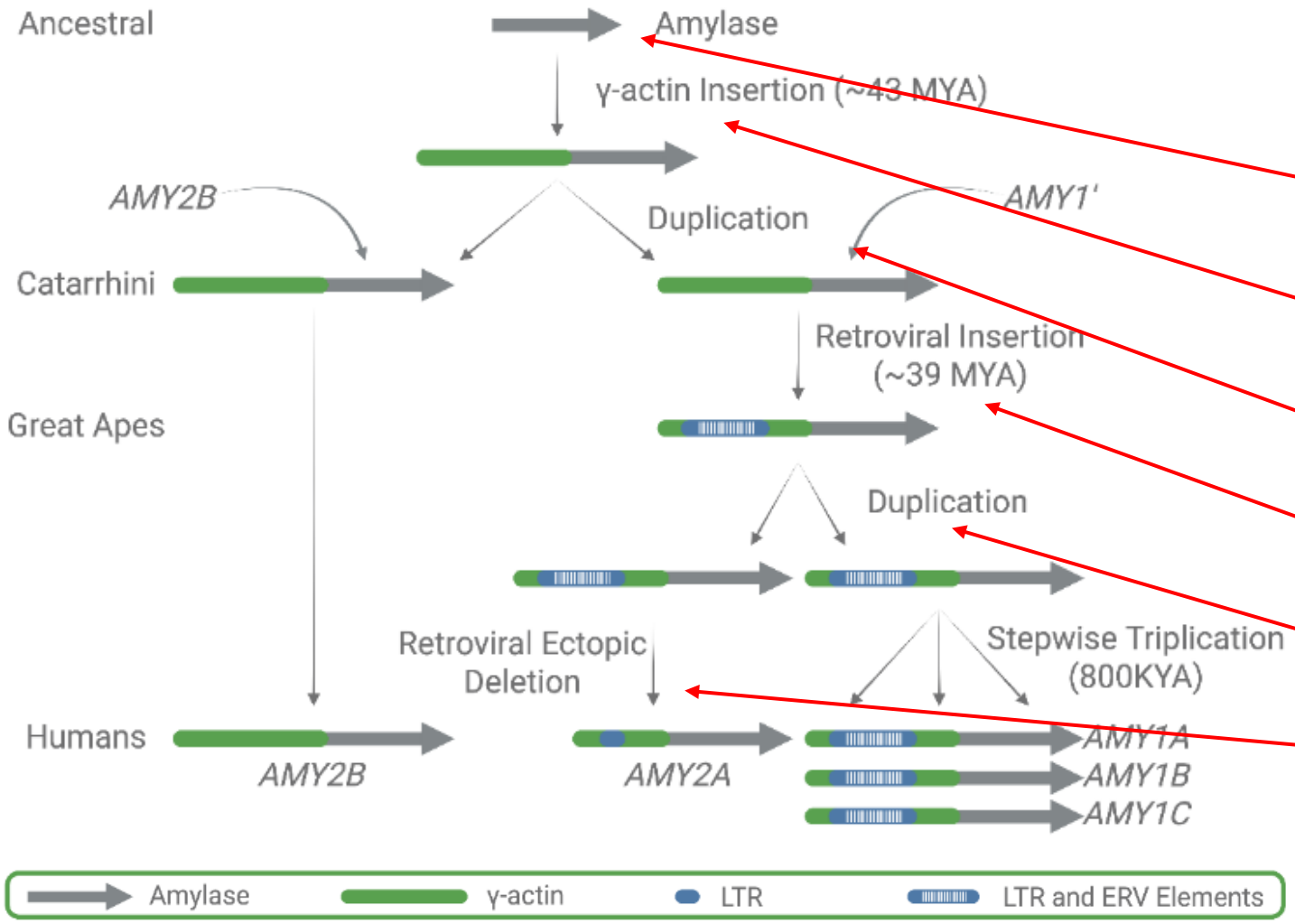
Feyza Yilmaz†, Charikleia Karageorgiou†, Kwondo Kim†, Petar Pajic, Kendra Scheer, Human Genome Structural Variation Consortium, Christine R. Beck, Ann-Marie Torregrossa, Charles Lee*, Omer Gokcumen*

Recurrent independent contraction and expansion



- All (n=11) but one of lemurs harbor a single *AMY* --> ancestral state
- An ancestral duplication of the amylase locus in the Catarrhini lineage
- A lineage-specific duplication in New World monkeys
- A loss of a single *AMY* gene in leaf-eating monkeys

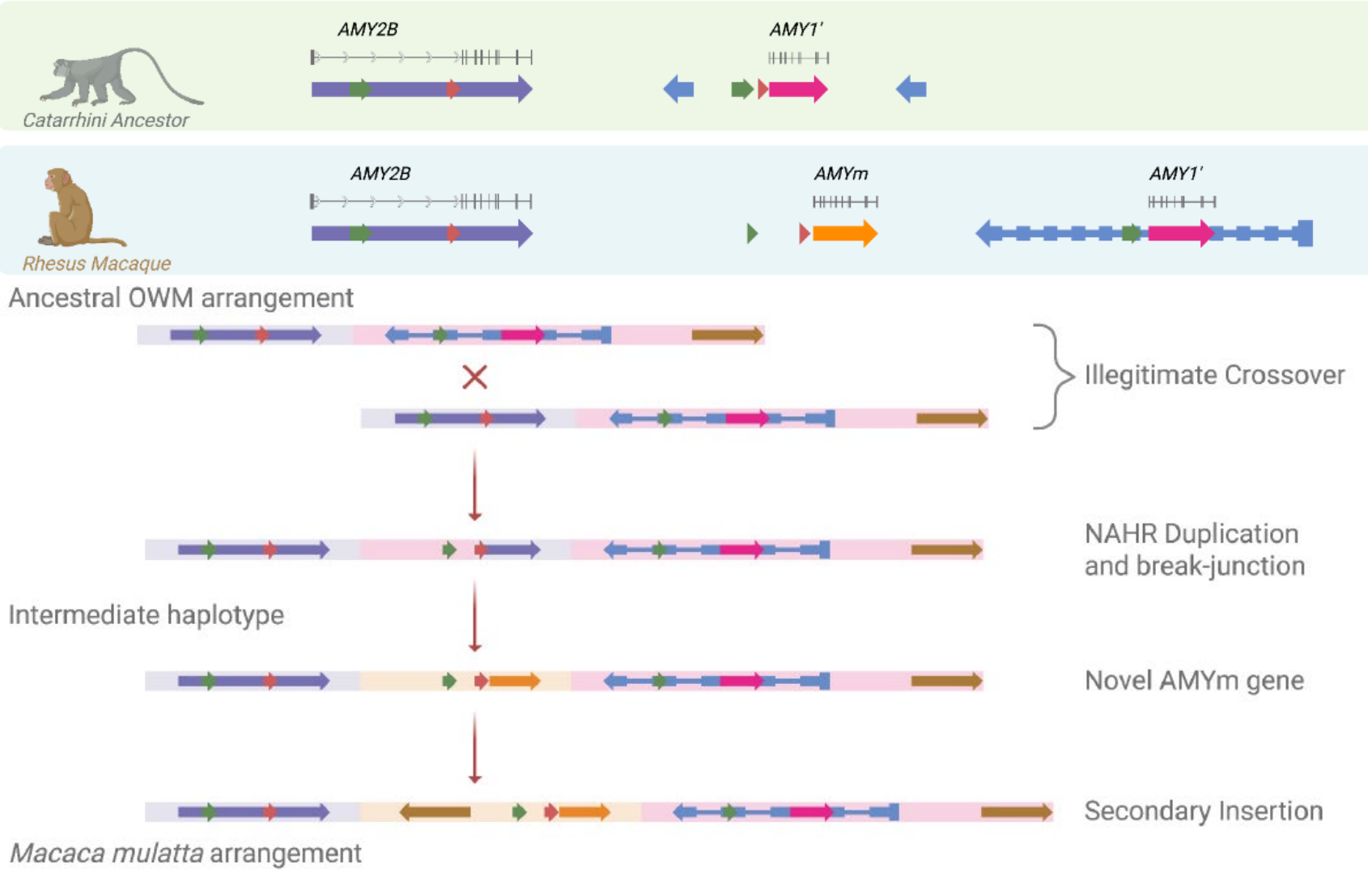
Reconstruction of the amylase duplications



- The primate ancestor possessed a single amylase gene, orthologous to human *AMY2B*
- In the Catarrhini ancestor, an insertion of the 3' untranslated region of a γ-actin pseudogene insertion occurred 5' upstream of the ancestral amylase gene
- The γ-actin-*AMY2B* duplicated, thereby generating a new amylase copy (*AMY1'*)
- In the great ape lineage, an endogenous retrovirus (ERV) inserted into the γ-actin region
- This γ-actin–ERV–*AMY1'* duplicated into *AMY1* and the precursor of *AMY2A* in great apes
- The progenitor of *AMY2A* underwent an ectopic **deletion of a portion of the ERV element**, leading to its current structure

Questions (from the article)

- What was the initial mutational driver of these duplications?
- What could be the adaptive and functional impact of these lineage-specific duplications?

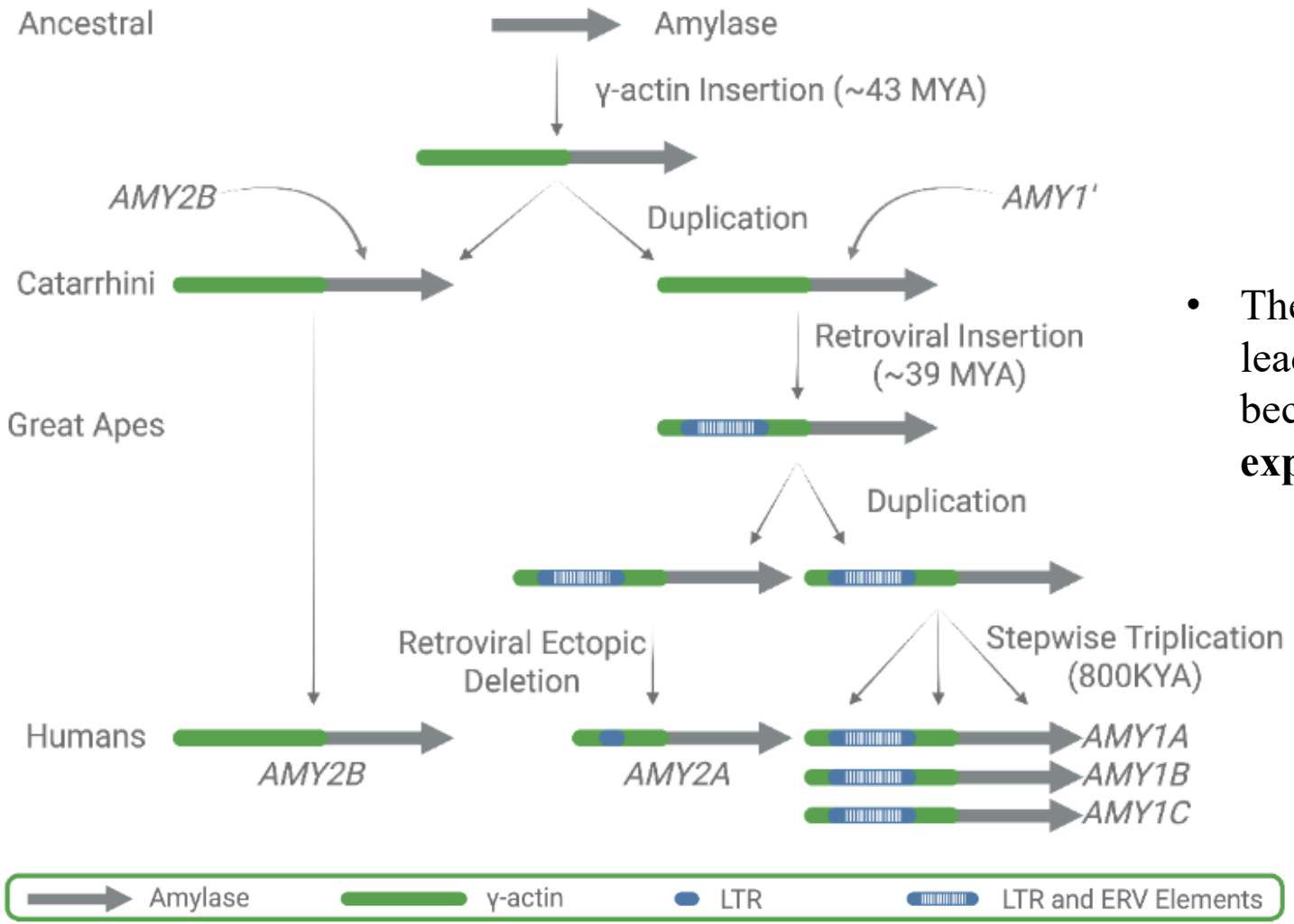


- The mutational mechanisms driving additional duplications, following the initial structural changes in the Catarrhini ancestors, have been characterized as **non-allelic homologous recombination (NAHR)**
- However, the origins of primary duplications, which **arose from a single-copy** ancestral haplotype, remain poorly understood.

2--> 3+ NAHR
1--> 2 ?????

Questions (from the article)

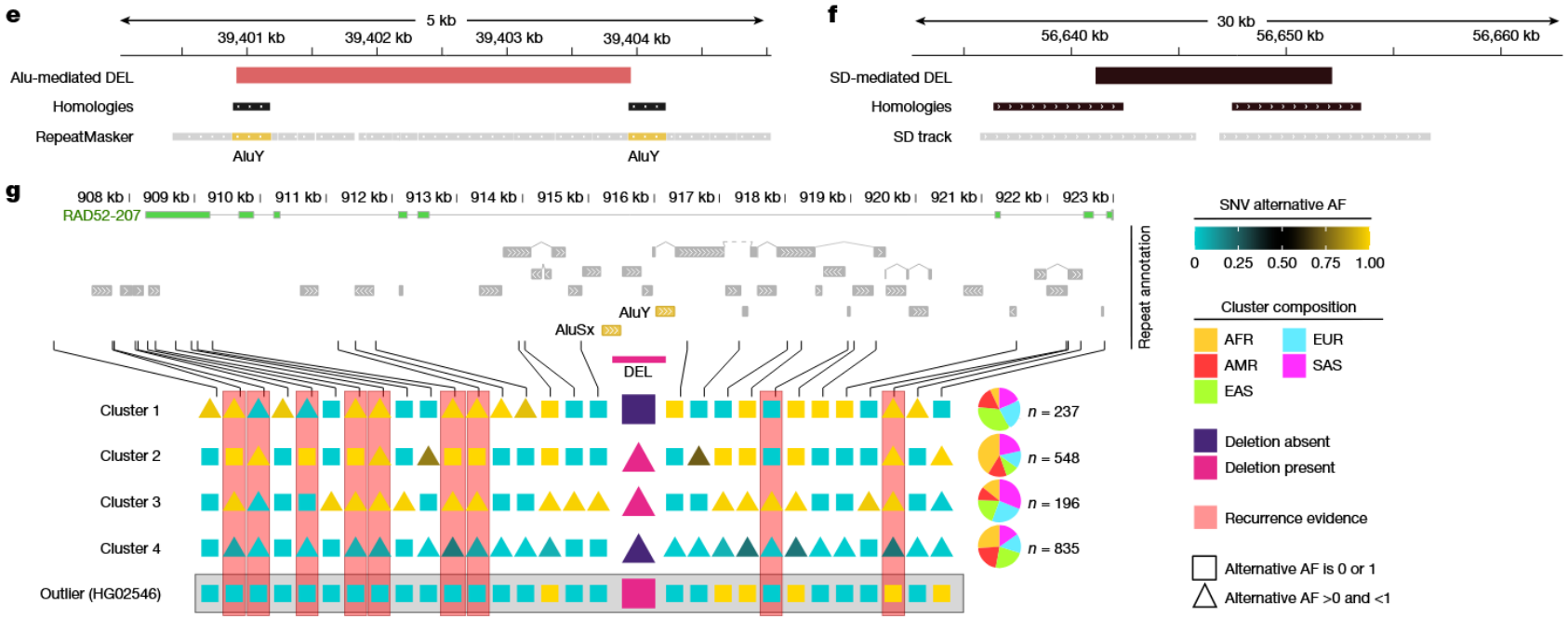
- What was the initial mutational driver of these duplications?
- What could be the adaptive and functional impact of these lineage-specific duplications?



- The observation that *AMY1'* represents the ancestral copy leading to great ape *AMY2A* and *AMY1* is remarkable because these genes have distinct functions with **specialized expression in pancreas and salivary glands**, respectively.

AMY1' --> *AMY2A* (pancreas) and *AMY1* (salivary glands) why?

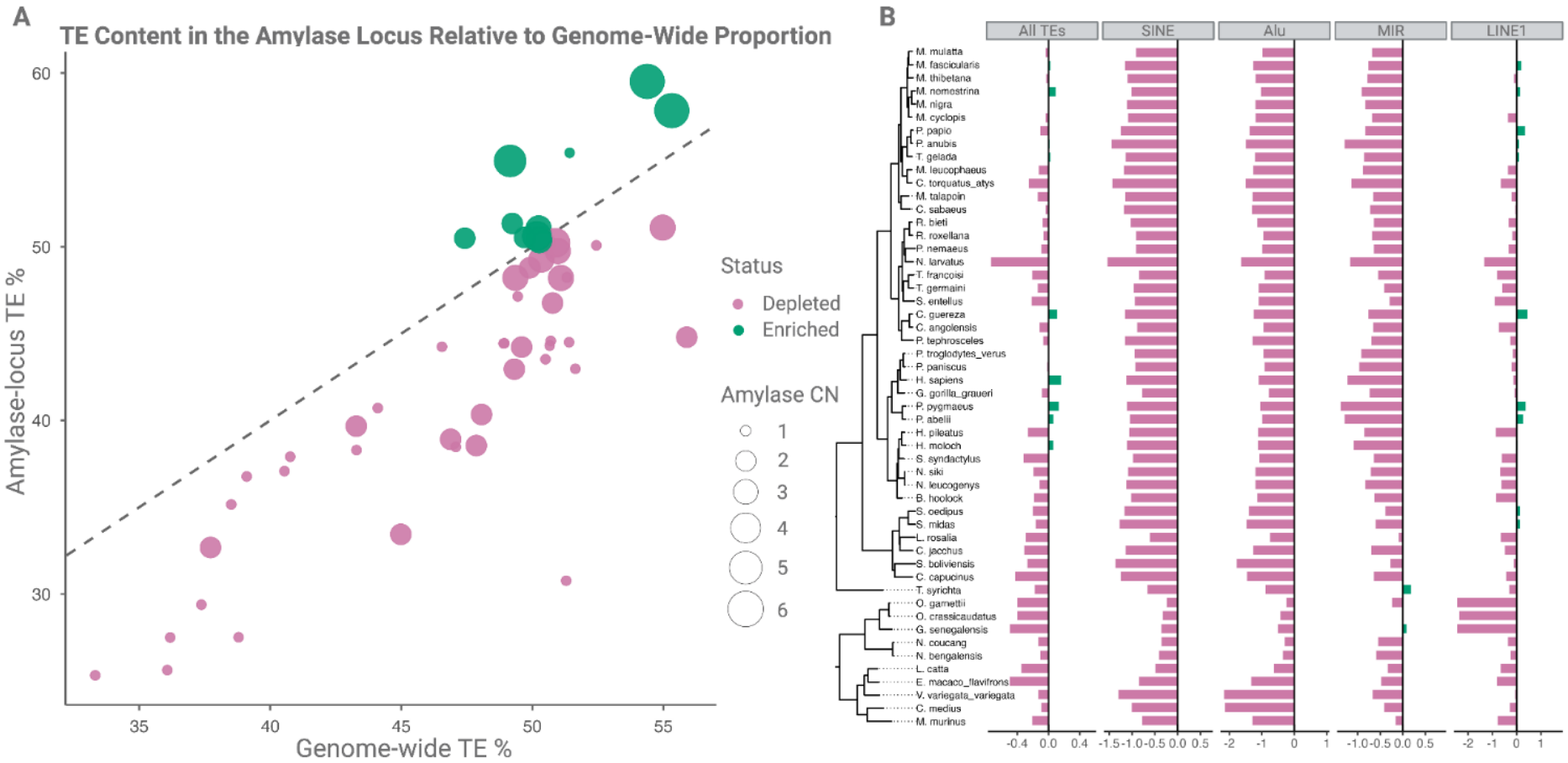
What was the initial mutational driver of these duplications?



Schloissnig et al., *Nature*, 2025

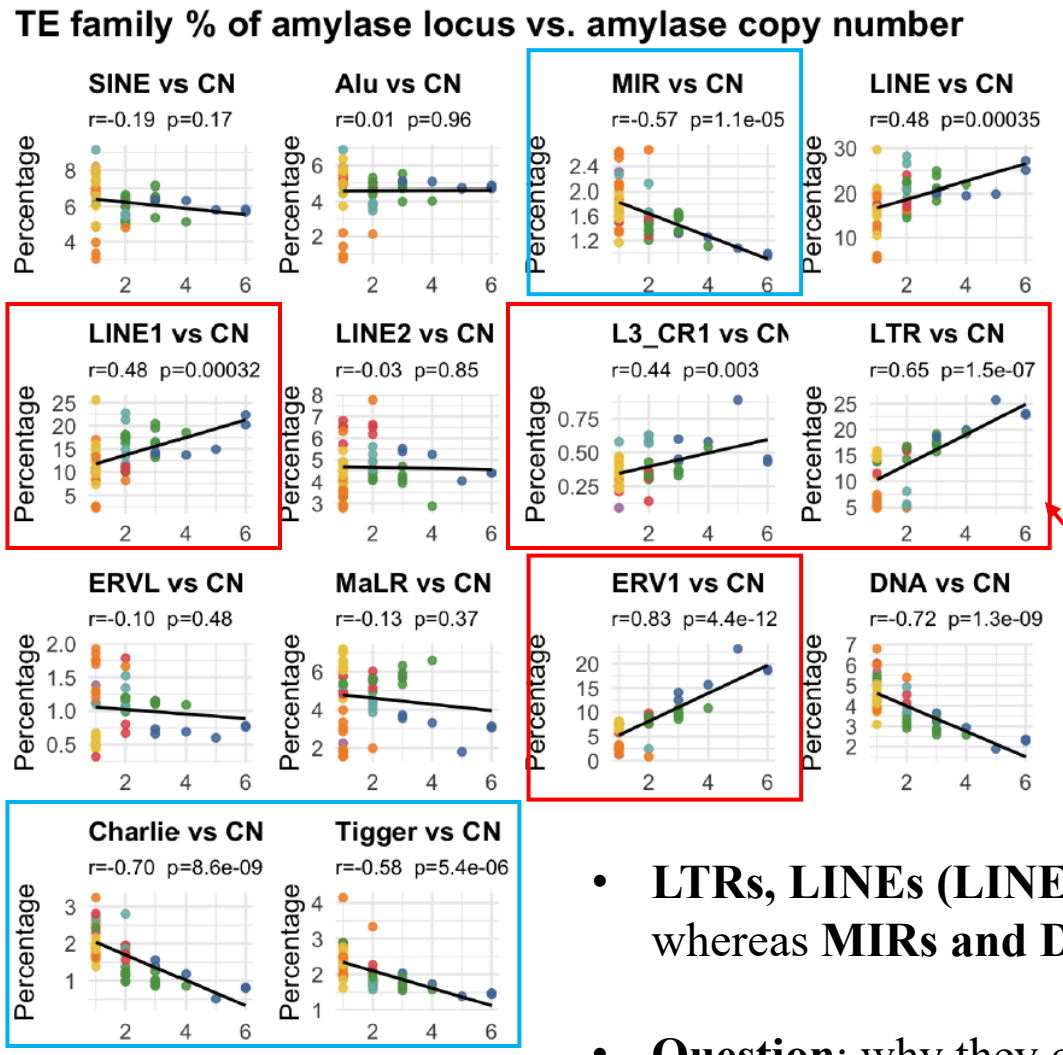
- Recent studies have shown that **TE-mediated rearrangements** can arise through diverse molecular mechanisms and induce **structural instability**
- The researchers hypothesized that **TEs might contribute to the formation of primary duplications in primates**, thereby predisposing the amylase locus to structural instability

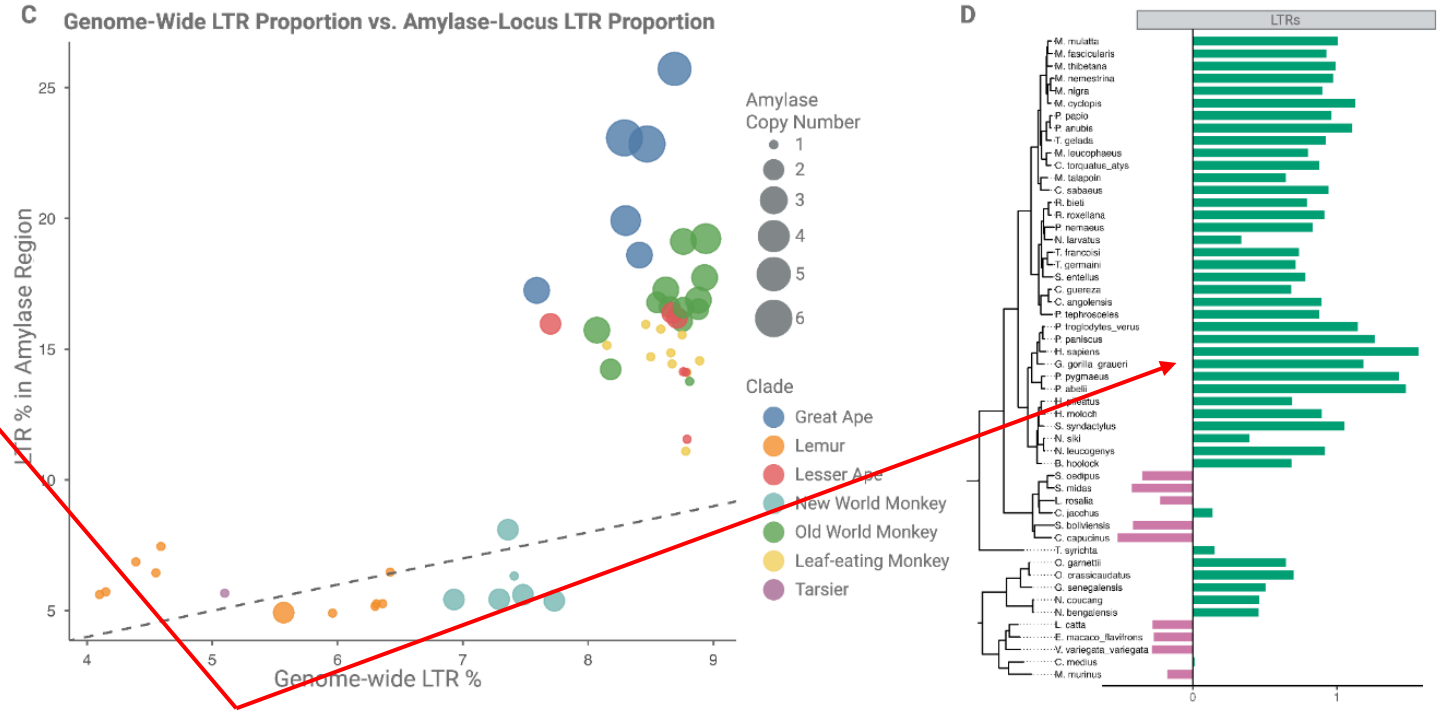
TE depletion? No correlation?



- Annotated transposable elements across 53 primate genomes with standardized primate repeat dataset
- However, they found that the **amylase locus is generally depleted in transposable elements** compared to genome-wide averages

LTRs correlated with the CN of *AMYs*

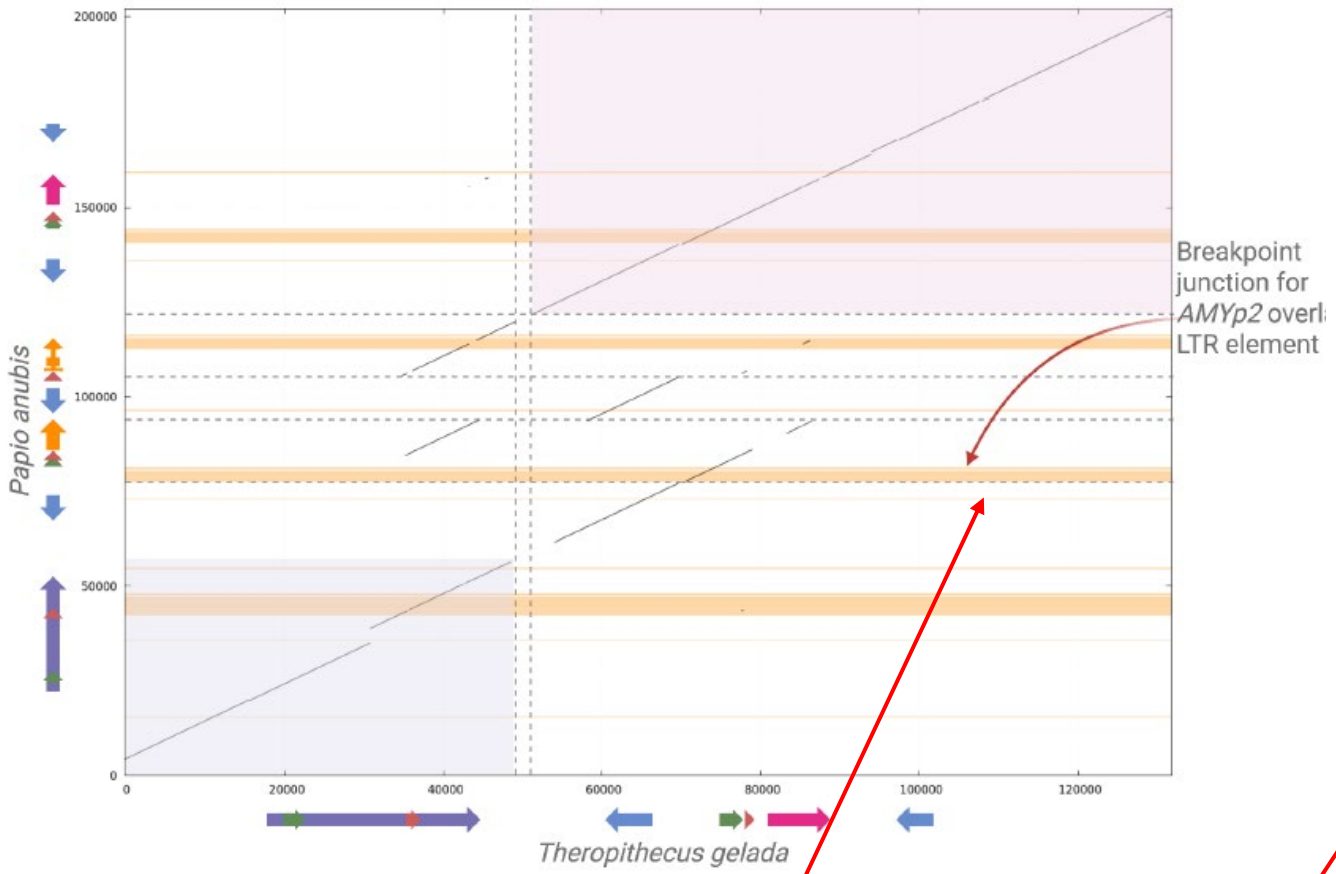


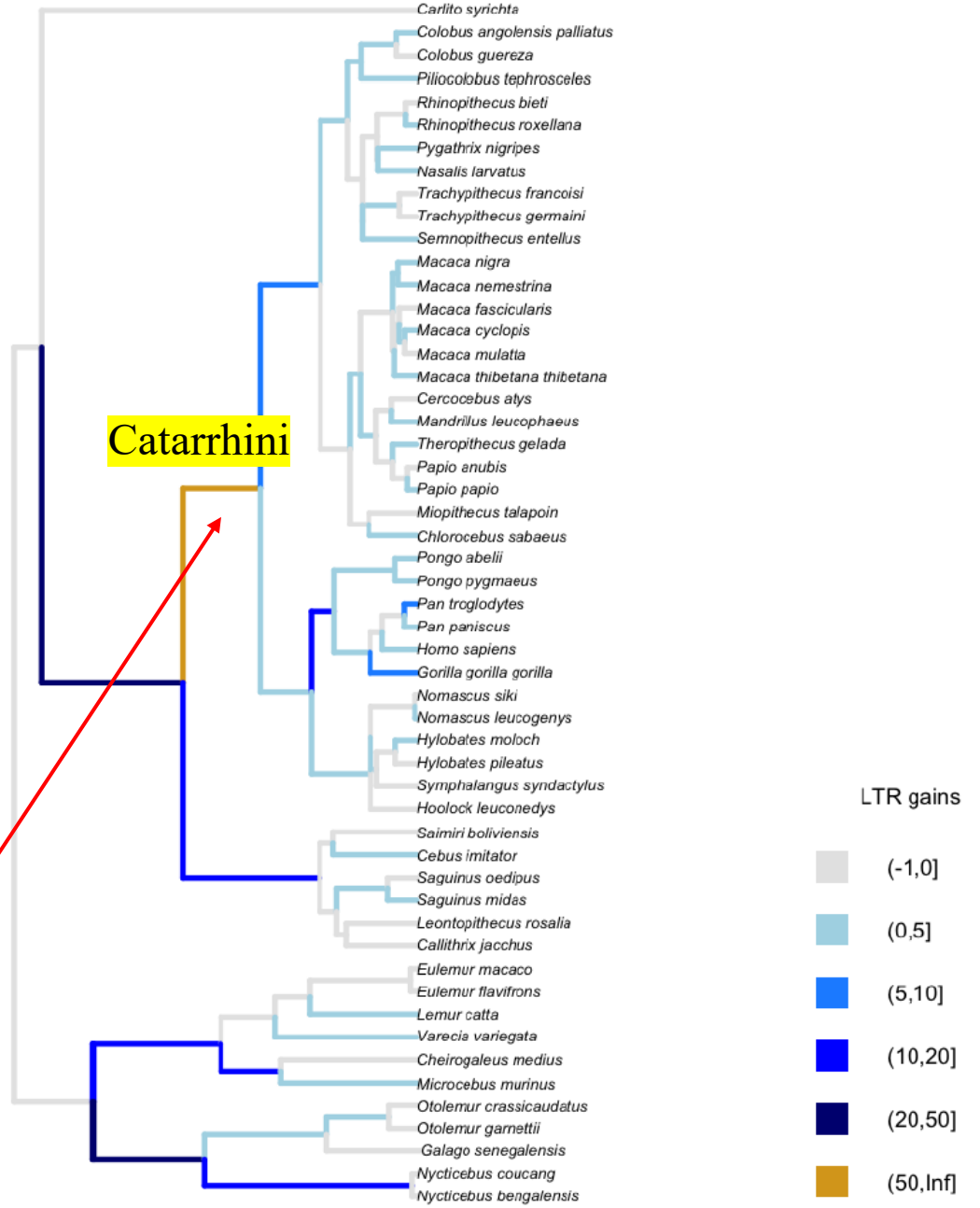


- **LTRs, LINEs (LINE1, L3/CR1) and ERV1 show positive correlations** with *AMY* CN, whereas **MIRs and DNA transposons (Charlie, Tigger) are negatively** correlated
- **Question**: why they only focused on LTRs? Why LINEs show positive correlations?

有问题？！选择性的放结果？

LTRs contributed to *AMYs* CNV



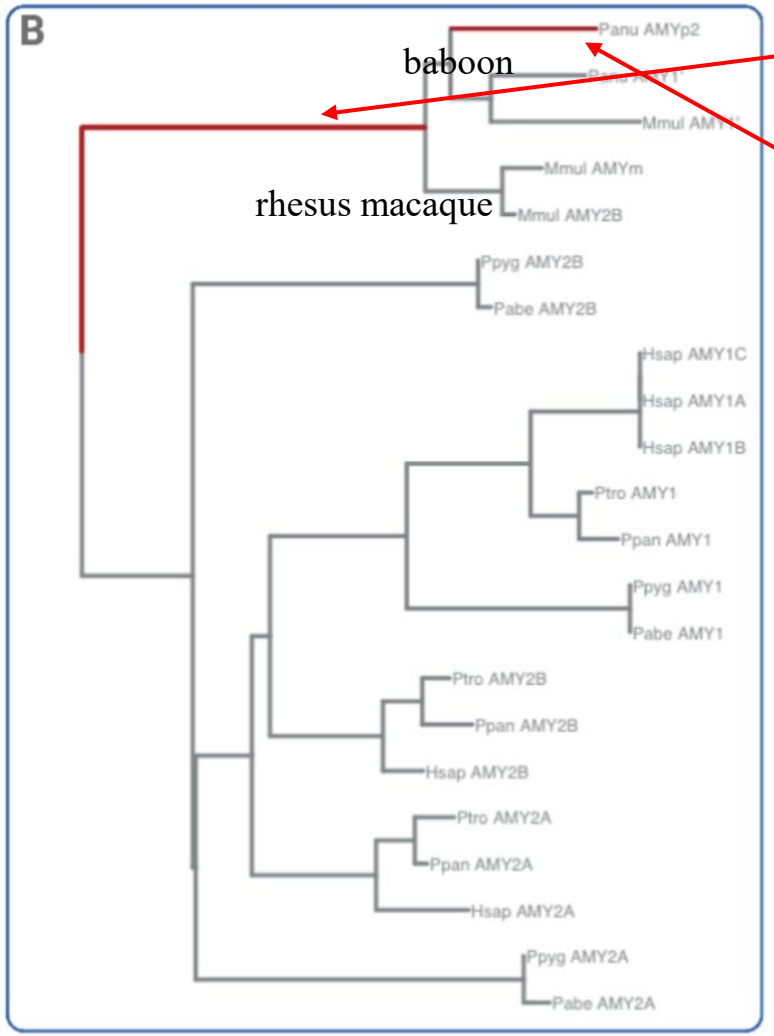


- **LTRs are proximal to the breakpoint junctions**
- Identified orthologous LTR insertions across primates
- **Elevated LTR gain events** on the branch leading to Catarrhini, coinciding with the **initial amylase duplication events** in this lineage

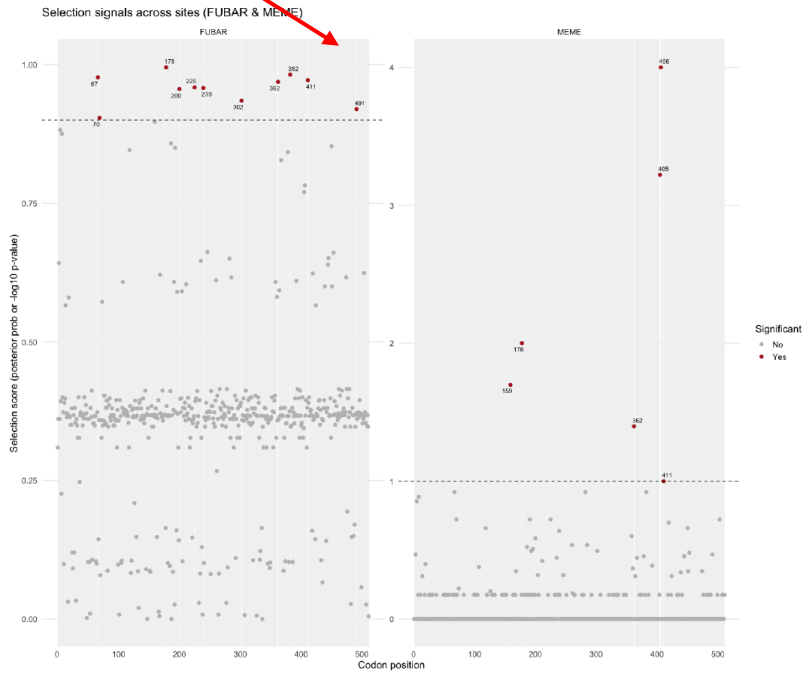
Weak! Just association!

What could be the adaptive and functional impact of these lineage-specific duplications?

- Function
- Regulation

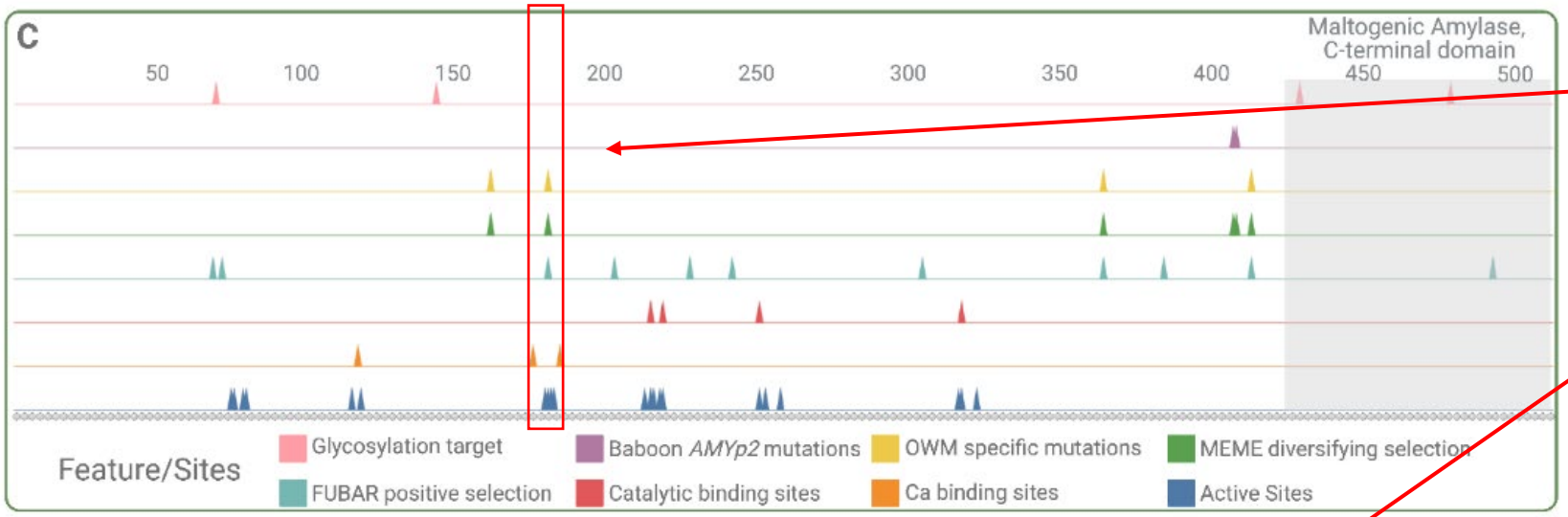


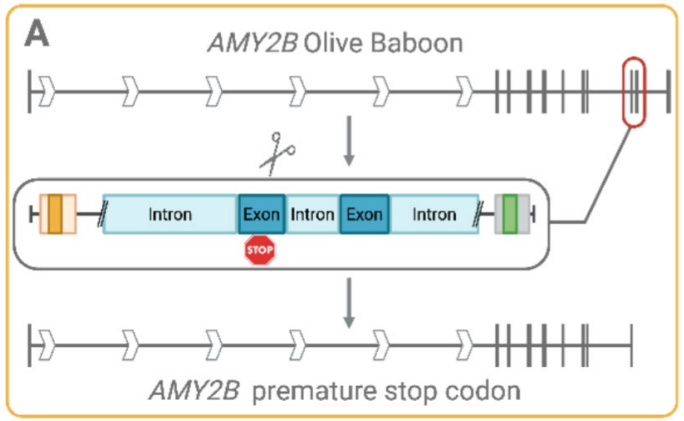
- Identified a significant **positive selection signal on the internal branch leading to baboon and rhesus macaque paralogs** compared to great ape paralogs (aBSREL, p = 0.038)
- Detected strong evidence **for positive selection specifically on *AMYp2* in olive baboons**, supported by both aBSREL ($p < 10\text{-}6$) and RELAX analyses ($p < 0.0001$)
- Identified **six codons exhibiting episodic positive selection** that likely contribute to this overall selection signal



What could be the adaptive and functional impact of these lineage-specific duplications?

- Function
- Regulation





- Identified one strongly selected codon (p = 0.01, MEME) in *AMYp2* predicted as an **active site**, which may reflect **fine-tuned functional divergence** in the olive baboon lineage.
- Detected a **premature stop codon mutation** in the ancestral *AMY2B* gene copy within baboons
- **It is plausible here that the newly derived gene (*AMYp2*) functionally compensates for the observed loss of function in the ancestral gene (*AMY2B*)**

Interesting but weak! Only predicted in silico!

Take home messages

1. aBSREL (Adaptive Branch-Site Random-Effects Likelihood)

- **它的作用**：检测**某个特定的演化分支（Branch）**是否经历了爆发式的适应性进化（即正选择）。
- **分析原理**：它比较两种假设（模型）。零假设是该分支上的所有位点你要么是中性的，要么是受净化选择的（$\omega \leq 1$）。备择假设是允许该分支上一部分位点的 $\omega > 1$。aBSREL 会遍历树上的每一个分支，进行似然比检验（Likelihood-ratio tests），并使用 Holm-Bonferroni 方法校正多重比较。
- **文章发现**：在通往狒狒和恒河猴旁系同源基因的内部节点分支上发现了显著的正选择信号（$p = 0.038$），并且在橄榄狒狒特有的 AMYp2 基因分支上发现了极强的正选择信号（$p < 10^{-6}$）。

4. RELAX

- **它的作用**：检测在特定的演化分支上，自然选择的强度是**放松了（Relaxed）还是加强了（Intensified）**。
- **分析原理**：作者将狒狒的 AMYp2 设定为“前景分支”，其余设定为“背景分支”。RELAX 会拟合一个缩放参数 K。如果 $K > 1$，说明选择压力（无论是正选择还是净化选择）被强化了；如果 $K < 1$，说明选择压力放松了（更趋向于中性进化）。
- **文章发现**：AMYp2 基因不仅受到正选择（aBSREL），而且其选择强度也发生了显著改变（$p < 0.0001$）。结合 AMY2B 的提前终止（假基因化），作者推测 AMYp2 可能正在发生新功能化（Neofunctionalization），以补偿丢失的功能。

2. MEME (Mixed-Effects Model of Evolution)

- **它的作用**：检测**某个特定的位点（密码子）是否在部分演化分支**上经历了偶发性的多样化选择（Episodic diversifying selection）。
- **分析原理**：传统的位点模型假设一个位点在整棵树上受到的选择压力是一致的，但这不符合实际（比如某个氨基酸可能只在人类分支受正选择，在其他分支很保守）。MEME 允许 ω 值在不同的分支上发生变化，从而能够极其敏锐地捕捉到那些只在特定谱系中发生过短暂正选择的位点。
- **文章发现**：在 AMYp2 基因中鉴定出一个受到强烈选择的位点（第 178 位密码子，苏氨酸突变为丝氨酸，FDR 校正 $p \leq 0.10$），该位点恰好预测为酶的活性位点。

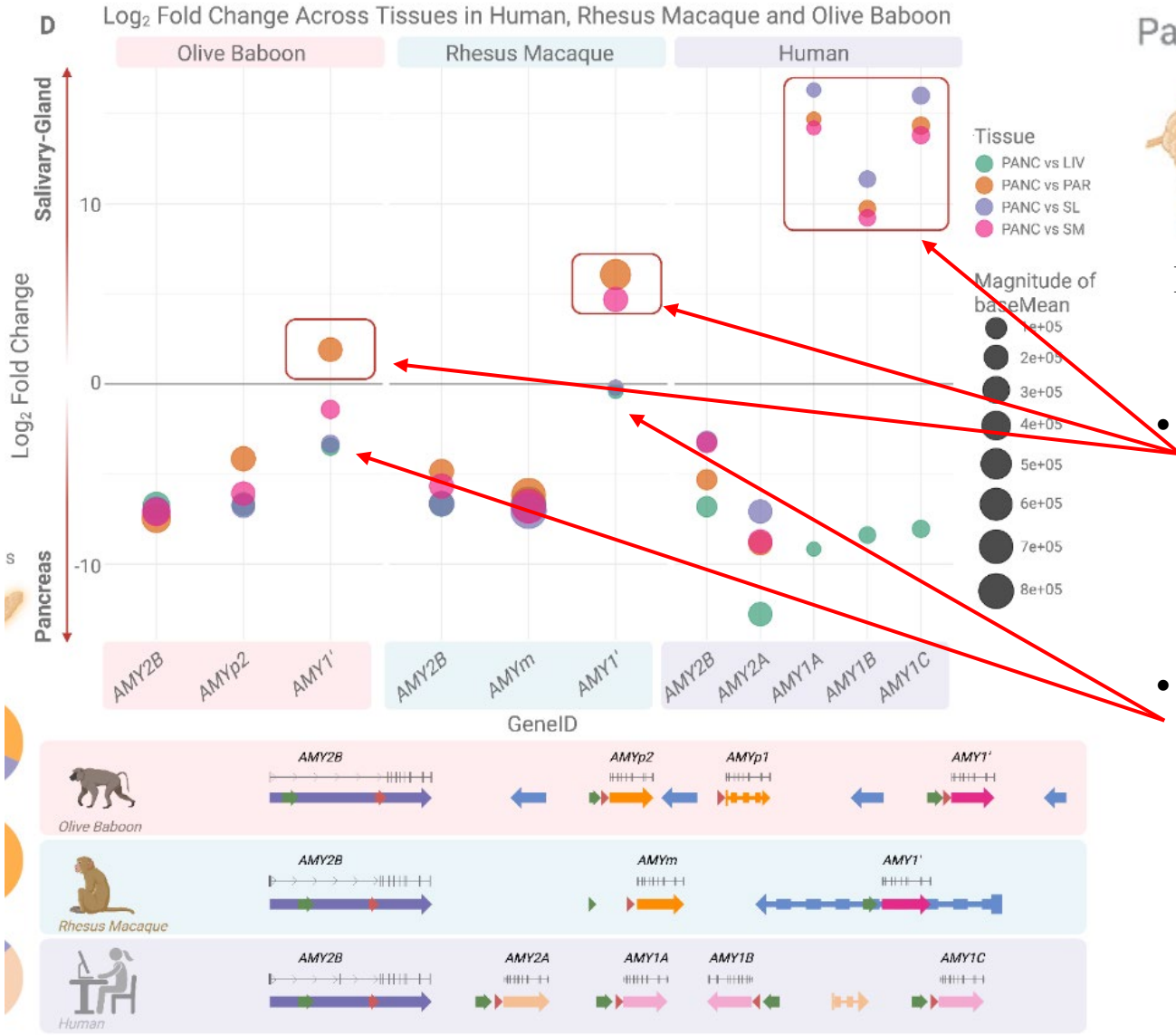
3. FUBAR (Fast, Unconstrained Bayesian AppRoximation)

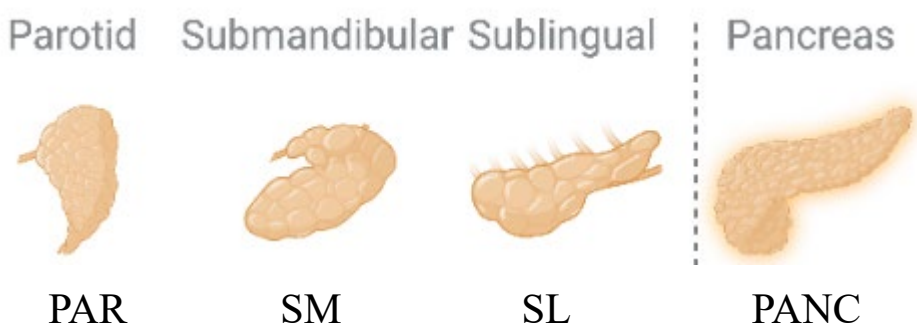
- **它的作用**：检测**某个特定的位点**是否在**整个演化树**上经历了普遍的、持续的正选择（Pervasive selection）。
- **分析原理**：采用贝叶斯马尔可夫链蒙特卡洛（MCMC）框架，假设选择压力在整棵树上是恒定的，并计算 $\omega > 1$ 的后验概率。
- **文章发现**：作者将后验概率大于 0.90 的密码子视为受到正选择的位点。这种持续的正选择通常比 MEME 发现的偶发性选择更少见。

Gemini 3.1 pro

What could be the adaptive and functional impact of these lineage-specific duplications?

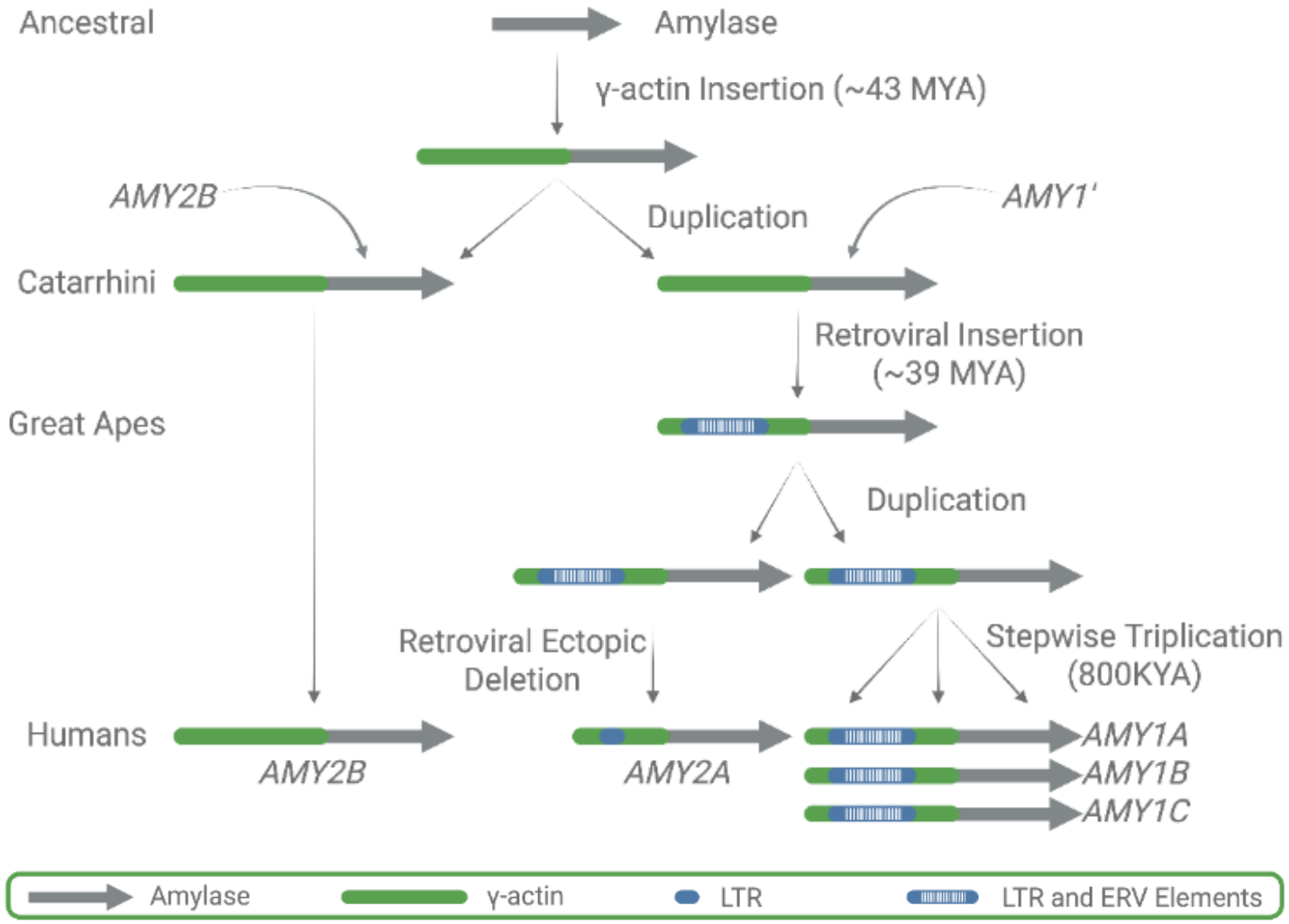
- Function
- Regulation

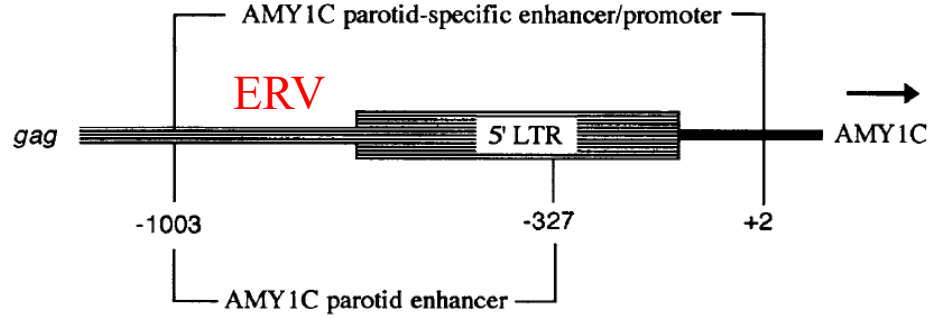




- Transcriptomic analysis revealed that, similar to humans, **the last gene (*AMY1'*)** in the amylase cluster consistently shows **elevated expression in salivary tissues** relative to the other paralogs
- The ancestral *AMY1'* gene had already **acquired expression in both the pancreas and salivary glands**

Why *AMY2A* in humans had no expression in salivary glands?



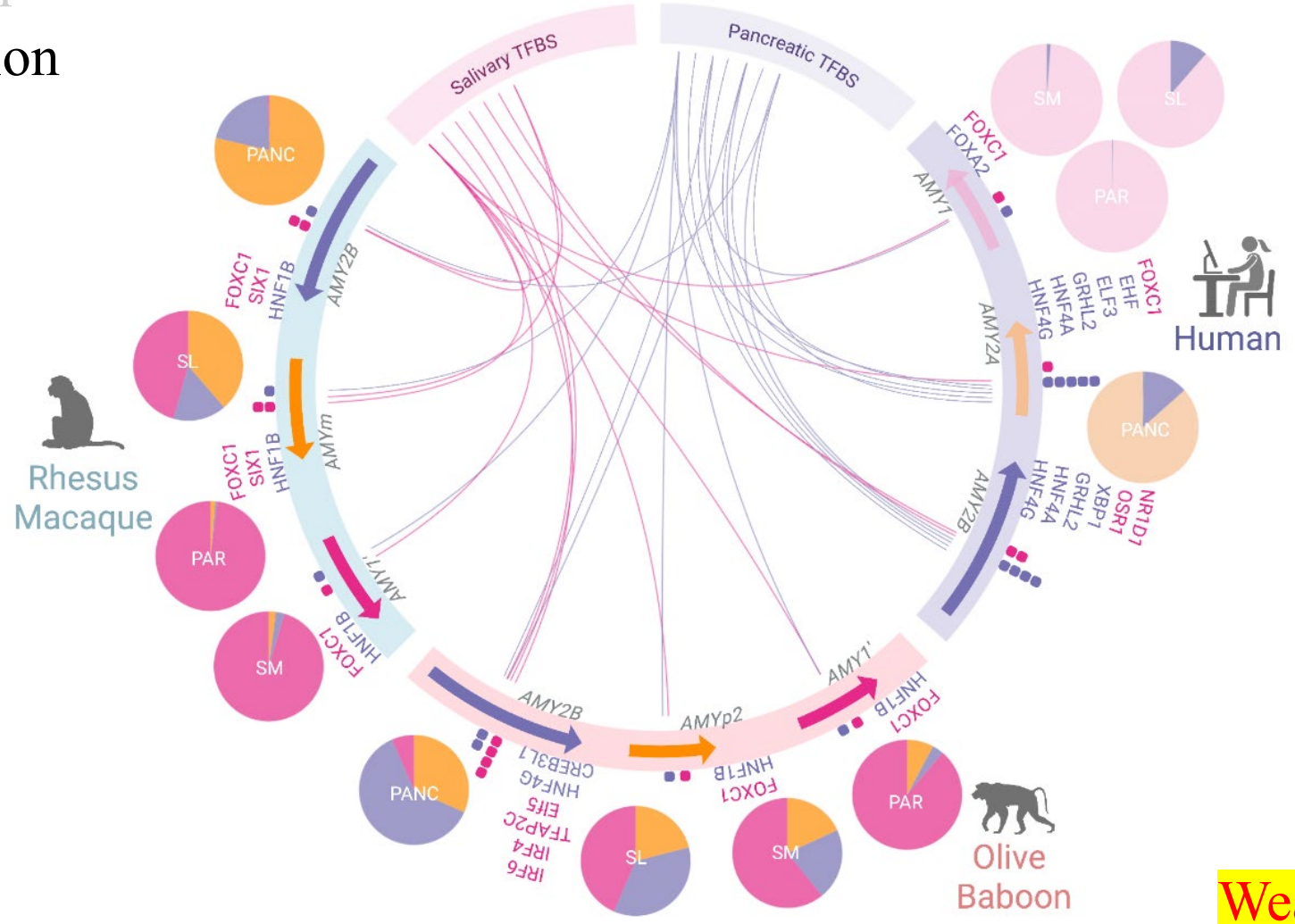


Meisler et al., *Critical Reviews in Oral Biology & Medicine*, 1993

- Following duplication in the great ape lineage, **subfunctionalization** occurred: *AMY1* retained salivary gland expression, while *AMY2A* lost salivary expression and became restricted to the pancreas. **This shift may have been facilitated by an ERV insertion.**
- **Question**: if the ERV is essential for salivary gland expression, why *AMY1'* without ERV insertion has expression in both tissues?

What could be the adaptive and functional impact of these lineage-specific duplications?

- Function
- Regulation



- Examined transcription factor binding sites across paralogs and species --> tissue-specific motifs in promoters
- Paralogs with known tissue-specific expression did not show consistent enrichment of salivary- or pancreas-biased motifs
- The presence or absence of salivary gland or pancreas-biased transcription factor binding sites alone does not fully account for the observed tissue-specific expression patterns, highlighting the potentially combinatorial nature of transcriptional regulation

Weak! No any robust conclusions!

Limitations of the study (from the article)

- …
- Second, **genome assembly quality can influence TE annotation**, particularly in repetitive regions. We showed that **LTR annotations are broadly consistent between short-read and long-read assemblies** for the same species and that the significant correlation between assembly N50 and **genome-wide TE content is primarily driven by satellite sequences rather than LTRs.**
- ...

$$\text{Raw LAI} = \frac{\text{Intact LTR retrotransposon length}}{\text{Total LTR sequence length}} \times 100$$

These findings suggest that a more continuous genome assembly would result in more intact LTR elements being identified. **Thus, the amount of identifiable intact LTR elements, in turn, can indicate the assembly quality of the intergenic and repetitive sequence space**

Conflicts!

Comments

- Learned how to tell story for similar cases (e.g., my own study)
- Learned phylogenetic generalized least-squares (PGLS) models
- Learned a few selection signals
- Learned the statements to boast the significance of the article
- MBE/GBE/GIGA-SCIENCE… is welcoming~

Thanks and Q.A.