

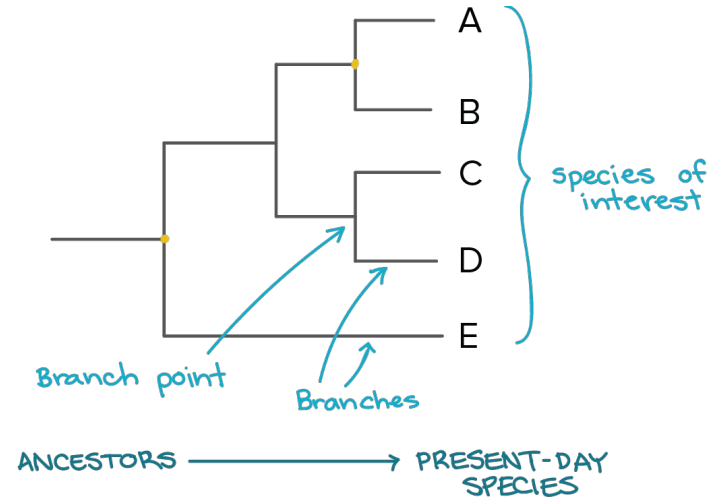
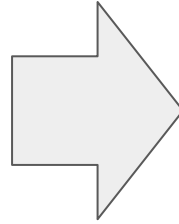
# Phylogenetic assessment of alignments reveals neglected tree signal in gaps

Christophe Dessimoz and Manuel Gil

Reported by Lev Mazaev

# Множественное выравнивание

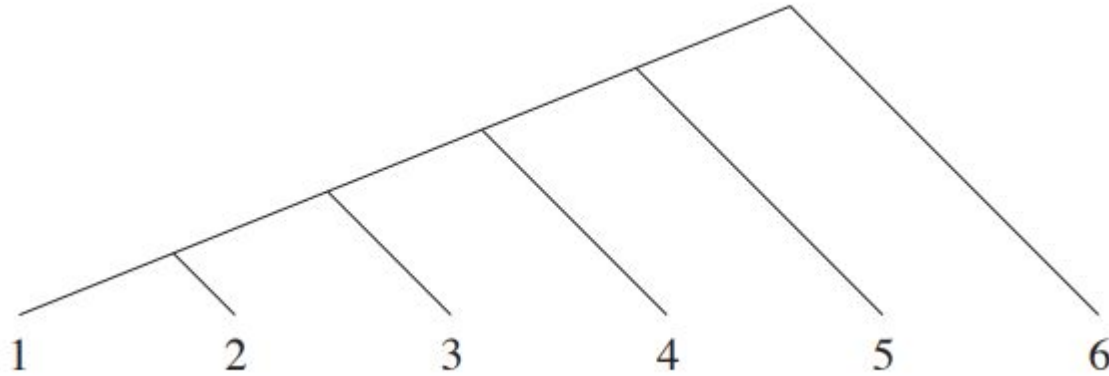
*Sequence1*    **-TCAGGA-TGAAC----**  
*Sequence2*    **ATCACGA-TGAACC---**  
*Sequence3*    **ATCAGGAATGAATCC--**  
*Sequence4*    **-TCACGATTGAATCGC-**  
*Sequence5*    **-TCAGGAATGAATCGCM**



Оценка точности выравнивания:

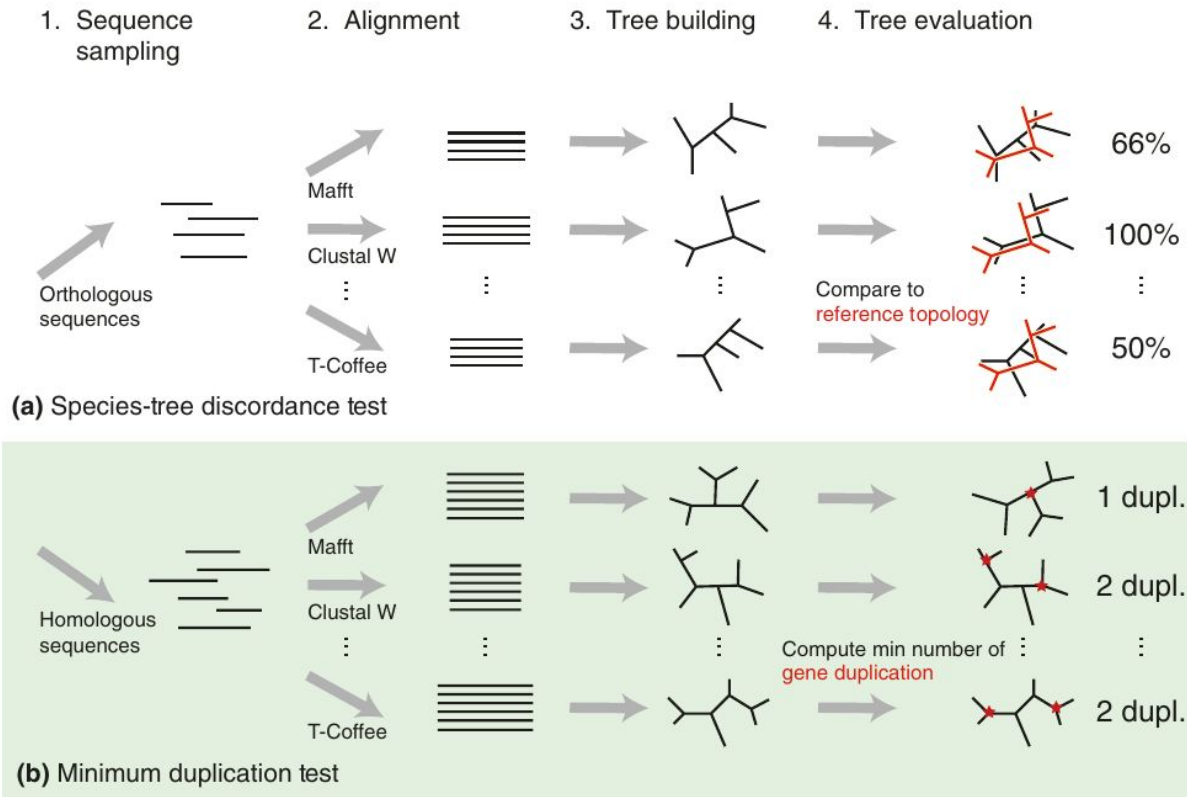
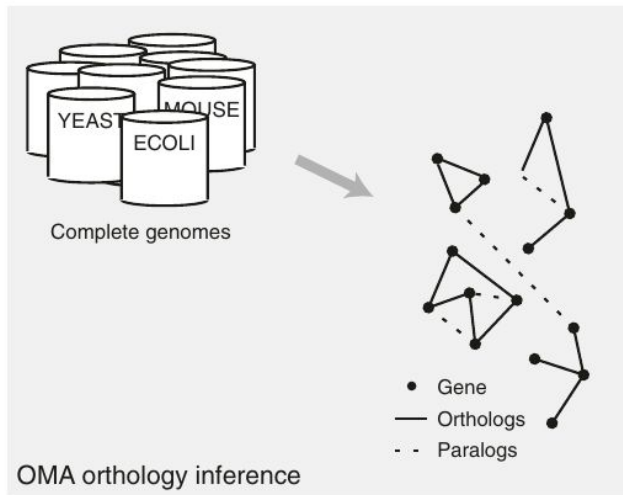
- Из структурного выравнивания
- Симуляция

Нужен альтернативный метод:

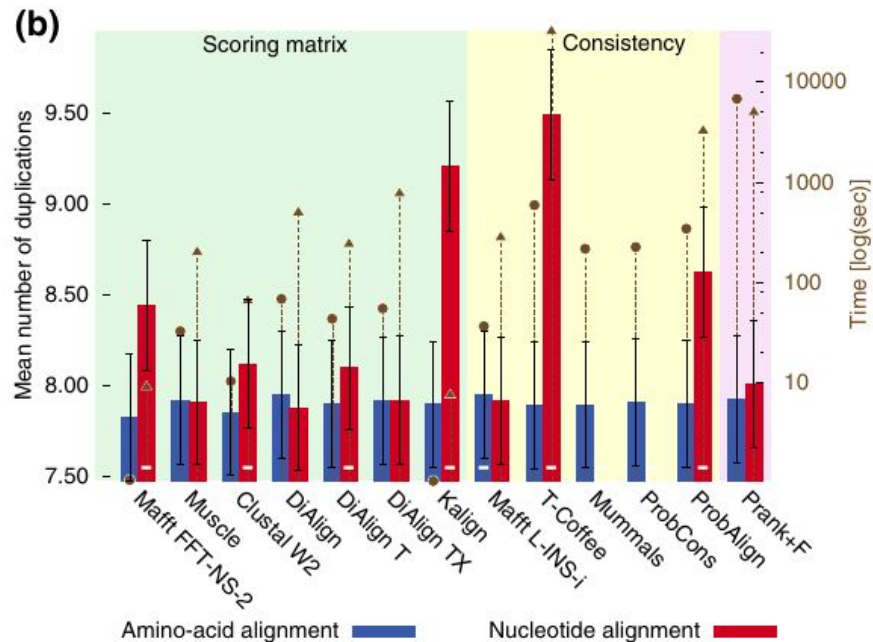
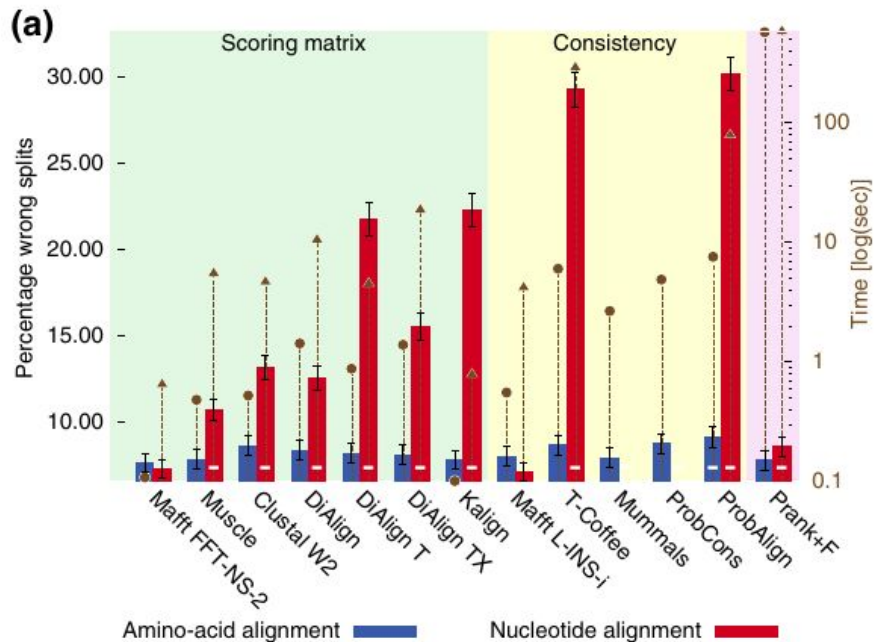


1. Лучший подход к выравниванию?
2. Есть ли информация в пропусках?
3. Влияние вариативности выравнивания?

# Оценка точности выравнивания на основе филогении



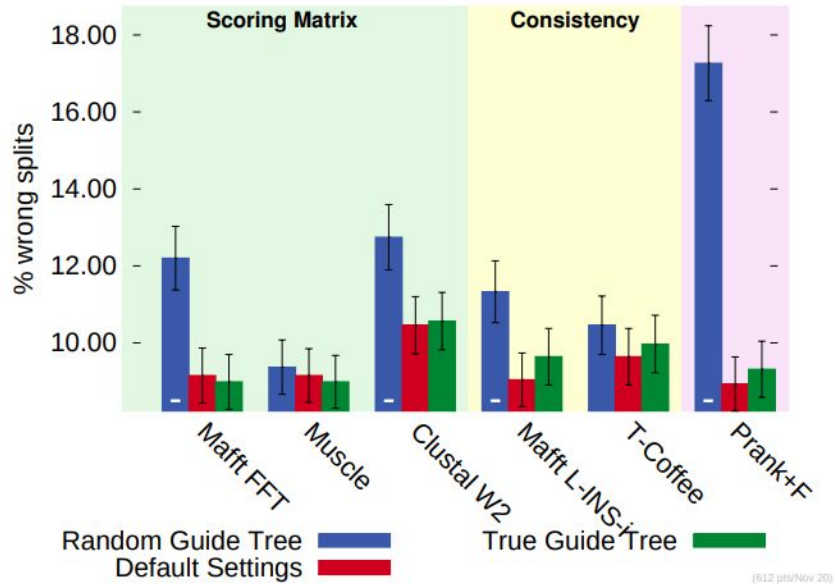
# Сравнение методов выравнивания



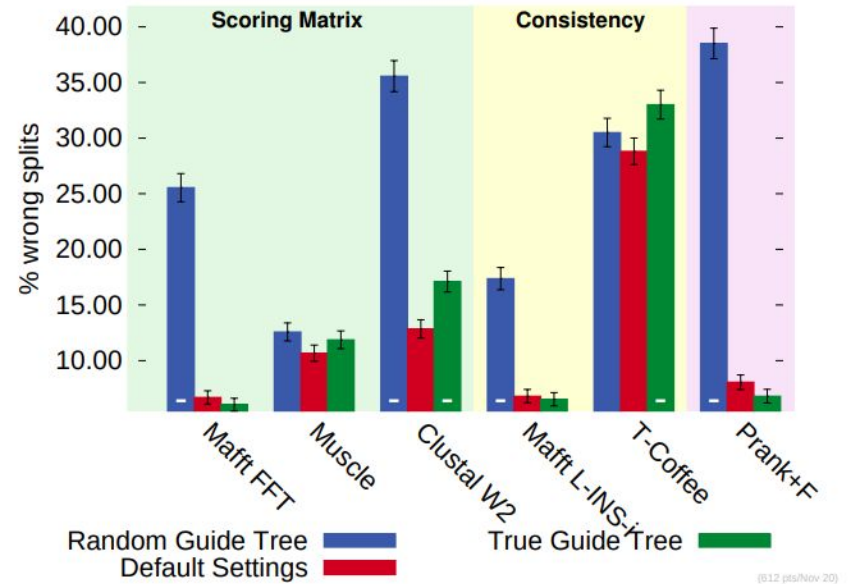
Аминокислоты: Mafft-FFT-NS-2, DiAlign TX, Prank

Нуклеотиды: Mafft L-INS-i

# Использование вспомогательного дерева



(a) Eukaryotes, Amino-acids

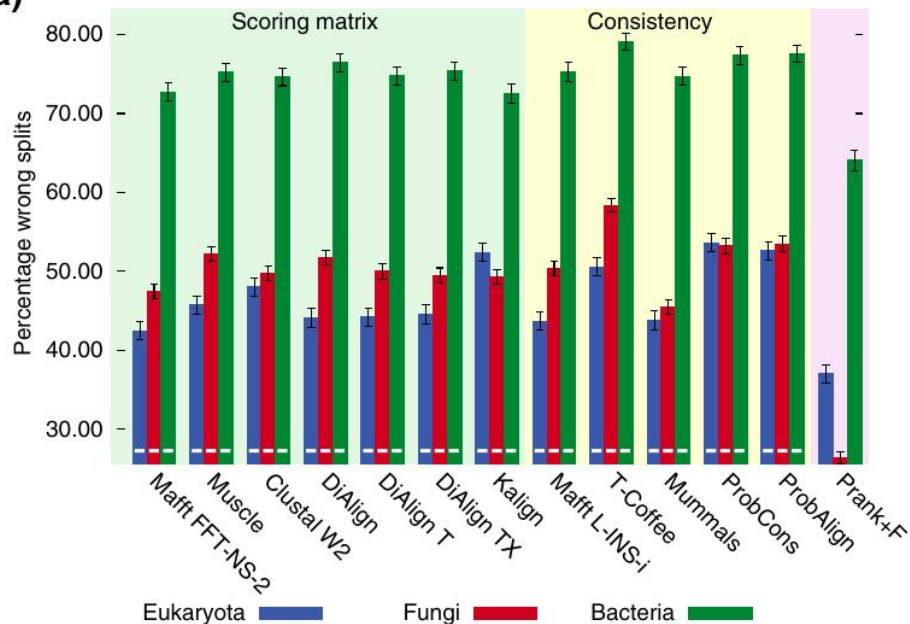


(b) Eukaryotes, Nucleotides

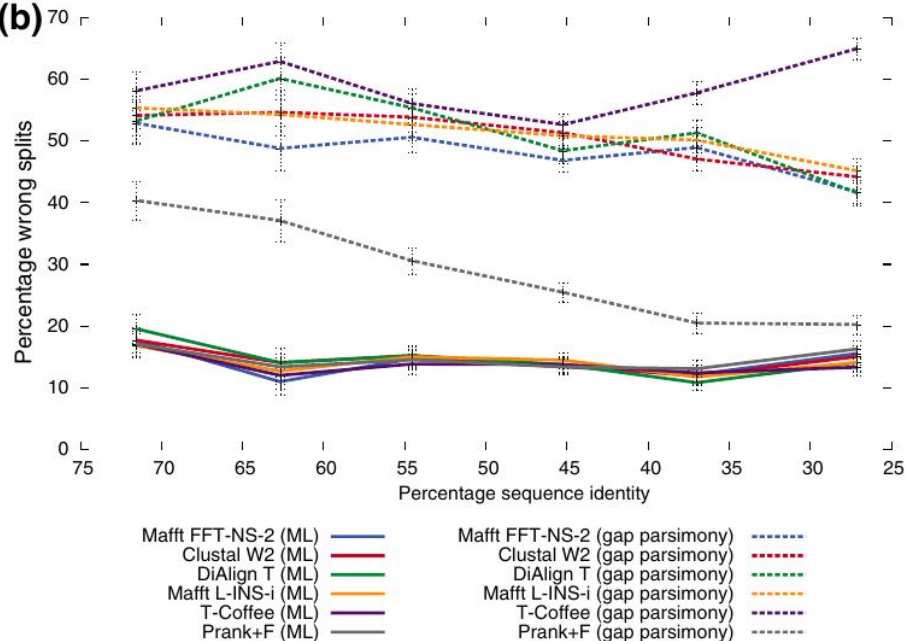
Muscle и T-Coffee нечувствительны

# Пропуски несут в себе филогенетический сигнал

(a)

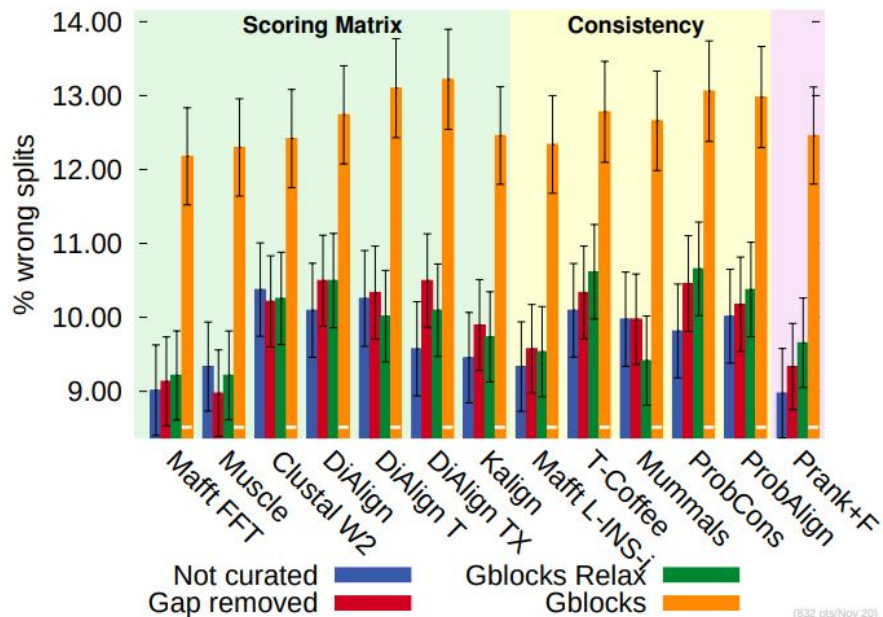


(b)

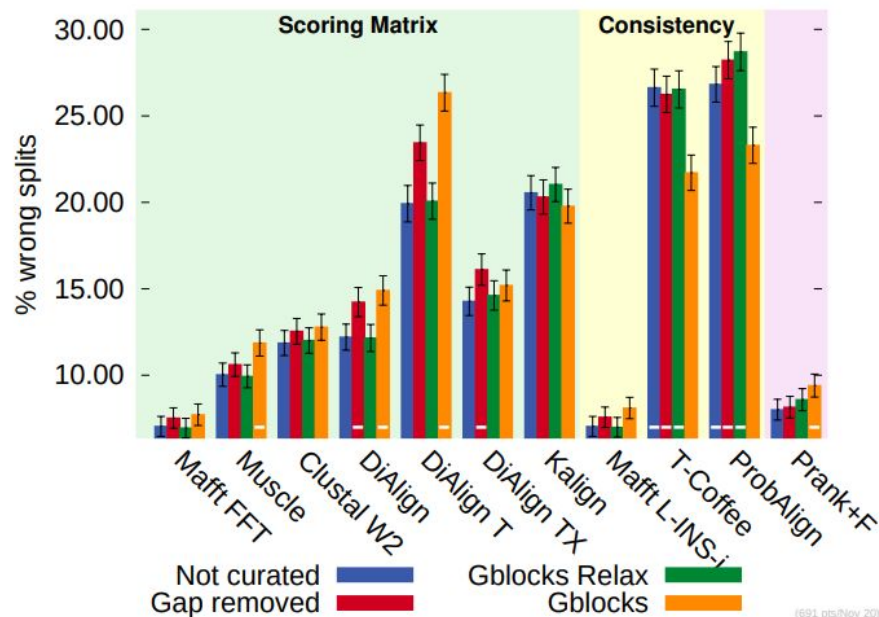


Prank на пропусках близок к ML на обычных выравниваниях

# Исключение пропусков и вариативных участков



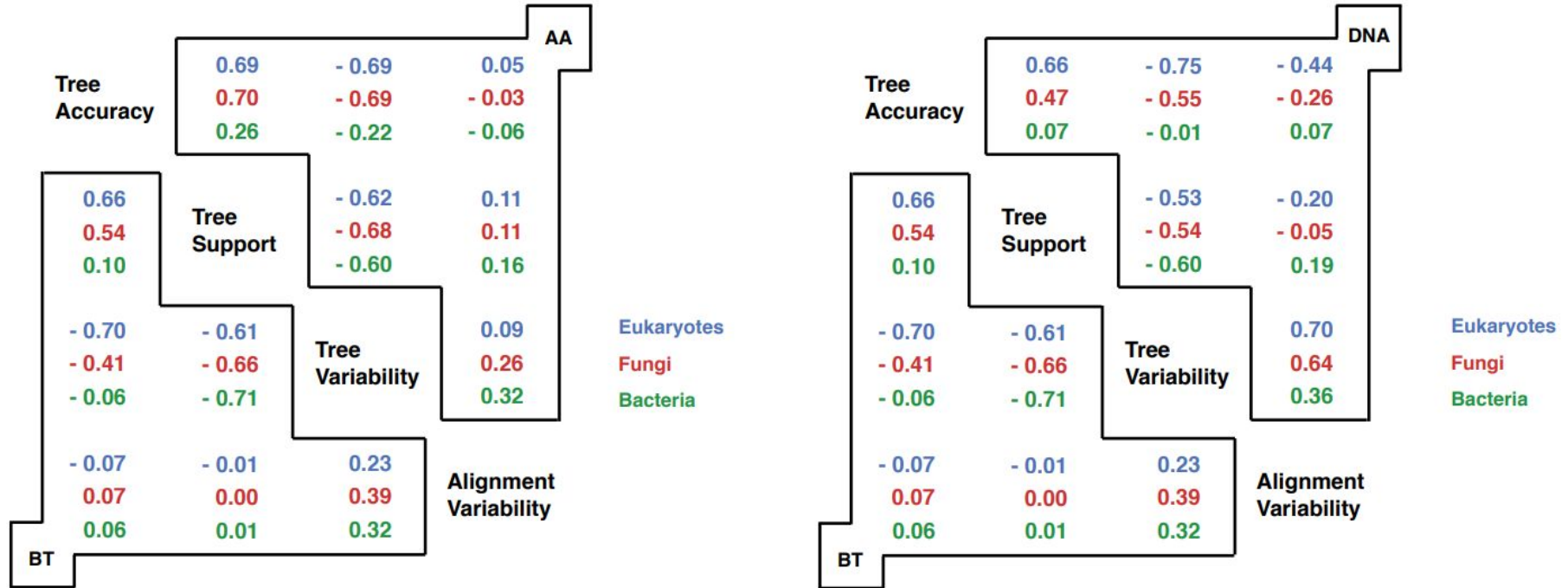
(a) Eukaryotes, Amino-acids



(b) Eukaryotes, Nucleotides



# Влияние вариативности выравнивания

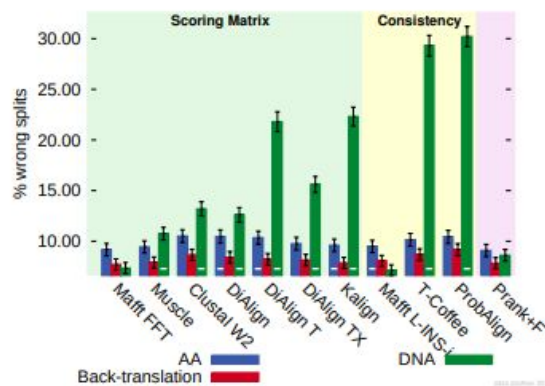


Вариативность выравнивания и точность дерева слабо коррелируют

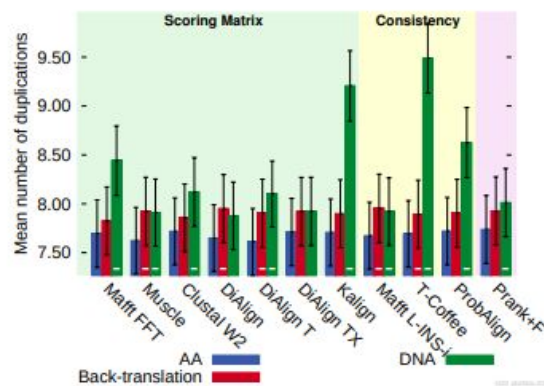
# Заключение

- Использование деревьев для оценки предпочтительнее структурных данных
- Консистентные методы выравнивания не лучше методов, основанных на матрицах
- Пропуски содержат в себе значимый филогенетический сигнал, который не учитывается программами выравнивания
- Низкая корреляция между вариативностью выравнивания и точностью дерева: нет необходимости в использовании более, чем одного метода

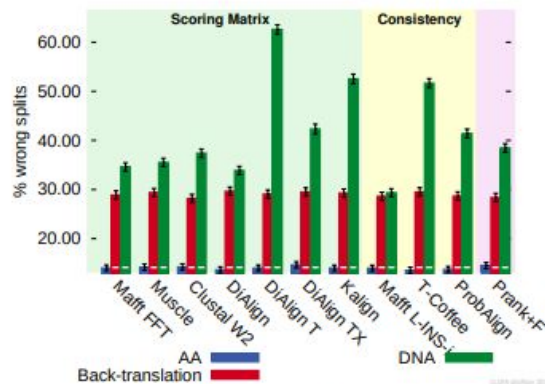
# Comparison between amino-acid, back-translated, and nucleotide data



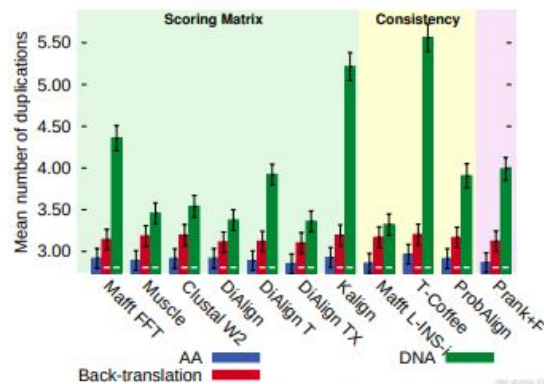
(a) Species-Tree Discordance, Eukaryotes



(b) Minimum Duplication Test, Metazoa

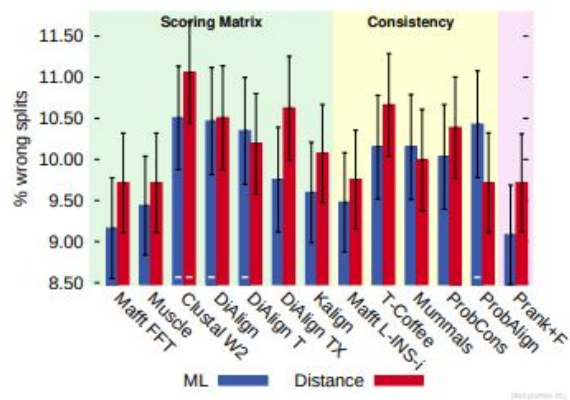


(c) Species-Tree Discordance, Fungi

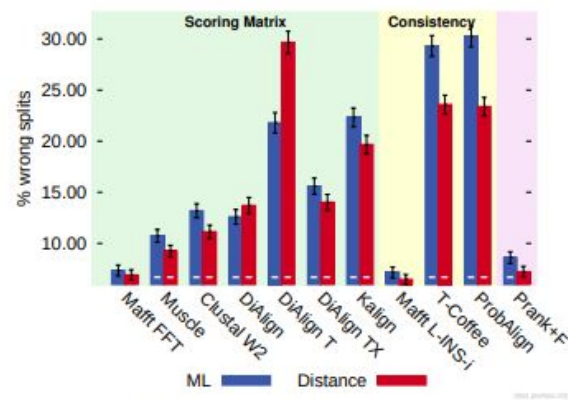


(d) Minimum Duplication Test, Fungi

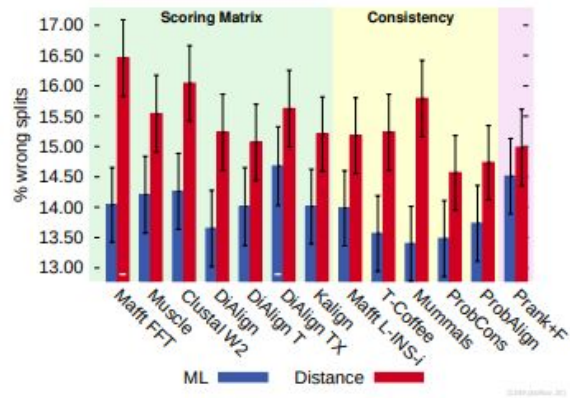
# Insensitivity to tree building method: species-tree discordance test



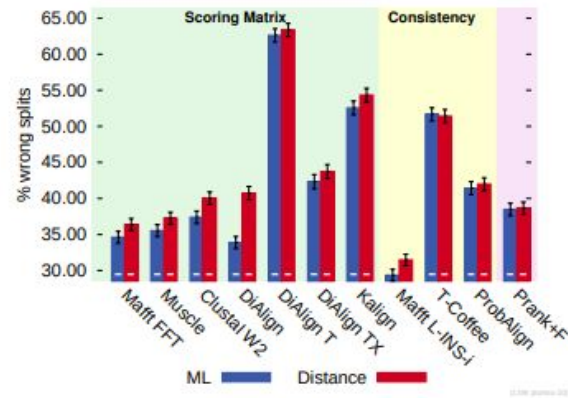
(a) Eukaryotes, Amino-acids



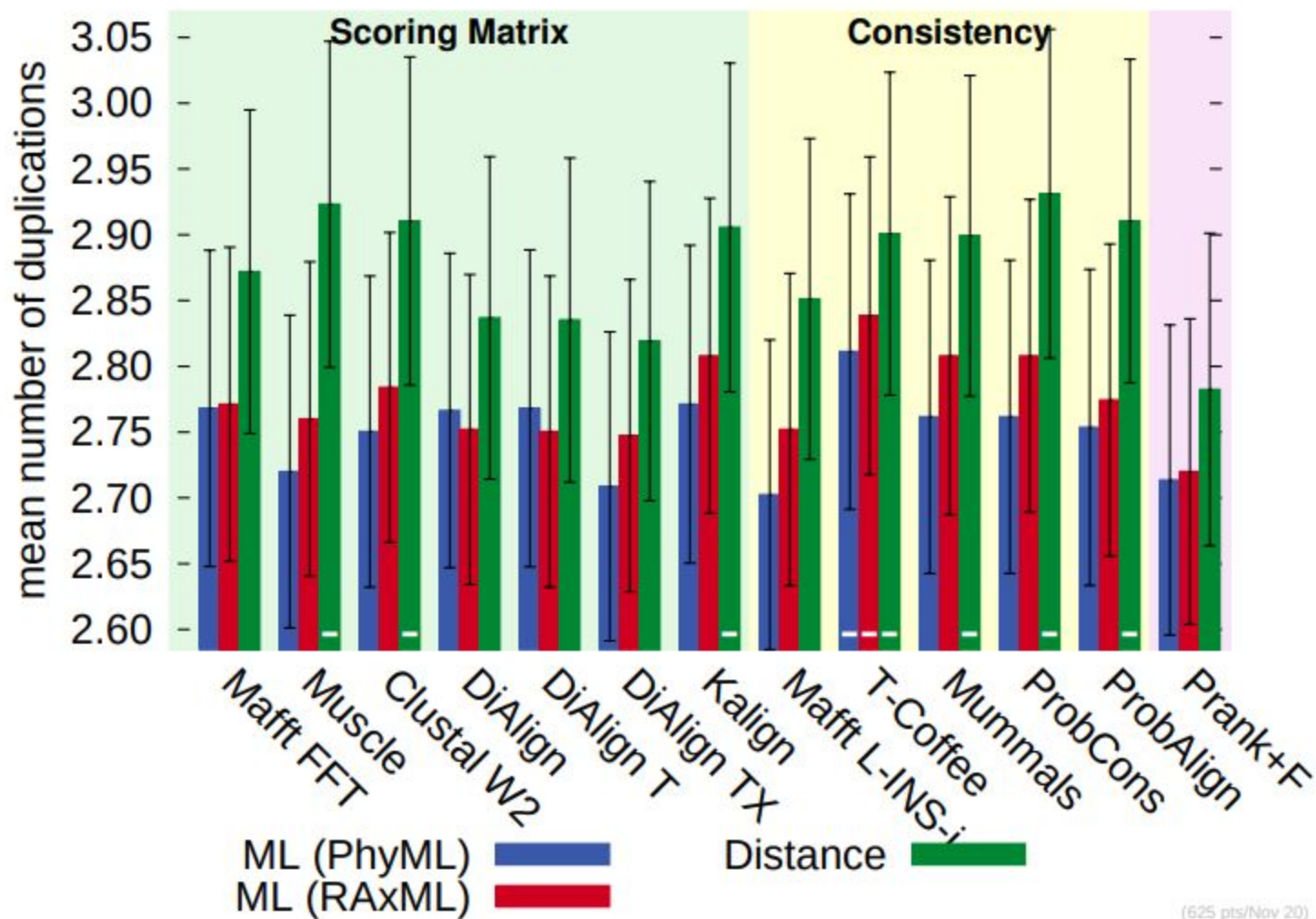
(b) Eukaryotes, Nucleotides



(c) Fungi, Amino-acids



(d) Fungi, Nucleotides



$$d = \frac{z(r_S(BS,TA)) - z(-r_S(AV,TA))}{\sqrt{(n_{BS,TA} - 3)^{-1} + (n_{AV,TA} - 3)^{-1}}}$$