

## Анализ выравниваний HTR2A

Прежде всего, хотелось бы отметить, что mafft и muscle получили одинаковые выравнивания. Поэтому далее они будут рассматриваться как одно и то же выравнивание.

### 1. Water.

Сравнение двух полученных в результате выравнивания последовательностей. Желтым обозначены участки в гене человека, а зеленым - в гене мыши.

Тексты разные! В текстах присутствуют различающиеся фрагменты. Их всего: 303 .

Результат сравнения:

[Экспортировать в Word](#)

```
ATGGAAATTCTTTGTGAAGAATACTTCTTCTGAGCTCAACTTCCAAACTCCCTTAATGCAA
TTAAATGATGACATCTAGGCTCTACAGCTTAATGACTTAACTCTAGGAGAAAGCTAACACTTC
GATAGCATTTGAAGTGGACAGTCTGAETCTGAAAACTGAACCAACCTTCTCTGTGAAGGGTG
CTCTCACCCGTACATGTTCTCTCCCTTCTTCATCTCCAGGAAAAAACTGGTCTGCTTTAATG
ACAGCTGTAAGTGATTATTCTACATTTGCTGGAAACATACTGTGCATCATGGCAGTGTCC
CTAGAGAAAAAGCTGCAGAATGCCACCACTATTTCTGTATGTCACCTGCCATAGCTGAT
ATGCTGCTGGGTTTCTTGTCTATGCCCGTGTCCATGTTAACCATCTGTATG
GGTACCGGTGGCTCTGCCAGCAAGCT
TGTGCACTCTGGATTACCTGGAGTGTCTCTTCTCCACGGCTCCATCATGCACCTCTG
CGCCATCTCTCTGGACCGCTACGTGCTATCCAGAATCCCATCCACAGCCGCTTCAA
CTCCAGAACTAAAGCTCTCTGAAAAATCAATTGCTGTCTGGACCATATCATAG
GTATACCATGCCAATCCAGTCTTG
GGCTACAGGAGTATTGTAAGGTCTTAAAGGAGGGAGTGTCTTCTCGCCGATGAACT
TTGCTCTGATGCTCTTTTGTGTCATTTTCACTCCCTTAACCAICATGGTATCACCT
ACTTCTACTATCAAGTCACTCAGAAAGAAGCTACCTTGTGTGTAGAGTGATCTTGAGCA
CATCGGGCCAAATAGCTCTCTTTCAGCTTCTCCCTCAGAGTTCTTGTCTATCAGAAAAGC
TCTTCACGCGGTCTATCCATAGGAGAGCCAGGGCTCTTACAGGCAGGAGGAGGACTATGCAGT
CCATCAGCAATGAGCAAAAAGCATGCAAGGTGCTGGGCATCGCTTCTCTCTGTTTGTG
TGATGTGGTGCCCTTCTTCTATCACAACTATCATGGCCGTCATCTGCAAAAGATCCTGCA
ATGAGGAATGTCATTGGGAGCCCTGCTCAATGTGTTGTGTGGATCGGTTATCTCTCTCAG
CATGTCAACCCACTAGTCTATCTACAGTGTTCAACTAAGTACCTATAGGTCAAGCCTTTTCACGGT
ATGATTCAGTGTGAGTACAAGGAAGAACAAAGCCATCTGCAGTTAATTTAGTGAACACATA
TACCGGACCTTTGGCCCTACAAGTCTAGCTCAAGCTTCAAGCTGGGACAAAAAAGAACTCAAAGG
AAGATGCCAATGATCAGACAGATAATGACTGCTCATGGTTCATAGGAAAGGAGGACATCTGG
AAGAGGCTCTGTAAGAGACAATAAGAAAGCGGAGTGAATGAAAAGGTGAGCTGTGTGTGA
```

Проанализировав это выравнивание, можно заметить, что в нем довольно часто встречаются mismatch'и. Также у человека мы можем увидеть несколько рядов N - то есть это может быть любой нуклеотид из A/C/G/T. В том же месте у мыши мы можем наблюдать gap'ы. Можно сказать, что это участок, специфичный для человека, либо же это говорит о делеции у мыши или же инсерции у человека.

Если мы сравним изначальную последовательность гена мыши и выровненную, то увидим, что из изначальной последовательности из конца и начала вырезана часть последовательности, а также для лучшей совместимости вставлены пропуски.

При этом изначальная последовательность гена человека и выровненная остались без изменений.

Алгоритмы отработали примерно одинаково, но всё же отличия есть. Water не изменил ген человека, а mafft и muscle практически не изменили ген мыши. И если сравнивать оба выравнивания, то у гена человека mafft и muscle добавили gap'ы в самом начале и в конце, а также они сделали много

инсерций в начале и конце гена мыши. А середина у обоих выравниваний совпадает.

Слева сравнение выровненных генов человека, а справа - генов мыши.

```
ATGGATATTTCTTTGTGAAGAAAATACCTTCTTTGAGCTCAACTACAAACTCCCTAATGC
AA
TTAAATGATGACACCAGGCTCTACAGTAATGACTTTAACTCTGGAGAAGCTAACACTT
CT
GATGCATTAACTGGACAGTCGACTCTGAAAATCGAACCAACCTTTCTGTGAAGGGT
GC
CTCTACCGCTCGTGTCTCTCTTACTTTCATCTCCAGGAAAAAACTGGTCTGCTTTAC
TG
ACAGCCGTAGTGATTATTCTAACTATTGTGGAAACATACCTCGTCATCATGGCAGTGT
CC
CTAGAGAAAAAGCTGCAGAATGCCACCAACTATTTCCTGATGTCACCTTGCCATAGCTC
AT
ATGCTGCTGGGTTTCTTGTGTCATGCCCGTGTCCATGTTAACCATCCTGTATG
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GGTACCGGTGGCCTCTGCCGAGCAAG
CT
TTGTGCTGCTGGATTACCTGGACGTGCTCTTCTCCACGGCCTCCATCATGCACCTC
TG
CGCCATCTCGCTGGACCGCTACGTCGCCATCCAGAATCCATCCACCACAGCCGCTTC
AA
CTCCAGAACTAAGGCATTCTTGAAATCATTGCTGTTTGGACCATATCAGTAG
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GTATATCCATGCCAATACCAGTCTT
TG
GGCTACAGGACGATTTCGAAGGTCTTAAAGGAGGGAGTGTCTTACTCGCCGATGATAA
CT
TTGTCTGATCGGCTCTTTTGTGTCATTTTTCATTCCCTTAAACCATCATGGTGATCAC
CT
ACTTTCTAACTATCAAGTCACTCCAGAAAGAAGCTACTTTGTGTGAAGTATCTTGG
CA
CACGGGCCAAATTAGCTTCTTTCAGCTTCCTCCCTCAGAGTCTTTGTCTTCAGAAAA
GC
TCTTCCAGCGGTGATCCATAGGGAGCCAGGCTCTACACAGGCAGGAGGACTATGCA
GT
CCATCAGCAATGAGCAAAAGGCATGCAAGGTGCTGGCATCGTCTTCTTCTGTTGT
GG
TGATGTGGTGCCCTTTCTTCATCACAACATCATGGCCGTATCTGCAAGAGTCTCTG
CA
ATGAGGATGTCATTGGGGCCCTGCTCAATGTGTTTGTGATCGGTTATCTCTCTTC
AG
CAGTCAACCCACTAGTCTACACACTGTTCACAAGACCTATAGGTGAGCCTTTTCACC
GT
ATATTCACTGTCAGTACAAGGAAAAACAAAAACCAATGTCAGTTAATTTAGTGAACAC
AA
TACCGGCTTTGGCCTACAAGTCTAGCCAACTTCAAATGGGACAAAAAAGAATTCAAA
GC
AAGATGCCAAGACAACAGATAATGACTGCTCAATGGTTGCTCTAGGAAAGCAGCATTC
TG
AAGAGGCTTCTAAAGACAATAGCGACGGAGTGAATGAAAAGGTGAGCTGTGTGA
```

```
ACGAGCCACACAGGCTCCGCTTCAAGCATGCCCTAGCTCCAGGACGTAAAGATGAATGG
GAGCCCCGGCTATGACTCGCTAGTCTCTCCACACTTCATCTGCTACAACCTCCGGCTTAC
AC
ATGGAAATCTCTGTGAAGACAATATCTCCCTGAGCTCAATCCAAACTCCTTAATGC
AA
TTAGGTGACGACTCGAGGCTCTACCTAATGACTTCAACTCCAGGATGTAAACACTT
CC
GAAGCCTCGAACTGGACAATTGATGCTGAAAACAGAACCAACCTCTCTGCGAAGGGT
AC
CTCCCACCGACATGCCTCTCCATTTTCATCTCCAGGAAAAAACTGGTCTGCTTTAT
TG
ACAACTGTGCTGATTATCTCACCATTGCGGGAACATACTGGTATCATGGCAGTGT
CC
CTAGAGAAAAAGCTGCAGAATGCCACCAACTATTTCCTGATGTCACCTTGCCATAGCTC
AT
ATGCTGCTGGGTTTCTTGTGTCATGCCCGTGTCCATGTGTAACCATCCTGTATG
GGTACCGGTGGCCTTTGCCACGCAAG
CT
CTGTGCGCTGTGATTACCTGGATGTGCTCTTCTCCACGGCTCCATCATGCACCTC
TG
CGCCATCTCCCTGGACCGCTACGTGGCTATCCAGAACCCATTACCATAGCCGCTTC
AA
CTCCAGAACCAAGCCTTCTTGAAATCATTGCGGTGTGGACCATATCCGTAG
GTATATCCATGCCAATCCAGTCTT
CG
GGCTACAGGATGATTTCGAAGGTCTTAAAGGAAGGGAGCTGCCTGCTCGCCGATGACAA
CT
TTGTCTCATAGGCTCTTTTGTGGCATTTTTCATCCCCCTAACCATCATGGTGATCAC
CT
ACTTCTGACTATCAAGTCACTTCAGAAAGAAGCCACCTTGTGTGTGAGTGACCTCAG
CA
CTCGGGCCAAATATCTCTCTCAGCTTCTCTCCCTCAGAGTTCCTGTATCATGAAAAA
GC
TCTTCCAGCGGTCCATCCACAGAGAGCCAGGCTCTACGAGGCGGAAGGACGATGCA
GT
CCATCAGCAACGAGCAAAAGCATGCAAGGTGCTGGGCATCGTGTCTTCTGTGTTGT
TG
TAATGTGGTGCCCATCTTTCATCACCATATATGCGCCGTATCTGCAAGAATCCTG
CA
ATGAAAATGTCAATTGGAGCCCTGCTCAATGTGTTGTGCTGGATTGTTATCTCTCTC
AG
CCGTCAACCCACTGGTATATACGTTGTTCATAAACTTATAGGTCGCGCTTCTCACG
GT
ACATTCACTGCCAGTACAAGGAGACAGAAAGCCGCTGCAGTTAATTTAGTGAACAC
TA
TACCAACATTGGCCTACAAGTCTAGTCAGCTCCAGGTGGGACAAAAAAGAATCACA
GG
AAGATGCTGAGCCGACAGCTAATGACTGCTCCATGGTTACACTAGGGAACCAACACTC
GA
ACCGATCTCTGAGATGCCAGGGTGTGTGACAGTACAACCACTGTGTGGG
GACAGGATAGGAGCTTGAATATTAGTCTCTGACCACTGACCACTGACCACTAATGTCGG
CA
AGCATTTGACAGGAGTGTGTAAGCTTTAACAAGCAATGAGATCTTAAAT
TG
TGCCTTATGTACAATGTATGAGCCATAAAATCCATGTAATTAAGTCTGAT
GTCAATATGAAACCGGCTGCCATGTGTGATGGG
```