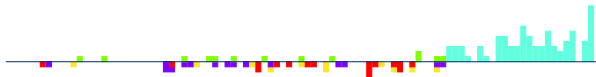


E: Evolutionary Excerpt

Problem Author: Ragnar Groot Koerkamp

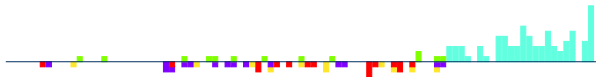


- **Problem:** Given two independent uniform random sequences over “ACTG” of length $n = 10^6$, find a common subsequence of length at least 500 000.

Statistics: 139 submissions, 13 accepted, 84 unknown

E: Evolutionary Excerpt

Problem Author: Ragnar Groot Koerkamp

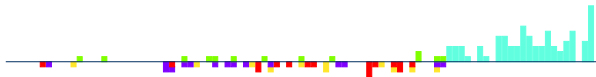


- **Problem:** Given two independent uniform random sequences over “ACTG” of length $n = 10^6$, find a common subsequence of length at least 500 000.
- Naive solution: run the Longest Common Subsequence algorithm. $\mathcal{O}(n^2)$ is too slow!

Statistics: 139 submissions, 13 accepted, 84 unknown

E: Evolutionary Excerpt

Problem Author: Ragnar Groot Koerkamp

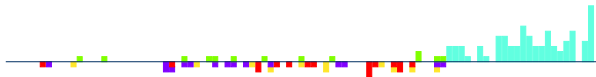


- **Problem:** Given two independent uniform random sequences over “ACTG” of length $n = 10^6$, find a common subsequence of length at least 500 000.
- Naive solution: run the Longest Common Subsequence algorithm. $\mathcal{O}(n^2)$ is too slow!
- Greedy: if the front two characters are the same, take it. Otherwise, remove the first character from the longer sequence. $\rightarrow$ length 400 000.

Statistics: 139 submissions, 13 accepted, 84 unknown

E: Evolutionary Excerpt

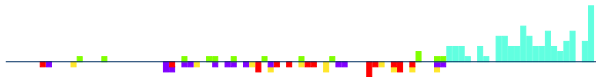
Problem Author: Ragnar Groot Koerkamp



- Greedy, second attempt: Instead of only comparing the front characters, we can compare the front character of each sequence with the first three or four characters of the other sequence, and use the first match we find. → length 531 000.

E: Evolutionary Excerpt

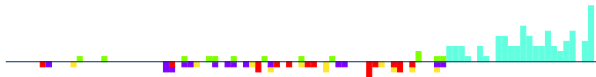
Problem Author: Ragnar Groot Koerkamp



- Greedy, second attempt: Instead of only comparing the front characters, we can compare the front character of each sequence with the first three or four characters of the other sequence, and use the first match we find. → length 531 000.
- LCS DP, but smarter: instead of computing the full n^2 DP table, we can only keep entries close to the diagonal. Keeping a diagonal of width $k = 10$ → length 624 000, $\mathcal{O}(nk)$.

E: Evolutionary Excerpt

Problem Author: Ragnar Groot Koerkamp



- Greedy, second attempt: Instead of only comparing the front characters, we can compare the front character of each sequence with the first three or four characters of the other sequence, and use the first match we find. $\rightarrow$ length 531 000.
- LCS DP, but smarter: instead of computing the full n^2 DP table, we can only keep entries close to the diagonal. Keeping a diagonal of width $k = 10 \rightarrow$ length 624 000, $\mathcal{O}(nk)$.
- Split the input in chunks of size $k \geq 7$, and run LCS for each chunk. $\rightarrow \mathcal{O}(nk)$, length 502 000 for $k = 7$, length 530 000 for $k = 10$. Probability of failure is less than 10^{-16} for $k = 7$, and less than 10^{-1000} from $k = 9$ onward.