

## Bài 4. Genome

DNA là thành phần cơ bản cấu tạo thành bộ genome của sinh vật. DNA bao gồm 4 loại khác nhau là {A,C,G,T}. Để nghiên cứu các sinh vật ở mức độ phân tử, người ta tiến hành giải mã bộ genome của chúng.

Để giải mã bộ genome của một sinh vật, máy giải mã thế hệ mới sẽ sinh ra  $N$  đoạn cơ sở, mỗi đoạn cơ sở là một dãy bao gồm 30 DNA. Các đoạn cơ sở sẽ được ghép nối với nhau để tạo thành một bộ genome hoàn chỉnh.

Ta nói một đoạn DNA  $X$  được bao phủ bởi một đoạn cơ sở  $Y$  nếu tồn tại một đoạn liên tiếp trong  $Y$  trùng với  $X$ . Giả sử  $k$  là số nguyên dương, một đoạn DNA  $X$  được gọi là *đoạn tin tưởng cấp  $k$*  nếu  $X$  được bao phủ bởi ít nhất  $k$  đoạn cơ sở.

**Yêu cầu:** Cho  $N$  đoạn cơ sở và số nguyên dương  $k$ , hãy tìm đoạn tin tưởng cấp  $k$  có độ dài lớn nhất.

**Dữ liệu:** Vào từ file văn bản GENOME.INP có cấu trúc như sau:

- Dòng đầu chứa hai số nguyên dương  $N$  và  $k$  ( $0 < k \leq N \leq 30000$ );
- Mỗi dòng trong  $N$  dòng tiếp theo chứa một đoạn cơ sở.

**Kết quả:** Đưa ra file văn bản GENOME.OUT một số nguyên là độ dài của đoạn tin tưởng tìm được. (Ghi số  $-1$  nếu không tồn tại đoạn tin tưởng cấp  $k$ ).

**Ví dụ:**

GENOME . INP	GENOME . OUT
4 3 AAAAAAAAATAAATAAAAAAAAAAAAATG AAAAAAAAAAAAAAAAAAAAATAAATGAAAA AAAAAAAAAAAAAAAAAAAAATGAAAAAAAA AAAAAAAAAAAAAATGAAAAAAGGGGAAAA	15

**Lưu ý:** 50% số test có  $N \leq 1000$  tương ứng với 50% tổng số điểm dành cho bài.

----- Hết -----