

宇野 毅明 (うの たけあき)

2003年: 東京工業大学 情報科学科卒業

2008年: 東京工業大学 システム科学専攻 博士取得
東京工業大学 経営工学専攻助手 (金融工学)

年齢・職種: 国立情報学研究所・准教授・40歳、総合研究大学院大学

研究分野: アルゴリズム理論とその応用

- グラフアルゴリズムを中心とした離散アルゴリズム
- 組合せ最適化とそれに関わる数理
- 列挙アルゴリズムと、頻出パターンマイニングへの応用

最近の研究: ゲノム情報学やデータマイニングで出てくる巨大なデータベースの基礎的(とはいっても非常に時間のかかる)な解析を超高速で行うアルゴリズムの開発

もつとー: 速い安い(単純)うまい

趣味(日課?): 子供と遊ぶこと

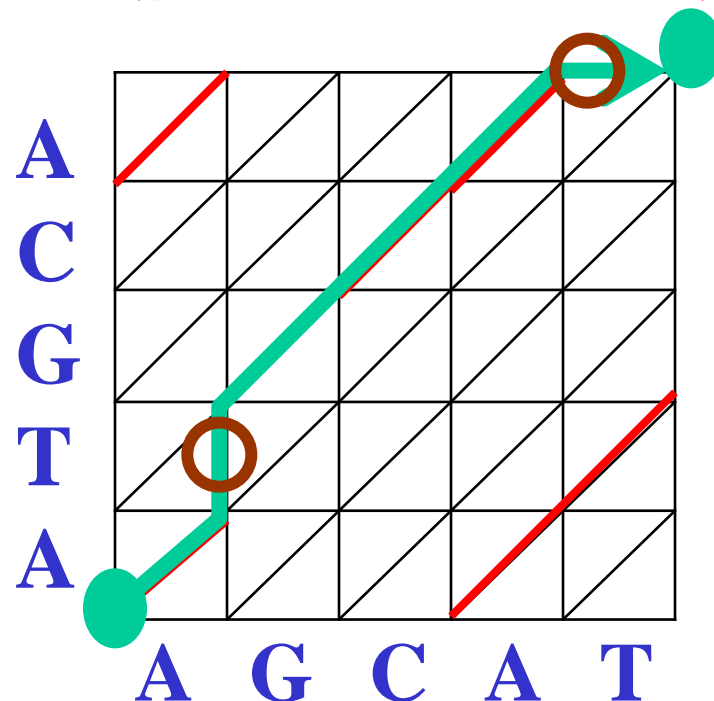
最短路問題

- 2つのゲノムに対して、グリッド型のネットワークを作る

ATGCA
AGCAT

横移動 : 左側に空白を挿入
縦移動 : 下側に空白を挿入
斜め移動 : 空白無し(同じ文字ならコスト0)

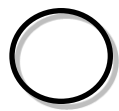
↓
○ ○
ATGCA*
A*GCAT



何がほんとに似ているか

- 生物系の研究者に聞くと、編集距離は(距離が大きいと)研究者の直感に対応していないようだ

→ 同じ編集距離(ハミング距離)でも、似てる感が大きく違う



ATGCATATATATATATATGCATGC
ATGCATGCGCGCGCGCGCGCATGC



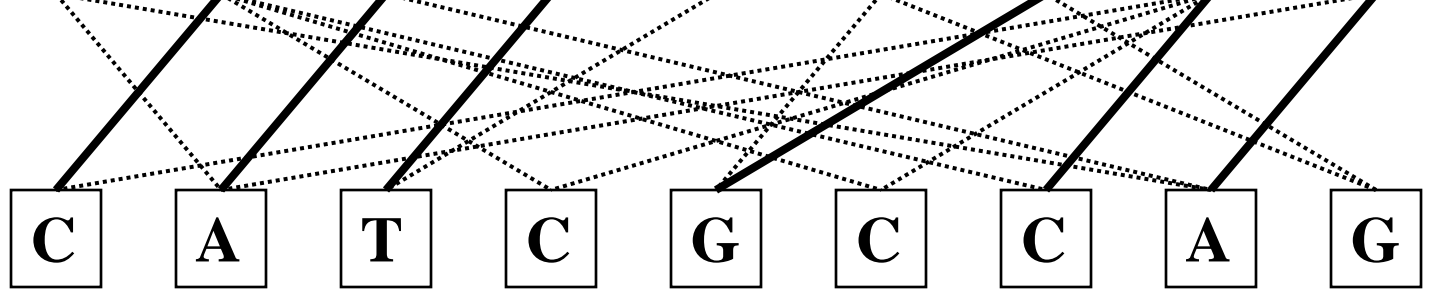
ATGCATGCATATATATATGCATGC
AAGGAAGGAAAAAAAAGGAAGG

- どうも、部分的に集中して似てるところが効いているようだ
- あまり短い部分だけが似ていても、意味がないようだ
- さらに、実用上は「似ている部分が見つければいい」ので、解の中に「似ていない部分が混ざっていてもいい」
(見つけ損ないはダメだが、偽者を見つけても大きな被害はない)

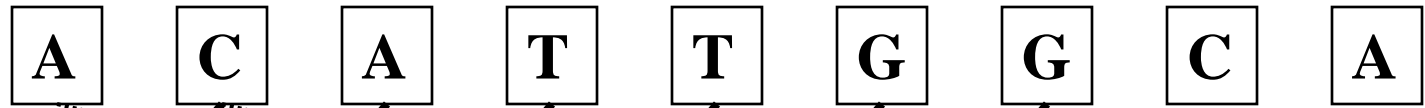
string (a)



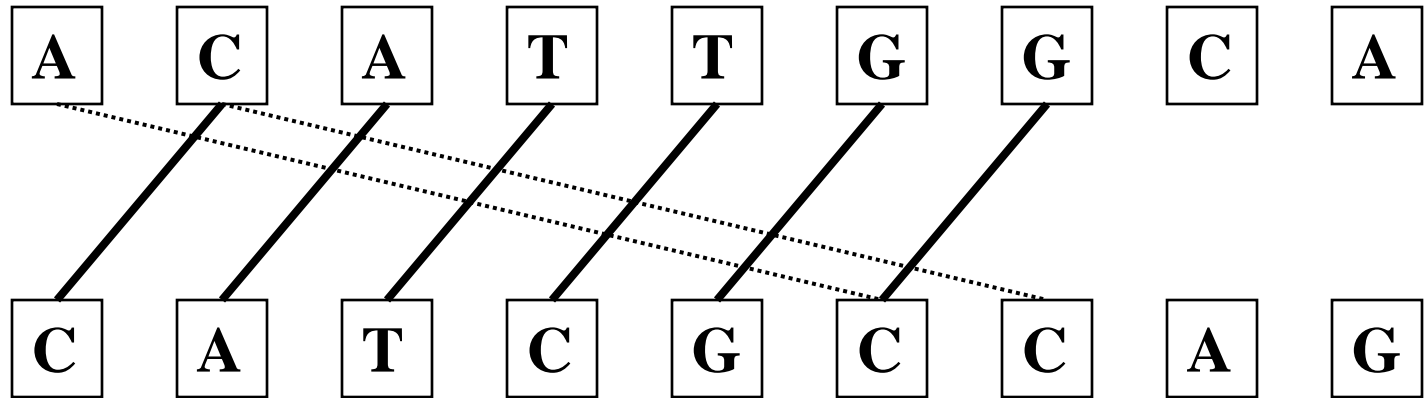
string (b)

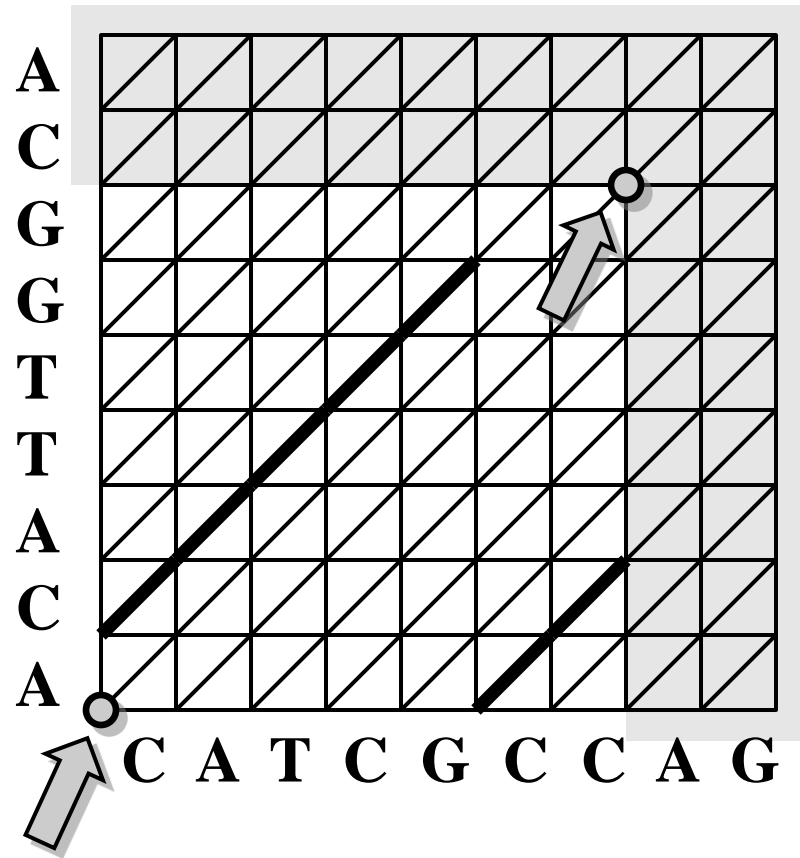
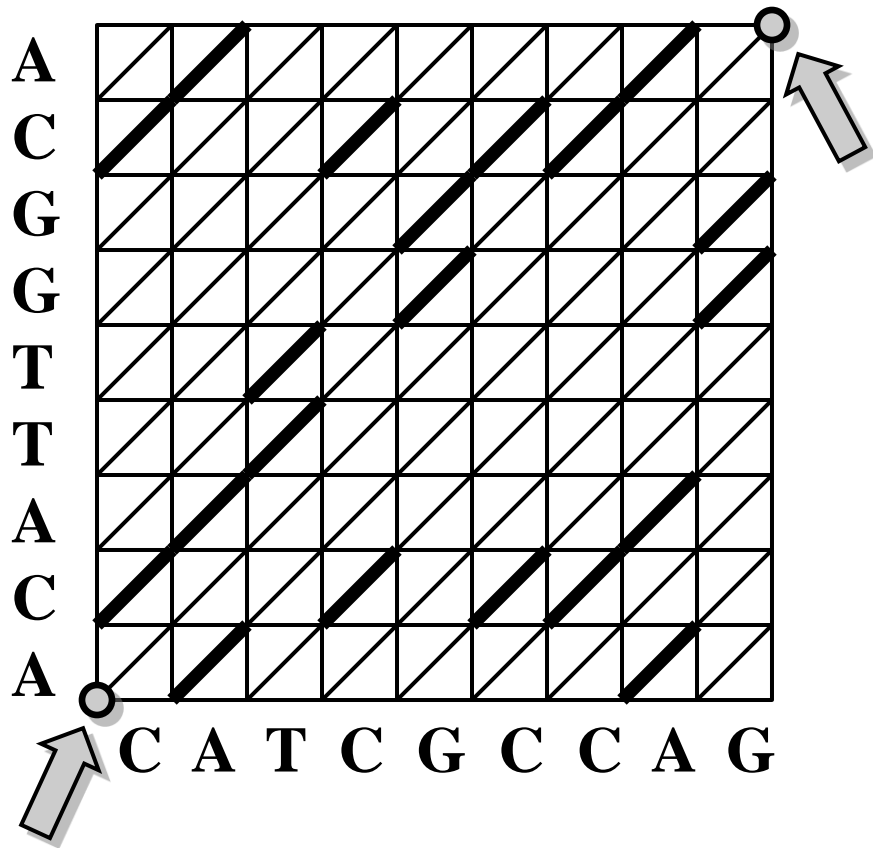


string (a)

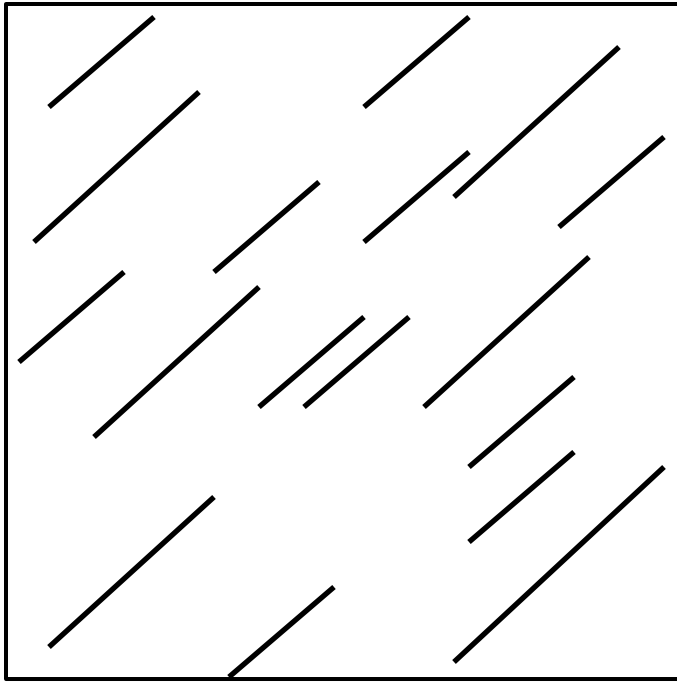


string (b)





•••
GGCA



ATCG•••

