

Parallel Application Systems in Genetic Information Processing

I C O T

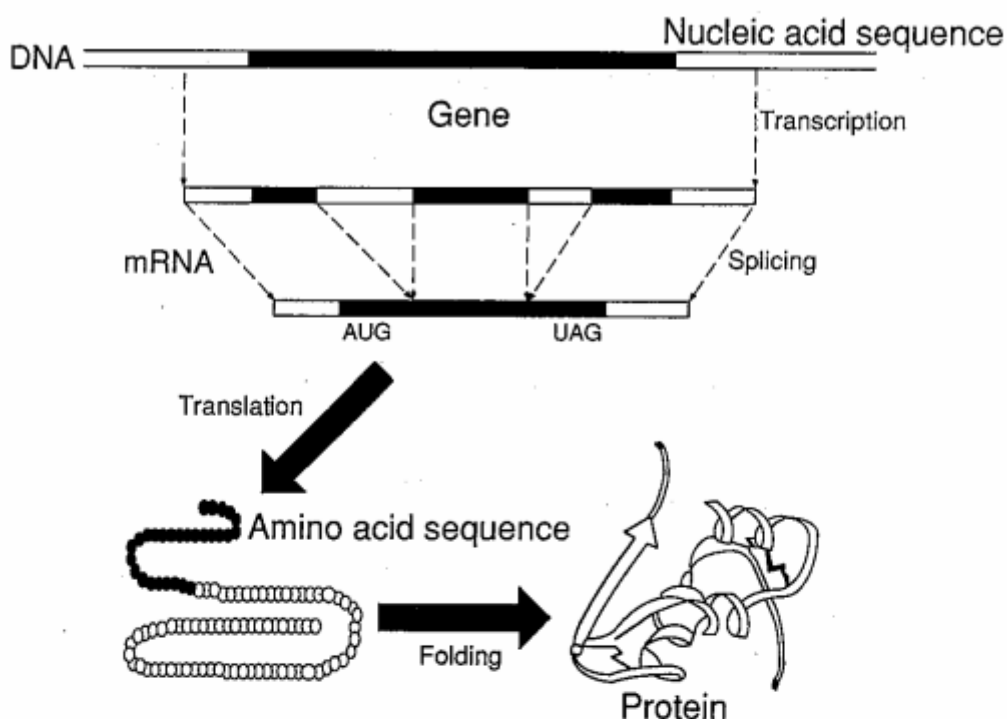
Second Research Department

MASATO ISHIKAWA

Contents

- Genetic sequence analysis
- Parallel iterative aligners
 - Parallel best-first search
 - Parallel hill-climbing search
 - Multi-group genetic algorithm
- Alignment workbench
 - Constraint-based alignment
 - Refinement tools

Gene and Protein



Multiple Sequence Alignment

Protein sequences

```

HTLV : PVLQLSPAELHSFTHCGQTALTLQGATTEASNILRSCHACRGGNPQHQMGRGHI
RSV  : QATFQAYPLREAKDLHTALHIGPRALSKACNISMQQAREVVTCPHCNSAPALEAGVN
MMTV : ISDP IHEATQAHTLHHLNAHTLRLLYK I TREQARDIVKACKQCVVATPVPHLGVN
SMRV : ILTAESAQESHALHHQNAALRFQFH I TREQAREIVKLCPCPDWGSAPQLGVN
M-MuLV : LHQLTHLSFSKMKALLERSHSPYYMLNRDRTLKNI TETCKACAQVNASKSAVKQGTR
copia : HEKLLHPG IQTKTLFGETYYFPNSQLL IQNI INECSI CNLAKTEHRNTDMPKTIT
    
```

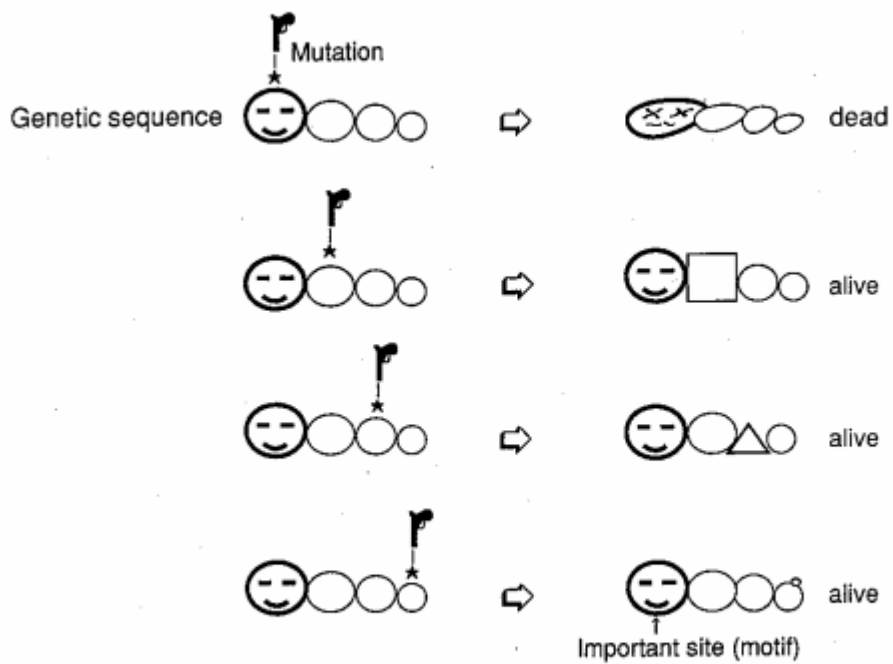


Sequence Alignment

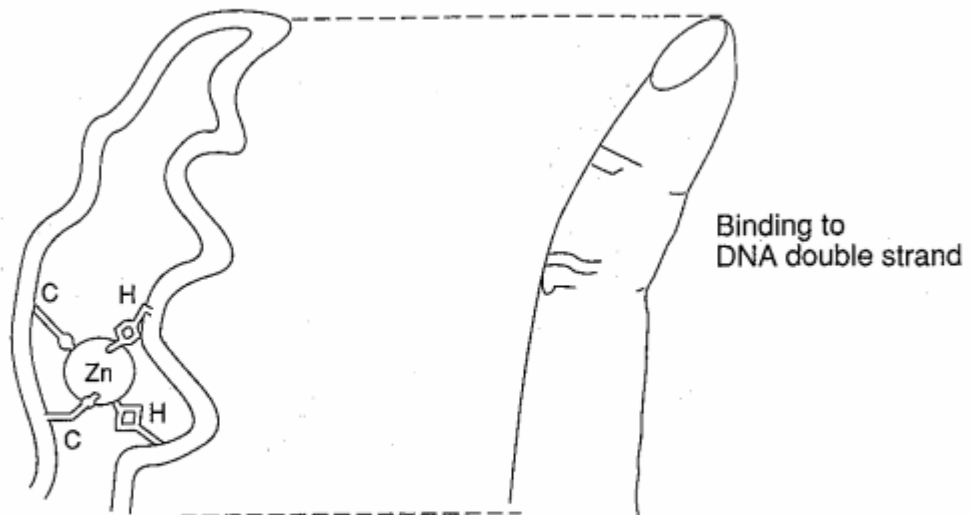
```

HTLV : -PVLQ---LSPA-ELHS-FTHCG---QTAL--TLQ---GATTTEA--SNILRSCHAC---RGGNPQHQMGRGHI---
RSV  : QATFQAYPLREAKDLHT-ALHIG---PRAL--SKA---CNISMQQA--REVVTCPHC-----NSAPALEAG-VN--
MMTV : --ISD--PIHEATQAHT-LHHLN---AHTL--RLL---YKITREQA--RDIVKACKQC---VVATPVPHL--G-VN--
SMRV : --ILT--ALESQESHA-LHHQN---AAAL--RFQ---FHTREQA--REIVKLCPCN---PDWGSAPQL--G-VN--
M-MuLV : -----LHQLTHLSFSKMKALLERSHSPYYMLNRDRTL-KNITETCKAC--AQVNASKSAVKQG-TR--
copia : -----HEKLLHPG IQTKTLF-GET---YYFPNSQLL IQNI INECSI CNLAKTEHRNTDM--P-TKTT
    
```

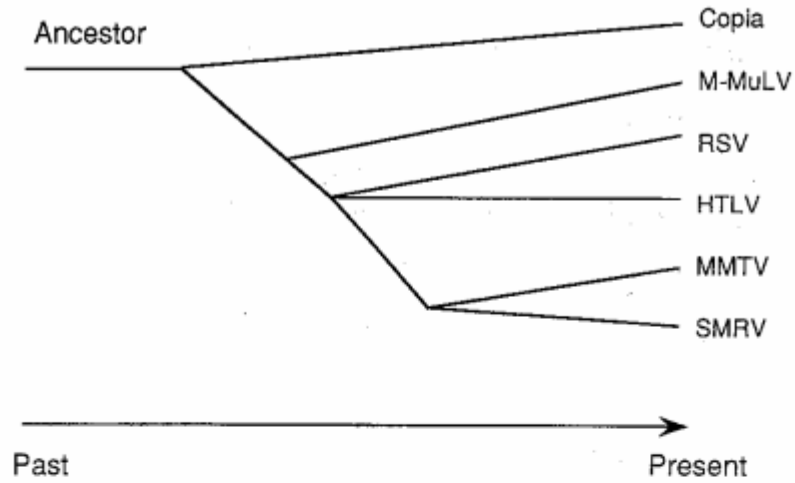
Principle of Sequence Alignment



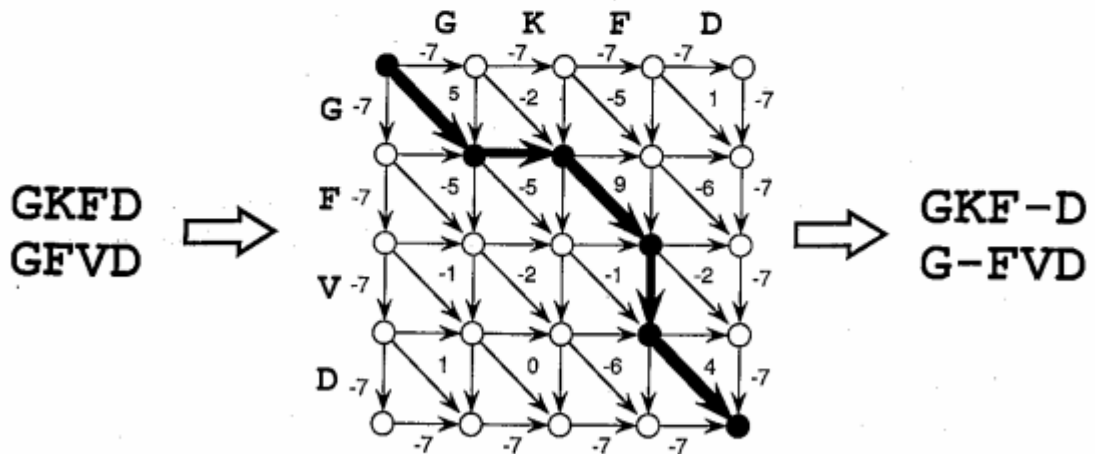
Zinc Finger Motif



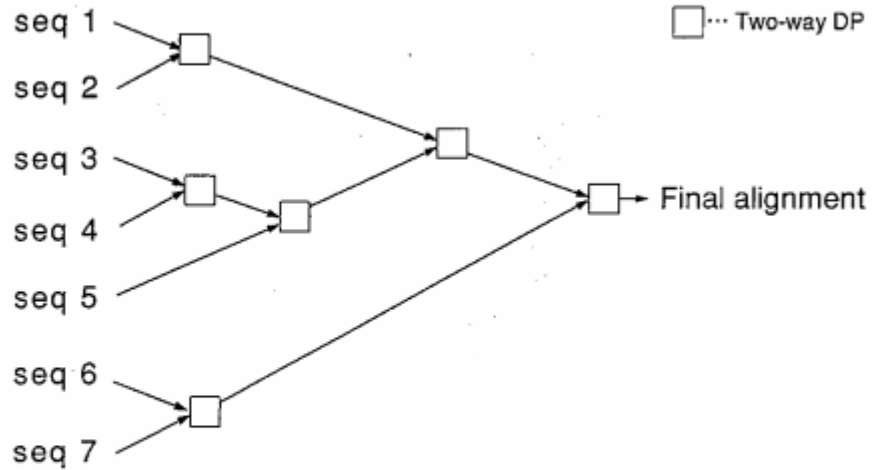
Phylogenetic Tree



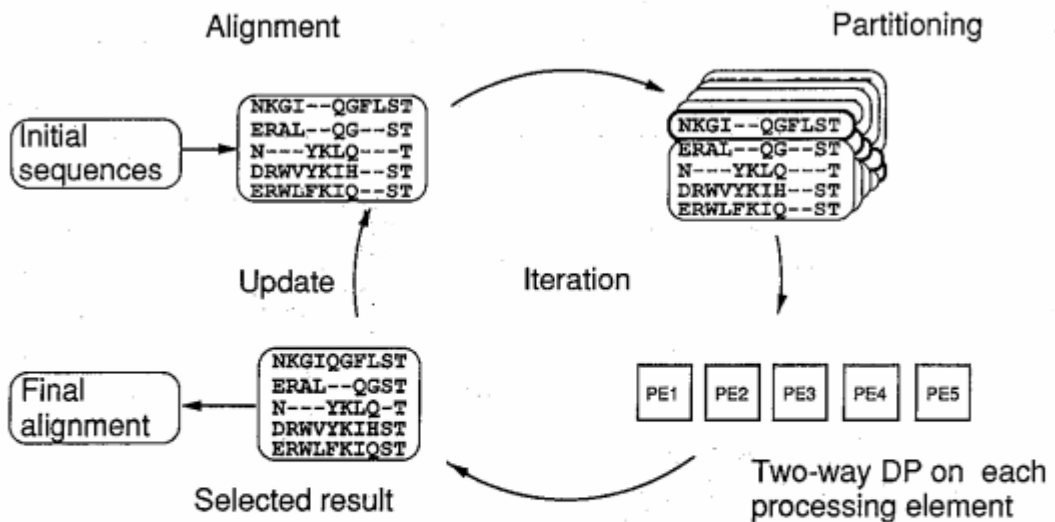
Two-way Dynamic Programming



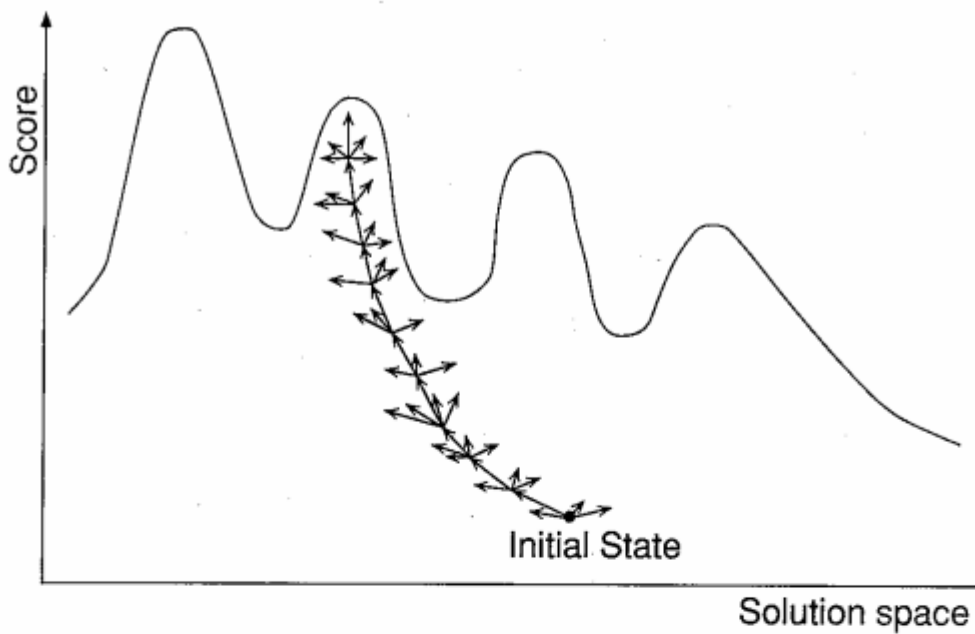
Tree-based Alignment Method



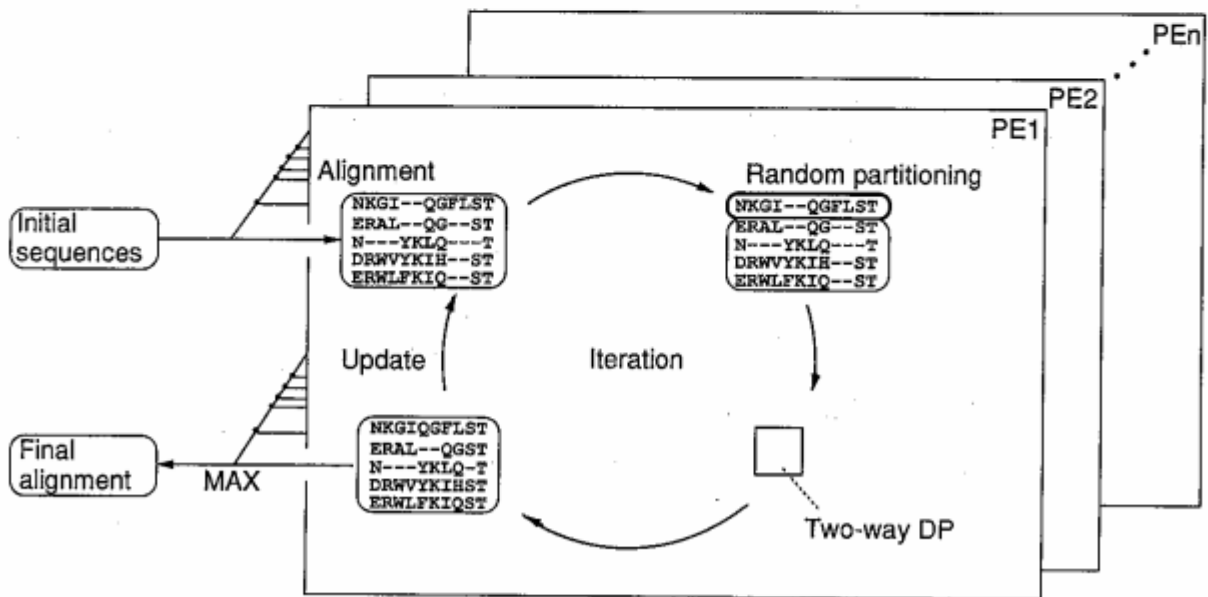
Parallel Best-first Iterative Aligner



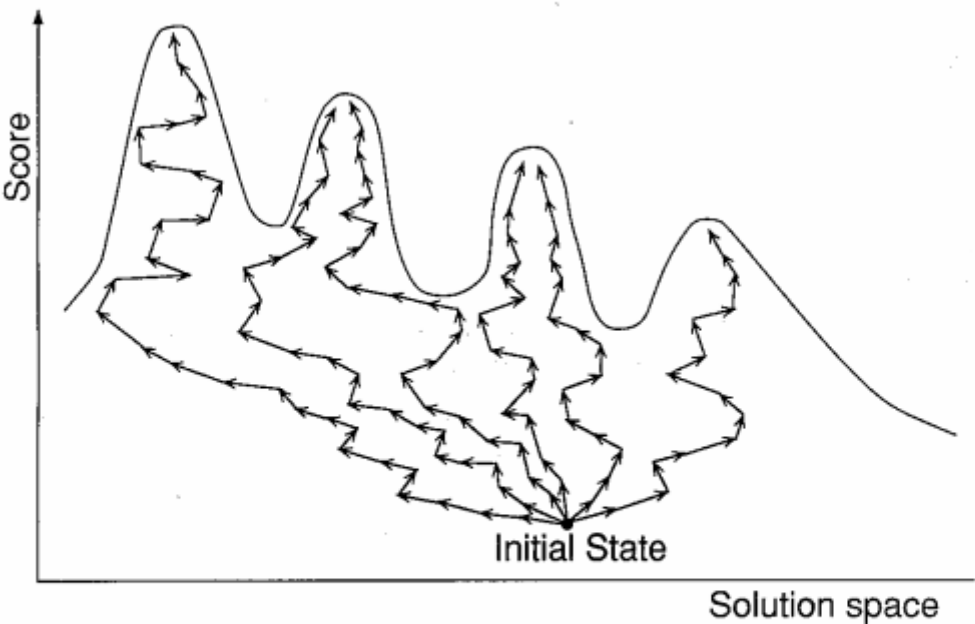
Best-first Search



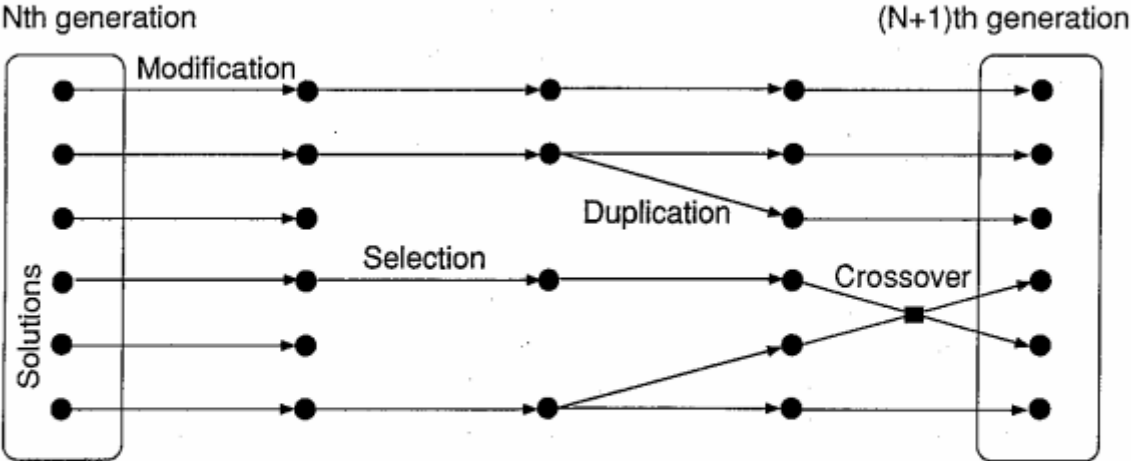
Parallel Hill-climbing Iterative Aligner



Parallel Hill-climbing Search

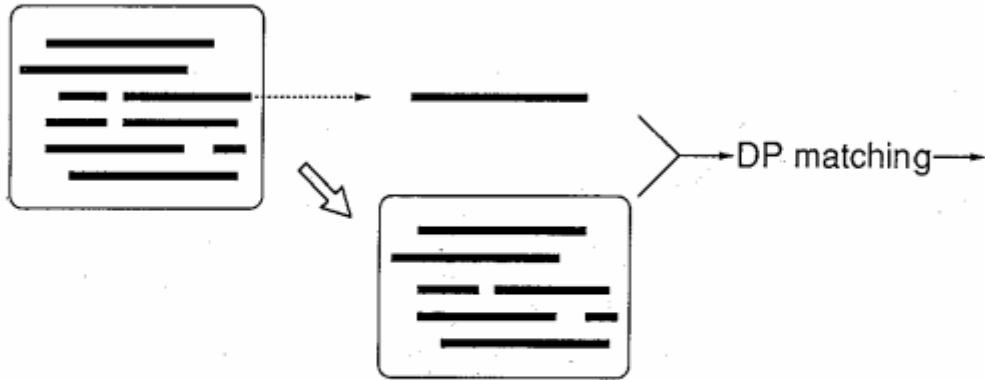


Genetic Algorithm (GA)



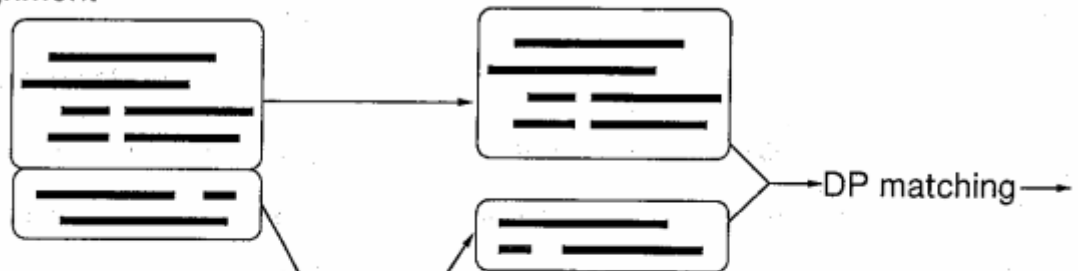
Modification of Alignment

Alignment

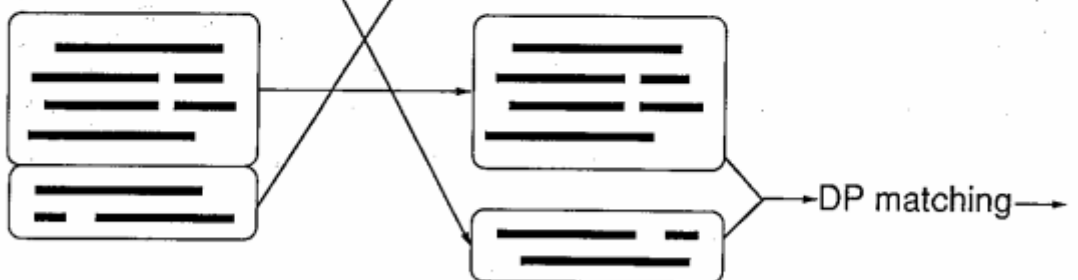


Crossover of Alignments

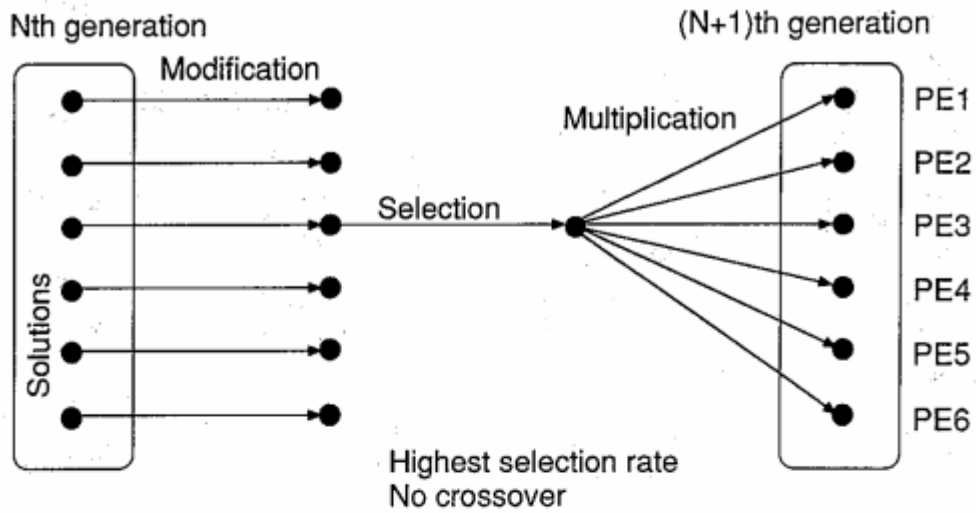
Alignment



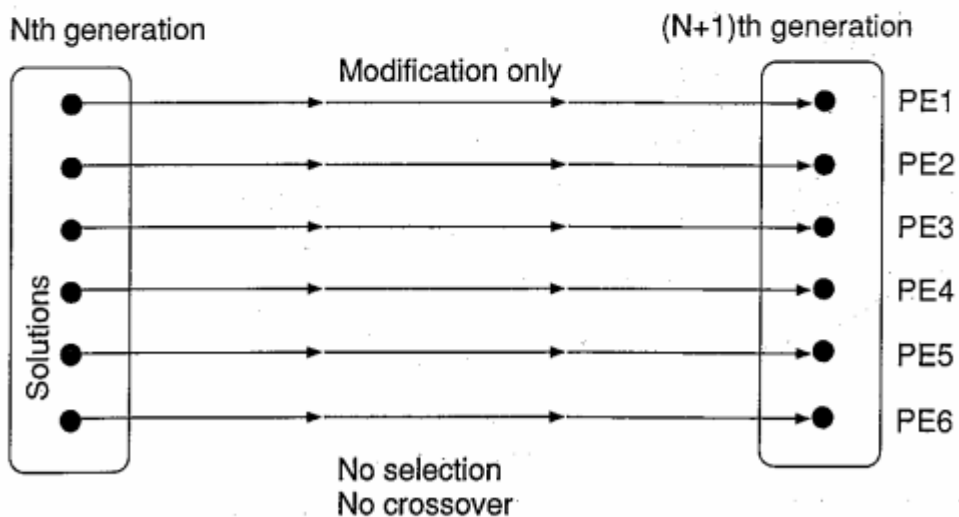
Alignment



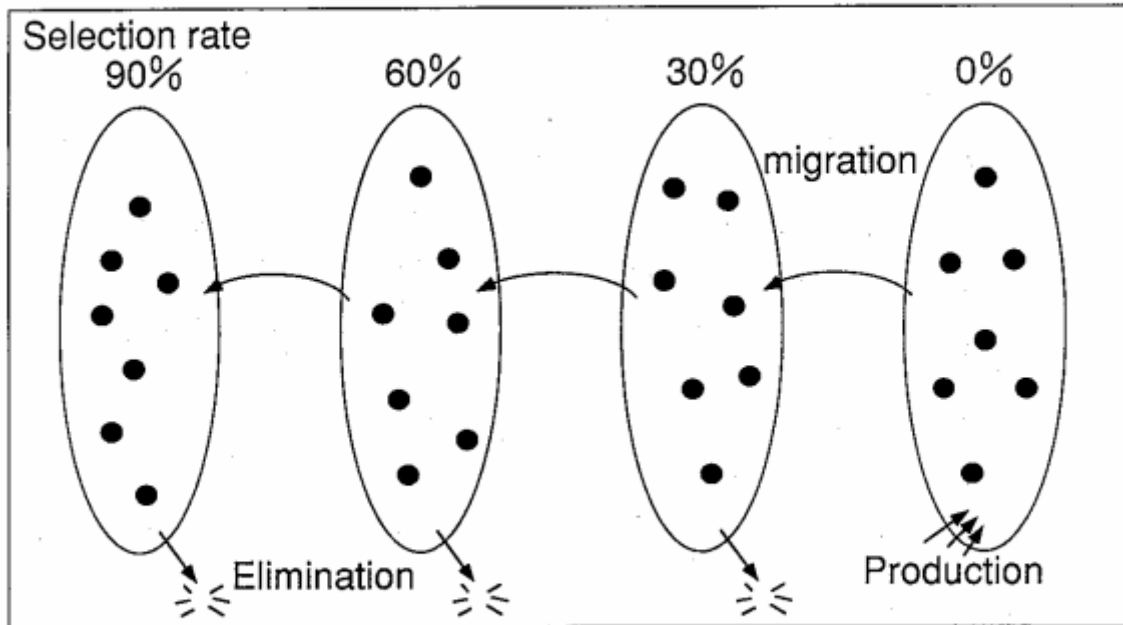
Best-first Search in GA



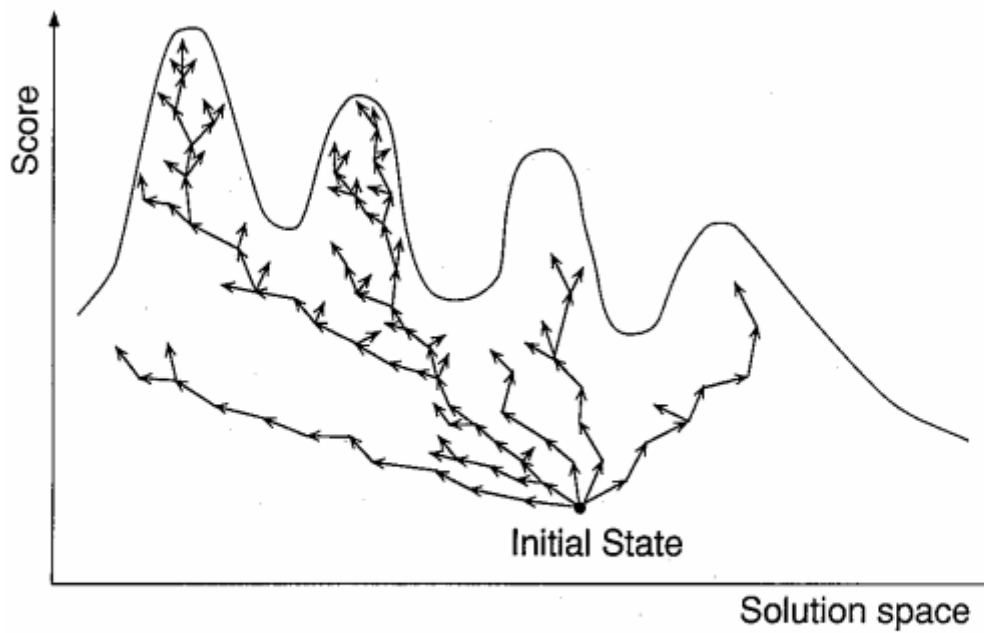
Parallel Hill-climbing in GA



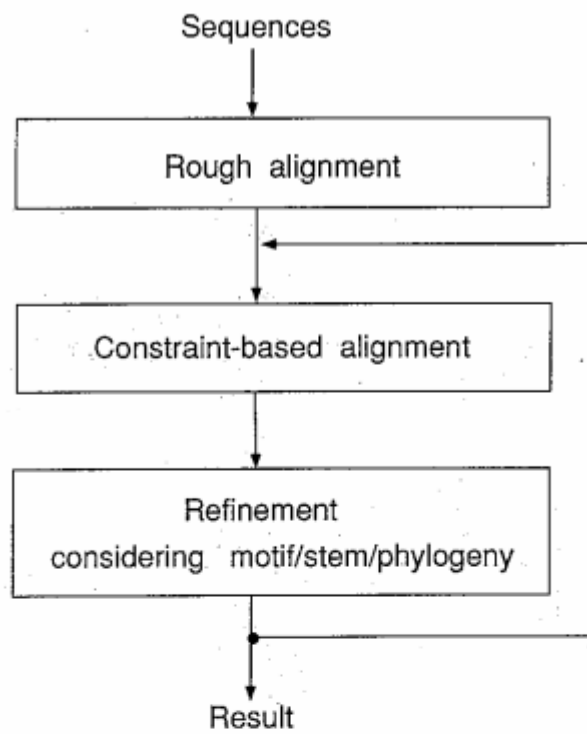
Multi-group GA Strategy



Multi-group GA Search



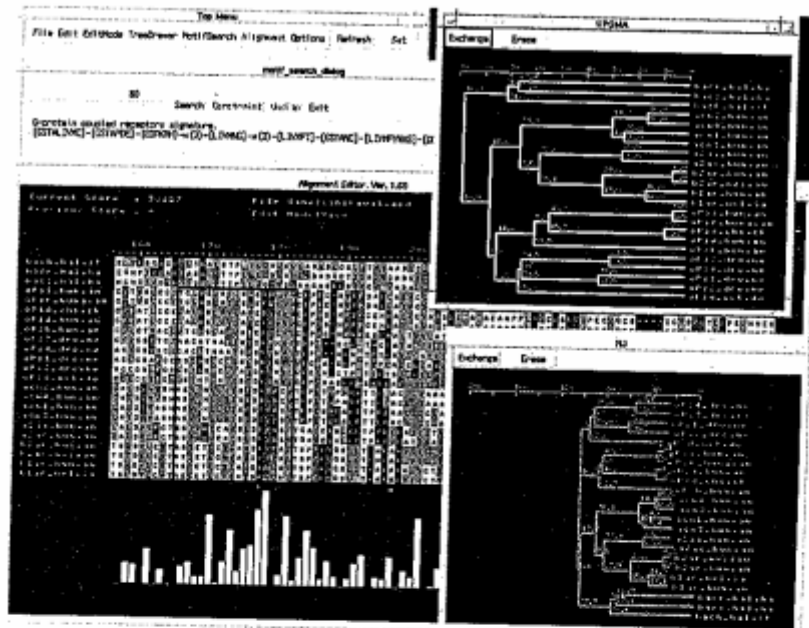
Alignment Process on Workbench



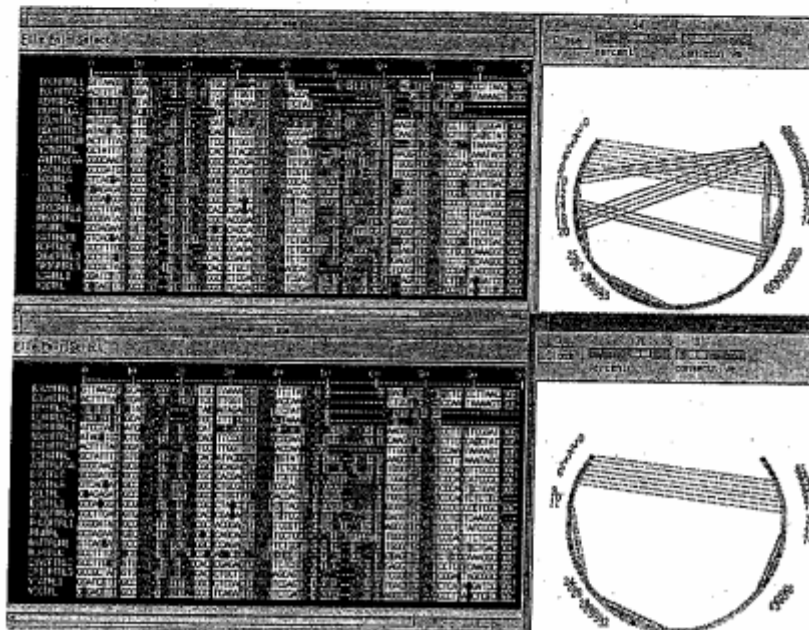
Constraint-based Alignment



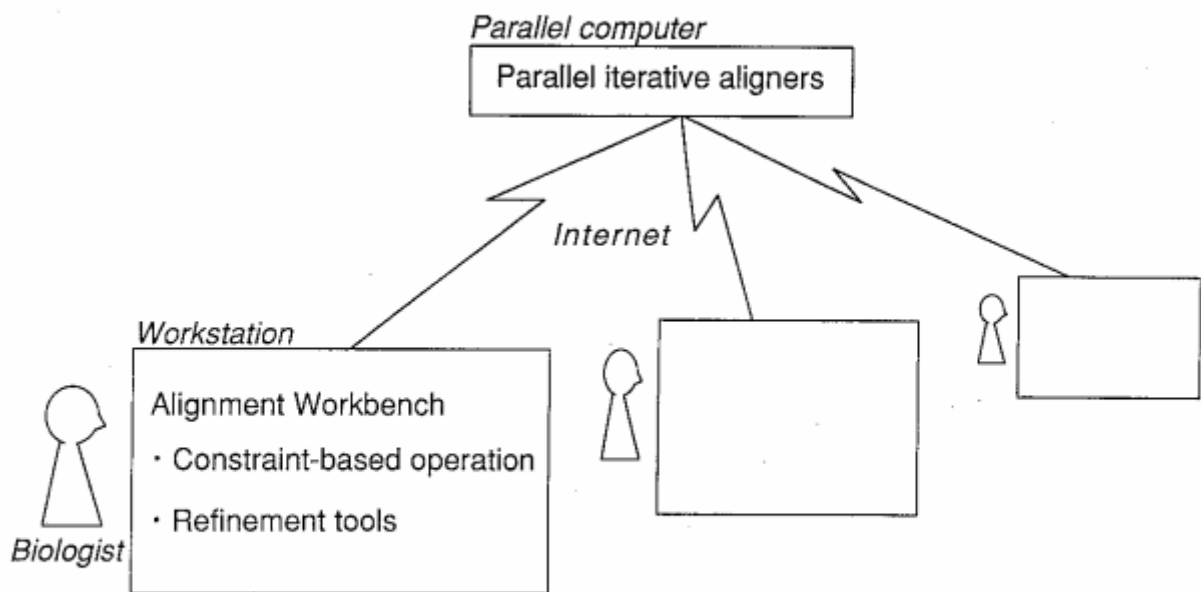
Sequence Motif Identification



RNA Stem Specification



Sequence Alignment Environment



FGCS Workshop 4

Fusion of Molecular Biology and Knowledge Processing

- Knowledge base construction in *Quixote*
- Stochastic learning on hidden markov model