Hannenhalli-Pevzner Theory

- Signed, unichromosomal genomes
- Operation: reversal (signed)
- Polynomial time algorithms for distance & operations

- First polynomial result for a realistic model of genome rearrangements

HP Theory

- Genomes modeled as “permutations”
- What they call “permutations” are **not** functions from E to E
- Rather, they are functions from P (positions) to E (extemities)
- P = {1, 2, 3, …, n}
- E = {1, 2, 3, …, n, -1, -2, -3, …, -n}
- $\pi : P \rightarrow E$

HP Theory

- Reversals are permutations on P
- $\rho : P \rightarrow P$
- Reversals are applied to the **right**:

  $\pi\rho$
- It is the only composition that makes sense
- Reversals do not always have small support

HP Theory

- Linear and circular cases are equivalent
- Extending the genome with 0 and n+1 essentially transforms the problem into a circular one
- In transforming π to σ, they fix σ:

  σ = 1 2 3 … n or σ(x) = x for all x

  σ is the “identity”
- The problem is then called “sorting” π

HP Theory

- Formula:

  $d(\pi) = b(\pi) - c(\pi) + h(\pi) + f(\pi)$

  $h(\pi)$ = number of hurdles of π

  $$f(\pi) = \begin{cases} 1 & \pi \text{ is a fortress} \\ 0 & \text{otherwise} \end{cases}$$

- Algorithms $O(n^4)$ and $O(n^5)$, later improved

Elementary HP Theory

- Oriented pairs
- Score
- Algorithm 1: perform oriented reversals with maximum score as long as possible

Elementary HP Theory

- After Algorithm 1, one ends up with a “positive permutation”
- Reduced “permutations”
- Framed intervals
- Hurdles: cutting and merging

Elementary HP Theory

- Algorithm 2:
- 2k hurdles:
  - merge two hurdles, nonconsecutive if possible
- 2k + 1 hurdles:
  - simple hurdle:
    - cut it
  - no simple hurdle:
    - merge two hurdles, nonconsecutive if possible

Elementary HP Theory

- Algorithm 2, simplified:
- 2k + 1 hurdles and simple hurdle:
  - cut it
- Else:
  - merge two hurdles, nonconsecutive if possible

Elementary HP Theory

- Final algorithm:
- **while** π is unsorted **do**
  - Algorithm 1
  - Reduce
  - Algorithm 2