On Reversal and Transposition Medians

Martin Bader | June 25, 2009

Genome Rearrangements

- During evolution, the gene order in a chromosome can change
- Gene order of two land snail mitochondrial DNAs

Cepaea nemoralis

$$\overrightarrow{cox1}\ \overrightarrow{V}\ \overrightarrow{rrnL}\ \overrightarrow{L1}\ \overrightarrow{A}\ \overrightarrow{nad6}\ \overrightarrow{P}\ \overrightarrow{nad5}\ \overrightarrow{nad1}\ \overrightarrow{nad4}\ \overrightarrow{L}\ \overrightarrow{cob}\ \overrightarrow{D}\ \overrightarrow{C}\ \overrightarrow{F}\ \overrightarrow{cox2}\ \overrightarrow{Y}\ \overrightarrow{W}\ \overrightarrow{G}\ \overrightarrow{H}$$

$$\overleftarrow{Q}\ \overleftarrow{L2}\ \overleftarrow{atp8}\ \overleftarrow{N}\ \overleftarrow{atp6}\ \overleftarrow{R}\ \overleftarrow{E}\ \overleftarrow{rrnS}\ \overleftarrow{M}\ \overleftarrow{nad3}\ \overleftarrow{S2}\ \overleftarrow{T}\ \overleftarrow{cox3}\ \overrightarrow{S1}\ \overrightarrow{nad4}\ \overrightarrow{I}\ \overrightarrow{nad2}\ \overrightarrow{K}$$

Albinaria coerulea

$$\overrightarrow{cox1}\ \overrightarrow{V}\ \overrightarrow{rrnL}\ \overrightarrow{L1}\ \overrightarrow{P}\ \overrightarrow{A}\ \overrightarrow{nad6}\ \overrightarrow{nad5}\ \overrightarrow{nad1}\ \overrightarrow{nad4}\ \overrightarrow{L}\ \overrightarrow{cob}\ \overrightarrow{D}\ \overrightarrow{C}\ \overrightarrow{F}\ \overrightarrow{cox2}\ \overrightarrow{Y}\ \overrightarrow{W}\ \overrightarrow{G}\ \overrightarrow{H}$$

$$\overleftarrow{Q}\ \overleftarrow{L2}\ \overleftarrow{atp8}\ \overleftarrow{N}\ \overleftarrow{atp6}\ \overleftarrow{R}\ \overleftarrow{E}\ \overleftarrow{rrnS}\ \overleftarrow{M}\ \overleftarrow{nad3}\ \overleftarrow{S2}\ \overrightarrow{S1}\ \overrightarrow{nad4}\ \overleftarrow{T}\ \overleftarrow{cox3}\ \overrightarrow{I}\ \overrightarrow{nad2}\ \overrightarrow{K}$$

Genome Rearrangements

- During evolution, the gene order in a chromosome can change
- Gene order of two land snail mitochondrial DNAs

Cepaea nemoralis

$$\overrightarrow{cox1}\ \overrightarrow{V}\ \overrightarrow{rrnL}\ \overrightarrow{L1}\ \overrightarrow{A}\ \overrightarrow{nad6}\ \overrightarrow{P}\ \overrightarrow{nad5}\ \overrightarrow{nad1}\ \overrightarrow{nad4}\ \overrightarrow{L}\ \overrightarrow{cob}\ \overrightarrow{D}\ \overrightarrow{C}\ \overrightarrow{F}\ \overrightarrow{cox2}\ \overrightarrow{Y}\ \overrightarrow{W}\ \overrightarrow{G}\ \overrightarrow{H}$$

$$\overleftarrow{Q}\ \overleftarrow{L2}\ \overleftarrow{atp8}\ \overleftarrow{N}\ \overleftarrow{atp6}\ \overleftarrow{R}\ \overleftarrow{E}\ \overleftarrow{rrnS}\ \overleftarrow{M}\ \overleftarrow{nad3}\ \overleftarrow{S2}\ \overleftarrow{T}\ \overleftarrow{cox3}\ \overrightarrow{S1}\ \overrightarrow{nad4}\ \overrightarrow{I}\ \overrightarrow{nad2}\ \overrightarrow{K}$$

Albinaria coerulea

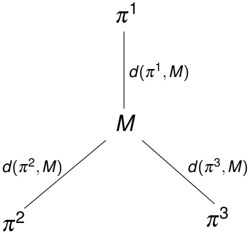
$$\overrightarrow{cox1}\ \overrightarrow{V}\ \overrightarrow{rrnL}\ \overrightarrow{L1}\ \overrightarrow{P}\ \overrightarrow{A}\ \overrightarrow{nad6}\ \overrightarrow{nad5}\ \overrightarrow{nad1}\ \overrightarrow{nad4}\ \overrightarrow{L}\ \overrightarrow{cob}\ \overrightarrow{D}\ \overrightarrow{C}\ \overrightarrow{F}\ \overrightarrow{cox2}\ \overrightarrow{Y}\ \overrightarrow{W}\ \overrightarrow{G}\ \overrightarrow{H}$$

$$\overleftarrow{Q}\ \overleftarrow{L2}\ \overleftarrow{atp8}\ \overleftarrow{N}\ \overleftarrow{atp6}\ \overleftarrow{R}\ \overleftarrow{E}\ \overleftarrow{rrnS}\ \overleftarrow{M}\ \overleftarrow{nad3}\ \overleftarrow{S2}\ \overrightarrow{S1}\ \overrightarrow{nad4}\ \overleftarrow{T}\ \overleftarrow{cox3}\ \overrightarrow{I}\ \overrightarrow{nad2}\ \overrightarrow{K}$$

- Reconstruct evolutionary events
- Use as distance measure
- Use for phylogenetic reconstruction

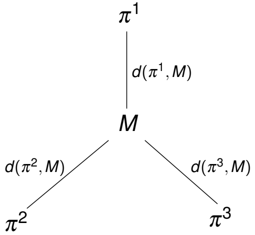
The Median Problem

- Given gene orders π^1, π^2, π^3
- Find M where $\sum_{i=1}^{3} d(\pi^i, M)$ is minimized



The Median Problem

- Given gene orders π^1, π^2, π^3
- Find M where $\sum_{i=1}^{3} d(\pi^i, M)$ is minimized



- NP-hard even for the most simple distance measures

Our contribution

- Exact algorithms for the Transposition Median Problem

  Exact algorithm for the weighted Reversal and Transposition Median Problem

  (Extension of Reversal Median solver, Caprara 2003)

- Improved exact algorithm for pairwise distances

  (Improvement of Christie 1998)

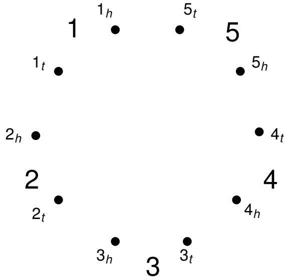
The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order

The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

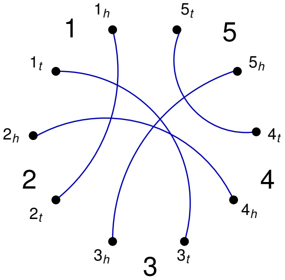
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

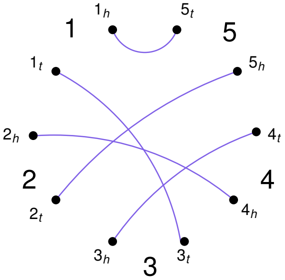
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

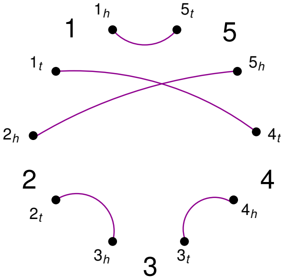
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

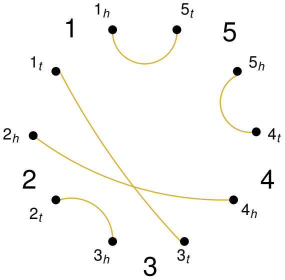
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

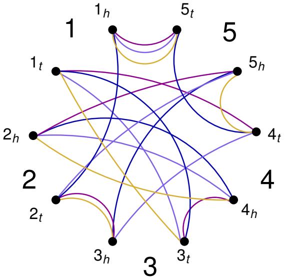
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



The Multiple Breakpoint Graph

- Edge-colored graph
- Contains neighborhood relations for each gene order
- Example:

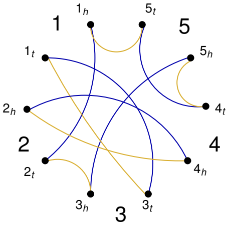
$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$
$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



Cycles and distances

- Edges of two colors form cycles

$$\pi^1 = (\overrightarrow{1}\ \overleftarrow{3}\ \overrightarrow{5}\ \overleftarrow{4}\ \overrightarrow{2})$$
$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$

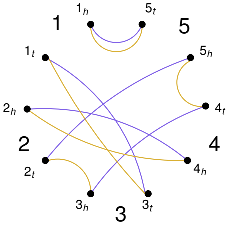


Cycles and distances

- Edges of two colors form cycles

$$\pi^2 = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{4}\ \overrightarrow{2}\ \overrightarrow{5})$$

$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$

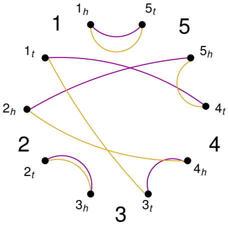


Cycles and distances

- Edges of two colors form cycles

$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$

$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$

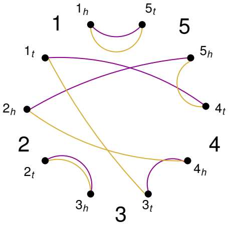


Cycles and distances

- Edges of two colors form cycles

$$\pi^3 = (\overrightarrow{1}\ \overleftarrow{4}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{5})$$

$$M = (\overrightarrow{1}\ \overleftarrow{3}\ \overleftarrow{2}\ \overrightarrow{4}\ \overrightarrow{5})$$



- Distances closely related to number of cycles

$$d_r = n - c + h + f$$

$$d_t \geq \frac{n - c_{odd}}{2}$$

$$d_w \geq \frac{w_t}{2}(n - c_{odd} - (2 - \frac{2w_r}{w_t})c_{even})$$

Sketch of the algorithm

- Solve Cycle Median Problem
- Verify solution

Sketch of the algorithm

- Solve Cycle Median Problem
  - Start with empty M
  - Subsequently add edges
  - Estimate lower bound for partial solution
  - Continue with partial solution with least lower bound (branch and bound)
  - NEW: Consider cycle lengths
- Verify solution

Sketch of the algorithm

- Solve Cycle Median Problem
  - Start with empty M
  - Subsequently add edges
  - Estimate lower bound for partial solution
  - Continue with partial solution with least lower bound (branch and bound)
  - NEW: Consider cycle lengths
- Verify solution
  Calculate edge weights
  - … either by an exact algorithm for pairwise distances
  - … or by an approximation algorithm (faster)

Experiments

- Create random input
  - Start with *id* of size n ($n = 37$ and $n = 100$)
  - Create 3 sequences of operations of length r ($2 \leq r \leq 15$)
  - Use these sequences to obtain π^1, π^2, and π^3

Experiments

- Create random input
  - Start with *id* of size n ($n = 37$ and $n = 100$)
  - Create 3 sequences of operations of length r ($2 \leq r \leq 15$)
  - Use these sequences to obtain π^1, π^2, and π^3
- Testing
  - Most inputs could be solved within a few seconds
  - Verifying solutions with approximation algorithm is very accurate
  - Much faster than previous algorithm for the Transposition Median Problem (Yue et al. 2008)

Conclusion

We presented an algorithm that ...

- can solve the TMP and wRTMP exactly
- is fast enough for practical use
- is FREE SOFTWARE (GPL v3.0)

$\Rightarrow$ download it from

`http://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.inst.190/Mitarbeiter/bader/phylo-1.0.1.tar.gz`

Conclusion

We presented an algorithm that ...

- can solve the TMP and wRTMP exactly
- is fast enough for practical use
- is FREE SOFTWARE (GPL v3.0)

$\Rightarrow$ download it from

`http://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.inst.190/Mitarbeiter/bader/phylo-1.0.1.tar.gz`

Thanks for your attention!