Cancer Phylogenetics Pipeline

(i) Alignment

(ii) Somatic variant calling

SNV cluster

(x, y)
copy number state (# maternal, # paternal copies)

SNV count

variant allele frequency

(iii) SNV clustering

(iv) Phylogeny inference

(v) Downstream analysis
Cancer Phylogenetics Pipeline

(i) Alignment
(ii) Somatic variant calling
(iii) SNV clustering
(iv) Phylogeny inference
(v) Downstream analysis

- single nucleotide variant (SNV)
- copy number state (# maternal, # paternal copies)

SNV cluster

variant allele frequency

SNV count
Outline

• Metastasis
• Maximum parsimony
• Problem statement
• Complexity
• Algorithm & results
• Problem variants

Reading:


Tumorigenesis: (i) Cell Mutation

Clonal Theory of Cancer
[Nowell, 1976]
Tumorigenesis: (i) Cell Mutation, (ii) Cell Division

Clonal Theory of Cancer
[Nowell, 1976]
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

Primary Tumor

Brain Metastasis

Liver Metastasis
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

Question:
How to reconstruct pattern of metastasis?
Key Challenge in Computational Biology

Translating a biological problem into computer science

- Formulating a combinatorial problem
- Analyzing complexity & combinatorial structure
- Designing an algorithm
- Interpreting solutions and validating the algorithm

Biological question

9
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

Cell Tree
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

Goal: Given phylogenetic tree $T$, find parsimonious vertex labeling $\ell$ with fewest migrations
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

**Goal:** Given phylogenetic tree $T$, find *parsimonious* vertex labeling $\ell$ with fewest migrations

**Question:** Find parsimonious vertex labeling
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

**Goal:** Given phylogenetic tree $T$, find *parsimonious* vertex labeling $\ell$ with fewest migrations
Outline

- Metastasis
- Maximum parsimony
- Problem statement
- Complexity
- Algorithm & results
- Problem variants

Reading:


Character-Based Phylogeny Reconstruction: Criterion

**Parsimony**: minimize number of changes on edges of tree

(a) Parsimony Score = 3

(b) Parsimony Score = 2
A Small and a Large Problem

**Small Maximum Parsimony Phylogeny Problem:**
Given $m \times n$ matrix $A = [a_{i,j}]$ and tree $T$ with $m$ leaves, find assignment of character states to each internal vertex of $T$ with minimum parsimony score.

**Large Maximum Parsimony Phylogeny Problem:**
Given $m \times n$ matrix $A = [a_{i,j}]$, find a tree $T$ with $m$ leaves labeled according to $A$ and an assignment of character states to each internal vertex of $T$ with minimum parsimony score.

**Question:** Are both problems easy (i.e. in P)?
Small Maximum Parsimony Phylogeny Problem

Key observations: (1) Characters can be solved independently. (2) Optimal substructure in subtrees.
Small Maximum Parsimony Phylogeny Problem:
Given rooted tree $T$ whose leaves are labeled by $\sigma : L(T) \rightarrow \Sigma$, find assignment of states to each internal vertex of $T$ with minimum parsimony score.
Recurrence for Small Maximum Parsimony Problem

**Small Maximum Parsimony Phylogeny Problem:**
Given rooted tree $T$ whose leaves are labeled by $\sigma : L(T) \rightarrow \Sigma$, find assignment of states to each internal vertex of $T$ with minimum parsimony score.

Let $\mu(v, s)$ be the minimum number of mutations in the subtree rooted at $v$ when assigning state $s$ to $v$.

Let $\delta(v)$ be the set of children of $v$. 

---
Recurrence for Small Maximum Parsimony Problem

Small Maximum Parsimony Phylogeny Problem:
Given rooted tree $T$ whose leaves are labeled by $\sigma : L(T) \rightarrow \Sigma$, find assignment of states to each internal vertex of $T$ with minimum parsimony score.

Let $\mu(v, s)$ be the minimum number of mutations in the subtree rooted at $v$ when assigning state $s$ to $v$.

$$c(s, t) = \begin{cases} 
0, & \text{if } s = t \\
1, & \text{if } s \neq t,
\end{cases}$$

$$\mu(v, s) = \min \begin{cases} 
\infty, & \text{if } v \in L(T) \text{ and } s \neq \sigma(v), \\
0, & \text{if } v \in L(T) \text{ and } s = \sigma(v), \\
\sum_{w \in \delta(v)} \min_{t \in \Sigma} \{c(s, t) + \mu(w, t)\}, & \text{if } v \not\in L(T).
\end{cases}$$

Let $\delta(v)$ be the set of children of $v$. 

Let $\delta(v)$ be the set of children of $v$. 

$$\mu(v, s) = \min \begin{cases} 
\infty, & \text{if } v \in L(T) \text{ and } s \neq \sigma(v), \\
0, & \text{if } v \in L(T) \text{ and } s = \sigma(v), \\
\sum_{w \in \delta(v)} \min_{t \in \Sigma} \{c(s, t) + \mu(w, t)\}, & \text{if } v \not\in L(T).
\end{cases}$$
Filling out DP Table and Traceback

**Fill out μ**

\[
\text{Fill}(T, v, \delta, \Sigma) \quad O(|\Sigma|^2)
\]

if \( v \in \mathcal{L}(T) \) then

for \( s \in \Sigma \)

if \( s = \delta(v) \) then

\[ \mu(v, s) = 0 \]

else

\[ \mu(v, s) = \infty \]

else

for \( w \in \delta(v) \) // children

\[ \text{Fill}(T, w, \delta, \Sigma) \]

\[ \mu(v, s) = 0 \]

for \( w \in \delta(v) \)

\[ \mu(v, s) = \min_{t \in \Sigma} \delta c(s, t) + \mu(w, t) \]

**Backtrace** \((T, v, \mu)\)

if \( v = r(T) \)

\[ \delta(r(T)) = \arg \min_{s \in \Sigma} \delta \mu(r(T), s) \]

else

let \( u \) be the parent of \( v \) and let \( s \) be the state

\[ \delta(v) = \arg \min_{t \in \Sigma} \delta c(s, t) + \mu(v, t) \]

for \( w \in \delta(v) \)

\[ \text{Backtrace}(T, w, \mu) \]

Let \( r(T) \) be the root vertex
Outline

• Metastasis
• Maximum parsimony
• Problem statement
• Complexity
• Algorithm & results
• Problem variants

Reading:


Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

Cell Tree

primary tumor $P$

metastasis $M_2$

metastasis $M_1$

Phylogenetic Tree

Instance of small maximum parsimony problem

Vertex labeling $\ell$

Instance of small maximum parsimony problem

Goal: Given phylogenetic tree $T$, find parsimonious vertex labeling $\ell$ with fewest migrations

Minimum Migration Analysis in Ovarian Cancer


- Instance of the maximum parsimony small phylogeny problem [Fitch, 1971; Sankoff, 1975]
Minimum Migration Analysis in Ovarian Cancer


- Instance of the maximum parsimony small phylogeny problem [Fitch, 1971; Sankoff, 1975]

$\mu^* = 13$

$m = 7$ anatomical sites

**Diagram Notes:**
- **ROv**: Right Ovary
- **LOv**: Left Ovary
- **LFTB**: Left Fallopian Tube
- **RFTA**: Right Fallopian Tube
- **ApC**: Appendix
- **Om**: Omentum
- **SBwl**: Small Bowel

**Migration Graph $G$:**
- Nodes represent anatomical sites.
- Edges indicate possible migration paths.
Minimum Migration History is \textbf{Not} Unique

- Enumerate all minimum-migration vertex labelings in the backtrace step

\(\mu^* = 13\)
Comigrations: Simultaneous Migrations of Multiple Clones

- Multiple tumor cells migrate simultaneously through the blood stream [Cheung et al., 2016]
- Second objective: number $\gamma$ of comigrations is the number of multi-edges in migration graph $G$

$\mu^* = 13$
$\gamma = 10$

Migration Graph $G$

Clone Tree $T$

$^\dagger$ Not necessarily true in the case of directed cycles
Comigrations: Simultaneous Migrations of Multiple Clones

- Multiple tumor cells migrate simultaneously through the blood stream [Cheung et al., 2016]
- Second objective: number $\gamma$ of comigrations is the number of multi-edges in migration graph $G^+$

$\mu^* = 13$
$\gamma = 7$

$\mu^* = 13$
$\gamma = 11$

$\mu^* = 13$
$\gamma = 10$

$\mu^* = 13$
$\gamma = 11$

$\mu^* = 13$
$\gamma = 7$

- Not necessarily true in the case of directed cycles

ApC  Appendix
LFTB  Left Fallopian Tube
LOv  Left Ovary
RFTA  Right Fallopian Tube
ROv  Right Ovary
SBwl  Small Bowel
Om  Omentum
Tradeoffs between Migrations, Comigrations and Migration Pattern

<table>
<thead>
<tr>
<th></th>
<th>single-source seeding (S)</th>
<th>multi-source seeding (M)</th>
<th>reseeding (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>monoclonal (m)</td>
<td>tree</td>
<td>directed acyclic graph</td>
<td>directed graph</td>
</tr>
<tr>
<td>polyclonal (p)</td>
<td>multi-tree</td>
<td>directed acyclic multi-graph</td>
<td>directed multi-graph</td>
</tr>
</tbody>
</table>

Migration patterns can be distinguished using two different criteria. First, by the number of clones that migrate between two anatomical sites: with monoclonal (m) seeding, every metastasis is seeded by a single clone, whereas with polyclonal (p) seeding, multiple clones migrate from one anatomical site to another. Second, by the migration topology: with single-source seeding (S), every metastasis is seeded from a single anatomical site, with multi-source seeding (M), a metastasis may be seeded by multiple anatomic sites, and with reseeding (R), clones migrate back and forth between anatomical sites, resulting in cycles of seeding.

The migration pattern affects the migration number \( \mu \) and the comigration number \( \min \). Suboptimal solutions are shown with the taxonomies of migration patterns between anatomical sites.

Figure 1: Tradeoffs between Migrations, Comigrations and Migration Pattern

\[
\begin{align*}
\mu & = \mu_{\text{min}} \\
\text{comigration number} \gamma & = \gamma_{\text{min}}
\end{align*}
\]

\( m-1 \) and \( m-1 \) are lower bounds on the migration and comigration number, respectively.
Constrained Multi-objective Optimization Problem

**Parsimonious Migration History (PMH):** Given a phylogenetic tree $T$ and a set $\mathcal{P} \subseteq \{S, M, R\}$ of allowed migration patterns, find vertex labeling $\ell$ with minimum migration number $\mu^*(T)$ and smallest comigration number $\hat{\gamma}(T)$.

---

Outline

- Metastasis
- Maximum parsimony
- Problem statement
- Complexity
- Algorithm & results
- Problem variants

Reading:


Results [El-Kебir, WABI 2018]

**Parsimonious Migration History (PMH):** Given a phylogenetic tree $T$ and a set $\mathcal{P} \subseteq \{S, M, R\}$ of allowed migration patterns, find vertex labeling $\ell$ with minimum migration number $\mu^*(T)$ and smallest comigration number $\hat{\gamma}(T)$.

**Theorem 1:** PMH is NP-hard when $\mathcal{P} = \{S\}$

**Theorem 2:** PMH is fixed parameter tractable in the number $m$ of locations when $\mathcal{P} = \{S\}$
PMH is NP-hard when $\mathcal{P} = \{\mathcal{S}\}$

**3-SAT:** Given $\varphi = \bigwedge_{i=1}^{k}(\varphi_{i,1} \lor \varphi_{i,2} \lor \varphi_{i,3})$ with variables $\{x_1, ..., x_n\}$ and $k$ clauses, find $\phi : [n] \to \{0,1\}$ satisfying $\varphi$

$$\Sigma = \{x_1, ..., x_n, \neg x_1, ..., \neg x_n, c_1, ..., c_k, \bot\}$$
PMH is NP-hard when $\mathcal{P} = \{S\}$

**3-SAT:** Given $\varphi = \bigwedge_{i=1}^{k}(y_{i,1} \lor y_{i,2} \lor y_{i,3})$ with variables $\{x_1, \ldots, x_n\}$ and $k$ clauses, find $\phi : [n] \rightarrow \{0,1\}$ satisfying $\varphi$

Three ideas:

1. Ensure that $(x, \neg x) \in E(G)$ or $(\neg x, x) \in E(G)$
2. Ensure that $\ell^*(r(T)) = \bot$
3. Ensure that $\varphi$ is satisfiable if and only if $\ell^*$ encodes a satisfying truth assignment
**Lemma:** Let $B > 10k + 1$ and $A > 2Bn + 27k$.
Then, $\varphi$ is satisfiable if and only if $\mu^*(T) = (B + 1)n + 25k$
PMH is NP-hard when $\mathcal{P} = \{S\}$

$\varphi = (x_1 \lor x_2 \lor \neg x_3) \land (\neg x_1, \neg x_2, \neg x_3)$

$k = 2, n = 3$

$B = 10k + 2 = 22$

$A = 2Bn + 27k + 1 = 187$

$\Sigma = \{x_1, x_2, x_3, \neg x_1, \neg x_2, \neg x_3, c_1, c_2, \bot\}$

Lemma: Let $B > 10k + 1$ and $A > 2Bn + 27k$.

Then, $\varphi$ is satisfiable if and only if $\mu^*(T) = (B + 1)n + 25k$

$\mu^*(T) = (B + 1)n + 25k = 23 \cdot 3 + 50 \cdot 2 = 119$
PMH is FPT in number $m$ of locations when $\mathcal{P} = \{S\}$

Lemma: If there exists labeling $\ell$ consistent with $\hat{G}$ then
\[
d_T(u,v) \geq d_{\hat{G}}(\text{lca}_{\hat{G}}(u), \hat{\ell}(v)) \quad \forall u, v \in V(T) \text{ such that } u \preceq_T v. \quad (1)
\]

\[
\ell^*(v) = \begin{cases} 
\text{LCA}_{\hat{G}}(r(T)), & \text{if } v = r(T), \\
\sigma(\ell^*(\pi(v)), \text{LCA}_{\hat{G}}(v)), & \text{if } v \neq r(T),
\end{cases}
\]

where $\sigma(s, t) = s$ if $s = t$ and otherwise $\sigma(s, t)$ is the unique child of $s$ that lies on the path from $s$ to $t$ in $\hat{G}$.

Lemma: If (1) holds then $\ell^*$ is a minimum migration labeling consistent with $\hat{G}$.
PMH is FPT in number \( m \) of locations when \( \mathcal{P} = \{S\} \)

Lemma: If there exists labeling \( \ell \) consistent with \( \hat{G} \) then
\[
d_T(u, v) \geq d_{\hat{G}}(\text{lca}_{\hat{G}}(u), \hat{\ell}(v)) \quad \forall u, v \in V(T) \text{ such that } u \preceq_T v. \quad (1)
\]

\[
\ell^*(v) = \begin{cases} 
\text{lca}_{\hat{G}}(r(T)), & \text{if } v = r(T), \\
\sigma(\ell^*(\pi(v)), \text{lca}_{\hat{G}}(v)), & \text{if } v \neq r(T),
\end{cases}
\]

where \( \sigma(s, t) = s \) if \( s = t \) and otherwise \( \sigma(s, t) \) is the unique child of \( s \) that lies on the path from \( s \) to \( t \) in \( \hat{G} \).

Lemma: If (1) holds then \( \ell^* \) is a minimum migration labeling consistent with \( \hat{G} \).
Outline

• Metastasis
• Maximum parsimony
• Problem statement
• Complexity
• Algorithm & results
• Problem variants

Reading:


Simulations

Available on: https://github.com/elkebir-group/PMH-S
Outline

• Metastasis
• Maximum parsimony
• Problem statement
• Complexity
• Algorithm & results
• Problem variants

Reading:


Resolving Clone Tree Ambiguities

Polytomy

Clone Tree $T$

PMH Problem
Parsimonious Migration History

Allowed patterns $\mathcal{P}$
$\{S, M, R\}$

PMH-TR Problem
Parsimonious Migration History with Tree Resolution

Polyclonal single source seeding (pS)

Monoclonal single source seeding (mS)

$x = 4$
$y = 2$

$x = 2$
$y = 2$
Resolving Clone Tree Ambiguities

### Parsimonious Migration History with Tree Inference

**PMH-TI Problem**

- **$F$**
  - $F^{-} = \begin{bmatrix} 0.82 & 0 & 0 \\ 0.88 & 0.24 & 0.64 \\ 0.73 & 0 & 0.73 \end{bmatrix}$
  - $F^{+} = \begin{bmatrix} 0.86 & 0 & 0 \\ 0.92 & 0.30 & 0.68 \\ 0.78 & 0 & 0.77 \end{bmatrix}$

- **$P$**
  - $P = \begin{bmatrix} 0.82 & 0 & 0 \\ 0.88 & 0.24 & 0.64 \\ 0.73 & 0 & 0.73 \end{bmatrix}$

- **$\hat{F}$**
  - **Resolved Polytomy**

- **$\mu = 2$**
  - **$\gamma = 2$**

**Monoclonal Single Source Seeding (mS)**
Clone Tree

Migration Graph

Label ancestral vertices by anatomical sites

Resolve clone tree ambiguities

Cell Division and Mutation History

Cell Migration History

mutation

migration

Standard Phylogenetic Techniques

Sample Tree

Mutation Matrix

$P$

$M_1$

$M_2$

inferred extant clones

unobserved clones

Sequencing and Mutation Calling

Tumor Phylogenetic Techniques

Mutation Matrix

$P$

$M_1$

$M_2$

*homoplasy

inferred extant clones

Sequencing and Mutation Calling

Tumor Phylogenetic Techniques

Mutation Matrix

$P$

$M_1$

$M_2$

*homoplasy

inferred extant clones

Sequencing and Mutation Calling

Tumor Phylogenetic Techniques

Mutation Matrix

$P$

$M_1$

$M_2$

*homoplasy

inferred extant clones

Sequencing and Mutation Calling

Tumor Phylogenetic Techniques

Mutation Matrix

$P$

$M_1$

$M_2$

*homoplasy

inferred extant clones

Sequencing and Mutation Calling
MACHINA accurately infers clone trees and migration histories on simulated data.
Applying MACHINA to Metastatic Breast Cancer


- Triple negative, basal-like breast cancer presenting with Stage IIA disease
- Treated with neoadjuvant AC-T achieving stable disease, followed by mastectomy and radiation
- After 17 months, patient presented with Stage IV disease with 7 distant metastases
- Died of disease in 25 months
- Six tumors for WGS: primary, rib, kidney, brain, liver, and lung

Reported

\( (\mu, \gamma) = (12, 6) \)

\[ (\mu, \gamma) = (5, 5) \]

Resolved monoclonal single-source seeding (mS)
Polyclonal single-source seeding (pS)
Monoclonal multi-source seeding (mM)
Polyclonal multi-source seeding (pM)

\[ (\mu, \gamma) = (5, 5) \]

_ resolved polytomy
_ resolved polytomy

\[ (\mu, \gamma) = (12, 6) \]

\( \gamma = 1 \)

\( \mu = 5 \)

\( \gamma = 5 \)

\( \mu = 5 \)

\[ (\mu, \gamma) = (5, 5) \]

\( (\mu, \gamma) = (12, 6) \)

\( \gamma = 1 \)

\( \mu = 5 \)

\( \gamma = 5 \)

\( \mu = 5 \)

\[ (\mu, \gamma) = (5, 5) \]

\( (\mu, \gamma) = (12, 6) \)

\( \gamma = 1 \)

\( \mu = 5 \)

\( \gamma = 5 \)

\( \mu = 5 \)