Topics of today:

1. Formalizing the number of occurrences (ϕ) of families/adjacencies/telomeres

2. Revisiting breakpoint and SCJ double distances

3. SCJ median, halving and guided halving

4. Breakpoint median, halving and guided halving

Occurrences of families

Given a family f and a genome $\mathbb{G}$, let $\phi(f, \mathbb{G})$ be the number of occurrences of family f in $\mathbb{G}$.

If $\mathbb{S}$ is singular, then $\phi(f, \mathbb{S}) = 1$ for each $f \in \mathcal{F}(\mathbb{S})$.

If $\mathbb{D}$ is duplicated, then $\phi(f, \mathbb{D}) = 2$ for each $f \in \mathcal{F}(\mathbb{D})$.

If $\mathbb{C}_1$ and $\mathbb{C}_2$ are canonical, then $\mathcal{F}_\star = \mathcal{F}(\mathbb{C}_1) = \mathcal{F}(\mathbb{C}_2)$ and $\phi(f, \mathbb{C}_1) = \phi(f, \mathbb{C}_2) = 1$ for each $f \in \mathcal{F}_\star$.

If $\mathbb{B}_1$ and $\mathbb{B}_2$ are balanced, then $\mathcal{F}_\star = \mathcal{F}(\mathbb{B}_1) = \mathcal{F}(\mathbb{B}_2)$ and $\phi(f, \mathbb{B}_1) = \phi(f, \mathbb{B}_2)$ for each $f \in \mathcal{F}_\star$.

Occurrences of adjacencies

Given an adjacency xy and a canonical genome $\mathbb{C}$, let $\phi(xy, \mathbb{C}) = \begin{cases} 1, & xy \in \alpha(\mathbb{C}), \\ 0, & xy \notin \alpha(\mathbb{C}). \end{cases}$

Given an adjacency xy and a duplicated genome $\mathbb{D}$,

let $\phi(xy, \mathbb{D})$ be the number of occurrences of adjacencies of type $x_i y_j$ in $\alpha(\mathbb{D})$.

Note that $\phi(xy, \mathbb{D}) \in \{0, 1, 2\}$.

Given an adjacency xy and k genomes $\mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_k$,

let $\phi(xy, \mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_k) = \phi(xy, \mathbb{G}_{1..k}) = \sum_{i=1}^{k} \phi(xy, \mathbb{G}_i)$.

Occurrences of telomeres

Given a telomere x and a canonical genome $\mathbb{C}$, let $\phi(x, \mathbb{C}) = \begin{cases} 1, & x \in \gamma(\mathbb{C}), \\ 0, & x \notin \gamma(\mathbb{C}). \end{cases}$

Given a telomere x and a duplicated genome $\mathbb{D}$,

let $\phi(x, \mathbb{D})$ be the number of occurrences of telomeres of type $x_{[i]}$ in $\gamma(\mathbb{D})$.

Note that $\phi(x, \mathbb{D}) \in \{0, 1, 2\}$.

Given a telomere x and k genomes $\mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_k$,

let $\phi(x, \mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_k) = \phi(x, \mathbb{G}_{1..k}) = \sum_{i=1}^{k} \phi(x, \mathbb{G}_i)$.

Breakpoint model - distance and double distance

Breakpoint distance of canonical genomes $\mathbb{C}_1$ and $\mathbb{C}_2$, with $n = |\mathcal{G}_\star|$, $a = |\alpha_\star|$ and $t = |\gamma_\star|$:

$$d_{\mathrm{BP}}(\mathbb{C}_1, \mathbb{C}_2) = n - a - \frac{t}{2}.$$

Breakpoint double distance of sing-dup-canonical genomes $\mathbb{S}$ and $\mathbb{D}$, with $\mathcal{G}_\star = \mathcal{G}(\mathbb{S}) \cap \mathcal{G}(\mathbb{D})$ and $n = |\mathcal{G}_\star|$:

$$d_{\mathrm{BP}}^2(\mathbb{S}, \mathbb{D}) = d_{\mathrm{BP}}(2 \cdot \mathbb{S}, \mathbb{D}) = n' - a' - \frac{t'}{2} = 2n - a' - \frac{t'}{2},$$

where $n' = |\mathcal{G}(2 \cdot \mathbb{S}) \cap \mathcal{G}(\mathbb{D})| = 2|\mathcal{G}_\star| = 2n$, $a' = |\alpha(2 \cdot \mathbb{S}) \cap \alpha(\mathbb{D})|$ and $t' = |\gamma(2 \cdot \mathbb{S}) \cap \gamma(\mathbb{D})|$.

Since it is possible to find a matching that fulfills each candidate adjacency/telomere between $2 \cdot \mathbb{S}$ and $\mathbb{D}$:

$$a' = \sum_{xy \in \alpha(\mathbb{S})} \phi(xy, \mathbb{D}) \quad \text{and}$$

$$t' = \sum_{x \in \gamma(\mathbb{S})} \phi(x, \mathbb{D})$$

SCJ model - distance and double distance

SCJ distance of canonical genomes $\mathbb{C}_1$ and $\mathbb{C}_2$, with $n = |\mathcal{G}_\star|$ and $a = |\alpha_\star|$:

$$d_{\mathrm{SCJ}}(\mathbb{C}_1, \mathbb{C}_2) = 2n - 2a - \kappa(\mathbb{C}_1) - \kappa(\mathbb{C}_2).$$

SCJ double distance of sing-dup-canonical genomes $\mathbb{S}$ and $\mathbb{D}$, with $\mathcal{G}_\star = \mathcal{G}(\mathbb{S}) \cap \mathcal{G}(\mathbb{D})$ and $n = |\mathcal{G}_\star|$:

$$d_{\mathrm{SCJ}}^2(\mathbb{S}, \mathbb{D}) = d_{\mathrm{SCJ}}(2 \cdot \mathbb{S}, \mathbb{D}) = 2n' - 2a' - \kappa(2 \cdot \mathbb{S}) - \kappa(\mathbb{D}) = 4n - 2a' - 2\kappa(\mathbb{S}) - \kappa(\mathbb{D})$$

where $n' = |\mathcal{G}(2 \cdot \mathbb{S}) \cap \mathcal{G}(\mathbb{D})| = 2|\mathcal{G}_\star| = 2n$ and $a' = |\alpha(2 \cdot \mathbb{S}) \cap \alpha(\mathbb{D})|$.

Since it is possible to find a matching that fulfills each candidate adjacency between $2 \cdot \mathbb{S}$ and $\mathbb{D}$:

$$a' = \sum_{xy \in \alpha(\mathbb{S})} \phi(xy, \mathbb{D})$$

SCJ median of canonical genomes

Given three canonical genomes $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$, find another canonical genome $\mathbb{M}$ that minimizes the sum:

$$s_{\text{SCJ}}(\mathbb{M}) = d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_1) + d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_2) + d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_3)$$

Recall that:

$$
\begin{aligned}
d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_i) &= & |\alpha(\mathbb{M}) \setminus \alpha(\mathbb{C}_i)| & + & |\alpha(\mathbb{C}_i) \setminus \alpha(\mathbb{M})| \\
&= & \sum_{xy \in \alpha(\mathbb{M})} (1 - \phi(xy, \mathbb{C}_i)) & + & \sum_{xy \notin \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_i)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
s_{\text{SCJ}}(\mathbb{M}) &= & \sum_{xy \in \alpha(\mathbb{M})} [\, 1 - \phi(xy, \mathbb{C}_1)) \;+\; (1 - \phi(xy, \mathbb{C}_2)) \;+\; (1 - \phi(xy, \mathbb{C}_3)) \,] \\
& & +\; \sum_{xy \notin \alpha(\mathbb{M})} [\, \phi(xy, \mathbb{C}_1) \;+\; \phi(xy, \mathbb{C}_2) \;+\; \phi(xy, \mathbb{C}_3) \,] \\
&= & \sum_{xy \in \alpha(\mathbb{M})} (3 - \phi(xy, \mathbb{C}_{1\ldots 3})) \;+\; \sum_{xy \notin \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_{1\ldots 3}) \\
&= & \sum_{xy} [\, \phi(xy, \mathbb{M}) \cdot (3 - \phi(xy, \mathbb{C}_{1\ldots 3})) \;+\; (1 - \phi(xy, \mathbb{M})) \cdot \phi(xy, \mathbb{C}_{1\ldots 3}) \,] \\
&= & \sum_{xy} [\, 3 \cdot \phi(xy, \mathbb{M}) - \phi(xy, \mathbb{M}) \cdot \phi(xy, \mathbb{C}_{1\ldots 3}) \;+\; \phi(xy, \mathbb{C}_{1\ldots 3}) - \phi(xy, \mathbb{M}) \cdot \phi(xy, \mathbb{C}_{1\ldots 3}) \,] \\
&= & |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)| \;+\; \sum_{xy} [\, \phi(xy, \mathbb{M})(3 - 2 \cdot \phi(xy, \mathbb{C}_{1\ldots 3})) \,] \\
&= & |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)| \;+\; \sum_{xy \in \alpha(\mathbb{M})} (3 - 2 \cdot \phi(xy, \mathbb{C}_{1\ldots 3}))
\end{aligned}
$$

SCJ median of canonical genomes

$$s_{SCJ}(\mathbb{M}) \quad = \quad |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)| \quad + \quad \sum_{xy \in \alpha(\mathbb{M})}(3 - 2 \cdot \phi(xy, \mathbb{C}_{1..3}))$$

$$= \quad |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)| \quad + \quad \omega(\mathbb{M})$$

Since $|\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)|$ is given (does not depend on $\mathbb{M}$), for minimizing $s_{SCJ}(\mathbb{M})$ we need to minimize:

$$\omega(\mathbb{M}) \; = \; \sum_{xy \in \alpha(\mathbb{M})} \omega(xy) \; = \; \sum_{xy \in \alpha(\mathbb{M})} (3 - 2 \cdot \phi(xy, \mathbb{C}_{1..3}))$$

where $\omega(xy) = 3 - 2 \cdot \phi(xy, \mathbb{C}_{1..3}) \in \{-3, -1, +1, +3\}$.

For minimizing $\omega(\mathbb{M})$:

▶ Do not add to $\mathbb{M}$ any adjacency xz that have $\omega(xz) > 0$:
  this happens when $\phi(xz, \mathbb{C}_{1..3}) \leq 1$ (xz occurs in at most one genome among $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$).

▶ Add to $\mathbb{M}$ any adjacency xy that have $\omega(xy) < 0$:
  this happens when $\phi(xy, \mathbb{C}_{1..3}) \geq 2$ (xy occurs in at least two genomes among $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$).

▶ For $z \neq y$: $\omega(xz) > 0 \Leftrightarrow \omega(xy) < 0$.

There is no adjacency xy with $\omega(xy) = 0$. Therefore, the SCJ median problem has a unique solution:

$$\alpha(\mathbb{M}) = \{xy : \phi(xy, \mathbb{C}_{1..3}) \geq 2\}$$

SCJ median of canonical genomes - intuition

Let $\mathcal{F}_\star = \mathcal{G}_\star = \{1, 2, 3, ..., n\}$

and start with $\mathbb{M} = [1]\ [2]\ ...\ [n]$

$\alpha(\mathbb{M}) = \emptyset$ and $s_{\text{SCJ}}(\mathbb{M}) = |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + |\alpha(\mathbb{C}_3)|$

Effect of adding an adjacency xy to $\mathbb{M}$:

1. If xy is not present in any genome among $\{\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3\}$, then $\Delta s_{\text{SCJ}} = +3$.

2. If xy is present in exactly one genome among $\{\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3\}$, then $\Delta s_{\text{SCJ}} = +1$.
   ($\Delta d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_i) = -1$, but $2 \times \Delta d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_i) = +1$)

3. If xy is present in exactly two genomes among $\{\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3\}$, then $\Delta s_{\text{SCJ}} = -1$.
   ($2 \times \Delta d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_i) = -1$, but $\Delta d_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_i) = +1$)

4. If xy is present in all three genomes $\{\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3\}$, then $\Delta s_{\text{SCJ}} = -3$.

SCJ median of k canonical genomes

Given k canonical genomes $\mathbb{C}_1, \mathbb{C}_2, \ldots \mathbb{C}_k$, find another canonical genome $\mathbb{M}$ that minimizes the sum:

$$
\begin{aligned}
\mathbf{s}_{\text{SCJ}}(\mathbb{M}) \quad &= \quad \mathbf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_1) + \mathbf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_2) + \ldots + \mathbf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_k) \\
&= \quad |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + \ldots + |\alpha(\mathbb{C}_k)| \quad + \quad \omega(\mathbb{M})
\end{aligned}
$$

Analogously to the median of three genomes, we need to minimize:

$$
\omega(\mathbb{M}) \;=\; \sum_{xy \in \alpha(\mathbb{M})} \omega(xy)
$$

where $\omega(xy) = k - 2 \cdot \phi(xy, \mathbb{C}_{1..k}) \in \{-k, -k+2, \ldots, +k-2, +k\}$.

For minimizing $\omega(\mathbb{M})$:

- Do not add to $\mathbb{M}$ any adjacency xz that have $\omega(xz) > 0$:
  this happens when $\phi(xz, \mathbb{C}_{1..k}) < \frac{k}{2}$ (xz occurs in less than half of the genomes among $\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_k$).
- Add to $\mathbb{M}$ any adjacency xy that have $\omega(xy) < 0$:
  this happens when $\phi(xy, \mathbb{C}_{1..3}) > \frac{k}{2}$ (xy occurs in more than half of the genomes among $\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_k$).
- For $z \neq y$: $\omega(xz) > 0 \Leftrightarrow \omega(xy) < 0$.
- Any adjacency xy with $\omega(xy) = 0$ is optional (can be added to the median or not). If there is no such an adjacency (e.g., if k is odd), the SCJ median problem has a unique solution.

In general, the following set of adjacencies define a SCJ median of k genomes:

$$
\alpha(\mathbb{M}) = \left\{ xy : \phi(xy, \mathbb{C}_{1..k}) > \frac{k}{2} \right\}
$$

SCJ median of k canonical linear genomes

1. Compute the general SCJ median $\mathbb{M}$ as described above.

2. For each circular chromosome in $\mathbb{M}$, remove one adjacency xy with smallest weight $\omega(xy)$.

SCJ halving of a duplicated genome

Given a duplicated genome $\mathbb{D}$, find a singular genome $\mathbb{H}$ that minimizes the SCJ double distance:

$$d_{\text{SCJ}}^2(\mathbb{H}, \mathbb{D}) = d_{\text{SCJ}}(2 \cdot \mathbb{H}, \mathbb{D})$$

Therefore:

$$
\begin{aligned}
d_{\text{SCJ}}(2 \cdot \mathbb{H}, \mathbb{D}) \quad &= \quad |\alpha(2 \cdot \mathbb{H}) \setminus \alpha(\mathbb{D})| \quad + \quad |\alpha(\mathbb{D}) \setminus \alpha(2 \cdot \mathbb{H})| \\[6pt]
&= \quad \sum_{xy \in \alpha(\mathbb{H})}(2 - \phi(xy, \mathbb{D})) \quad + \quad \sum_{xy \notin \alpha(\mathbb{H})} \phi(xy, \mathbb{D}) \\[6pt]
&= \quad \sum_{xy} [\ \phi(xy, \mathbb{H}) \cdot (2 - \phi(xy, \mathbb{D})) \quad + \quad (1 - \phi(xy, \mathbb{H})) \cdot \phi(xy, \mathbb{D})\ ] \\[6pt]
&= \quad \sum_{xy} [\ 2 \cdot \phi(xy, \mathbb{H}) - \phi(xy, \mathbb{H}) \cdot \phi(xy, \mathbb{D}) \quad + \quad \phi(xy, \mathbb{D}) - \phi(xy, \mathbb{H}) \cdot \phi(xy, \mathbb{D})\ ] \\[6pt]
&= \quad |\alpha(\mathbb{D})| \quad + \quad \sum_{xy} [\ \phi(xy, \mathbb{H})(2 - 2 \cdot \phi(xy, \mathbb{D}))\ ] \\[6pt]
&= \quad |\alpha(\mathbb{D})| \quad + \quad \sum_{xy \in \alpha(\mathbb{H})}(2 - 2 \cdot \phi(xy, \mathbb{D}))
\end{aligned}
$$

SCJ halving of a duplicated genome

$$d_{\text{SCJ}}^2(\mathbb{H}, \mathbb{D}) = |\alpha(\mathbb{D})| + \sum_{xy \in \alpha(\mathbb{H})}(2 - 2 \cdot \phi(xy, \mathbb{D}))$$

$$= |\alpha(\mathbb{D})| + \omega(\mathbb{H})$$

Since $|\alpha(\mathbb{D})|$ is given (does not depend on $\mathbb{H}$), for minimizing $d_{\text{SCJ}}^2(\mathbb{H}, \mathbb{D})$ we need to minimize:

$$\omega(\mathbb{H}) = \sum_{xy \in \alpha(\mathbb{H})} \omega(xy) = \sum_{xy \in \alpha(\mathbb{H})} (2 - 2 \cdot \phi(xy, \mathbb{D}))$$

where $\omega(xy) = 2 - 2 \cdot \phi(xy, \mathbb{D}) \in \{-2, 0, +2\}$.

For minimizing $\omega(\mathbb{H})$:

▶ Do not add to $\mathbb{H}$ any adjacency xz that have $\omega(xz) > 0$:
   this happens when $\phi(xz, \mathbb{D}) = 0$ (xz does not occur in $\mathbb{D}$).

▶ Add to $\mathbb{H}$ any adjacency xy that have $\omega(xy) < 0$:
   this happens when $\phi(xy, \mathbb{D}) = 2$ (xy occurs twice in $\mathbb{D}$).

▶ For $z \neq y$: $\omega(xz) > 0 \Leftrightarrow \omega(xy) < 0$.

▶ Any adjacency xy with $\omega(xy) = 0$ (occurs once in $\mathbb{D}$) is optional (can be added to $\mathbb{H}$ or not).

Solution with the minimum number of adjacencies: $\alpha(\mathbb{H}) = \{xy : \phi(xy, \mathbb{D}) = 2\}$

SCJ aliquoting of a k-folded genome

Given a k-folded genome $\mathbb{K}$, find a singular genome $\mathbb{A}$ that minimizes the SCJ k-folded distance:

$$d_{\text{SCJ}}^{k}(\mathbb{A}, \mathbb{K}) = d_{\text{SCJ}}(k \cdot \mathbb{A}, \mathbb{K})$$

Therefore:

$$
\begin{aligned}
d_{\text{SCJ}}(k \cdot \mathbb{A}, \mathbb{K}) \quad &= \quad |\alpha(k \cdot \mathbb{A}) \setminus \alpha(\mathbb{K})| \quad + \quad |\alpha(\mathbb{K}) \setminus \alpha(k \cdot \mathbb{A})| \\
&= \quad \sum_{xy \in \alpha(\mathbb{A})}(k - \phi(xy, \mathbb{K})) \quad + \quad \sum_{xy \notin \alpha(\mathbb{A})} \phi(xy, \mathbb{K}) \\
&= \quad \sum_{xy}[\ \phi(xy, \mathbb{A}) \cdot (k - \phi(xy, \mathbb{K})) \quad + \quad (1 - \phi(xy, \mathbb{A})) \cdot \phi(xy, \mathbb{K})\ ] \\
&= \quad \sum_{xy}[\ k \cdot \phi(xy, \mathbb{A}) - \phi(xy, \mathbb{A}) \cdot \phi(xy, \mathbb{K}) \quad + \quad \phi(xy, \mathbb{K}) - \phi(xy, \mathbb{A}) \cdot \phi(xy, \mathbb{K})\ ] \\
&= \quad |\alpha(\mathbb{K})| \quad + \quad \sum_{xy}[\ \phi(xy, \mathbb{A})(k - 2 \cdot \phi(xy, \mathbb{K}))\ ] \\
&= \quad |\alpha(\mathbb{K})| \quad + \quad \sum_{xy \in \alpha(\mathbb{A})}(k - 2 \cdot \phi(xy, \mathbb{K}))
\end{aligned}
$$

The solution for the SCJ aliquoting problem of a k-folded genome is:

$$\alpha(\mathbb{A}) = \left\{ xy : \phi(xy, \mathbb{K}) > \frac{k}{2} \right\}$$

SCJ guided halving/aliquoting of a k-folded genome

Given a k-folded genome $\mathbb{K}$ and a canonical genome $\mathbb{C}$ find a canonical genome $\mathbb{A}$ that minimizes the sum:

$$\mathsf{ga}_{\mathsf{SCJ}}(\mathbb{A}) = \mathsf{d}^k_{\mathsf{SCJ}}(\mathbb{A}, \mathbb{K}) + \mathsf{d}_{\mathsf{SCJ}}(\mathbb{A}, \mathbb{C}) = \mathsf{d}_{\mathsf{SCJ}}(\mathbf{k} \cdot \mathbb{A}, \mathbb{K}) + \mathsf{d}_{\mathsf{SCJ}}(\mathbb{A}, \mathbb{C})$$

Therefore:

$$
\begin{aligned}
\mathsf{ga}_{\mathsf{SCJ}}(\mathbb{A})) \quad &= \quad |\alpha(\mathbf{k} \cdot \mathbb{A}) \setminus \alpha(\mathbb{K})| \;+\; |\alpha(\mathbb{K}) \setminus \alpha(\mathbf{k} \cdot \mathbb{A})| \;+\; |\alpha(\mathbb{A}) \setminus \alpha(\mathbb{C})| \;+\; |\alpha(\mathbb{C}) \setminus \alpha(\mathbb{A})| \\[4pt]
&= \quad \sum_{xy \in \alpha(\mathbb{A})}(k - \phi(xy, \mathbb{K})) + \sum_{xy \notin \alpha(\mathbb{A})} \phi(xy, \mathbb{K}) + \sum_{xy \in \alpha(\mathbb{A})}(1 - \phi(xy, \mathbb{C})) + \sum_{xy \notin \alpha(\mathbb{A})} \phi(xy, \mathbb{C}) \\[4pt]
&= \quad \sum_{xy \in \alpha(\mathbb{A})}(k + 1 - \phi(xy, \mathbb{K}, \mathbb{C})) \;+\; \sum_{xy \notin \alpha(\mathbb{A})} \phi(xy, \mathbb{K}, \mathbb{C}) \\[4pt]
&= \quad \sum_{xy}[\; \phi(xy, \mathbb{A}) \cdot (k + 1 - \phi(xy, \mathbb{K}, \mathbb{C})) \;+\; (1 - \phi(xy, \mathbb{A})) \cdot \phi(xy, \mathbb{K}, \mathbb{C}) \;] \\[4pt]
&= \quad \sum_{xy}[\; (k + 1) \cdot \phi(xy, \mathbb{A}) - \phi(xy, \mathbb{A}) \cdot \phi(xy, \mathbb{K}, \mathbb{C}) \;+\; \phi(xy, \mathbb{K}, \mathbb{C}) - \phi(xy, \mathbb{A}) \cdot \phi(xy, \mathbb{K}, \mathbb{C}) \;] \\[4pt]
&= \quad |\alpha(\mathbb{K})| \;+\; |\alpha(\mathbb{C})| \;+\; \sum_{xy}[\; \phi(xy, \mathbb{A})(k + 1 - 2 \cdot \phi(xy, \mathbb{K}, \mathbb{C})) \;] \\[4pt]
&= \quad |\alpha(\mathbb{K})| \;+\; \sum_{xy \in \alpha(\mathbb{A})}(k + 1 - 2 \cdot \phi(xy, \mathbb{K}, \mathbb{C}))
\end{aligned}
$$

The solution for the guided SCJ aliquoting problem of a k-folded genome is:

$$\alpha(\mathbb{A}) = \left\{ xy : \phi(xy, \mathbb{K}, \mathbb{C}) > \frac{k+1}{2} \right\}$$

Perfect matching and circular canonical genomes

For a given set $N = \{1, 2, \ldots, n\}$,

let G be a complete graph with vertices $V(G) = \{g^h : g \in N\} \cup \{g^t : g \in N\}$

A perfect matching M in G corresponds to $|M| = n$ adjacencies and, consequently, defines a circular canonical genome $\mathbb{C}$, with $\mathcal{F}(\mathbb{C}) = \mathcal{G}(\mathbb{C}) = N$ and $\alpha(\mathbb{C}) = M$.

SCJ median of k canonical circular genomes

Given k canonical circular genomes $\mathbb{C}_1, \mathbb{C}_2, \ldots \mathbb{C}_k$, find a canonical circular genome $\mathbb{M}$ that minimizes the sum:

$$
\begin{aligned}
\mathsf{s}_{\text{SCJ}}(\mathbb{M}) \quad &= \quad \mathsf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_1) + \mathsf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_2) + \ldots + \mathsf{d}_{\text{SCJ}}(\mathbb{M}, \mathbb{C}_k) \\
&= \quad |\alpha(\mathbb{C}_1)| + |\alpha(\mathbb{C}_2)| + \ldots + |\alpha(\mathbb{C}_k)| \quad + \quad \omega(\mathbb{M}) \\
&= \quad 3n \quad + \quad \omega(\mathbb{M})
\end{aligned}
$$

Again, we need to minimize $\omega(\mathbb{M}) = \sum_{xy \in \alpha(\mathbb{M})} \omega(xy)$, where $\omega(xy) = k - 2 \cdot \phi(xy, \mathbb{C}_{1..k})$.

1. Build the complete graph G with vertices $V(G) = \{g^h : g \in \mathcal{G}_\star\} \cup \{g^t : g \in \mathcal{G}_\star\}$
2. Assign weights to each edge xy of G: $\omega(xy) = k - 2 \cdot \phi(xy, \mathbb{C}_{1..k})$.

Perfect matching M in $G \Leftrightarrow$ Circular genome $\mathbb{M}$; with $\omega(M) = \omega(\mathbb{M})$

Minimum weight matching M gives a minimum weight circular SCJ median $\mathbb{M}$

Breakpoint median of canonical circular genomes

Given canonical circular genomes $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$, find a canonical circular genome $\mathbb{M}$ that minimizes the sum:

$$
\begin{aligned}
s_{\mathrm{BP}}(\mathbb{M}) &= & d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_1) &+ d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_2) &+ d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_k) \\
&= & n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_1) &+ n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_2) &+ n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_3) \\
&= & 3n &- \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_{1..3}) \\
&= & 3n &- \omega'(\mathbb{M})
\end{aligned}
$$

Now we need to maximize $\omega'(\mathbb{M}) = \sum_{xy \in \alpha(\mathbb{M})} \omega'(xy)$, where $\omega'(xy) = \phi(xy, \mathbb{C}_{1..3})$.

1. Build the complete graph G with vertices $V(G) = \{g^h : g \in \mathcal{G}_\star\} \cup \{g^t : g \in \mathcal{G}_\star\}$
2. Assign weights to each edge xy of G: $\omega'(xy) = \phi(xy, \mathbb{C}_{1..k})$.

Perfect matching M in G $\Leftrightarrow$ Circular genome $\mathbb{M}$; with $\omega'(M) = \omega'(\mathbb{M})$

Maximum weight matching M gives a minimum weight circular breakpoint median $\mathbb{M}$

Breakpoint median of canonical genomes

Given canonical genomes $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$, find a canonical genome $\mathbb{M}$ that minimizes the sum:

$$
\begin{aligned}
s_{\mathrm{BP}}(\mathbb{M}) \;=\;& d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_1) \;+\; d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_2) \;+\; d_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_k) \\[4pt]
=\;& n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_1) - \sum_{x \in \gamma(\mathbb{M})} \frac{\phi(x, \mathbb{C}_1)}{2} \;+\; n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_2) \\
& -\sum_{x \in \gamma(\mathbb{M})} \frac{\phi(x, \mathbb{C}_2)}{2} \;+\; n - \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_3) - \sum_{x \in \gamma(\mathbb{M})} \frac{\phi(x, \mathbb{C}_3)}{2} \\[4pt]
=\;& 3n \;-\; \sum_{xy \in \alpha(\mathbb{M})} \phi(xy, \mathbb{C}_{1..3}) \;-\; \sum_{x \in \gamma(\mathbb{M})} \frac{\phi(x, \mathbb{C}_{1..3})}{2} \\[4pt]
=\;& 3n \;-\; \omega'(\mathbb{M})
\end{aligned}
$$

Now we need to maximize $\omega'(\mathbb{M}) \;=\; \sum_{xy \in \alpha(\mathbb{M})} \omega'(xy) + \sum_{x \in \gamma(\mathbb{M})} \omega'(x)$,
where $\omega'(xy) = \phi(xy, \mathbb{C}_{1..3})$ and $\omega'(x) = \frac{\phi(x, \mathbb{C}_{1..3})}{2}$.

1. Build the complete graph G with vertices $V(G) = \{g^h : g \in \mathcal{G}_\star\} \cup \{g^t : g \in \mathcal{G}_\star\}$
2. Assign weights to each edge xy of G: $\omega'(xy) = \phi(xy, \mathbb{C}_{1..k})$.
3. Build the complete graph G_t with vertices $V(G_t) = \{t_{g^h} : g \in \mathcal{G}_\star\} \cup \{t_{g^t} : g \in \mathcal{G}_\star\}$
4. Assign weight 0 to each edge of G_t
5. Add one edge connecting each vertex x in G to the corresponding vertex t_x in G_t, with weight
   $\omega'(xt_x) = \frac{\phi(x, \mathbb{C}_{1..k})}{2}$

Perfect matching M in $G + G_t \Leftrightarrow$ Genome $\mathbb{M}$; with $\omega'(M) = \omega'(\mathbb{M})$

Maximum weight matching M gives a minimum weight breakpoint median $\mathbb{M}$

References

Multichromosomal median and halving problems under different genomic distances

(Eric Tannier, Chunfang Zheng and David Sankoff)

BMC Bioinformatics volume 10, Article number: 120 (2009)

SCJ: A Breakpoint-Like Distance that Simplifies Several Rearrangement Problems

(Pedro Feijão and João Meidanis)

TCBB volume 8 Number: 5 (2011)

On the Complexity of Rearrangement Problems under the Breakpoint Distance

(Jakub Ková č)

JCB volume 21, Number 1 (2014)