Supplementary Material: Genome Rearrangement with ILP

Tom Hartmann[1], Nicolas Wieseke[1], Roded Sharan[2] Martin Middendorf[1], and Matthias Bernt[1]

[1] Parallel Computing and Complex Systems Group, Institute of Computer Science, University Leipzig, Germany
{thartmann,wieseke,middendorf,bernt}@informatik.uni-leipzig.de
[2] Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
roded@post.tau.ac.il

1 Supplementary Proofs

Theorem 1. *Let* $x_1, x_2, N \in \mathbb{R}$ *such that* $N > |x_1 - x_2| \geq 0$ *and* $A_1, \ldots, A_n, B_1, \ldots, B_m$ *be binary variables with* $n + m > 0$ *such that* $n, m \in \mathbb{N}_0$. *Then the implication* $A_1 = 1 \wedge \ldots \wedge A_n = 1 \wedge B_1 = 0 \wedge \ldots \wedge B_m = 0 \Rightarrow x_1 = x_2$ *is satisfied by using the following constraints:*

$$N(1 - A_1) + \ldots + N(1 - A_n) + NB_1 + \ldots + NB_m + x_1 \geq x_2$$
$$N(1 - A_1) + \ldots + N(1 - A_n) + NB_1 + \ldots + NB_m + x_2 \geq x_1$$

Proof. *Assume* $A_i = 1$ *and* $B_j = 0$ *for all* $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$. *Then, the constraints imply* $x_1 \geq x_2$ *and* $x_2 \geq x_1$. *Therefore,* $x_1 = x_2$ *holds. Note that both constraints are always satisfied if there exist at least one* A_i *with* $A_i = 0$ *or one* B_j *with* $B_j = 1$. *This is true since* $N > |x_1 - x_2| \geq 0$ *holds by the assumption.*

The following corollaries follow directly from Theorem 1.

Corollary 1. *Implication* (5) *is satisfied by*

$$(1 - B^L_{el}) + (1 - B^M_{fl}) + O_{efl} \geq 1.$$

Corollary 2. *Implication* (6) *is satisfied by*

$$(1 - T_l) + (1 - B^M_{el}) + (1 - B^R_{fl}) + O_{efl} \geq 1.$$

Corollary 3. *Implication* (7) *is satisfied by*

$$(1 - U_l) + (1 - B^M_{el}) + (1 - B^R_{fl}) + O_{efl} \geq 1.$$

Corollary 4. *Implication* (8) *is satisfied by*

$$(1 - I_l) + (1 - B^M_{el}) + (1 - B^R_{fl}) + e \geq f,$$
$$(1 - I_l) + (1 - B^M_{el}) + (1 - B^R_{fl}) + f \geq e.$$

Corollary 5. *Implication* (10) *is satisfied by*

$$I_l \leq C_l^L,$$
$$(1 - I_l) \geq C_l^R.$$

Corollary 6. *Implication* (11) *is satisfied by*

$$(1 - T_l) \geq C_l^L,$$
$$(1 - T_l) \geq C_l^R.$$

Corollary 7. *Implication* (12) *is satisfied by*

$$(1 - U_l) + 1 \geq C_l^L + C_l^R$$
$$(1 - U_l) + C_l^L + C_l^R \geq 1.$$

Corollary 8. *Implication* (13) *is satisfied by*

$$(1 - B_{el}^X) \geq L_{el}^X.$$

Corollary 9. *Implication* (14) *is satisfied by*

$$(1 - O_{efl}) + (1 - B_{fl}^X) + L_{fl}^X \geq 1.$$

Corollary 10. *Implication* (15) *is satisfied by*

$$(1 - O_{fel}) + (1 - B_{fl}^X) \geq L_{el}^X.$$

Corollary 11. *Implication* (16) *is satisfied by*

$$B_{el}^X \leq R_{el}^X.$$

Corollary 12. *Implication* (17) *is satisfied by*

$$(1 - O_{fel}) + (1 - B_{fl}^X) + R_{el}^X \geq 1.$$

Corollary 13. *Implication* (18) *is satisfied by*

$$(1 - O_{efl}) + (1 - B_{fl}^X) \geq R_{el}^X.$$

Corollary 14. *Implication* (19) *is satisfied by*

$$(1 - R_{el}^L) + (1 - L_{el}^M) + W_{el}^L \geq 1.$$

Corollary 15. *Implication* (20) *is satisfied by*

$$(1 - R_{el}^L) + L_{el}^M \geq W_{el}^L.$$

Corollary 16. *Implication* (21) *is satisfied by*

$$(1 - L_{el}^M) + R_{el}^L \geq W_{el}^L.$$

Corollary 17. *Implication* (22) *is satisfied by*

$$(1 - R^M_{el}) + (1 - L^R_{el}) + W^R_{el} \geq 1.$$

Corollary 18. *Implication* (23) *is satisfied by*

$$(1 - R^M_{el}) + L^R_{el} \geq W^R_{el}.$$

Corollary 19. *Implication* (24) *is satisfied by*

$$(1 - L^R_{el}) + R^M_{el} \geq W^R_{el}.$$

Corollary 20. *Implication* (25) *is satisfied by*

$$(1 - C^X_l) + (1 - W^X_{el}) + (1 - W^X_{fl}) + O_{efl+1} \geq (1 - O_{efl}),$$
$$(1 - C^X_l) + (1 - W^X_{el}) + (1 - W^X_{fl}) + (1 - O_{efl}) \geq O_{efl+1}.$$

Corollary 21. *Implication* (26) *is satisfied by*

$$(1 - W^L_{el}) + (1 - W^R_{fl}) + O_{efl+1} \geq (1 - O_{efl}),$$
$$(1 - W^L_{el}) + (1 - W^R_{fl}) + (1 - O_{efl}) \geq O_{efl+1}.$$

Corollary 22. *Implication* (27) *is satisfied by*

$$W^L_{el} + W^R_{el} + O_{efl+1} \geq O_{efl},$$
$$W^L_{el} + W^R_{el} + O_{efl} \geq O_{efl+1}.$$

Corollary 23. *Implication* (28) *is satisfied by*

$$W^L_{fl} + W^R_{fl} + O_{efl+1} \geq O_{efl},$$
$$W^L_{fl} + W^R_{fl} + O_{efl} \geq O_{efl+1}.$$

Corollary 24. *Implication* (29) *is satisfied by*

$$C^X_l + (1 - W^X_{el}) + (1 - W^X_{fl}) + O_{efl+1} \geq O_{efl},$$
$$C^X_l + (1 - W^X_{el}) + (1 - W^X_{fl}) + O_{efl} \geq O_{efl+1}.$$

Corollary 25. *Implication* (30) *is satisfied by*

$$(1 - C^X_l) + (1 - W^X_{el}) + S_{el+1} \geq (1 - S_{el}),$$
$$(1 - C^X_l) + (1 - W^X_{el}) + (1 - S_{el}) \geq S_{el+1}.$$

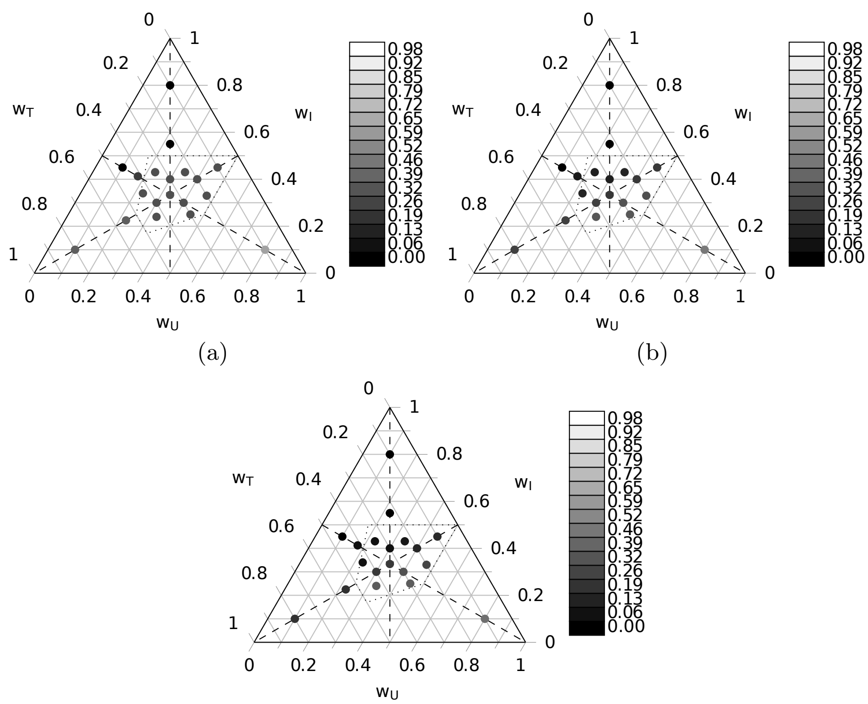
Corollary 26. *Implication* (31) *is satisfied by*

$$C^X_l + (1 - W^X_{el}) + S_{el+1} \geq S_{el},$$
$$C^X_l + (1 - W^X_{el}) + S_{el} \geq S_{el+1}.$$

Corollary 27. *Implication* (32) *is satisfied by*

$$W^L_{el} + W^R_{el} + S_{el+1} \geq S_{el},$$
$$W^L_{el} + W^R_{el} + S_{el} \geq S_{el+1}.$$

2 Supplementary Figures



(a) (b) (c)

Fig. S1:
Fraction of inversions of all scenarios which are obtained by solving (a) Π_1, (b) Π_2, and (c) Π_3. Only instances that have been solved in less than 90000s are included. The dotted polygon shows the all-type area and dashed lines show weighting schemes (w_I, w_T, w_U) such that $w_T = w_U$ (repectively $w_I = w_T$ and $w_I = w_U$).

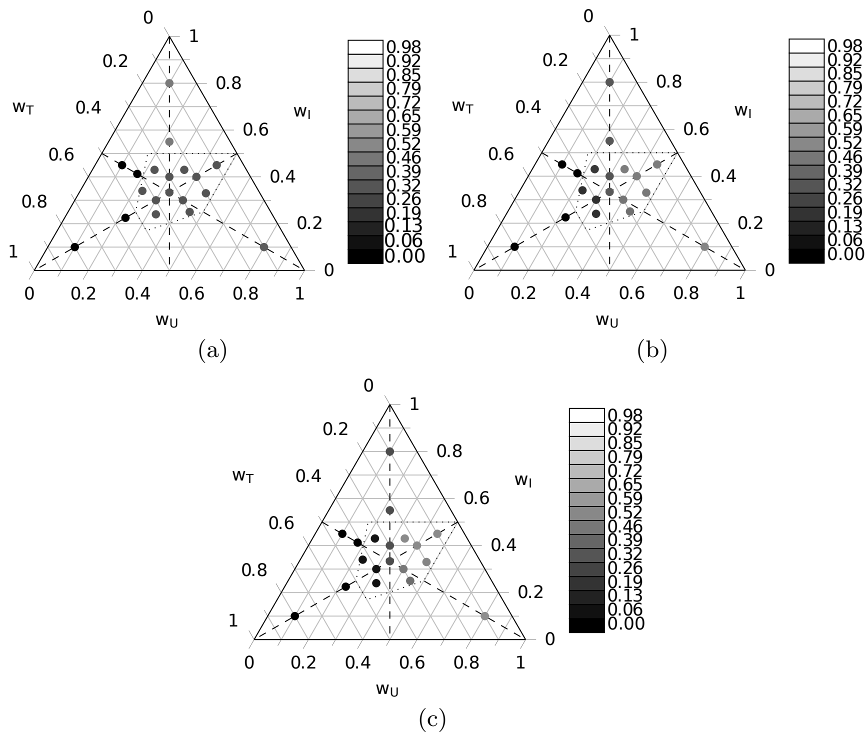


Fig. S2:

Fraction of transpositions of all scenarios which are obtained by solving a) Π_1, b) Π_2, and c) Π_3. Only instances that have been solved in less than 90000s are included. The dotted polygon shows the all-type area and dashed lines show weighting schemes (w_I, w_T, w_U) such that $w_T = w_U$ (repectively $w_I = w_T$ and $w_I = w_U$).

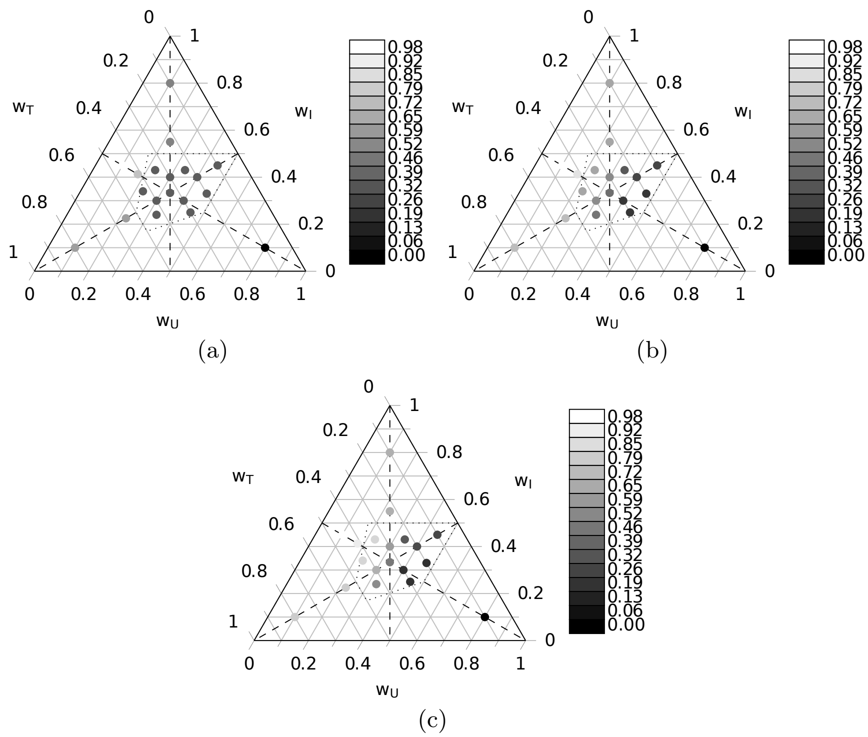


Fig. S3:
Fraction of inverse transpositions of all scenarios which are obtained by solving (a) Π_1, (b) Π_2, and (c) Π_3. Only instances that have been solved in less than 90000s are included. The dotted polygon shows the all-type area and dashed lines show weighting schemes (w_I, w_T, w_U) such that $w_T = w_U$ (repectively $w_I = w_T$ and $w_I = w_U$).

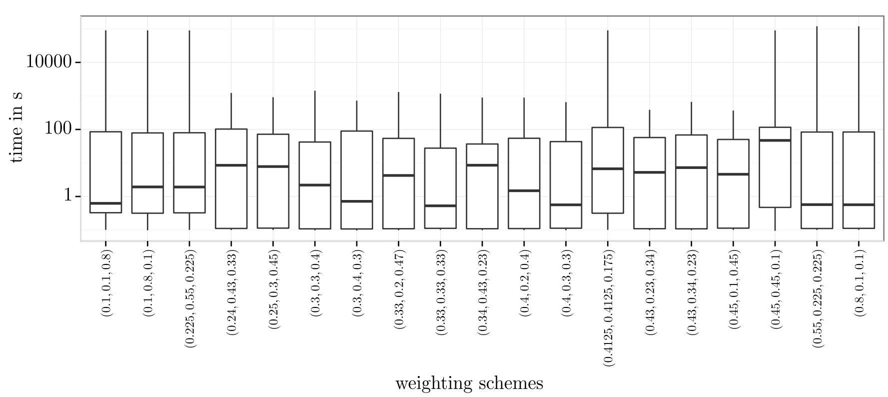


Fig. S4:
Runtime for all weighting schemes which were used to solve Π_i with $i \in [1:3]$.

3 Supplementary Tables

Constants	**Description**
n	number of genes of π
d	distance $d(\pi, \iota)$ between π and ι
l	$l \in [0 : d]$
w_I	cost of an inversion
w_T	cost of a transposition
w_U	cost of an inverse transposition
binary Variables	**Description**
O_{efl}	$O_{efl} = 1 \Leftrightarrow$ gene e is to the left of gene f in permutation π_l
S_{el}	$S_{el} = 1 \Leftrightarrow e$ has a negative sign in π_l
I_l	$I_l = 1 \Leftrightarrow \rho_l$ is an inversion with $\rho_l(\pi_l) = \pi_{l+1}$
T_l	$T_l = 1 \Leftrightarrow \rho_l$ is a transposition with $\rho_l(\pi_l) = \pi_{l+1}$
U_l	$U_l = 1 \Leftrightarrow \rho_l$ is an inverse transposition with $\rho_l(\pi_l) = \pi_{l+1}$
C_l^X	$C_l^X = 1 \Leftrightarrow X$-th interval gets inverted from π_l to π_{l+1},$X \in \{L, R\}$
B_{el}^X	$B_{el}^X = 1 \Leftrightarrow \pi_l(e)$ is the X-th bounding element, $X \in \{L, M, R\}$
L_{el}^X	$L_{el}^X = 1 \Leftrightarrow \pi_l(e)$ is left of X-th bounding element, $X \in \{M, R\}$
R_{el}^X	$R_{el}^X = 1 \Leftrightarrow \pi_l(e)$ is right of X-th bounding element, $X \in \{L, M\}$
W_{el}^X	$W_{el}^X = 1 \Leftrightarrow \pi_l(e)$ is between B_l^L and B_l^M or B_l^M and B_l^R with $X \in \{L, R\}$, respectively

Table S1:
Variables that are used in the ILP formulation.