基因组一般移位排序问题的多项式时间算法

尹 晓 朱大铭
(山东大学计算机科学与技术学院 济南 250101)

摘 要 基因组移位排序在基因组重组排序计算研究中占有重要位置.交互型移位和非交互型移位均为移位的特殊形式.目前见到的多种移位排序算法均是针对交互型移位而得到的,未见基因组一般移位排序计算的研究结果.文中讨论包括交互型移位和非交互型移位的一般移位排序问题的求解方法,给出该问题的一个多项式时间算法.算法的关键在于将一般移位排序问题在线性时间内归约为交互型移位排序问题,利用交互型移位排序的算法来求解一般移位排序.作者的算法证实了 Ozery-Flato 等关于一般移位排序问题可以多项式时间解决的猜测.

关键词 算法;基因组重组;移位;移位距离;计算生物学
中图法分类号 TP18 **DOI 号**: 10.3724/SP.J.1016.2010.00785

Polynomial-Time Algorithm for Sorting Genomes by Generalized Translocations

YIN Xiao ZHU Da-Ming
(*School of Computer Science and Technology, Shandong University, Jinan* 250101)

Abstract Sorting genomes by translocations plays an important role in the research of genome rearrangement. Translocation is a prevalent rearrangement event in the evolution of multi-chromosomal species which exchanges ends between two chromosomes. Translocations include reciprocal translocations and non-reciprocal translocations. Translocation sorting problem asks to find a shortest sequence of translocations to transform one genome into another. Several polynomial algorithms have been presented, all of them only allowing reciprocal translocations. Thus they can only be applied to a pair of genomes having the same set of chromosome ends. Such a restriction can be removed if non-reciprocal translocations are also allowed. In this paper, the authors study for the problem of sorting by generalized translocations, which allows both reciprocal and non-reciprocal translocations, and present a polynomial-time algorithm for this problem, in which the problem of sorting by generalized translocations is reduced in linear time to the problem of sorting by reciprocal translocations. This algorithm confirms Ozery-Flato's conjecture that sorting by generalized translocations could be solved in polynomial time.

Keywords algorithm; genome rearrangement; translocation; translocation distance; computational biology

1 引 言

基因组重组是基因组改变基因排列顺序的生化过程.基因组重组的生化过程非常复杂,但可将其归为 3 种基本操作,即反转(reversal)、移位(translocation)和转位(transposition).基因组重组排序要求计算将一个基因组转化为另一个基因组的最短重组

收稿日期:2009-03-02;最终修改稿收到日期:2009-09-30.本课题得到国家自然科学基金(60573024,60970003,60603007)和教育部博士点基金(20090131110009)资助.**尹 晓**,女,1982 年生,博士研究生,主要研究方向为算法分析与设计、计算生物学. E-mail: xyin@mail.sdu.edu.cn. **朱大铭**,男,1964 年生,教授,博士生导师,主要研究领域为算法分析与设计、计算生物学.

操作序列,从而推断生命的演化过程.基因组重组排序多年来一直是计算生物学的热点问题.有向基因组反转和移位排序多项式算法的设计成功,应是20世纪90年代人们用"计算"解决分子生物学问题的典型贡献[1-2],有关反转和转位排序问题的算法与复杂性主要研究结果见文献[3-7].本文讨论基因组的移位排序算法.

移位交换基因组中两条染色体的前缀或后缀.若交换的两部分都不为空,则为交互型移位,否则为非交互型移位.所谓交互型移位排序,即只通过交互型移位对基因组进行排序.Hannenhalli最先给出了交互型移位排序问题的多项式时间算法,时间复杂性为$O(n^3)$[2],其中n为基因组中基因的个数.之后,朱等将文献[2]算法的时间复杂度改进为$O(n^2\log n)$[8].刘等进一步将算法的时间复杂度改进为$O(n^2)$[9].最近Ozery-Flato等将交互型移位排序算法的时间复杂度改进为$O(n\sqrt{n\log n})$[10],这是目前最好的交互型移位排序算法.

交互型移位排序要求源基因组和目标基因组必须具有相同的尾基因集合.然而,在实际的基因组重组排序实例中,源基因组和目标基因组未必具有相同的尾基因集合.可以认为,两个基因组具有不同尾基因集合,是因非交互型移位形成的.非交互型移位有3种形式:合并(fussion)、分裂(fission)与分裂-合并(fission-fussion).合并将两条染色体连接为一条,因而消除两个尾基因;分裂将一条染色体断为两条染色体,从而产生两个新的尾基因;分裂-合并先将一条染色体分裂为两段,再将其中一段与另一条染色体合并.

Hannenhalli设计了考虑反转、交互型移位和非交互型移位的多项式算法[11]. Tesler将Hannenhalli的算法实现为基因组排序软件[12],供同行测试和使用.但在实际基因组比较中,人们还需经常考虑只包括交互型移位和非交互型移位的基因组重组排序结果.Ozery-Flato曾在文献[13]中猜测,包括交互型移位和非交互型移位的移位排序问题可以多项式时间解决,但至今未见有人给出该问题的算法与复杂性研究结果.

我们将包括交互型移位和非交互型移位的移位排序问题称作一般的移位排序问题.本文讨论一般移位排序问题的求解方法,给出该问题的一个多项式时间算法.本文算法证实了Ozery-Flato的猜测.下面将交互型移位排序问题简称为SRT(Sorting by Reciprocal Translocations),将一般移位排序问题简称为SGT(Sorting by Generalized Translocations).

2 预备知识

基因组是染色体的集合,每条染色体表示为一个整数序列,其中的整数表示基因.在整数前增加一个符号±,表示基因在染色体中的相对方向.例如,$\{(3,-5),(2,4,-6),(-1,7)\}$表示一个包含7个基因3条染色体的基因组.一个基因在同一个基因组中只出现一次,因此若基因组A含有n个基因,则采用整数集$\{1,2,\cdots,n\}$表示A中的所有基因.

设$S=(x_1,x_2,\cdots,x_l)$为染色体中的一段基因序列,$-S=(-x_l,-x_{l-1},\cdots,-x_1)$表示$S$的逆序.给定两条染色体$X$和$Y$,若$X=Y$或$X=-Y$,则$X$与$Y$实际表示同一条染色体,称$X$与$Y$等价.若基因组$A$中的每条染色体$X$,在基因组$B$中都存在一条染色体$Y$与$X$等价,则称$A$与$B$等价,记作$A=B$.

给定两条染色体$X=(X_1,X_2)$,$Y=(Y_1,Y_2)$,其中X_1,X_2,Y_1,Y_2表示基因序列.移位将X中的基因序列X_1或X_2,与Y中的基因序列Y_1或Y_2交换.移位有两种类型:P-P移位和P-S移位.P-P移位将$(X_1,X_2),(Y_1,Y_2)$变为$(Y_1,X_2),(X_1,Y_2)$;P-S移位将$(X_1,X_2),(Y_1,Y_2)$变为$(-Y_2,X_2),(Y_1,-X_1)$.若X_1,X_2,Y_1,Y_2都不为空,则该移位为交互型移位.

除了交互型移位外,还有3种非交互型移位:分裂、合并、分裂-合并.分裂将一条染色体断开,形成两条染色体.X的分裂将$X=(X_1,X_2)$变为(X_1)和(X_2).合并将两条染色体首尾连接,形成一条染色体.染色体X,Y合并后可得到$(X,Y),(-X,Y),(X,-Y)$或$(-X,-Y)$.分裂-合并先将一条染色体分裂为两段,再将其中一段连接到另一条染色体上.$X=(X_1,X_2)$与Y的分裂-合并有4种可能的结果:$(X_1),(Y,X_2)$;$(X_1),(-X_2,Y)$;$(X_1,Y),(X_2)$或$(Y,-X_1),(X_2)$.3种非交互型移位都可看作是移位时交换的某段基因序列为空的特殊情况.例如,X分裂为(X_1)和(X_2)可看作是(X_1,X_2)与"空染色体"$(\varnothing,\varnothing)$移位得到$(X_1,\varnothing)$和$(\varnothing,X_2)$.$X,Y$合并为$(X,Y)$可看作是$(X,\varnothing)$与$(\varnothing,Y)$移位得到$(X,Y)$和$(\varnothing,\varnothing)$;$X=(X_1,X_2)$与$Y$分裂-合并得到$(X_1),(Y,X_2)$,可看作是$(X_1,X_2)$与$(Y,\varnothing)$移位

得到$(X_1,\varnothing),(Y,X_2)$.

设有一条染色体 $X=(x_1,x_2,\cdots,x_m)$，将 x_1 和 $-x_m$ 称作 X 的尾基因. X 的尾基因集合记为 $Tails(X)=\{x_1,-x_m\}$. 基因组 A 的尾基因集合记为 $Tails(A)=\bigcup_{X\in A}Tails(X)$. 例如，$Tails(\{(1,-3,-4),(-5,-2,6)\})=\{1,4,-5,-6\}$. 若基因组 A、B 满足 $Tails(A)=Tails(B)$，则称 A 与 B 是同尾基因组. 交互型移位排序问题要求源基因组和目标基因组必须是同尾基因组.

2.1　同尾基因组的断点图

给定一条染色体 $X=(x_1,x_2,\cdots,x_m)$，将 X 的每个基因 x_i 都用一个有序的点对 $(l(x_i),r(x_i))$ 代替：若 x_i 的符号为＋，则 $l(x_i)=x_i^t,r(x_i)=x_i^h$；若 x_i 的符号为－，则 $l(x_i)=x_i^h,r(x_i)=x_i^t$. 由 X 可得到一个点序列 $l(x_1)r(x_1)l(x_2)r(x_2)\cdots l(x_m)r(x_m)$. 下面我们也将染色体 X 的点序列直接称作染色体 X. 对于 $1\leq i\leq m-1$，称 $r(x_i)$ 与 $l(x_{i+1})$ 在 X 中相邻. 若 X 是基因组 A 的染色体且点 u,v 在 X 中相邻，则称 u,v 在 A 中相邻.

给定同尾基因组 A、B，用以下方法构造断点图 $G(A,B)$：点集为 $V=\{u:u=x^t$ 或 x^h，其中 x 为 A 中的基因$\}$. 设 $u,v\in V$，若 u,v 在 A 中相邻，则在 u,v 之间连黑边；若 u,v 在 B 中相邻，则连灰边. 设 A（或 B）的基因数为 n，染色体数为 N，则 $G(A,B)$ 的黑边和灰边数目都为 $n-N$.

设 $g=(u,v)$ 为 $G(A,B)$ 中的一条灰边，若 u,v 在 A 中位于同一条染色体，则称 g 为体内灰边；否则，称 g 为体间灰边. $G(A,B)$ 中点的度都为 0 或 2，因而 $G(A,B)$ 可唯一地分解为黑边和灰边交替且顶点互不重叠的圈. 只有一条黑边的圈称作短圈，否则称作长圈. 当 $A=B$ 时，$G(A,B)$ 由 $n-N$ 个短圈构成.

2.2　最小子排列和偶隔离带

给定基因组 A 的染色体 X 上的一段基因序列 $I=x_i,x_{i+1},\cdots,x_j$，则 I 在 $G(A,B)$ 中对应点序列 $l(x_i)r(x_i)\,l(x_{i+1})\,r(x_{i+1})\cdots l(x_j)\,r(x_j)$. 定义 $V(I)=\bigcup_{i\leq k\leq j}\{x_k^t,x_k^h\}$，$IN(I)=V(I)\backslash\{l(x_i),r(x_j)\}$.

设 $I=x_i,x_{i+1},\cdots,x_j$ 为 A 的染色体 X 的一段基因序列，$I'=x_i,permutation(x_{i+1},\cdots,x_{j-1}),x_j$ 为 B 的染色体 Y 的一段基因序列，其中 $permutation(x_{i+1},\cdots,x_{j-1})$ 表示 $x_{i+1},\cdots,x_{j-1}$ 的一个重排列. 若 $permutation(x_{i+1},\cdots,x_{j-1})\neq x_{i+1},\cdots,x_{j-1}$，则称 I 为基因组 A 的一个子排列. 将不包含任意其它更小子排列的子排列称为一个最小子排列，简记为 MSP(Minimum Sub Permutation).

子排列 I 在 $G(A,B)$ 中表现为由点集 $IN(I)$ 支撑的一个子图，满足：(1) 该子图中不存在边 (u,v) 使 $u\in IN(I)$ 而 $v\notin IN(I)$；(2) 该子图至少包含一个长圈. 下面也直接将（最小）子排列 I 在 $G(A,B)$ 中对应的子图称为（最小）子排列.

若 $G(A,B)$ 满足：(1) MSP 数为偶数；(2) 所有的 MSP 均被同一个子排列所包含，则称 $G(A,B)$ 包含一个偶隔离带.

2.3　交互型移位距离

同尾基因组 A 和 B 的交互型移位距离即从 A 转化为 B 的最少交互型移位次数，记作 $d_r(A,B)$. 令 $c(A,B)$ 和 $s(A,B)$ 分别表示 $G(A,B)$ 中的圈数和 MSP 数. 定义参数 $f(A,B)$，

$$f(A,B)=\begin{cases}2, & G(A,B)\text{包含偶隔离带}\\ 1, & s(A,B)\text{为奇数}\\ 0, & \text{否则}\end{cases}.$$

定理 1(Hannenhalli)[2]. 同尾基因组 A、B 的交互型移位距离 $d_r(A,B)=n-N-c(A,B)+s(A,B)+f(A,B)$.

3　一般移位排序

下面我们讨论 A、B 不一定是同尾基因组的一般情况. 将交互型移位和非交互型移位统称为移位. 基因组移位排序问题描述如下：

实例——基因组 A、B，A 和 B 具有相同的基因集合 $\{1,2,\cdots,n\}$.

目标——计算将 A 转化为 B 的最少移位次数和最短移位序列.

将 A 转化为 B 的最少移位次数称为 A 与 B 的移位距离，记为 $d(A,B)$.

3.1　加帽基因组

给定两个基因组 $A=\{X_1,X_2,\cdots,X_M\}$，$B=\{Y_1,Y_2,\cdots,Y_N\}$，其中 $X_i(1\leq i\leq M)$ 和 $Y_j(1\leq j\leq N)$ 分别为 A 和 B 的染色体. 首先给 A、B 添加基因，构造同尾基因组 $\hat{A}$、$\hat{B}$. 所添加的基因集合为 $Caps=\{C_1,C_2,\cdots,C_{2n}\}$，其中对于 $1\leq k\leq 2n$，$C_k=\{n+k\}$. 将 $Caps$ 中的元素均称作帽子. $\hat{A}$、$\hat{B}$ 的构造方法如下：给 A 的每条染色体 X_i 的左右两端分别添加帽子 C_{2i-1} 和 C_{2i}，并添加 $n-M$ 条由两个帽子构成的“空染色体”，得到基因组 $\hat{A}=\{\hat{X}_1,\hat{X}_2,\cdots,\hat{X}_n\}$，其

中，当 $1\leq i\leq M$ 时，$\hat{X}_i=(C_{2i-1},X_i,C_{2i})$；当 $M<i\leq n$ 时，$\hat{X}_i=(C_{2i-1},C_{2i})$. 给 B 的每条染色体 Y_i 的左右两端分别添加 $(-1)^{j+1}C_j$ 和 $(-1)^kC_k$，并添加 $n-N$ 条“空染色体”，得到基因组 $\hat{B}=\{\hat{Y}_1,\hat{Y}_2,\cdots,\hat{Y}_n\}$，其中，当 $1\leq i\leq N$ 时，$\hat{Y}_i=((-1)^{j+1}C_j,Y_i,(-1)^kC_k)$；当 $N<i\leq n$ 时，$\hat{Y}_i=((-1)^{j+1}C_j,(-1)^kC_k)$，$1\leq j,k\leq 2n$ 且每个帽子在 $\hat{B}$ 中只出现一次. 我们将 $\hat{A}$、$\hat{B}$ 分别称为 A、B 的加帽基因组. B 共有 $(2n)!/2^{n-N}(n-N)!$ 个加帽基因组. 将 B 的加帽基因组的集合记作 Γ. 例如：$A=\{(3,4),(2,1)\}$，$B=\{(1),(2),(3,4)\}$，A 的加帽基因组 $\hat{A}=\{(5,3,4,6),(7,2,1,8),(9,10),(11,12)\}$，$B$ 的其中一个加帽基因组为 $\hat{B}=\{(5,1,6),(7,2,8),(9,3,4,10),(11,12)\}$.

3.2 用交互型移位模拟一般移位

设 $X=(X_1,X_2)$，$Y=(Y_1,Y_2)$ 为 A 中的两条染色体. 其中，X_1,X_2,Y_1,Y_2 表示基因序列，并且可以为空. 相应地，$\hat{A}$ 中一定存在两条 X,Y 添加帽子后的染色体 $\hat{X}=(\hat{X}_1,\hat{X}_2)=(C_k,X_1,X_2,C_{k+1})$，$\hat{Y}=(\hat{Y}_1,\hat{Y}_2)=(C_l,Y_1,Y_2,C_{l+1})$. 其中，$\hat{X}_1=(C_k,X_1)$，$\hat{X}_2=(X_2,C_{k+1})$，$\hat{Y}_1=(C_l,Y_1)$，$\hat{Y}_2=(Y_2,C_{l+1})$. 显然，$\hat{X}_1,\hat{X}_2,\hat{Y}_1,\hat{Y}_2$ 都不为空，因而对 X,Y 做一次交互型移位、合并、分裂或分裂-合并一定等价于对 $\hat{X}$，$\hat{Y}$ 做一次交互型移位.

反之，设 ρ 为作用于 $\hat{A}$ 的染色体 $\hat{X}$，$\hat{Y}$ 的一个交互型移位. 用 $\rho_{pp}(\hat{X},\hat{Y},X_1X_2,Y_1Y_2)$ 表示在 $\hat{X}$ 的 X_1,X_2 和 $\hat{Y}$ 的 Y_1,Y_2 之间断开做 P-P 移位，用 $\rho_{ps}(\hat{X},\hat{Y},X_1X_2,Y_1Y_2)$ 表示在 $\hat{X}$ 的 X_1,X_2 和 $\hat{Y}$ 的 Y_1,Y_2 之间断开做 P-S 移位. 有如下情况：

(1) $\hat{X}$，$\hat{Y}$ 都不是空染色体. 设 $\hat{X}=(C_k,X_1,X_2,C_{k+1})$，$\hat{Y}=(C_l,Y_1,Y_2,C_{l+1})$，其中 X_1,X_2,Y_1,Y_2 为非空的基因序列. 相应地，A 中一定存在两条染色体 $X=(X_1,X_2)$，$Y=(Y_1,Y_2)$.

若 $\rho=\rho_{pp}(\hat{X},\hat{Y},X_1X_2,Y_1Y_2)$，则得到的新染色体为 (C_k,X_1,Y_2,C_{l+1}) 和 (C_l,Y_1,X_2,C_{k+1}). 显然，去掉帽子后，ρ 等价于 X,Y 的交互型移位 $\rho_{pp}(X,Y,X_1X_2,Y_1Y_2)$. 同样，$\rho_{ps}(\hat{X},\hat{Y},X_1X_2,Y_1Y_2)$ 等价于 $\rho_{ps}(X,Y,X_1X_2,Y_1Y_2)$.

若 $\rho=\rho_{pp}(\hat{X},\hat{Y},C_kX_1,C_lY_1)$，则得到新染色体 (C_l,X_1,X_2,C_{k+1}) 和 (C_k,Y_1,Y_2,C_{l+1}). 这时 ρ 只是在 $\hat{X},\hat{Y}$ 之间交换帽子 C_k,C_l，并没有改变染色体 X,Y. 我们将这种操作称作换帽. 同样，$\rho_{pp}(\hat{X},\hat{Y},X_2C_{k+1},Y_2C_{l+1})$，$\rho_{ps}(\hat{X},\hat{Y},C_kX_1,Y_2C_{l+1})$ 和 $\rho_{ps}(\hat{X},\hat{Y},X_2C_{k+1},C_lY_1)$ 都是换帽操作.

若 $\rho=\rho_{pp}(\hat{X},\hat{Y},C_kX_1,Y_2C_{l+1})$，得到新染色体 $(C_l,Y_1,Y_2,X_1,X_2,C_{k+1})$ 和 (C_k,C_{l+1}). 这时 ρ 等价于 X,Y 的合并. 同样，$\rho_{pp}(\hat{X},\hat{Y},X_2C_{k+1},C_lY_1)$，$\rho_{ps}(\hat{X},\hat{Y},C_kX_1,C_lY_1)$ 和 $\rho_{ps}(\hat{X},\hat{Y},X_2C_{k+1},Y_2C_{l+1})$ 都等价于 X,Y 的合并.

若 $\rho=\rho_{pp}(\hat{X},\hat{Y},X_1X_2,C_lY_1)$，则得到新染色体 $(C_k,X_1,Y_1,Y_2,C_{l+1})$ 和 (C_l,X_2,C_{k+1}). 这时 ρ 等价于 X,Y 的分裂-合并. 同样地，$\rho_{ps}(\hat{X},\hat{Y},X_1X_2,C_lY_1)$，$\rho_{pp}(\hat{X},\hat{Y},X_1X_2,Y_2C_{l+1})$，$\rho_{ps}(\hat{X},\hat{Y},X_1X_2,Y_2C_{l+1})$，$\rho_{pp}(\hat{X},\hat{Y},C_kX_1,Y_1Y_2)$，$\rho_{ps}(\hat{X},\hat{Y},C_kX_1,Y_1Y_2)$，$\rho_{pp}(\hat{X},\hat{Y},X_2C_{k+1},Y_1Y_2)$ 和 $\rho_{ps}(\hat{X},\hat{Y},X_2C_{k+1},Y_1Y_2)$ 都等价于 X,Y 的分裂-合并.

(2) $\hat{X}$ 不是空染色体，$\hat{Y}$ 是空染色体. 设 $\hat{X}=(C_k,X_1,X_2,C_{k+1})$，$\hat{Y}=(C_l,C_{l+1})$，其中 X_1,X_2 不为空. 相应地，A 中一定存在一条染色体 $X=(X_1,X_2)$.

若 $\rho=\rho_{pp}(\hat{X},\hat{Y},X_1X_2,C_lC_{l+1})$，则得到新染色体 (C_k,X_1,C_{l+1})，和 (C_l,X_2,C_{k+1}). 去掉帽子后，ρ 等价于 X 的分裂. 同样地，$\rho_{ps}(\hat{X},\hat{Y},X_1X_2,C_lC_{l+1})$ 也等价于 X 的分裂.

$\rho_{pp}(\hat{X},\hat{Y},C_kX_1,C_lC_{l+1})$，$\rho_{pp}(\hat{X},\hat{Y},X_2C_{k+1},C_lC_{l+1})$，$\rho_{ps}(\hat{X},\hat{Y},C_kX_1,C_lC_{l+1})$，$\rho_{ps}(\hat{X},\hat{Y},X_2C_{k+1},C_lC_{l+1})$ 都是换帽操作.

(3) $\hat{X}$，$\hat{Y}$ 都是空染色体. 设 $\hat{X}=(C_k,C_{k+1})$，$\hat{Y}=(C_l,C_{l+1})$.

$\rho_{pp}(\hat{X},\hat{Y},C_kC_{k+1},C_lC_{l+1})$ 和 $\rho_{ps}(\hat{X},\hat{Y},C_kC_{k+1},C_lC_{l+1})$ 都是换帽操作.

根据以上讨论可知，A 的交互型移位、分裂、合并、分裂-合并都对应于 $\hat{A}$ 的一个交互型移位，$\hat{A}$ 的交互型移位则对应一个换帽操作或者 A 的一个交互型移位、分裂、合并或分裂-合并. 下面我们证明 A 和 B 的移位距离 $d(A,B)$ 恰好等于 $\hat{A}$ 与 B 的所有加帽基因组的交互型移位距离的最小值.

定理 2. $d(A,B)=\min\limits_{\hat{B}\in\Gamma}d_r(\hat{A},\hat{B})$.

证明. 对 A 做移位 ρ 后得到的新基因组记为 $A\cdot\rho$. 假设 A 转化为 B 的最短移位序列为 $\rho_1,\rho_2,\cdots,\rho_k$，即 $A\cdot\rho_1\cdot\rho_2\cdot\cdots\cdot\rho_k=B$. 对于 $1\leq i\leq k$，ρ_i 都对应基因组 $\hat{A}$ 的一个交互型移位 $\hat{\rho}_i$，即 $\hat{A}\cdot\hat{\rho}_1\cdot\hat{\rho}_2\cdot\cdots\cdot\hat{\rho}_k=\hat{B}$，$\hat{B}\in\Gamma$. 因而 $d(A,B)=k\geq d_r(\hat{A},\hat{B})\geq\min\limits_{\hat{B}\in\Gamma}d_r(\hat{A},\hat{B})$.

设 $d_r(\hat{A},\hat{B}_{OPT})=\min\limits_{\hat{B}\in\Gamma}d_r(\hat{A},\hat{B})$，假设 $\hat{A}$ 转化为 $\hat{B}_{OPT}$ 的最短交互型移位序列为 $\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_l$，即 $\hat{A}\cdot$

$\hat{\rho}_1 \cdot \hat{\rho}_2 \cdot \cdots \cdot \hat{\rho}_t = \hat{B}_{OPT}$. 对于 $1 \leq i \leq t$，若 $\hat{\rho}_i$ 不是换帽操作，则对应 A 的一个移位. 设 $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_t$ 中的非换帽操作数为 s，将这 s 个操作重新编号为 $\hat{\rho}'_1, \hat{\rho}'_2, \cdots, \hat{\rho}'_s$，$s \leq t$. 对于 $1 \leq i \leq s$，$\hat{\rho}'_i$ 对应 A 的一个移位 ρ_i，即 $A \cdot \rho_1 \cdot \rho_2 \cdot \cdots \cdot \rho_s = B$. 因而 $d(A, B) \leq s \leq t = \min_{\hat{B} \in \Gamma} d_r(\hat{A}, \hat{B})$.　　证毕.

根据定理 1 和定理 2，可用如下方法计算 A 和 B 的移位距离与最短移位序列：针对 B 的加帽基因组集合 Γ 中的每个基因组 $\hat{B}$，使用解答 SRT 问题的算法计算 $d_r(\hat{A}, \hat{B})$，选择使 $d_r(\hat{A}, \hat{B})$ 达到最小的加帽基因组 $\hat{B}_{OPT}$，则可由 $\hat{A}$ 转化为 $\hat{B}_{OPT}$ 的交互型移位距离与最短交互型移位序列，得到由 A 转化为 B 的移位距离与最短移位序列. 这种计算方法关于 n 是指数时间的. 下面我们将与 $\hat{A}$ 的交互型移位距离最小的 B 的加帽基因组总记作 $\hat{B}_{OPT}$. 我们将在断点图的基础上构造部分图，利用部分图的性质和定理 1 的交互型移位距离公式，设计一个在 n 的多项式时间内计算 $\hat{B}_{OPT}$ 的算法，从而给出一般移位排序的多项式算法.

4　基因组移位排序算法

4.1　部分图

在 Γ 中任选一个 B 的加帽基因组 $\hat{B}$，构造断点图 $G(\hat{A}, \hat{B})$. 在 $G(\hat{A}, \hat{B})$ 中，每个帽子都对应一个度为 0 的点和一个度为 2 的点. 将对应于帽子的度为 2 的点称作 A-cap. 每个 A-cap 都通过灰边与另一个 A-cap 或 B 的尾基因的点邻接. 若 A-cap 通过灰边与 A-cap 邻接，则表示 $\hat{B}$ 中的一条空染色体. 将与 A-cap 邻接的 B 的尾基因的点称作 B-tail. 在 $G(\hat{A}, \hat{B})$ 中去掉以 A-cap 为端点的灰边，得到的图称作 $G(\hat{A}, \hat{B})$ 的部分图. 将这个部分图记作 $G^P(A, B)$. 例如，给定基因组 $A = \{(2,1,3,5), (6,8,-7,4,10,-9)\}$，$B = \{(1,2,3,4), (5,6,7), (8,9,10)\}$，断点图 $G(\hat{A}, \hat{B})$ 如图 1(a) 所示，部分图 $G^P(A, B)$ 如图 1(b) 所示.

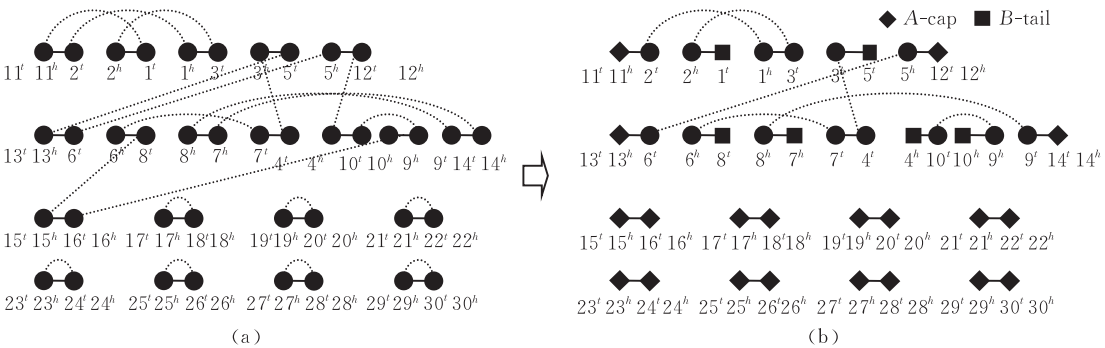


图 1　部分图

$G^P(A, B)$ 与构造断点图时选择哪个 B 的加帽基因组无关. $G^P(A, B)$ 由黑边和灰边交错的圈和路径构成. 若一条路径只有一条黑边，则称之为短路径，否则称作长路径. 给 $G^P(A, B)$ 添加 $n+N$ 条灰边可将其变为完整的断点图，其中 $n-N$ 条灰边连接两个 A-cap，$2N$ 条灰边连接一个 A-cap 和一个 B-tail. 每种添加灰边的方式都确定了一个 B 的加帽基因组. 根据定理 2，要计算 $\hat{B}_{OPT}$，只需找一种添加灰边的方法，使得 $G^P(A, B)$ 添加灰边后得到的断点图的交互型移位距离最小.

将 $G^P(A, B)$ 中的路径分类：两端都为 A-cap 的路径为 AA 路径；一端为 A-cap 一端为 B-tail 的路径为 AB 路径；两端都为 B-tail 的路径为 BB 路径. $G^P(A, B)$ 中的 AA 路径数与 BB 路径数之差为 $n-N$. $G^P(A, B)$ 中圈和路径的总数记作 $l(A, B)$，BB 路径数记作 $p(A, B)$.

4.2　半最小子排列

设 I 为 $\hat{A}$ 中的一段基因序列，若点集 $IN(I)$ 支撑的 $G^P(A, B)$ 的子图满足以下条件，则称该子图为一个真子排列：(1) 不存在边 (u, v) 使 $u \in IN(I)$，$v \notin IN(I)$；(2) 至少包含一个长圈；(3) 不包含路径. 若一个真子排列不包含其它真子排列，则称之为真最小子排列，简记为 RMSP (Real Minimum Sub Permutation). 若 $IN(I)$ 支撑的 $G^P(A, B)$ 的子图满足以下条件，则称该子图为一个半子排列：(1) 不存在边 (u, v) 使 $u \in IN(I)$，$v \notin IN(I)$；(2) 至少包含一个长圈或长路径；(3) 不包含 AA 路径和 BB 路径. 若一个半子排列不包含其它半子排列和真子排列，则称之为半最小子排列，简记为 SMSP (Semi Minimum Sub Permuation).

例如,图 1(b)所示部分图中由点集$\{11^h, 2^t, 2^h, 1^t, 1^h, 3^t\}$支撑的子图是一个 SMSP. SMSP 和 RMSP 的区别在于:SMSP 一定包含 AB 路径,而 RMSP 不包含任何路径. 若将一个 SMSP 中的 AB 路径都用灰边闭合,则该 SMSP 变为 RMSP. 将 $G^P(A,B)$中的 SMSP 数记作 $sm(A,B)$,RMSP 数记作 $r(A,B)$.

设 S_1, S_2 是 $G^P(A,B)$中的两个 SMSP,若(1) S_1和S_2位于同一半子排列中;(2) S_1和 S_2所在的染色体上没有 RMSP,则称 S_1, S_2 相互依赖. 例如,图 2(a)中S_1, S_2相互依赖. 设P_1, P_2分别是 $G^P(A,B)$中的 AA 路径和 BB 路径,若(1) P_1和 P_2的所有顶点均位于同一染色体 X;(2)除 P_1和 P_2外,X 上没有其它路径;(3) X 上没有 RMSP 或体间灰边的端点,则称 P_1, P_2相互依赖. 例如,图 2(b)中 P_1和 P_2相互依赖.

图 2 相互依赖的 SMSP 和路径

性质 1. 设 S_1和 S_2为两个相互依赖的 SMSP,添加一条灰边连接 S_1中 AB 路径的 A-cap 和 S_2中 AB 路径的B-tail,则 S_1和 S_2合并为一个新的 SMSP.

性质 2. 设 P_1为一条 AA 路径,P_2为一条 BB 路径,P_1和 P_2相互依赖. 若添加一条灰边连接 P_1和 P_2,则会产生一个新的 SMSP.

引入参数$ds(A,B)$. 若$G^P(A,B)$满足:$sm(A,B)=2$ 且这两个 SMSP 相互依赖,则定义 $ds(A,B)=1$;否则定义 $ds(A,B)=0$.

引理 1. 若 $G^P(A,B)$满足 $ds(A,B)=0$,则总能添加$\lfloor sm(A,B)/2 \rfloor$ 条灰边,使 $G^P(A,B)$中减少 $2\times\lfloor sm(A,B)/2 \rfloor$个 SMSP.

证明. 在 $G^P(A,B)$中添加$\lfloor sm(A,B)/2 \rfloor$ 条灰边减少 $2\times\lfloor sm(A,B)/2 \rfloor$个 SMSP 的方法如下:将 $G^P(A,B)$中相互依赖的 SMSP 对的数目记作 v. 若 $v=1$,可知除了相互依赖的一对 SMSP 外,$G^P(A,B)$中还存在其它 SMSP(否则 $ds(A,B)=1$). 在相互依赖的一对 SMSP 中取一个 SMSP,记作 S_1. 另取一个 SMSP,记作 S_2. 添加一条灰边将 S_1中的 AB 路径的 A-cap 与 S_2中 AB 路径的 B-tail 连接. 若 $v>1$,将相互依赖的 SMSP 对记为 PS_1, $PS_2, \cdots, PS_v$. 对于 $1\leq i\leq v-1$,添加一条灰边连接 PS_i中一个 SMSP 的 AB 路径的 A-cap 与 PS_{i+1}中一个 SMSP 的 AB 路径的 B-tail. 这时图中任意两个 SMSP 都不相互依赖. 任取两个 SMSP,添加灰边连接一个 SMSP 的 AB 路径的 A-cap 与另一个 SMSP 的 AB 路径的 B-tail,直到图中的 SMSP 数小于 2 为止.

上述过程中每次添加灰边所连接的两个 SMSP 都不相互依赖,因而每次都会减少 2 个 SMSP,因此总共减少了 $2\times\lfloor sm(A,B)/2 \rfloor$个 SMSP . 证毕.

引理 2. 总能添加$p(A,B)$条灰边将$G^P(A,B)$中的每一条 BB 路径都与一条 AA 路径连接,且添加灰边后不产生新的 SMSP.

证明. 在 $G^P(A,B)$中添加 $p(A,B)$条灰边的方法如下:将 $G^P(A,B)$中相互依赖的路径对的数目记作 w. 若 $w=1$,可知除了相互依赖的路径对中的 AA 路径外,$G^P(A,B)$中一定还存在其它 AA 路径. 添加一条灰边将相互依赖的路径对中的 BB 路径与路径对外的一条 AA 路径连接. 若$w>1$,将$G^P(A,B)$中相互依赖的路径对记为 $PP_1, PP_2, \cdots, PP_w$. 对于 $1\leq i\leq w-1$,添加一条灰边连接 PP_i中的 BB 路径与 PP_{i+1}中的 AA 路径. 这时图中任意一对 AA 路径和 BB 路径都不相互依赖. 继续添加灰边将其余的 BB 路径与任意 AA 路径连接,直到图中不存在 BB 路径为止.

上述过程共添加了 $p(A,B)$条灰边,且每条灰边连接的 BB 路径与 AA 路径都不相互依赖,因而不会产生新的 SMSP. 证毕.

4.3 计算移位距离

在 $G^P(A,B)$中添加灰边后形成的断点图可能包含偶隔离带,下面考察可能导致偶隔离带形成的部分图结构.

(1)若 $G^P(A,B)$中有偶数个 RMSP,且所有的 RMSP 都被同一真子排列包含,则称 $G^P(A,B)$包含一个真偶隔离带. 图 3(a)所示结构就是一个真偶隔离带.

(2)若 $G^P(A,B)$中有偶数个 RMSP,所有的 RMSP 不能被同一真子排列包含,但都被同一个半子排列包含,则称 $G^P(A,B)$包含一个半偶隔离带. 图 3(b)所示结构为一个半偶隔离带.

半偶隔离带一定包含一条长的 AB 路径,而真

偶隔离带不包含任何路径. 若半偶隔离带中的 AB 路径都闭合，则变为真偶隔离带.

(3) 若 $G^P(A,B)$ 包含真偶隔离带，$sm(A,B)\geq 2$，并且有两个 SMSP 与所有的 RMSP 都被同一个半子排列所包含，则称 $G^P(A,B)$ 包含强偶隔离带. 图 3(c)所示结构为一个强偶隔离带.

(4) 若 $G^P(A,B)$ 有奇数个 RMSP，所有的 RMSP 都位于同一条染色体 X，并被同一个真子排列或半子排列包含. 若 X 上存在一个 SMSP，该 SMSP 与所有的 RMSP 都被同一个半子排列所包含，则称 $G^P(A,B)$ 包含弱偶隔离带.

若将弱偶隔离带中的 SMSP 闭合为 RMSP，则它会变为真偶隔离带或半偶隔离带. 例如，图 3(d)和(e)所示的两种结构都是弱偶隔离带，(d)中的 SMSP 闭合后变为真偶隔离带，(e)中的 SMSP 闭合后变为半偶隔离带.

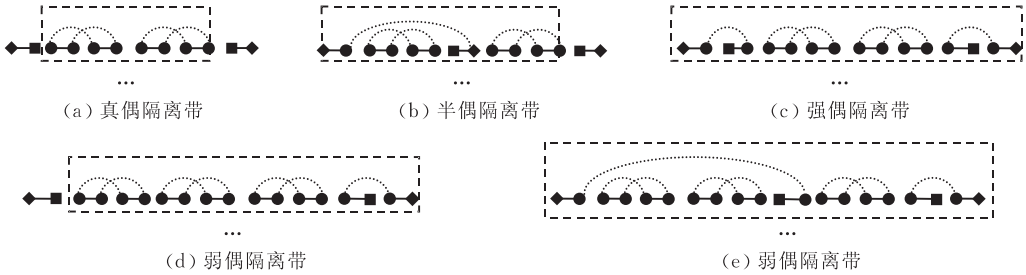


图 3　可能导致偶隔离带形成的部分图结构

注意，$G^P(A,B)$ 最多只能包含一个真偶隔离带、半偶隔离带、强偶隔离带或弱偶隔离带.

引入参数 $o(A,B)$ 和 $\delta(A,B)$. $o(A,B)$ 的取值方法定义为：如果 $r(A,B)$ 为奇数，则 $o(A,B)=1$；否则 $o(A,B)=0$. $\delta(A,B)\in\{0,1,2\}$，$\delta(A,B)$ 的取值方法较复杂，我们逐条给出如下：

若 $G^P(A,B)$ 满足下列条件之一，则 $\delta(A,B)=2$.

(α1) 包含真偶隔离带，且 $sm(A,B)=0$；

(α2) 包含强偶隔离带，且 $sm(A,B)=2$.

若 $G^P(A,B)$ 满足下列条件之一，则 $\delta(A,B)=1$.

(β1) 包含强偶隔离带，$sm(A,B)>2$；

(β2) 包含真偶隔离带，不包含强偶隔离带，$sm(A,B)>0$；

(β3) 包含弱偶隔离带，且 $sm(A,B)=1$；

(β4) 包含半偶隔离带，且 $sm(A,B)$ 为偶数；

(β5) 包含半偶隔离带，$sm(A,B)=1$，且唯一的 SMSP 与半偶隔离带被同一半子排列包含；

(β6) $ds(A,B)=1$ 且 $o(A,B)=0$.

其它情况下，$\delta(A,B)=0$.

下面在表示部分图 $G^P(A,B)$ 满足 $\delta(A,B)$ 取值条件时，直接使用前面给出的条件编号. 我们将给出移位距离的精确下界，并证明移位距离可以达到这个下界，从而给出基因组一般移位排序算法.

假设在 $G^P(A,B)$ 中添加 $n+N$ 条灰边 $g_1, g_2, \cdots, g_{n+N}$ 得到 $G(\hat{A},\hat{B})$，即 $G^P(A,B)=G_0\xrightarrow{g_1}G_1\xrightarrow{g_2}\cdots\xrightarrow{g_{n+N}}G_{n+N}=G(\hat{A},\hat{B})$. G_i 的各项参数 l_i、p_i、r_i、sm_i、ds_i、o_i、δ_i 与在 $G^P(A,B)$ 中的定义相同. 对于参数 ϕ，令 $\Delta\phi_i=\phi_i-\phi_{i-1}$，设

$$\Delta_i=(l_i-p_i-r_i-\lceil(sm_i+o_i)/2\rceil-\delta_i)-(l_{i-1}-p_{i-1}-r_{i-1}-\lceil(sm_{i-1}+o_{i-1})/2\rceil-\delta_{i-1}).$$

引理 3.　对于 i，$1\leq i\leq n+N$，任何情况下添加灰边 g_i 都满足 $\Delta\delta_i\geq -1$.

证明.　因为 $\delta\in\{0,1,2\}$，若 $\delta_{i-1}\in\{0,1\}$，显然有 $\Delta\delta_i\geq -1$，因而只需证明当 $\delta_{i-1}=2$ 时 $\delta_i>0$ 即可. 当 $\delta_{i-1}=2$ 时，有以下两种情况：

(1) G_{i-1} 满足条件(α1). 若 g_i 连接一对相互依赖的 AA 路径和 BB 路径，则会产生一个 SMSP，这时 G_i 满足条件(β2)，$\delta_i=1$. 其它情况下 g_i 都不会增加 SMSP，G_i 仍满足条件(α1)，$\delta_i=2$.

(2) G_{i-1} 满足条件(α2). 若 g_i 将一个 SMSP 变为 RMSP，则 G_i 满足条件(β3)，$\delta_i=1$；若 g_i 将两个 SMSP 中的 AB 路径连接，则这两个 SMSP 都被破坏，$sm_i=0$，因而 G_i 满足条件(α1)，$\delta_i=2$；若 g_i 将一个 SMSP 中的 AB 路径与其它路径连接，则会减少一个 SMSP. 这时 G_i 满足条件(β2)，$\delta_i=1$；若 g_i 连接一对相互依赖的 AA 路径和 BB 路径，则会产生一个新的 SMSP. 这时 G_i 满足条件(β1)，$\delta_i=1$. 其它情况下 g_i 都不会增加或减少 SMSP，G_i 仍满足(α2)，$\delta_i=2$.

上述情况都有 $\delta_i>0$，因而结论得证.　　证毕.

引理 4.　$c(\hat{A},\hat{B})-s(\hat{A},\hat{B})-f(\hat{A},\hat{B})\leq l(A,B)-p(A,B)-r(A,B)-\lceil(sm(A,B)+o(A,B))/2\rceil-\delta(A,B)$.

证明. 只需证明对于 $i,1\leq i\leq n+N$,都有 $\Delta_i\leq 0$,则结论得证.这是因为 $c(\hat{A},\hat{B})=l_{n+N}$, $s(\hat{A},\hat{B})=r_{n+N}$, $p_{n+N}=sm_{n+N}=0$, $f(\hat{A},\hat{B})=o_{n+N}+\delta_{n+N}=\lceil(sm_{n+N}+o_{n+N})/2\rceil+\delta_{n+N}$,因而 $c(\hat{A},\hat{B})-s(\hat{A},\hat{B})-f(\hat{A},\hat{B})=l_{n+N}-p_{n+N}-r_{n+N}-\lceil(sm_{n+N}+o_{n+N})/2\rceil-\delta_{n+N}\leq l_0-p_0-r_0-\lceil(sm_0+o_0)/2\rceil-\delta_0=l(A,B)-p(A,B)-r(A,B)-\lceil(sm(A,B)+o(A,B))/2\rceil-\delta(A,B)$.在下面的证明中,对于固定的 i,略去参数下标"i".根据 g_i 分为下述多种情况讨论.

情况 1. g_i 闭合一条 AB 路径.若 g_i 不改变 SMSP 的数目,则 $\Delta l=\Delta p=\Delta r=\Delta o=\Delta sm=0$.但 g_i 有可能将半偶隔离带变为真偶隔离带,因而 $\Delta\delta\geq 0$,显然 $\Delta\leq 0$;若 g_i 将一个 SMSP 变为 RMSP,则 $\Delta l=\Delta p=0$, $\Delta r=1$, $\Delta sm=-1$. Δo 可能为 1 或 -1,下面分别讨论:

情况 1.1. 若 $\Delta o=1$,则 $\Delta\lceil(sm+o)/2\rceil=0$,又由引理 3 可知 $\Delta\delta\geq -1$,因而 $\Delta\leq 0$.

情况 1.2. 若 $\Delta o=-1$,则 $\Delta\lceil(sm+o)/2\rceil=-1$.若要证明 $\Delta\leq 0$,只需证明 $\Delta\delta\geq 0$ 即可.若 $\delta_{i-1}=0$,显然 $\Delta\delta\geq 0$.由 $\Delta r=1$ 和 $\Delta o=-1$ 可知 r_{i-1} 为奇数,因而 $\delta_{i-1}\neq 2$.若 $\delta_{i-1}=1$,由 r_{i-1} 为奇数可知 G_{i-1} 只能满足条件(β3).添加 g_i 后,唯一的 SMSP 变为 RMSP,G_i 包含真偶隔离带或半偶隔离带且 $sm_i=0$,因而满足条件(α1)或(β4),显然 $\Delta\delta\geq 0$.上述情况都满足 $\Delta\delta\geq 0$,因而 $\Delta\leq 0$.

情况 2. g_i 连接一条 AB 路径的 A-cap 和另一条 AB 路径的 B-tail,$\Delta l=-1$, $\Delta p=\Delta r=\Delta o=0$. g_i 最多可减少两个 SMSP,因而 Δsm 可能为 0、-1 或 -2,下面分别讨论:

情况 2.1. 若 $\Delta sm=0$,则 $\Delta\lceil(sm+o)/2\rceil=0$.又由引理 3 可知 $\Delta\delta\geq -1$,因而 $\Delta\leq 0$.

情况 2.2. 若 $\Delta sm=-1$,则 $\Delta\lceil(sm+o)/2\rceil=0$ 或 -1.若 $\Delta\lceil(sm+o)/2\rceil=0$,显然 $\Delta\leq 0$.若 $\Delta\lceil(sm+o)/2\rceil=-1$,要证明 $\Delta\leq 0$,只需证明 $\Delta\delta\geq 0$ 即可.若 $\delta_{i-1}=0$,显然 $\Delta\delta\geq 0$.由 $\Delta\lceil(sm+o)/2\rceil=-1$ 和 $\Delta(sm+o)=-1$ 可知 $sm_{i-1}+o_{i-1}$ 为奇数,因而 $\delta_{i-1}\neq 2$.若 $\delta_{i-1}=1$,由 $sm_{i-1}+o_{i-1}$ 为奇数可知 G_{i-1} 只能满足条件(β1)、(β2)或(β5).若 G_{i-1} 满足条件(β1)或(β2),减少一个 SMSP 后,G_i 可能满足条件(β1)、(β2)、(α1)或(α2),显然 $\Delta\delta\geq 0$;若 G_{i-1} 满足条件(β5),减少一个 SMSP 后,G_i 满足(β4),$\Delta\delta=0$.以上情况都满足 $\Delta\delta\geq 0$,因而 $\Delta\leq 0$.

情况 2.3. 若 $\Delta sm=-2$,则 $\Delta\lceil(sm+o)/2\rceil=-1$.若要证明 $\Delta\leq 0$,只需证明 $\Delta\delta\geq 0$.若 $\delta_{i-1}=0$,显然 $\Delta\delta\geq 0$.若 $\delta_{i-1}=1$,又由 $\Delta sm=-2$ 可知 $sm_{i-1}\geq 2$,因而 G_{i-1} 可能满足条件(β1)、(β2)或(β4).若 G_{i-1} 满足条件(β1)或(β2),减少两个 SMSP 后,G_i 可能满足条件(β1)、(β2)、(α1)或(α2),显然 $\Delta\delta\geq 0$;若 G_{i-1} 满足(β4),减少两个 SMSP 后,G_i 仍然满足(β4),$\Delta\delta=0$.若 $\delta_{i-1}=2$,由 $sm_{i-1}\geq 2$ 可知 G_i 满足条件(α2),减少两个 SMSP 后 G_i 满足条件(α1),$\Delta\delta=0$.以上情况都有 $\Delta\delta\geq 0$,因而 $\Delta\leq 0$.

情况 3. g_i 连接一条 AA 路径和一条 BB 路径.若这两条路径不相互依赖,则 SMSP 的数目不变,$\Delta l=\Delta p=-1$, $\Delta r=\Delta o=\Delta sm=0$.但 g_i 有可能产生半偶隔离带,因而 $\Delta\delta\geq 0$,可得 $\Delta\leq 0$.若这两条路径相互依赖,则会产生一个新的 SMSP, $\Delta l=\Delta p=-1$, $\Delta r=\Delta o=0$, $\Delta sm=1$. $\Delta\lceil(sm+o)/2\rceil$ 可能为 0 或 1,下面分别讨论:

情况 3.1. 若 $\Delta\lceil(sm+o)/2\rceil=1$,又由引理 3 可知 $\Delta\delta\geq -1$,因而 $\Delta\leq 0$.

情况 3.2. 若 $\Delta\lceil(sm+o)/2\rceil=0$.要证明 $\Delta\leq 0$,只需证明 $\Delta\delta\geq 0$.若 $\delta_{i-1}=0$,显然 $\Delta\delta\geq 0$.由 $\Delta\lceil(sm+o)/2\rceil=0$ 和 $\Delta(sm+o)=1$ 可知 $sm_{i-1}+o_{i-1}$ 为奇数,因而 $\delta_{i-1}\neq 2$.若 $\delta_{i-1}=1$,由 $sm_{i-1}+o_{i-1}$ 为奇数可知 G_{i-1} 可能满足条件(β1)、(β2)或(β5).若 G_i 满足条件(β1)或(β2),增加一个 SMSP 后 G_i 仍然满足条件(β1)或(β2),$\Delta\delta=0$;若 G_i 满足条件(β5),增加一个 SMSP 后,G_i 满足条件(β4),$\Delta\delta=0$.上述情况都有 $\Delta\delta\geq 0$,因而 $\Delta\leq 0$.

情况 4. g_i 连接一条 AB 路径的 B-tail 和一条 AA 路径(或 g_i 连接一条 AB 路径的 A-cap 和一条 BB 路径,或 g_i 连接一条 AB 路径的 A-cap 和一条 AA 路径). g_i 最多减少一个 SMSP,$\Delta l=-1$, $\Delta p=\Delta r=\Delta o=0$, Δsm 可能为 0 或 -1,根据情况 2.1 和情况 2.2 可知 $\Delta\leq 0$.

情况 5. g_i 闭合一条 AA 路径.所有参数保持不变,$\Delta=0$.

情况 6. g_i 连接两条 AB 路径的 A-cap. g_i 最多减少两个 SMSP,$\Delta l=-1$, $\Delta p=1$, $\Delta r=0$, $\Delta sm\geq -2$, $\Delta o=0$,又由引理 3 可知 $\Delta\delta\geq -1$,因而 $\Delta\leq 0$.

情况 7. g_i 连接两条 AA 路径. $\Delta l=-1$, $\Delta p=\Delta r=\Delta sm=\Delta o=\Delta\delta=0$, $\Delta=-1$. 证毕.

若添加到 G_{i-1} 中的一条灰边 g_i 满足 $\Delta_i=0$,则将 g_i 称作有效灰边.下面我们证明总能找到 $n+N$

条有效灰边依次添加到$G^P(A,B)$中，将其变为断点图$G(\hat{A},\hat{B}_{OPT})$. 在下面引理和定理的证明中，对于固定的“i”，略去参数下标.

引理 5.　设S为G_{i-1}中的一个 SMSP 且S包含 2 条AB路径. 若g_i为闭合S中一条AB路径的灰边，那么g_i一定是有效灰边.

证明.　因为S中包含两条AB路径，闭合其中一条AB路径后，S在G_i中仍然是 SMSP，即G_i中 SMSP 的数目不变，其它参数也不变，因而$\Delta=0$.

证毕.

根据引理 5，下面可假设G_{i-1}中的 SMSP 都只包含一条AB路径.

引理 6.　若G_{i-1}满足$ds_{i-1}=1$，则总能找到一条有效灰边g_i，使得添加g_i后的图G_i满足$ds_i=0$.

证明.　因为$ds_{i-1}=1$，可知G_{i-1}中存在一对相互依赖的 SMSP，记作S_1,S_2. 下面给出找满足上述条件的有效灰边g_i的方法：

(1) 若$o_{i-1}=1$，则$\delta_{i-1}=0$. 令g_i为闭合S_1(或S_2)中的AB路径的灰边，则S_1(或S_2)变为 RMSP，$ds_i=0$. 因为新的 RMSP 与原有的 RMSP 位于不同的染色体，因而$\delta_i=0$. g_i令$\Delta l=\Delta p=0,\Delta r=1,\Delta o=\Delta sm=-1,\Delta\delta=0$，因而$\Delta=0$.

(2) 若$o_{i-1}=0$，则G_{i-1}满足条件(β6)，$\delta_{i-1}=1$.

若$r_{i-1}>0$，令g_i为闭合S_1(或S_2)的AB路径的灰边，则S_1(或S_2)变为 RMSP 且该 RMSP 与原有的 RMSP 位于不同的染色体，$ds_i=\delta_i=0$. $\Delta l=\Delta p=0,\Delta r=\Delta o=1,\Delta sm=\Delta\delta=-1,\Delta=0$.

若$r_{i-1}=0$，令g_i为连接S_1中的A-cap 和S_2中的B-tail 的灰边. g_i将S_1,S_2合并为一个 SMSP，$ds_i=\delta_i=0$. $\Delta l=-1,\Delta p=\Delta r=\Delta o=0,\Delta sm=\Delta\delta=-1,\Delta\lceil(sm+o)/2\rceil=0,\Delta=0$.　证毕.

引理 7.　若G_{i-1}包含真偶隔离带且$sm_{i-1}>0$，则总能找到一条有效灰边g_i，使得添加g_i后的图G_i满足：(1) $ds_i=0$；(2) G_i不包含真偶隔离带或者G_i满足条件(α1).

证明.　G_{i-1}可能满足条件(α2)、(β1)或(β2). 下面对这 3 种情况分别给出找满足上述条件的有效灰边g_i的方法：

(1) 若G_{i-1}满足(α2)，设强偶隔离带中的两个 SMSP 为S_1,S_2. 令g_i为连接S_1的AB路径的A-cap 和S_2的AB路径的B-tail 的灰边. 添加g_i后，S_1,S_2不再是 SMSP. 显然$ds_i=0$并且G_i满足条件(α1). $\Delta l=-1,\Delta p=\Delta r=\Delta o=0,\Delta sm=-2,\Delta\delta=0$，因而$\Delta=0$.

(2) 若G_{i-1}满足(β1)或(β2)，用以下方法选择一个 SMSP，记作S：若G_{i-1}中存在一对相互依赖的 SMSP，则令S为其中一个 SMSP；否则，令S为任意一个 SMSP. 令g_i为闭合S中AB路径的灰边，则S变为 RMSP，$\delta_i=0$. 上述选择S的方式可以保证$ds_i=0$. r_i为奇数，显然G_i不包含真偶隔离带. $\Delta l=\Delta p=0,\Delta r=\Delta o=1,\Delta sm=\Delta\delta=-1$，因而$\Delta=0$.

证毕.

引理 8.　若G_{i-1}包含弱偶隔离带且G_{i-1}中至少有一条AA路径，则总能找到一条有效灰边g_i，使得添加g_i后的图G_i满足：(1) $ds_i=0$；(2) G_i不包含真偶隔离带，弱偶隔离带或半偶隔离带.

证明.　下面分 3 种情况给出找满足上述条件的有效灰边g_i的方法：

(1) 若$sm_{i-1}=1$，可知G_{i-1}满足条件(β3)，$\delta_{i-1}=1$. 设S为G_{i-1}中唯一的 SMSP. 令g_i为连接S的AB路径的B-tail 与一条AA路径的灰边. 添加g_i后S不再是 SMSP，显然$ds_i=sm_i=0$且G_i不再包含弱偶隔离带. 因r_i为奇数，因而G_i也不包含真偶隔离带或半偶隔离带. $\Delta l=-1,\Delta p=\Delta r=\Delta o=0,\Delta sm=\Delta\delta=-1$，可得$\Delta=0$.

(2) 若$sm_{i-1}=2$，则$\delta_{i-1}=0$. 设这两个 SMSP 为S_1,S_2. 令g_i为连接S_1的AB路径的A-cap 和S_2的AB路径的B-tail 的灰边. 添加g_i后S_1和S_2不再是 SMSP，显然$ds_i=sm_i=0$且G_i不包含真偶隔离带、弱偶隔离带或半偶隔离带. $\Delta l=-1,\Delta p=\Delta r=\Delta o=0,\Delta sm=-2,\Delta\delta=0,\Delta=0$.

(3) 若$sm_{i-1}>2$，则G_{i-1}中至少存在一个与弱偶隔离带不同染色体的 SMSP. 用以下方法选择一个 SMSP，记作S：若G_{i-1}中存在一对相互依赖的 SMSP，则令S为其中一个 SMSP；否则，令S为任意一个与弱偶隔离带不同染色体的 SMSP. 令g_i为闭合S的AB路径的灰边，g_i将S变为 RMSP. 选择S的方式可以保证$ds_i=0$. 在G_i中新的 RMSP 与其它 RMSP 不位于同一染色体，因而G_i不包含真偶隔离带、弱偶隔离带或半偶隔离带. $\Delta l=\Delta p=0,\Delta r=1,\Delta o=\Delta sm=-1,\Delta\delta=0,\Delta=0$.　证毕.

引理 9.　若G_{i-1}包含半偶隔离带，$ds_{i-1}=0$且G_{i-1}中至少有一条AA路径，则总能找到一条有效灰边g_i，使得添加g_i后的图G_i满足：(1) $ds_i=0$；(2) G_i中不包含真偶隔离带或半偶隔离带.

证明.　设P为半偶隔离带中的长AB路径.

找满足上述条件有效灰边 g_i 的方法如下：

(1) 若 sm_{i-1} 为偶数，则 G_{i-1} 满足条件(β4)，$\delta_{i-1}=1$. 令 g_i 为连接 P 的 B-tail 与一条 AA 路径的灰边. SMSP 的数目不变，因而 $ds_i=ds_{i-1}=0$. 因为一条染色体上只有两个 A-cap，因而添加 g_i 后，所有的 RMSP 一定不被同一半子排列包含，即 G_i 不包含半偶隔离带. $\Delta l=-1$，$\Delta p=\Delta r=\Delta o=\Delta sm=0$，$\Delta\delta=-1$，因而 $\Delta=0$.

(2) 若 sm_{i-1} 为奇数，下面分两种子情况讨论：

(2.1) 若 $sm_{i-1}=1$，设 S 为 G_{i-1} 中唯一的 SMSP.

若 G_{i-1} 满足条件(β5)，则 $\delta_{i-1}=1$. 令 g_i 为闭合 S 中的 AB 路径的灰边. 添加 g_i 后 S 变为 RMSP，r_i 为奇数且 $sm_i=0$，因而 $ds_i=0$ 且 G_i 不包含半偶隔离带. g_i 满足 $\Delta l=\Delta p=0$，$\Delta r=\Delta o=1$，$\Delta sm=\Delta\delta=-1$，$\Delta=0$. 若 G_{i-1} 不满足条件(β5)，则 $\delta_{i-1}=0$. 令 g_i 为连接 S 中 AB 路径的 A-cap 与 P 的 B-tail 的灰边. 添加 g_i 后 $sm_i=ds_i=0$. 由于在 G_{i-1} 中半偶隔离带与 S 不被同一个半子排列包含(否则，G_{i-1} 满足条件(β5))，因而添加 g_i 后所有的 RMSP 不再被同一半子排列包含，即 G_i 不包含半偶隔离带. $\Delta l=-1$，$\Delta p=\Delta r=\Delta o=0$，$\Delta sm=-1$，$\Delta\delta=0$，可得 $\Delta=0$.

(2.2) 若 $sm_{i-1}>1$，则 $\delta_{i-1}=0$. 这时 G_{i-1} 中一定存在一个与半偶隔离带不同染色体的 SMSP，按以下方法选择一个 SMSP，记为 S：若 G_{i-1} 中有一对相互依赖的 SMSP，则令 S 为 SMSP 对中的一个 SMSP；否则，令 S 为任意一个与半偶隔离带不同染色体的 SMSP. 令 g_i 为连接 S 中 AB 路径的 A-cap 与 P 的 B-tail 的灰边. 显然 g_i 是体间灰边，因而 G_i 不再包含半偶隔离带. 选择 S 的方法可以保证 $ds_i=0$. $\Delta l=-1$，$\Delta p=\Delta r=\Delta o=0$，$\Delta sm=-1$，$\Delta\delta=0$，可得 $\Delta=0$.

上述情况中 g_i 均没有闭合路径 P，因而 G_i 也不会包含真偶隔离带. 证毕.

引理 10. 总能找到 $p(A,B)$ 条灰边将 $G^P(A,B)$ 中的每一条 BB 路径都与一条 AA 路径连接，添加灰边后不产生新的 SMSP，也不产生半偶隔离带.

证明. 由引理 2 可知，存在 $p(A,B)$ 条灰边将每一条 BB 路径都与一条 AA 路径连接，不会产生新的 SMSP. 下面我们修改引理 2 的添加灰边的方法，使之也不产生半偶隔离带. 按照引理 2 添加灰边产生半偶隔离带只有一种情况：$G^P(A,B)$ 中包含偶数个 RMSP，这些 RMSP 位于同一条染色体 X；X 上没有体间灰边的端点；X 上有一条 AA 路径和一条 BB 路径，分别记作 P_1 和 P_2. 当且仅当添加灰边连接 P_1 和 P_2 时会产生半偶隔离带. 这种情况下，因 $n>M$，可知 $G^P(A,B)$ 中一定有一条位于 X 之外的短 AA 路径，记作 P_3. 添加灰边将 P_2 与 P_3 连接，即可保证不会产生半偶隔离带. 证毕.

性质 3. 若 G_{i-1} 满足以下 3 个条件：(C_1) $ds_{i-1}=0$；(C_2) G_{i-1} 不包含弱偶隔离带或半偶隔离带；(C_3) G_{i-1} 不包含真偶隔离带或者 G_{i-1} 满足条件(α1)，在 G_{i-1} 中添加一条灰边 g_i，若 g_i 令 $\Delta ds=0$ 且 g_i 不在 G_i 中产生新的 SMSP 和半偶隔离带，则 g_i 一定满足 $\Delta\delta=0$.

定理 3. 总能找到 $n+N$ 条有效灰边添加到 $G^P(A,B)$ 中，使其变为断点图 $G(\hat{A},\hat{B}_{OPT})$.

证明. 假设已经在 $G^P(A,B)$ 中添加了 $i-1$ 条灰边得到 G_{i-1}.

(S_1) 若 G_{i-1} 中存在包含两条 AB 路径的 SMSP，则添加灰边闭合其中一条 AB 路径. 由引理 5 可知，添加的灰边都是有效灰边.

(S_2) 若 G_{i-1} 满足 $ds_{i-1}=1$，则根据引理 6，总能找到一条有效灰边 g_i，使得 G_i 满足 $ds_i=0$.

(S_3) 若 G_{i-1} 包含真偶隔离带且 $sm_{i-1}>0$，则根据引理 7，总能找到一条有效灰边 g_i，使得添加 g_i 后的图 G_i 满足：(1) $ds_i=0$；(2) G_i 中不包含真偶隔离带或者 G_i 满足条件(α1).

(S_4) 若 G_{i-1} 包含弱偶隔离带，因为 $n>M$ 并且前面的步骤中都没有减少 AA 路径，可知 G_{i-1} 至少包含一条 AA 路径，根据引理 8，总能找到一条有效灰边 g_i，使得添加 g_i 后的图 G_i 满足：(1) $ds_i=0$；(2) G_i 中不包含真偶隔离带、弱偶隔离带或半偶隔离带.

(S_5) 若 G_{i-1} 包含半偶隔离带且 $ds_{i-1}=0$，因为 $n>M$ 并且前面的步骤中都没有减少 AA 路径，可知 G_{i-1} 至少包含一条 AA 路径，根据引理 9，总能找到一条有效灰边 g_i，使得添加 g_i 后的图 G_i 满足：(1) $ds_i=0$；(2) G_i 中不包含真偶隔离带或半偶隔离带. 又因为无论如何添加灰边半偶隔离带都不会转化为弱偶隔离带，因而 G_i 中也一定不包含弱偶隔离带.

先按照(S_1) 在 $G^P(A,B)$ 中添加灰边，使部分图中每个 SMSP 只包含一条 AB 路径. 再按照(S_2～S_5) 在部分图中添加灰边，设得到的新部分图仍为

G_{i-1}，则 G_{i-1} 一定满足性质 3 中的条件(C_1)、(C_2)和(C_3)．下面的步骤总假设 G_{i-1} 满足这 3 个条件，并在证明中直接利用性质 3．

(S_6)若 $sm_{i-1}\geq 2$，由于 $ds_{i-1}=0$，根据引理 1 总能找到 $\lfloor sm_{i-1}/2\rfloor$ 条灰边减少 $2*\lfloor sm_{i-1}/2\rfloor$ 个 SMSP．其中每条灰边都满足 $\Delta ds=0$ 且不会产生新的 SMSP 和半偶隔离带，因而都满足 $\Delta l=-1$，$\Delta p=\Delta r=\Delta o=0$，$\Delta sm=-2$，$\Delta\delta=0$，可得 $\Delta=0$．

(S_7)若 $sm_{i-1}=1$，设唯一的一个 SMSP 为 S．因为 G_{i-1} 中存在 SMSP，显然 $n>M$．又因为前面的步骤都没有减少 AA 路径，因而 G_{i-1} 中一定有一条 AA 路径．若 $o_{i-1}=0$，添加 g_i 将 S 中的 AB 路径的 B-tail 与一条 AA 路径连接，则 S 不再是 SMSP．g_i 满足 $\Delta l=-1$，$\Delta p=\Delta r=\Delta o=0$，$\Delta sm=-1$，$\Delta\delta=0$，因而 $\Delta=0$．若 $o_{i-1}=1$，添加 g_i 闭合 S 中的 AB 路径，则 S 变为 RMSP，g_i 满足 $\Delta l=\Delta p=0$，$\Delta r=1$，$\Delta o=\Delta sm=-1$，$\Delta\delta=0$，因而 $\Delta=0$．

(S_8)若 $p_{i-1}>0$，即 G_{i-1} 中存在 BB 路径，因为前面的步骤都不会改变 BB 路径和 AA 路径的数目，根据引理 10，总能找到 p_{i-1} 条灰边将 G_{i-1} 中的每条 BB 路径都与一条 AA 路径连接．该过程中不产生新的 SMSP 和半偶隔离带，因而每条灰边都满足 $\Delta l=\Delta p=\Delta r=\Delta o=\Delta sm=\Delta\delta=0$，$\Delta=0$．

(S_9)若 G_{i-1} 中既没有 SMSP 也没有 BB 路径，添加灰边 g_i 闭合任意一条路径，g_i 不改变任何参数，因而 $\Delta=0$．

循环利用上述方法就能添加 $n+N$ 条有效灰边到 $G^P(A,B)$ 中，得到断点图 $G(\hat{A},\hat{B}_{OPT})$．　　证毕．

定理 3 的证明过程给出了寻找 $n+N$ 条有效灰边的方法，根据该方法在 $G^P(A,B)$ 中添加 $n+N$ 条灰边可将其变为断点图 $G(\hat{A},\hat{B}_{OPT})$．根据定理 2，$\hat{A}$ 和 $\hat{B}_{OPT}$ 的交互型移位距离恰好等于 A，B 的移位距离．用 SRT 的算法对 $\hat{A}$ 和 $\hat{B}_{OPT}$ 进行交互型移位排序，$\hat{A}$ 和 $\hat{B}_{OPT}$ 的最短交互型移位序列即可模拟出 A 和 B 的最短移位序列．完整的基因组移位排序算法 *Generalized_Sorting*(A,B) 给出如下．

算法 1.　*Generalized_Sorting*(A,B).

1. 任选一个 B 的加帽基因组 $\hat{B}$，构造断点图 $G(\hat{A},\hat{B})$
2. 构造部分图 $G=G^P(A,B)$
3. 对于 G 中所有包含两条 AB 路径的 SMSP，添加灰边闭合其中一条 AB 路径
4. if G 中有 2 个 SMSP，并且这两个 SMSP 相互依赖
5. 　按引理 6 找一条有效灰边 g 添加到 G 中，即 $G=G+\{g\}$
6. if G 包含真偶隔离带，且 G 中的 SMSP 数大于 0
7. 　按引理 7 找一条有效灰边 g，$G=G+\{g\}$
8. if G 包含弱偶隔离带
9. 　按引理 8 找一条有效灰边 g，$G=G+\{g\}$
10. if G 包含半偶隔离带
11. 　按引理 9 找一条有效灰边 g，$G=G+\{g\}$
12. while G 中有路径
13. 　if G 中的 SMSP 数不少于 2
14. 　　令 sm 为 G 中的 SMSP 数，按引理 1 找 $\lfloor sm/2\rfloor$ 条有效灰边 $g_1,g_2,\cdots,g_{\lfloor sm/2\rfloor}$，$G=G+\{g_1,g_2,\cdots,g_{\lfloor sm/2\rfloor}\}$
15. 　else if G 中只有一个 SMSP
16. 　　if G 中的 RMSP 数为偶数，
17. 　　令 g 为连接 SMSP 中的 AB 路径的 B-tail 和一条 AA 路径的灰边，$G=G+\{g\}$
18. 　　else
19. 　　令 g 为闭合 SMSP 中的 AB 路径的灰边，$G=G+\{g\}$
20. 　else if G 中有 BB 路径
21. 　　令 p 为 G 中的 BB 路径数，根据引理 10 找 p 条有效灰边 $g_1,g_2,\cdots,g_p$，$G=G+\{g_1,g_2,\cdots,g_p\}$
22. 　else
23. 　　令 g 为闭合任意一条路径的灰边，$G=G+\{g\}$
24. endwhile
25. 根据 G 计算 B 的最优加帽基因组 $\hat{B}_{OPT}$
26. 用 SRT 的算法对 $\hat{A}$ 和 $\hat{B}_{OPT}$ 进行交互型移位排序
27. 用 $\hat{A}$ 和 $\hat{B}_{OPT}$ 的交互型移位序列模拟 A 和 B 的移位序列

下面我们分析算法 *Generalized_Sorting*(A,B) 的时间复杂度．步 1～2 构造断点图和部分图需 $O(n)$时间，步 3～24 添加 $n+N$ 条有效灰边可在$O(n)$时间内完成，步 25 计算 $\hat{B}_{OPT}$ 需 $O(n)$时间．步 26 采用 SRT 的算法对$\hat{A}$和$\hat{B}_{OPT}$排序可在$O(n\sqrt{n\log n})$时间内完成[10]．因而 *Generalized_Sorting*(A,B)的时间复杂度为 $O(n\sqrt{n\log n})$．

参　考　文　献

[1] Hannenhalli S, Pevzner P A. Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals. Journal of the ACM, 1999, 46(1): 1-27

[2] Hannenhalli S. Polynomial-time algorithm for computing translocation distance between genomes. Discrete Applied Mathematics, 1996, 71(1-3): 137-151

[3] Kaplan H, Shamir R, Tarjan R E. Faster and simpler algorithm for sorting signed permutations by reversals//Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, Louisiana, United States, 1997: 344-351

[4] Tannier E, Bergeron A, Sagot M. Advances on sorting by reversals. Discrete Applied Mathematics, 2007, 155(6-7): 881-888

[5] Bafna V, Pevzner P A. Sorting by transpositions. SIAM Journal on Discrete Mathematics, 1998, 11(2): 224-240

[6] Hartman T, Shamir R. A simpler and faster 1.5-approximation algorithm for sorting by transpositions. Information and Computation, 2006, 204(2): 275-290

[7] Elias I, Hartman T A. 1.375-approximation algorithm for sorting by transpositions. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2006, 3(4): 369-379

[8] Zhu Da-Ming, Ma Shao-Han. An improved algorithm for the Translocation sorting problem of genomes. Chinese Journal of Computers, 2002, 25(2): 189-196(in Chinese)
(朱大铭，马绍汉. Translocation 排序问题的改进多项式算法. 计算机学报，2002，25(2)：189-196)

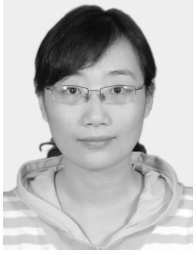
[9] Liu Xiao-Wen, Zhu Da-Ming, Ma Shao-Han, Li Zi-Mao, Wang Lu-Sheng. An $O(n^2)$ algorithm for sorting oriented genomes by translocations. Chinese Journal of Computers, 2004, 27(10): 1354-1360(in Chinese)
(刘晓文，朱大铭，马绍汉，李子茂，王鲁生. 有向基因组移位排序问题的 $O(n^2)$ 快速算法. 计算机学报，2004，27(10)：1354-1360)

[10] Ozery-Flato M, Shamir R. An $O(n^{3/2}\sqrt{\log n})$ algorithm for sorting by reciprocal translocations //Proceedings of the 17th Annual Symposium on Combinatorial Pattern Matching. Barcelona, Spain, 2006: 258-269

[11] Hannenhalli S, Pevzner P A. Transforming men into mice: Polynomial algorithm for genomic distance problem//Proceedings of the 36th Annual Symposium on Foundations of Computer Science. Milwaukee, Wisconsin, 1995: 581-592

[12] Tesler G. Efficient algorithms for multi-chromosomal genome rearrangements. Journal of Computer and System Sciences, 2002, 65(3): 587-609

[13] Ozery-Flato M, Shamir R. Sorting by reciprocal translocations via reversals theory. Journal of Computational Biology, 2007, 14(4): 408-422



YIN Xiao, born in 1982, Ph. D. candidate. Her research interests include algorithm design and analysis, bioinformatics.

ZHU Da-Ming, born in 1964, Ph. D., professor. His research interests include algorithm design and analysis, bioinformatics.

Background

The sorting of genome rearrangement becomes an important research area of computational biology and bioinformatics. It asks to compute a shortest sequence of rearrangement operations that transforms one genome into another. There are three classical genome rearrangement operations: Reversal, Translocation and Transposition. This paper focus on the problem of sorting signed genomes by translocations. The result of this paper is a part of National Natural Science Foundation of China project "Genome rearrangement and comparison algorithms and complexity". The project involves computing the genome rearrangement distance between two signed or unsigned genomes. The distance is viewed as a better estimation of the affinity relations among species, and it is helpful to guess the real biological evolution process.

For the problem of sorting unsigned genomes by reciprocal translocations, the authors proved it to be NP-hard in 2007 and presented a ratio－1.75 polynomial-time approximation algorithm in 2006 and a ratio－1.5 approximation algorithm in 2008. The problem of sorting signed genomes by reciprocal translocations was first solved in $O(n^3)$ time. The authors improved the algorithm so that the time complexity is improved to $O(n^2 \log n)$ in 2000 and $O(n^2)$ in 2004. In 2006, this algorithm is improved to $O(n\sqrt{n\log n})$. Translocations include reciprocal translocations and non-reciprocal translocations. All of those algorithms only allow reciprocal translocations. Thus they can only be applied to a pair of genomes having the same set of chromosome ends. Such a restriction can be removed if non-reciprocal translocations are also allowed. This paper presents a polynomial-time algorithm for sorting signed genomes by generalized translocations including reciprocal and non-reciprocal translocations, which confirms Ozery-Flato's conjecture that sorting by generalized translocations could be solved in polynomial time.