Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Chapter 1. Introduction

Hung-Hsun Chen

Department of Applied Mathematics
National Chiao Tung University, Taiwan

October 15, 2012

Outline

1 1.1 A Minimalist Introduction to Molecular Evolution

2 1.2 Birth of the Combinatorics of Genome Rearrangements

3 1.3 Statement of the Problem

4 1.4 Scope of This Survey

5 1.5 Overview of the Models

6 1.6 Organization of the Book

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Introduction

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

In this book, we will start with a short introduction to molecular evolution, for conceptual and historical purpose.

This introduction is not necessary to understand the combinatorial problems and their solutions, but it allows us to place them in their context and explain why they are important, independently of their mathematical value.

Introduction

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

In this book, we will start with a short introduction to molecular evolution, for conceptual and historical purpose.

This introduction is not necessary to understand the combinatorial problems and their solutions, but it allows us to place them in their context and explain why they are important, independently of their mathematical value.

The "molecules of heredity", the support of genetic information, are present in every cell of all living organisms(bacteria, plants, animal, etc.).

Definition 1.1

- *DNA: a double-stranded molecule in which each strand is a long succession of nucleotides.*
- *Chromosome: Each molecule is called a chromosome, are made of DNA.*
- *Genome: the set of all chromosomes.*

The "molecules of heredity", the support of genetic information, are present in every cell of all living organisms(bacteria, plants, animal, etc.).

Definition 1.1

- *DNA: a double-stranded molecule in which each strand is a long succession of nucleotides.*
- *Chromosome: Each molecule is called a chromosome, are made of DNA.*
- *Genome: the set of all chromosomes.*

The "molecules of heredity", the support of genetic information, are present in every cell of all living organisms(bacteria, plants, animal, etc.).

Definition 1.1

- *DNA: a double-stranded molecule in which each strand is a long succession of nucleotides.*
- *Chromosome: Each molecule is called a chromosome, are made of DNA.*
- *Genome: the set of all chromosomes.*

The "molecules of heredity", the support of genetic information, are present in every cell of all living organisms(bacteria, plants, animal, etc.).

Definition 1.1

- *DNA: a double-stranded molecule in which each strand is a long succession of nucleotides.*
- *Chromosome: Each molecule is called a chromosome, are made of DNA.*
- *Genome: the set of all chromosomes.*

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

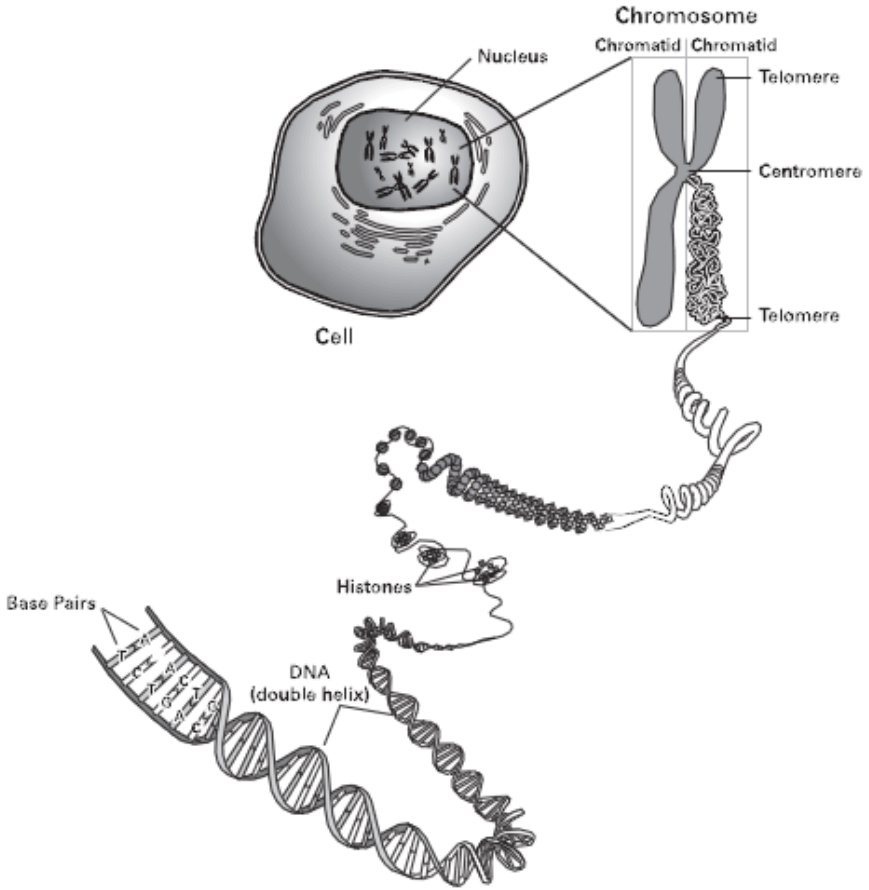


Figure 1.1
A chromosome and a fragment of a DNA molecule
Source: National Institutes of Health, National Human Genome Research Institute, Division of Intramural Research

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Remark 1.2

Nucleotides can be of four types - A, C, G, T.

An A on one strand is always coupled with a T and a C on one strand is always coupled with G.

A - T C - G

Definition 1.3

Complementary: the sequence on one strand determines the sequence on the other one.

Because of complementary, a DNA molecule is usually represented as a single sequence, but the organization in two strands will often be crucial for our purpose.

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Remark 1.2

Nucleotides can be of four types - A, C, G, T.

An A on one strand is always coupled with a T and a C on one strand is always coupled with G.

$$A - T \qquad C - G$$

Definition 1.3

Complementary: the sequence on one strand determines the sequence on the other one.

Because of complementary, a DNA molecule is usually represented as a single sequence, but the organization in two strands will often be crucial for our purpose.

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Remark 1.2

Nucleotides can be of four types - A, C, G, T.

An A on one strand is always coupled with a T and a C on one strand is always coupled with G.

$$A - T \qquad C - G$$

Definition 1.3

Complementary: the sequence on one strand determines the sequence on the other one.

Because of complementary, a DNA molecule is usually represented as a single sequence, but the organization in two strands will often be crucial for our purpose.

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Remark 1.2

Nucleotides can be of four types - A, C, G, T.

An A on one strand is always coupled with a T and a C on one strand is always coupled with G.

$$A - T \qquad C - G$$

Definition 1.3

Complementary: the sequence on one strand determines the sequence on the other one.

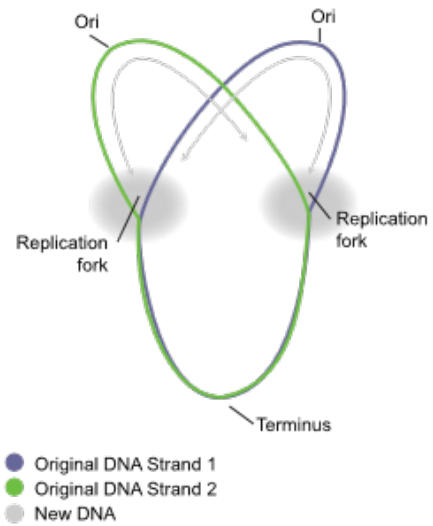
Because of complementary, a DNA molecule is usually represented as a single sequence, but the organization in two strands will often be crucial for our purpose.

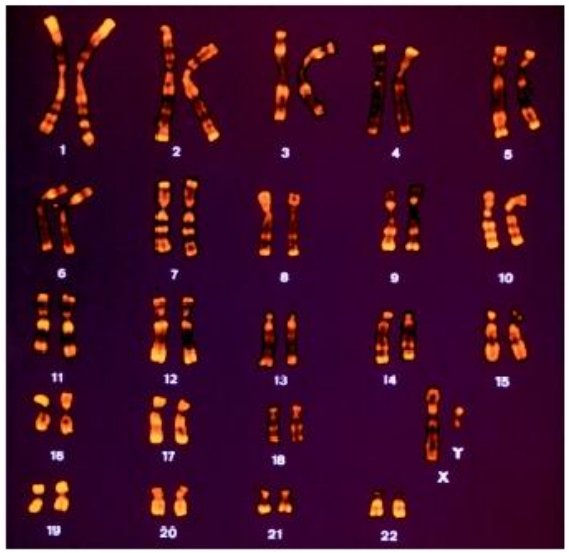
Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

There are two types of chromosomes:

- circular: the sequence forms a circle and has no ends. Ex: bacteria.
- linear: the sequence has two ends, called telomeres. Ex: Plants, animals.



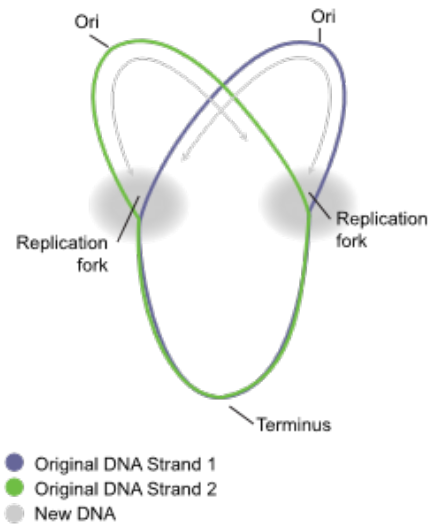


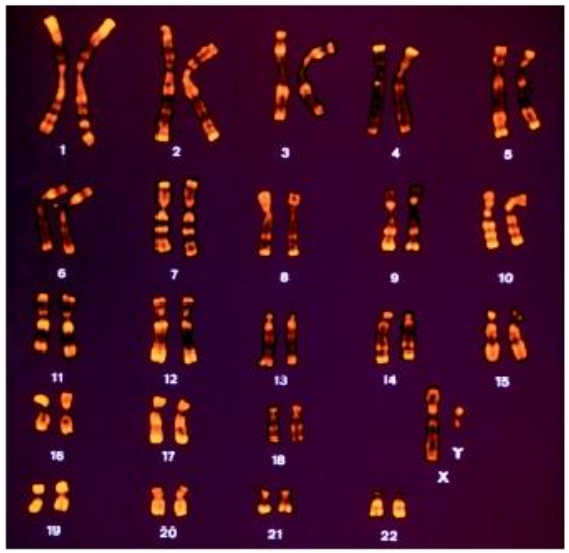
Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

There are two types of chromosomes:

- circular: the sequence forms a circle and has no ends. Ex: bacteria.
- linear: the sequence has two ends, called telomeres. Ex: Plants, animals.



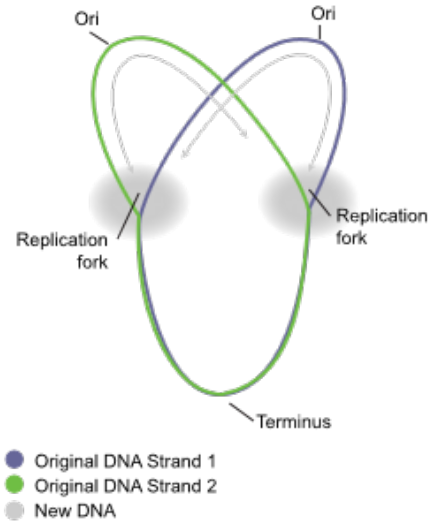


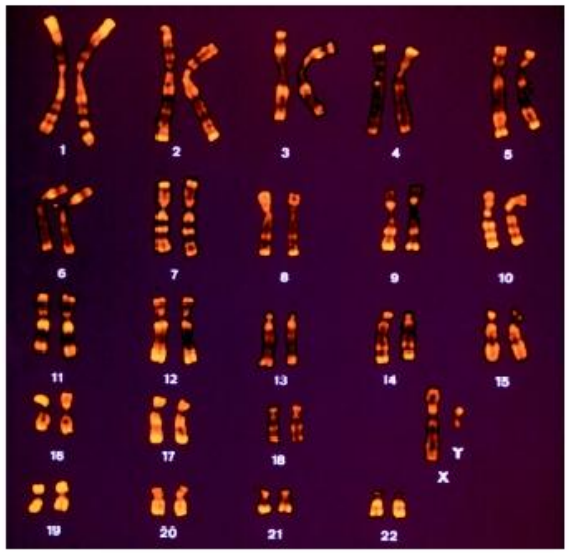
Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

There are two types of chromosomes:

- circular: the sequence forms a circle and has no ends. Ex: bacteria.
- linear: the sequence has two ends, called telomeres. Ex: Plants, animals.





Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Remark 1.4

- *A segment of DNA is a part of this molecule made of consecutive nucleotides.*
- *A gene is a segment of DNA that contains the information needed to construct the other molecules in the cell.*

Introduction

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Remark 1.4

- *A segment of DNA is a part of this molecule made of consecutive nucleotides.*
- *A gene is a segment of DNA that contains the information needed to construct the other molecules in the cell.*

Introduction

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Remark 1.4

- *A segment of DNA is a part of this molecule made of consecutive nucleotides.*
- *A gene is a segment of DNA that contains the information needed to construct the other molecules in the cell.*

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Evolution

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

A DNA molecule may evolve on different scale as follows:

- point mutations: at the level of nucleotides.
- rearrangements(or structural variations): on a sequence.

Remark 1.5

- *Detecting the events of point mutation is the goal of sequence alignment.*
- *Detecting the events of rearrangements is the goal of genome rearrangement problem.*

(for a presentation of this topic, see, for example, Setubal and Meidanis [333] or Jones and Pevzner[224]).

Point Mutation

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Point mutation:

- substitutions: one nucleotide is replaced with another.
  Ex: GTGCGTACACT becomes GTGCCTACACT.
- insertions: a nucleotide is added to the sequence.
  Ex: GTGCTACACT becomes GTGCCTACACT.
- deletions: a nucleotide is removed from the sequence.
  Ex: GTGCGTACACT becomes GTGCTACACT.

Point Mutation

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Point mutation:

- substitutions: one nucleotide is replaced with another.
  Ex: GTGCGTACACT becomes GTGCCTACACT.
- insertions: a nucleotide is added to the sequence.
  Ex: GTGCTACACT becomes GTGCCTACACT.
- deletions: a nucleotide is removed from the sequence.
  Ex: GTGCGTACACT becomes GTGCTACACT.

Point Mutation

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Point mutation:

- substitutions: one nucleotide is replaced with another.
  Ex: GTGCGTACACT becomes GTGCCTACACT.
- insertions: a nucleotide is added to the sequence.
  Ex: GTGCTACACT becomes GTGCCTACACT.
- deletions: a nucleotide is removed from the sequence.
  Ex: GTGCGTACACT becomes GTGCTACACT.

Point Mutation

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Point mutation:

- substitutions: one nucleotide is replaced with another.
  Ex: GTGCGTACACT becomes GTGCCTACACT.
- insertions: a nucleotide is added to the sequence.
  Ex: GTGCTACACT becomes GTGCCTACACT.
- deletions: a nucleotide is removed from the sequence.
  Ex: GTGCGTACACT becomes GTGCTACACT.

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Deletions
  Ex: CCGTGCGTACACTGC becomes CCGTACTGC.

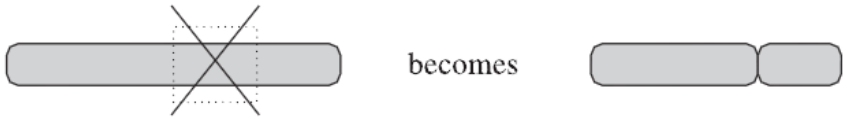


Figure 1.2
Deletion of the dotted region in a chromosome

- Transpositions
  Ex: CCGTGCGTACACTGC becomes CCGTACTGCGCGTAC.

Figure 1.3
Transposition of the dotted region in a chromosome

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Deletions
  Ex: CCGTGCGTACACTGC becomes CCGTACTGC.

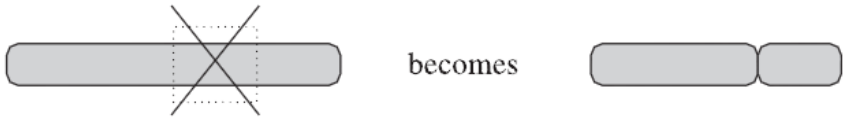


Figure 1.2
Deletion of the dotted region in a chromosome

- Transpositions
  Ex: CCGTGCGTACACTGC becomes CCGTACTGCGCGTAC.

Figure 1.3
Transposition of the dotted region in a chromosome

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Duplications
  - tandem duplications
    Ex: CCGTGCGTACACTGC becomes CCGTGCGTACGCGTACACTGC

Figure 1.5
Tandem duplication of the dotted region in a chromosome

  - retrotranspositions
    insert a copy of a gene at an arbitrary location in the genome.
  - whole genome duplications
    copy either the whole genome or some of its chromosomes.

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Duplications
  - tandem duplications
    Ex: CCGTGCGTACACTGC becomes CCGTGCGTACGCGTACACTGC

Figure 1.5
Tandem duplication of the dotted region in a chromosome

  - retrotranspositions
    insert a copy of a gene at an arbitrary location in the genome.
  - whole genome duplications
    copy either the whole genome or some of its chromosomes.

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Duplications
  - tandem duplications
    Ex: CCGTGCGTACACTGC becomes CCGTGCGTACGCGTACACTGC

Figure 1.5
Tandem duplication of the dotted region in a chromosome

  - retrotranspositions
    insert a copy of a gene at an arbitrary location in the genome.
  - whole genome duplications
    copy either the whole genome or some of its chromosomes.

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Duplications
  - tandem duplications
    Ex: CCGTGCGTACACTGC becomes CCGTGCGTACGCGTACACTGC

Figure 1.5
Tandem duplication of the dotted region in a chromosome

  - retrotranspositions
    insert a copy of a gene at an arbitrary location in the genome.
  - whole genome duplications
    copy either the whole genome or some of its chromosomes.

Rearrangement

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

- Inversions or reversals

CCGTGCGTACACTGC becomes CCGT GTACGC ACTGC

Figure 1.4
Reversal of the underlined segment, resulting in the boxed segment

- Reciprocal translocation

Ex:

$$\begin{array}{l} CCGTGCGTACA\underline{CTGC} \\ ACTGCCCGTG\underline{CGTAC} \end{array} \Rightarrow \begin{array}{l} CCGTGCGTACA\underline{CGTAC} \\ ACTGCCCGTG\underline{CTGC} \end{array}$$

Figure 1.6
Reciprocal translocation of the dotted regions in two chromosomes

- Fusion
  Ex:

$$\frac{CGTAC}{CTGC} \Rightarrow CGTACCTGC$$

Figure 1.7
Fusion of two chromosomes

- Fission: One chromosome splits into two (this is the inverse of a fusion). Ex:

$$CGTACCTGC \Rightarrow \frac{CGTAC}{CTGC}$$

- Horizonal, or lateral, transfer
  a segment of the genome is copied from one genome to another.

- Fusion

  Ex:

$$\frac{CGTAC}{CTGC} \Rightarrow \underline{CGTACCTGC}$$

Figure 1.7
Fusion of two chromosomes

- Fission: One chromosome splits into two (this is the inverse of a fusion). Ex:

$$\underline{CGTACCTGC} \Rightarrow \frac{CGTAC}{CTGC}$$

- Horizonal, or lateral, transfer

  a segment of the genome is copied from one genome to another.

- Fusion
  Ex:

$$\frac{CGTAC}{CTGC} \Rightarrow \underline{CGTACCTGC}$$

Figure 1.7
Fusion of two chromosomes

- Fission: One chromosome splits into two (this is the inverse of a fusion). Ex:

$$\underline{CGTACCTGC} \Rightarrow \frac{CGTAC}{CTGC}$$

- Horizonal, or lateral, transfer
  a segment of the genome is copied from one genome to another.

- Fusion
  Ex:

$$\frac{CGTAC}{CTGC} \Rightarrow \underline{CGTACCTGC}$$

Figure 1.7
Fusion of two chromosomes

- Fission: One chromosome splits into two (this is the inverse of a fusion). Ex:

$$\underline{CGTACCTGC} \Rightarrow \frac{CGTAC}{CTGC}$$

- Horizonal, or lateral, transfer
  a segment of the genome is copied from one genome to another.

Definition 1.6

Two segments are said to be homologous if

- *derive from a common ancestor, and*
- *distinguished by a replication event. or a duplication event.*

Definition 1.6

Two segments are said to be homologous if

- *derive from a common ancestor, and*
- *distinguished by a replication event. or a duplication event.*

Definition 1.6

Two segments are said to be homologous if

- *derive from a common ancestor, and*
- *distinguished by a replication event. or a duplication event.*

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

- In 1936, Dobzhansky and Sturtevant proposed for the first time to use the degree of disorder between the organization of genes in two different genomes as an indicator of an evolutionary.([145, 146])
- In 1941, Sturtevant and Novitsku formulated the problem of minimizing the number of inversions that may explain the differences in arrangements between two species.([343])

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

- In 1936, Dobzhansky and Sturtevant proposed for the first time to use the degree of disorder between the organization of genes in two different genomes as an indicator of an evolutionary.([145, 146])
- In 1941, Sturtevant and Novitsku formulated the problem of minimizing the number of inversions that may explain the differences in arrangements between two species.([343])

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

All these studies were based on the parsimony criterion, which makes molecular biologists often prefer explanation of differences between genomes that involve as few mutation as possible.

This principle makes the connection with combinatorial optimization possible.

In order to overcome the limit of nine genes stated by Sturtevant and Novitski[343], Watterson[369] proposed to represent the relative positions of genes in different genomes as permutations.

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

All these studies were based on the parsimony criterion, which makes molecular biologists often prefer explanation of differences between genomes that involve as few mutation as possible.

This principle makes the connection with combinatorial optimization possible.

In order to overcome the limit of nine genes stated by Sturtevant and Novitski[343], Watterson[369] proposed to represent the relative positions of genes in different genomes as permutations.

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

All these studies were based on the parsimony criterion, which makes molecular biologists often prefer explanation of differences between genomes that involve as few mutation as possible.

This principle makes the connection with combinatorial optimization possible.

In order to overcome the limit of nine genes stated by Sturtevant and Novitski[343], Watterson[369] proposed to represent the relative positions of genes in different genomes as permutations.

History of genome rearrangements

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

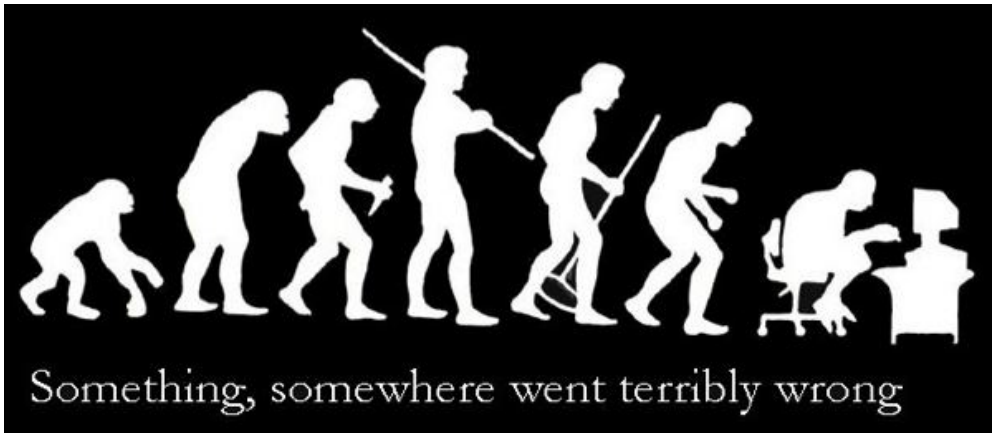
1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

In order to propose an evolutionary scenario between two species, one had to solve the problem of transforming one circular permutation into another with a minimum number of inversions.



Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The genome rearrangement problem is formulated in its most general form as follows:

given a set of genomes and a set of possible evolutionary events, find a shortest set of events transforming those genomes into one another.

Questions:

1. What "genome" means here, and what events are?
2. What "shortest" means here?

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The genome rearrangement problem is formulated in its most general form as follows:

given a set of genomes and a set of possible evolutionary events, find a shortest set of events transforming those genomes into one another.

Questions:

1. What "genome" means here, and what events are?
2. What "shortest" means here?

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The genome rearrangement problem is formulated in its most general form as follows:

given a set of genomes and a set of possible evolutionary events, find a shortest set of events transforming those genomes into one another.

Questions:

1. What "genome" means here, and what events are?
2. What "shortest" means here?

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The genome rearrangement problem is formulated in its most general form as follows:

given a set of genomes and a set of possible evolutionary events, find a shortest set of events transforming those genomes into one another.

Questions:

1. What "genome" means here, and what events are?
2. What "shortest" means here?

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Q1. What "genome" means here, and what events are?

It makes the diversity of the problem. Miscellaneous models have been proposed, depending on various parameters.

Q2. What "shortest" means here?

- the number of events.
- least weight if events are weighted.

The length (or weight) of an optimal sequence of events transforming one genome into another is called the distance between the two genomes.

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Q1. What "genome" means here, and what events are?

It makes the diversity of the problem. Miscellaneous models have been proposed, depending on various parameters.

Q2. What "shortest" means here?

- the number of events.
- least weight if events are weighted.

The length (or weight) of an optimal sequence of events transforming one genome into another is called the distance between the two genomes.

Statement of the Problem

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Q1. What "genome" means here, and what events are?

It makes the diversity of the problem. Miscellaneous models have been proposed, depending on various parameters.

Q2. What "shortest" means here?

- the number of events.
- least weight if events are weighted.

The length (or weight) of an optimal sequence of events transforming one genome into another is called the distance between the two genomes.

Scope of This Survey

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The motivation for this book is twofold:

- those problems deserve at least to mentioned here, since they are closely related to genome rearrangement problems.
- the study of related problems or variants of out problems may provide insight on the original problems we are interested in.

Scope of This Survey

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The motivation for this book is twofold:

- those problems deserve at least to mentioned here, since they are closely related to genome rearrangement problems.
- the study of related problems or variants of out problems may provide insight on the original problems we are interested in.

Scope of This Survey

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

The motivation for this book is twofold:

- those problems deserve at least to mentioned here, since they are closely related to genome rearrangement problems.
- the study of related problems or variants of out problems may provide insight on the original problems we are interested in.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Depending on the assumptions that are made on the data, or the events we want to study, different models can be used.

Model:

- Permutation.
- Partially ordered set.
- Strings.
- Disjoint cycle decomposition of permutations.
- Disjoint sets of paths and cycles.
- Collections of sets.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Depending on the assumptions that are made on the data, or the events we want to study, different models can be used.

Model:

- Permutation.
- Partially ordered set.
- Strings.
- Disjoint cycle decomposition of permutations.
- Disjoint sets of paths and cycles.
- Collections of sets.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Assumption:

1. the order of genes in each genome is known.
2. all genomes share the same set of genes.
3. all genomes contain a single copy of each gene.
4. all genomes consist of a single chromosome

If each gene can be assigned a unique number and is found exactly ones in each genome, we use the model of permutations.

Models

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Model by partial order sets:
If genomes have been only partially sequenced. (Ch.5)

Model by Strings:
If genes do not appear exactly once in each genome: due to duplications and deletions.

A great part of living organisms have a genome that consists of several chromosomes(ex: animals), and permations as we have presented them are no longer a realistic model in this case.

- disjoint cycle decomposition of permutations:
  if chromosomes are circular.
- disjoint sets of paths and cycles:
  if chormosomes are linear.

Models

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Model by partial order sets:
If genomes have been only partially sequenced. (Ch.5)

Model by Strings:
If genes do not appear exactly once in each genome: due to duplications and deletions.

A great part of living organisms have a genome that consists of several chromosomes(ex: animals), and permations as we have presented them are no longer a realistic model in this case.

- disjoint cycle decomposition of permutations:
  if chromosomes are circular.
- disjoint sets of paths and cycles:
  if chormosomes are linear.

Models

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Model by partial order sets:
If genomes have been only partially sequenced. (Ch.5)

Model by Strings:
If genes do not appear exactly once in each genome: due to duplications and deletions.

A great part of living organisms have a genome that consists of several chromosomes(ex: animals), and permations as we have presented them are no longer a realistic model in this case.

- disjoint cycle decomposition of permutations:
  if chromosomes are circular.
- disjoint sets of paths and cycles:
  if chormosomes are linear.

Models

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Model by partial order sets:
If genomes have been only partially sequenced. (Ch.5)

Model by Strings:
If genes do not appear exactly once in each genome: due to duplications and deletions.

A great part of living organisms have a genome that consists of several chromosomes(ex: animals), and permations as we have presented them are no longer a realistic model in this case.

- disjoint cycle decomposition of permutations:
  if chromosomes are circular.
- disjoint sets of paths and cycles:
  if chormosomes are linear.

Models

Chapter 1. Introduction
Hung-Hsun Chen
1.1 A Minimalist Introduction to Molecular Evolution
1.2 Birth of the Combinatorics of Genome Rearrangements
1.3 Statement of the Problem
1.4 Scope of This Survey
1.5 Overview of the Models
1.6 Organization

Model by partial order sets:
If genomes have been only partially sequenced. (Ch.5)

Model by Strings:
If genes do not appear exactly once in each genome: due to duplications and deletions.

A great part of living organisms have a genome that consists of several chromosomes(ex: animals), and permations as we have presented them are no longer a realistic model in this case.

- disjoint cycle decomposition of permutations:
  if chromosomes are circular.
- disjoint sets of paths and cycles:
  if chormosomes are linear.

Organization of the Book

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Organization of the Book

I Duplication-free models: permutations.

II Models handing duplications: strings.

III Multi-chromosomal models.

IV Median problem.

V A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Organization of the Book

- I Duplication-free models: permutations.
- II Models handing duplications: strings.
- III Multi-chromosomal models.
- IV Median problem.
- V A collection of summaries.
- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Organization of the Book

Organization of the Book

I. Duplication-free models: permutations.

II. Models handing duplications: strings.

III. Multi-chromosomal models.

IV. Median problem.

V. A collection of summaries.

- Appendix A: Graph theory.
- Appendix B: Algorithmic theory of complexity.

Reference

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

[74] H.-J BÖCKENHAUER AND D.BONGARTZ, eds., Algorithmic Aspects of Bioinformatics, Natural Computing Series, Springer-Verlag, 2007. Original German edition published by Teubner.

[145] T.DOBZHANSKY AND A.H.STURTEVANT, Inversions in the third chromosome of wild races of Drosophila pseudoobscura, and their use in the study of the history of the species, Proceedings of the National Academy of Sciences (USA), 22, no.7 (1936), 448-450.

[146] T.DOBZHANSKY AND A.H.STURTEVANT, Inversions in the chromosomes of Drosophila pseudoobscura, Genetics, 23 (1938), pp. 28-64.

[150] R.DURRETT, Genome rearrangement: Recent progress and open problems, 2003. http://www.math.cornel.edu/ durrett/FGR.

Reference

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

[165] K.ERIKSSON, Statistical and combinatorial aspects of comparative genomics, Scandinavian Journal of Statistics, 31, no.2 (2004), pp. 203-216.

[182] O.GASCUEL, ed., Mathematics of Evolution and Phylogeny, Oxford University Press, 2005.

[197] S.HANNENHALLI AND P.A.PEVZNER, Towards a computational theory of genome rearrangements, in Computer Science Today: Recent Trends and Developments, Jan van Leeuwen, ed., vol. 1000 of Lecture Notes in Computer Science, Springer-Verlag, 1995, pp. 184-202.

[222] T.JIANG, Y.XU, AND M.Q.ZHANG, eds., Current Topics in Computational Molecular Biology, MIT Press, 2002.

Reference

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

[224] N.C.JONES AND P.A. PEVZNER, An Introduction to Bioinformatics Algorithms, MIT Press, 2004.

[247] Z.LI, L.WANG, AND K.ZHANG, Algorithmic approaches for genome rearrangement: A review, IEEE Transactions on Sustems, Man and Cybernetics, part C, 36, no.5 (2006), pp. 636-648.

[274] B.M.E.MORET, S.K.WYMAN, D.A. BADER, T.WARNOW, AND M.YAN, A new implementation and detailed study of breakpoint analysis, in Proceedings of the Sixth Pacific Symposium on Biocomputing (PSB), World Scientific Press, 2001, pp. 583-594.

[289] J.D.PALMER AND L.A.HERBON, Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence, Journal of Molecular Evolution, 28, no. 1-2 (1988), pp. 87-97.

Reference

[296]P.A. PEVZNER, Computational Molecular Biology: An Algorithmic Approach, MIT Press, 2000.

[301] P.POPESCU, H.HAYES, AND B.DUTRILLAUX, Techniques in Animal Cytogenetics, Springer-Verlag, 2000.

[322] D.SANKOFF AND J.H.NADEAU, eds., Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and the Evolution of Gene Families, vol. 1 of Computational Biology, Kluwer Academic Pree, 2000.

Reference

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

[333] J.MEIDANIS AND J.SETUBAL, Introduction to Computational Molecular Biology, PWS Publishing, 1997.

[343] A.H.STURTEVANT AND E.NOVITSKI, The homologies of the chromosome elements in the genus Drosophila, Genetics, 26, no. 5 (1941), pp. 517-541.

[360] C.-W. TSENG AND M.V.ZELKOWITZ, eds., Computational Biology and Bioinformatics, vol. 68 of Advances in Computers, Elevier Academic Press, 2006.

Wikipedia (http://www.wikipedia.org/).

The End

Chapter 1. Introduction

Hung-Hsun Chen

1.1 A Minimalist Introduction to Molecular Evolution

1.2 Birth of the Combinatorics of Genome Rearrangements

1.3 Statement of the Problem

1.4 Scope of This Survey

1.5 Overview of the Models

1.6 Organization

Thank for your attention.